Supplementary Material for
A Bayesian model for sparse graphs with flexible degree distribution and overlapping community structure

Juho Lee
University of Oxford
AITRICS

Lancelot F. James
HKUST

Seungjin Choi
POSTECH

François Caron
University of Oxford

A On scale-free networks

A recent article by Broido and Clauset (2018) (BC) raised concerns about the claim that many real-world networks are scale-free. BC performed a statistical analysis on a large number of networks to test whether the degree distribution follows a power-law distribution or some alternative distributions. One of the alternative degree distributions considered is a power-law distribution with an exponential cut-off, which was shown to provide a better fit for a majority of the datasets considered. BC conclude in their article that scale-free networks are rare.

While we agree with the authors that there is a need for rigorous statistical testing of the scale-free hypothesis, and that scale-free networks may indeed by more rare than originally thought, we do not think that the conclusion of the authors is supported by their experiments, except if one considers a very narrow definition of a scale-free network. As pointed out by Barabasi (2018) in a blog post discussing their article, scale-freeness is an asymptotic property: as the sample size goes to infinity, the degree distribution converges to a power-law (up to a slowly varying function, see Definition 2.1). Degree distributions of finite-size graphs may still depart significantly from a pure power-law distribution.

A salient example is given by the class of networks introduced by Caron and Fox (2017), which are known to be scale-free with exponent between 1 and 2 for some values of the parameters. As shown in Figure 1, while the degree distribution is asymptotically power-law, any finite-size graph exhibits an exponential cut-off, which shifts to the right as the sample size increases. Therefore, any statistical test on a fixed-n graph is likely to reject the pure power-law hypothesis although the network model is indeed scale-free.

For reference, empirical degree distributions for the IG-NR and GIG-NR are also plotted in Figure 1. The IG-NR is scale-free, and each finite-n distribution is additionally close to a pure power-law distribution. The GIG-NR is not scale-free, and the asymptotic degree distribution is a power-law distribution with exponential cut-off.

B Background material on regular variation

In this section we give some definitions and properties of slowly and regularly varying functions, see the books of Bingham et al. (1989), Mikosch (1999) and Resnick (2007) for reference.

Definition B.1. *A positive function $f : \mathbb{R}_+ \to \mathbb{R}_+$ is regularly varying at infinity with index $\alpha \in \mathbb{R}$ if*

$$\lim_{t\to\infty} \frac{f(ct)}{f(t)} = c^\alpha$$

for all $c > 0$.

If $\alpha = 0$, the function is said to be slowly varying. Examples of slowly varying include constant functions, functions converging to a strictly positive constant, logarithms, etc. If a function f is regularly varying with index α, then there exists a slowly varying function ℓ such that

$$f(x) \sim \ell(x)x^\alpha$$

as $x \to \infty$.

Definition B.2. *A non-negative random variable X with cdf F_X is said to be regularly varying with exponent $\alpha \geq 0$ if*

$$1 - F_X(x) \sim \ell(x)x^{-\alpha} \tag{1}$$

as x tends to infinity, where ℓ is a slowly varying function at infinity. If $\alpha > 0$ and F_X is absolutely continuous with density f_X, where f_X is ultimately monotone, then

$$f_X(x) \sim \alpha\ell(x)x^{-\alpha-1} \tag{2}$$

as $x \to \infty$.

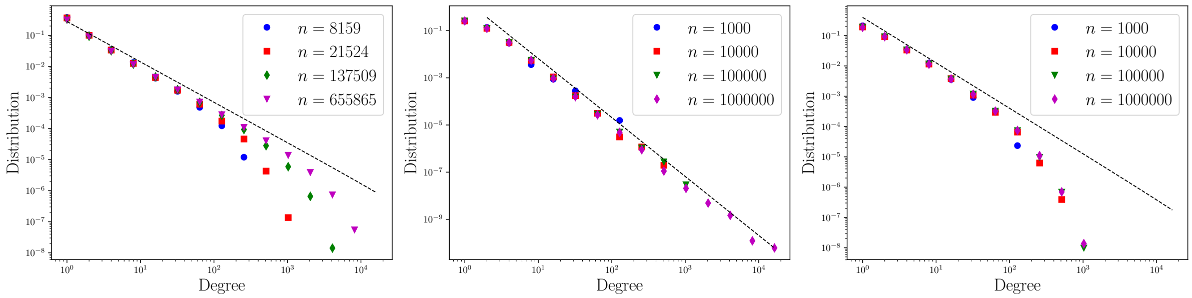


Figure 1: Empirical degree distribution for networks for growing sizes generated from the (left) GGP model, (middle) IG-NR model and (right) GIG-NR model. Both the GGP and IG-NR models are scale-free networks, with asymptotically power-law degree distributions.

