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

Abstract

We consider a non-projective class of inhomogeneous random graph models with interpretable parameters and a number of interesting asymptotic properties. Using the results of Bollobás et al. (2007), we show that i) the class of models is sparse and ii) depending on the choice of the parameters, the model is either scale-free, with power-law exponent greater than 2, or with an asymptotic degree distribution which is power-law with exponential cut-off. We propose an extension of the model that can accommodate an overlapping community structure. Scalable posterior inference can be performed due to the specific choice of the link probability. We present experiments on five different real-world networks with up to 100,000 nodes and edges, showing that the model can provide a good fit to the degree distribution and recovers well the latent community structure.

1 Introduction

Simple graphs are composed of a set of vertices with undirected connections between them. The graph may represent a set of friendship relationships between individuals, a physical infrastructure network, or a protein-protein interaction network. Defining flexible and realistic statistical graph models is of great importance in order to perform link prediction or for uncovering interpretable latent structure, and has been the subject of a large body of work in recent years, see e.g. (Newman, 2009; Kolaczyk, 2009; Goldenberg et al., 2010).

Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS) 2019, Naha, Okinawa, Japan. PMLR: Volume 89.

Our objective is to develop a class of models with interpretable parameters and realistic asymptotic properties. Of particular interest for this paper are the notions of sparsity and scale-freeness. A sequence of graphs is said to be sparse if the number of edges scales subquadratically with the number of nodes. The degree of a node is the number of connections of that node. The sequence of graphs is said to be scale-free if the proportion of nodes of degree k is approximately $k^{-\eta}$ when the number n of nodes is large, where the exponent η is greater than 1. That is, for large n, the degree distribution behaves like a power-law. These notions of sparsity and scale-freeness have received a lot of attention in the network literature in the past years (Barabási and Albert, 1999; Newman, 2009; Orbanz and Roy, 2015; Barabási, 2016; Caron and Fox, 2017); some authors argued that they are desirable properties of random graph models, and that many networks exhibit this scale-free behavior, usually with an exponent $\eta > 2$. Other authors have recently challenged the scale-free assumption, showing that a power-law distribution with exponential cut-off provides a good fit to many real-world networks (Newman, 2009; Broido and Clauset, 2018), see the supplementary material for more discussion about testing for network scale-freeness. Besides these global asymptotic properties, we are also interested in capturing some latent structure in graphs. Individuals may belong to some latent communities, and their level of affiliation to the community defines the probability that two nodes connect.

We propose a class of sparse graph models with overlapping community structure and well-specified asymptotic degree distributions. The graph can either be scale-free with exponent $\eta > 2$, or non-scale-free, with asymptotic degree distribution being a power-law distribution with exponential cut-off. The construction builds on inhomogeneous random graphs, a class of models exhibiting degree heterogeneity. This class of models has been studied extensively in the applied probability literature (Aldous, 1997; Chung and Lu, 2002b; Bollobás et al., 2007; van der Hofstad, 2016), but has been left unexplored for the statistical analysis of real-world networks. In Section 2 we provide a formal description of sparsity and scale-freeness for sequences of graphs. In Section 3 we describe the rank-1 inhomogeneous random graphs, and present their sparsity property and asymptotic degree distribution. The model is then extended in Section 4 in order to accommodate a latent community structure. Posterior inference is discussed in Section 5. In Section 6 we discuss the relative merits and drawbacks of our approach compared to other random graph models. Section 7 provides an illustration of the approach on several real-world networks, showing that the model can provide a good fit to the empirical degree distribution and recover the latent community structure.

Notations. Throughout the article, $X_n \overset{p}{\to} X$ denotes convergence in probability, and $a_n \sim b_n$ indicates $\lim_{n\to\infty} a_n/b_n \to 1$.

2 Sparse and scale-free networks

We first provide a formal definition of sparsity and scale-freeness, as there is no general agreement on the definition of a scale-free network and these notions are core to the results of this paper.

Let $(G_n)_{n\geq 1}$ be a sequence of simple random graphs of size n where $G_n = (V_n, E_n)$, $V_n = \{1, \ldots, n\}$ is the set of vertices and E_n the set of edges. Denote $|E_n|$ the number of edges. The graph is said to be sparse if $\mathbb{E}(|E_n|)/n^2 \to 0$. Let $N_k^{(n)}$ be the number of nodes of degree k in G_n. We now formally give the definition of a scale-free network informally introduced in Section 1.

Definition 2.1. *A random graph sequence $(G_n)_{n\geq 1}$ is said to be scale-free with exponent η iff there exists a slowly varying function ℓ and $\eta > 1$ such that, for each $k = 1, 2, \ldots$*

$$\frac{N_k^{(n)}}{n} \overset{p}{\to} \pi_k \tag{1}$$

as n tends to infinity, where

$$\pi_k \sim \ell(k) k^{-\eta} \text{ as } k \to \infty. \tag{2}$$

Background definitions and properties of slowly and regularly varying functions are given in the supplementary material. Intuitively, slowly varying functions are functions that vary more slowly than any power of x. The term scale-free comes from the fact that the asymptotic degree distribution satisfies some (asymptotic) scale-invariance. For any integer $m \geq 1$,

$$\lim_{k\to\infty} \frac{\pi_{mk}}{\pi_k} = m^{-\eta}. \tag{3}$$

The most classical case is when $\ell(k) = C$ is constant. In this case, the asymptotic degree distribution behaves as a pure power-law for k large. More generally, the scale-invariance property defined above will be satisfied for any slowly varying function ℓ, which can be e.g. logarithm, or iterated logarithm. Definition 2.1 is slightly more restrictive than the definition of a scale-free graph sequence in (van der Hofstad, 2016, Definition 1.4), which is implied from Definition 2.1 by properties of regularly varying functions (see supplementary material).

3 Rank-1 inhomogeneous random graphs

3.1 Definition

Let $(G_n)_{n\geq 1