Proposition B.1. *Let X be a regularly varying random variable with exponent $\alpha > 0$, and Y be a positive random variable, independent of X, with $\mathbb{E}(Y^{\alpha+\varepsilon}) < \infty$ for some $\varepsilon > 0$. Then $Z = XY$ is regularly varying with exponent $\alpha > 0$ and*

$$1 - F_Z(z) \sim \mathbb{E}(Y^\alpha)\ell(z)z^{-\alpha} \tag{3}$$

as z tends to infinity. If F_Z is absolutely continuous with ultimately monotone density f_Z, this implies

$$f_Z(z) \sim \alpha\mathbb{E}(Y^\alpha)\ell(z)z^{-\alpha-1}$$

as z tends to infinity.

Proposition B.2. *If f is a regularly varying function with index α, then*

$$\lim_{x\to\infty} \frac{\log f(x)}{\log x} = \alpha.$$

C Background on inhomogeneous random graph models

In this section, we review the general framework of inhomogeneous random graphs (IRG) presented in Bollobás et al. (2007). We start by introducing a *vertex space* and a *kernel* to define IRGs.

Definition C.1. *A **vertex space** $\mathcal{V}$ is a triplet $(\mathcal{S}, \mu, (\mathbf{x}_n)_{n\geq 1})$, where $\mathcal{S}$ is a separable metric space, μ is a Borel probability measure on $\mathcal{S}$, and $\mathbf{x}_n := (x_1, \ldots, x_n)$[1] is a random sequence of n points in $\mathcal{S}$ such that for each $n \geq 1$,*

$$\nu_n := \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i} \xrightarrow{p} \mu, \tag{4}$$

*where $\xrightarrow{p}$ denotes the convergence in probability. The pair $(\mathcal{S}, \mu)$ is called a **ground space**.*

[1] To be precise, we should write $\mathbf{x}_n = (x_1^{(n)}, \ldots, x_n^{(n)})$, but we omit the superscript for simplicity.

Definition C.2. *A **kernel** κ on a ground space $(\mathcal{S}, \mu)$ is a symmetric non-negative Borel measurable function on $\mathcal{S} \times \mathcal{S}$.*

Roughly speaking, a vertex space is a space of values assigned to vertices in a graph, such as vertex weights or popularities. Each vertex is associated with a point in $\mathcal{S}$, and these points are used to contruct edge probabilities between vertices through the kernel κ. Kernels should be further restricted to be in a class of functions satisfying some conditions, and we will explain those shortly after. Given a vertex space and a kernel, an IRG is defined with link function (edge probabilities)

$$p_{ij}^{(n)} = \frac{\kappa(x_i, x_j)}{n} \wedge 1. \tag{5}$$

All the following arguments will be explained with this choice of link function, but everything still holds with the following alternative choices of link functions (Bollobás et al., 2007, Remark 2.4).

$$p_{ij}^{(n)} = 1 - \exp\left(-\frac{\kappa(x_i, x_j)}{n}\right) \tag{6}$$

$$p_{ij}^{(n)} = \frac{\kappa(x_i, x_j)}{n + \kappa(x_i, x_j)}. \tag{7}$$

All these three functions are related to existing works on IRGs. The link function (5) is a generic version of Chung and Lu (2002), (6) is for Norros and Reittu (2006), and (7) is for Britton et al. (2006). We chose (6) for our model because of the computational efficiency in posterior inference.

Let G_n be a graphs generated from IRG described above with a vertex space $\mathcal{V}$ and a kernel κ. The kernel κ are assumed to be *graphical*, which is defined as follows.

Definition C.3. *A kernel κ on a vertex space $(\mathcal{S}, \mu, (\mathbf{x}_n))$ is **graphical** if the followings hold:*

(i) κ is continuous almost everywhere on $\mathcal{S} \times \mathcal{S}$.

(ii) $\kappa \in L^1(\mathcal{S} \times \mathcal{S}, \mu \times \mu)$.

(iii) Let E_n be the set of edges in G_n. Then,

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{E}(|E_n|) = \frac{1}{2}\iint_{\mathcal{S}\times\mathcal{S}} \kappa(x,y)d\mu(x)d\mu(y). \tag{8}$$

The first and second conditions are natural technical requirements. The third condition is related to the density of graphs. It requires κ to measure the density of the edges (Bollobás et al., 2007).

The following theorem characterizes the asymptotic degree distribution of IRGs.

Theorem C.1. *((Bollobás et al., 2007, Theorem 3.13)) Let κ be a graphical kernel on a vertex space $\mathcal{V}$. For any fixed $k \geq 0$,*

$$\frac{N_k^{(n)}}{n} \xrightarrow{p} \int_{\mathcal{S}} \frac{\lambda(x)^k}{k!} e^{-\lambda(x)} d\mu(x), \tag{9}$$

where $N_k^{(n)}$ is the number of vertices in with degree k in G_n, and

$$\lambda(x) := \int_{\mathcal{S}} \kappa(x,y)d\mu(y). \tag{10}$$

Hence, one can easily compute the asymptotic degree distribution of any IRG that fits into the framework once the corresponding vertex space and kernel is specified. This is what we do in the next two sections.

D Proof of Theorem 3.1 and Theorem 3.2

Theorem 3.1 and Theorem 3.2 in the main paper are directly obtained by showing that the Norros-Reittu IRG (NR-IRG) fits into the general framework discussed in Section C of this supplementary material. Actually, the NR-IRG has been discussed as an example of rank-1 IRGs, see (Bollobás et al., 2007, Section 16.4). More precisely, define a vertex space $\mathcal{V} = (\mathcal{S}, \mu, \mathbf{x}_n)$ with

$$\mathcal{S} = (0,\infty), \quad \mu = \mathcal{L}_{w_1}, \quad x_i = w_i\sqrt{\frac{\mathbb{E}(w_1)}{s^{(n)}/n}}, \tag{11}$$

where $\mathcal{L}_{w_1}$ denotes the law of w_1, and define a kernel

$$\kappa(x,y) = \frac{xy}{\mathbb{E}(w_1)}. \tag{12}$$

To see if this kernel is graphical, note that

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{n}\mathbb{E}(|E_n|) \\
&= \lim_{n\to\infty} \frac{1}{n}\mathbb{E}\left[\sum_{i<j}\left\{1 - \exp\left(-\frac{\kappa(x_i,x_j)}{n}\right)\right\}\right] \\
&= \lim_{n\to\infty} \frac{1}{n}\left[\sum_{i<j}\left\{1 - \exp\left(-\frac{w_i w_j}{s^{(n)}}\right)\right\}\right] \\
&\leq \lim_{n\to\infty} \frac{1}{n}\mathbb{E}\left[\sum_{i<j}\frac{w_i w_j}{s^{(n)}}\right] \\
&\leq \lim_{n\to\infty} \frac{1}{2}\mathbb{E}(s^{(n)}/n) = \mathbb{E}(w_1)/2 \\
&= \frac{1}{2}\iint_{\mathcal{S}^2} \kappa(x,y)d\mu(x)d\mu(y).
\end{aligned} \tag{13}$$

Hence, combined with Bollobás et al. (2007, Lemma 8.1), we get

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{E}(|E_n|) = \frac{1}{2}\iint_{\mathcal{S}^2} \kappa(x,y)d\mu(x)d\mu(y). \tag{14}$$

and the kernel κ is therefore graphical. The second part of Theorem 3.1 then follows from Bollobás et al. (2007, Proposition 8.9), and Theorem 3.2 follows from Theorem C.1 with

$$\lambda(x) = \int_{\mathcal{S}} \kappa(x,y)d\mu(y) = \int_0^\infty \frac{xy}{\mathbb{E}(w_1)} \mathrm{d}\mathcal{L}_{w_1}(y) = x. \tag{15}$$

E Proof of Theorem 4.1

Theorem 4.1 also follows by showing that the rank-c model fits into the general framework discussed in Section C. Define a vertex space $\mathcal{V} = (\mathcal{S}, \mu, \mathbf{x}_n)$ with

$$\mathcal{S} = (0,\infty)^{c+1}, \quad \mu = \mathcal{L}_{w_1}\mathcal{L}_{(v_{11},\dots,v_{1c})}, \tag{16}$$

where $\mathcal{L}_{w_1}$ and $\mathcal{L}_{(v_{11},\dots,v_{1c})}$ denote the laws of w_1 and $(v_{11},\dots,v_{1c})$, and

$$x_i = \left(w_i\sqrt{\frac{\mathbb{E}(w_1)}{s^{(n)}/n}}, v_{i1}\sqrt{\frac{\mathbb{E}(v_{11})}{r_1^{(n)}/n}}, \dots, v_{ic}\sqrt{\frac{\mathbb{E}(v_{1c})}{r_c^{(n)}/n}}\right). \tag{17}$$

Define a kernel on this space

$$\kappa(x,y) = \frac{x_{[1]}y_{[1]}}{\mathbb{E}(w_1)}\sum_{q=1}^{c}\frac{x_{[q+1]}y_{[q+1]}}{\mathbb{E}(v_{1q})}, \tag{18}$$

where $x_{[d]}$ denotes the dth component of x. To see if this kernel is graphical, note that

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{n}\mathbb{E}(|E_n|) \\
&= \lim_{n\to\infty} \mathbb{E}\Bigg[\frac{1}{n}\sum_{i<j}\Bigg\{1-\exp\Bigg(-\frac{w_iw_j}{s^{(n)}}\sum_{q=1}^{c}\frac{v_{iq}v_{jq}}{r_q^{(n)}/n}\Bigg)\Bigg\}\Bigg] \\
&\le \lim_{n\to\infty} \mathbb{E}\Bigg[\frac{1}{n}\sum_{i<j}\frac{w_iw_j}{s^{(n)}}\mathbb{E}\Bigg[\sum_{q=1}^{c}\frac{v_{iq}v_{jq}}{r_q^{(n)}/n}\Bigg]\Bigg]. \qquad (19)
\end{aligned}$$

Now note that

$$\begin{aligned}
\mathbb{E}\Bigg[\sum_{q=1}^{c}\frac{v_{iq}v_{jq}}{r_q^{(n)}/n}\Bigg] &= \frac{1}{n^2}\sum_{i',j'}\mathbb{E}\Bigg[\sum_{q=1}^{c}\frac{v_{i'q}v_{j'q}}{r_q^{(n)}/n}\Bigg] \\
&\le \mathbb{E}\Bigg[\sum_{q=1}^{c} r_q^{(n)}/n\Bigg] = \sum_{q=1}^{c}\mathbb{E}(v_{1q}). \qquad (20)
\end{aligned}$$

Plugging this into the above equation yields

$$\begin{aligned}
&\lim_{n\to\infty}\frac{1}{n}\mathbb{E}(|E_n|) \\
&\le \lim_{n\to\infty}\mathbb{E}\Bigg[\frac{1}{n}\sum_{i<j}\frac{w_iw_j}{s^{(n)}}\sum_{q=1}^{c}\mathbb{E}(v_{1q})\Bigg] \\
&\le \lim_{n\to\infty}\frac{1}{2}\mathbb{E}(s^{(n)}/n)\sum_{q=1}^{c}\mathbb{E}(v_{1q}) \\
&= \frac{1}{2}\mathbb{E}(w_1)\sum_{q=1}^{c}\mathbb{E}(v_{1q}) \\
&= \frac{1}{2}\iint \kappa(x,y)d\mu(x)d\mu(y). \qquad (21)
\end{aligned}$$

Hence, by Bollobás et al. (2007, Lemma 8.1), we get

$$\lim_{n\to\infty}\frac{1}{n}\mathbb{E}(|E_n|) = \frac{1}{2}\iint_{\mathcal{S}^2}\kappa(x,y)d\mu(x)d\mu(y). \qquad (22)$$

The second part of Theorem 3.1 then follows from Bollobás et al. (2007, Proposition 8.9), and Theorem 3.2 follows from Theorem C.1 with

$$\begin{aligned}
\lambda(x) &= \int_{\mathcal{S}}\kappa(x,y)d\mu(y) \\
&= x_{[1]}\sum_{q=1}^{c}x_{[q+1]}. \qquad (23)
\end{aligned}$$

F Details on posterior inferences

F.1 Posterior inference for the rank-1 model

The posterior ineference for rank-1 model is summarized in three steps.

1. Sample w via HMC (we use the transformation $w = e^{\hat{w}}$ and update $\hat{w}$).
2. Sample m from truncated Poisson distribution,

$$p(m_{ij}|w) = \frac{(\frac{w_iw_j}{s^{(n)}})^{m_{ij}}\exp(-\frac{w_iw_j}{s^{(n)}})\mathbb{1}_{\{m_{ij}>0\}}}{m_{ij}!(1-e^{-\frac{w_iw_j}{s^{(n)}}})}. \qquad (24)$$

3. Sample hyperparameters for $p(w)$ via Metropolis-Hastings.

We used the step size $\epsilon = 10^{-2}$ and the number of leapfrog steps $L = 20$ for all experiments.

Sampling hyperparameters for IG. In IG we have two hyperparameters $\alpha > 1$ and $\beta > 0$. We place a log-normal prior on $\alpha - 1$ and β.

$$\alpha = 1 + e^{\hat{\alpha}},\ \hat{\alpha}\sim\mathcal{N}(0,1) \qquad (25)$$

$$\beta = e^{\hat{\beta}},\ \hat{\beta}\sim\mathcal{N}(0,1). \qquad (26)$$

Then we updated $\hat{\alpha}$ and $\hat{\beta}$ via Metropolis-Hastings with proposal distribution $\mathcal{N}(\hat{\alpha}, 0.05)$ and $\mathcal{N}(\hat{\beta}, 0.05)$. We found that the initialization of $\hat{\alpha}$ and $\hat{\beta}$ was important to capture degree distributions. We initialized $\hat{\alpha}\sim\mathcal{N}(0, 0.1)$ and using the asymptotic relation

$$\frac{|E_n|}{n} \xrightarrow{p} \frac{\mathbb{E}(w_1)}{2} = \frac{\beta}{2(\alpha-1)}, \qquad (27)$$

set

$$\hat{\beta} = \log\frac{2(\alpha-1)|E_n|}{n}. \qquad (28)$$

Sampling hyperparameters for GIG. We have three parameters $\nu < 0$, $a > 0$, and $b > 0$ (we restricted $\nu < 0$ to get the positive power-law exponent). We placed log-normal priors on $-\nu, a$ and b.

$$\nu = -e^{\hat{\nu}},\ \hat{\nu}\sim\mathcal{N}(0,1) \qquad (29)$$

$$a = e^{\hat{a}},\ \hat{a}\sim\mathcal{N}(0,1) \qquad (30)$$

$$b = e^{\hat{b}},\ \hat{b}\sim\mathcal{N}(0,1). \qquad (31)$$

We updated $\hat{\nu}, \hat{a}$ and $\hat{b}$ via Metropolis-Hastings with proposal distributions $\mathcal{N}(\hat{\nu}, 0.05), \mathcal{N}(\hat{a}, 0.1)$ and $\mathcal{N}(\hat{b}, 0.05)$. We initialized $\nu\sim\mathrm{Unif}(-1,0)$ and $a\sim\mathrm{Unif}(0,10^{-3})$. b was initialized by solving

$$\frac{|E_n|}{n} = \frac{\mathbb{E}(w_1)}{2} = \frac{\sqrt{b}K_{\nu+1}(\sqrt{ab})}{\sqrt{a}K_\nu(\sqrt{ab})}, \qquad (32)$$

using a numerical root-finding algorithm.

F.2 Posterior inference for the rank−c model

The posterior inference for the rank−c model is quite similar to that of the rank-1 model.

1. Sample w via HMC (we use the transformation $w = e^{\hat{w}}$ and update $\hat{w}$).
2. Sample V via HMC (see below).
3. Sample M from multivariate truncated Poisson distribution,
$$p(M_{ij}|w, V) = \prod_{q=1}^{c} \frac{\lambda_{ijq}^{m_{ijq}} e^{-\lambda_{ijq}} \mathbb{1}_{\{\sum_{ijq'} m_{ijq'} > 0\}}}{1 - \exp(-\sum_{q'=1}^{c} \lambda_{ijq'})}, \tag{33}$$
where $\lambda_{ijq} = \frac{w_i w_j}{s^{(n)}} \frac{v_{iq} v_{jq}}{r_q^{(n)}/n}$.
4. Sample hyperparameters of $p(w)$ and $p(V)$.

Details on HMC for V. Each vector $(v_{i1}, \ldots, v_{ic})$ is a Dirichlet random variable such that $\sum_{q=1}^{c} v_{iq} = 1$, so transforming it to an unconstrained vector is quite tricky. We adapt the trick presented in Betancourt (2010). Let $v = (v_1, \ldots, v_c) \sim \text{Dir}(\gamma_1, \ldots, \gamma_c)$. Define i.i.d. beta random variables,

$$z_1 \sim \text{beta}(\tilde{\gamma}_1, \gamma_1) \tag{34}$$
$$z_2 \sim \text{beta}(\tilde{\gamma}_2, \gamma_2) \tag{35}$$
$$\vdots \tag{36}$$
$$z_{q-1} \sim \text{beta}(\tilde{\gamma}_{q-1}, \gamma_{q-1}), \tag{37}$$

where

$$\tilde{\gamma}_q = \sum_{t=q+1}^{c} \gamma_t. \tag{38}$$

Then, if we take a transform

$$v_q = (1 - z_q) \prod_{t=1}^{q-1} z_t \ \ (t < c), \tag{39}$$
$$v_c = \prod_{t=1}^{c-1} z_t, \tag{40}$$

we have $v \sim \text{Dir}(\gamma_1, \ldots, \gamma_c)$. The advantage of this transformation is that the Jacobian can be computed efficiently. By the chain rule, we have

$$\frac{\partial f(v)}{\partial z_q} = \frac{v_q}{z_q - 1} \frac{\partial f(v)}{\partial v_q} + \sum_{t=q+1}^{c} \frac{x_t}{z_q} \frac{\partial f(v)}{\partial v_t}. \tag{41}$$

We take another logistic transform on z_q to make it completely unconstrained.

$$z_q = \frac{1}{1 + e^{-\hat{z}_q}}, \tag{42}$$

Hence, the gradient for the unconstrained variable $\hat{z}_q$ is computed as

$$\frac{\partial f(v)}{\partial \hat{z}_q} = \frac{\partial f(v)}{\partial z_q} (z_q - z_q^2). \tag{43}$$

In our algorithm, HMC for V is done on the unconstrained variables $(\hat{z}_{i1}, \ldots, \hat{z}_{i,q-1})_{i=1}^{n}$.

Initialization and step sizes. Unlike Todeschini et al. (2016) where the model is initialized by running MCMC for the simplified model without communities to initialize w, we initialize the chain by running MCMC only for V while holding w fixed as $[1, \ldots, 1]^\top$. We found this helpful for the algorithm to discover better community structures. For this initialization, we ran HMC for V with $\epsilon = 10^{-1}$ and $L = 20$. After initialization, we ran HMC for w with $\epsilon = 5 \cdot 10^{-3}$ and $L = 20$, and ran HMC for V with $\epsilon = 2.5 \cdot 10^{-2}$ and $L = 20$. We decayed ϵ for V to $5 \cdot 10^{-3}$ after burn-in.

Sampling hyperparameters for $p(V)$. We assume log-normal prior distributions on the hyperparameters $\gamma_1, \ldots, \gamma_c$.

$$\gamma_q = e^{\hat{\gamma}_c}, \ \ \hat{\gamma}_q \sim \mathcal{N}(0, 1). \tag{44}$$

Then we updated $\hat{\gamma}_q$ via Metropolis-Hastings with proposal distribution $\mathcal{N}(\hat{\gamma}_q, 0.01)$. We initialized $\gamma_1 = \cdots = \gamma_c = 0.1$.

G Additional Figures

G.1 Empirical degree distributions and number of edges for rank$-c$ model

We first demonstrate the empirical degree distributions and number of edges of graphs generated from rank$-c$ model with $c = 5$. The results are presented in Fig. 2. As predicted from Theorem 4.1, the degree distribution and sparsity are not affected by the introduction of the community affiliation factors V.

G.2 Discovered community structures

We present the community structures discoverd by IG, GIG, CGGP and MMSB in Fig. 3. IG, GIG and CGGP discovered reasonable communities where the edge densities within communities are much higher than the edge densities between communities. However, MMSB completely failed to discover the communities. The fact that the models without degree heterogeneity fail to capture community structures has been reported in various works (Karrer and Newman, 2011; Gopalan et al., 2013; Todeschini et al., 2016), and our results confirm it.

References

Barabasi, L. (2018). All you need is love. Blog post.

Betancourt, M. (2010). Crusing the simplex: Hamiltonian Monte Carlo and the Dirichlet distribution. *arXiv:1010.3436*.

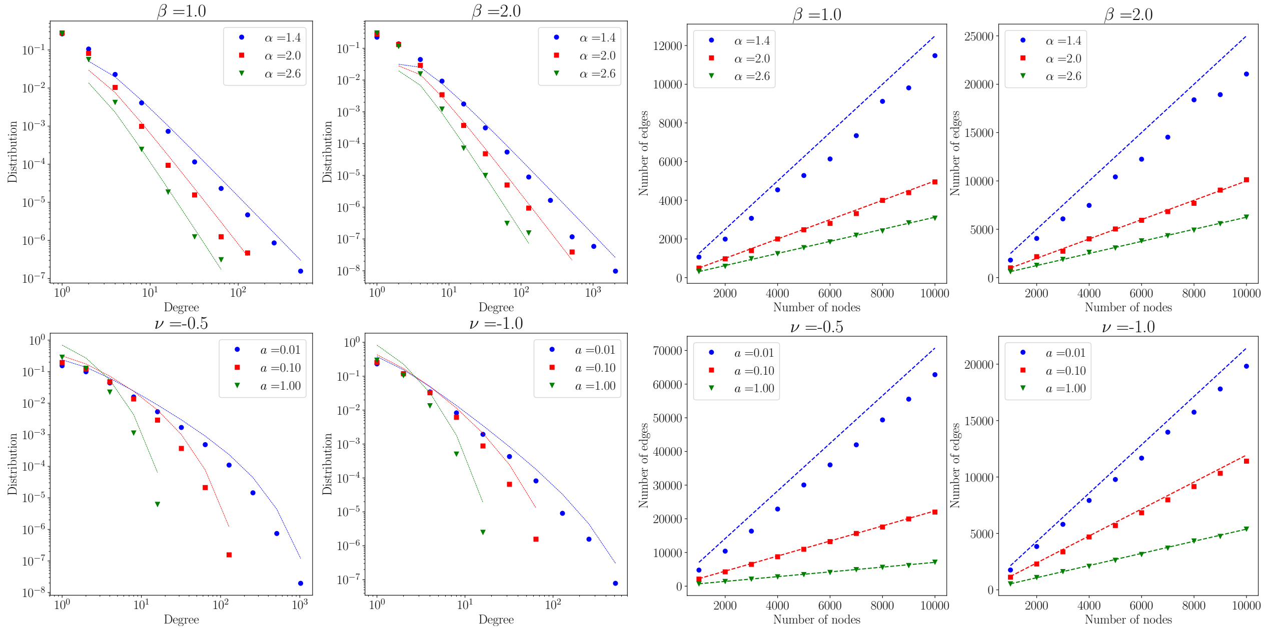
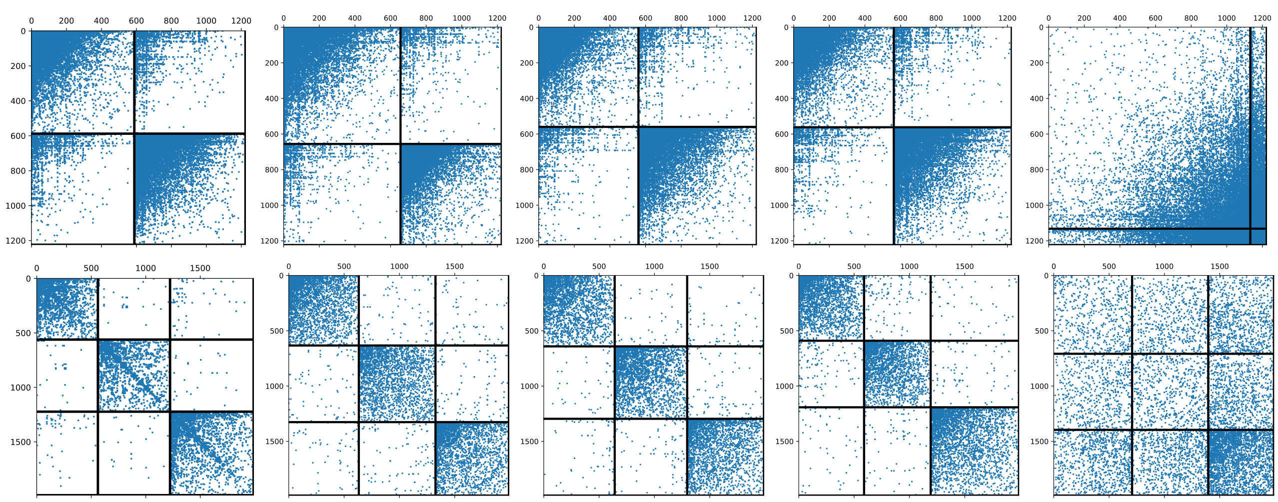


Figure 2: First row, first and second boxes: empirical degree distributions (dashed lines) of graphs with 10,000 nodes sampled from rank-c inverse gamma NR model compared to the theoretically expected asymptotic degree distribution (dotted lines), with various values of α and β. First row, third and fourth boxes: empirical number of edges (dashed lines) of graphs sampled from rank$-c$ inverse gamma NR model versus the number of nodes compared to the theoretically expected value of number of edges (dotted lines), with various values of α and β. Second row: the same figures for rank$-c$ GIG NR model with various values of ν and a with fixed $b = 2.0$. Best viewed magnified in color.



(a) Ground truth (b) IG-NR (c) GIG-NR (d) CGGP (e) MMSB

Figure 3: Group truth and latent communities discovered by the random graph models for (top row) `polblogs` and (bottom row) `DBLP`. The nodes within communities are sorted according to their degree for the ground truth, and according to their estimated node popularity parameter for the different models.

Bingham, N. H., Goldie, C. M., and Teugels, J. L. (1989). *Regular variation*, volume 27. Cambridge university press.

Bollobás, B., Janson, S., and Riordan, O. (2007). The phase transition in inhomogeneous random graphs. *Random Structures & Algorithms*, 31:3–122.

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

Broido, A. D. and Clauset, A. (2018). Scale-free networks are rare. *arXiv preprint arXiv:1801.03400*.

Caron, F. and Fox, E. B. (2017). Sparse graphs using exchangeable random measures. *Journal of the Royal Statistical Society B (discussion paper)*, 79:1295–1366.

Chung, F. and Lu, L. (2002). Connected components in random graphs with given expected degree sequences. *Annals of Combinatorics*, 6:125–145.

Gopalan, P., Wang, C., and Blei, D. M. (2013). Modeling overlapping communities with node popularities. In *Advances in Neural Information Processing Systems 26*.

Karrer, B. and Newman, M. E. J. (2011). Stochastic blockmodels and community structure in networks. *Physical Review E*, 83(1).

Mikosch, T. (1999). *Regular variation, subexponentiality and their applications in probability theory*. Eindhoven University of Technology.

Norros, I. and Reittu, H. (2006). On a conditionally Poissonian graph process. *Advances in Applied Probability*, 38(1):59–75.

Resnick, S. I. (2007). *Heavy-tail phenomena: probabilistic and statistical modeling*. Springer Science & Business Media.

Todeschini, A., Miscouridou, X., and Caron, F. (2016). Exchangeable random measures for sparse and modular graphs with overlapping communities. *arXiv:1602.0211*.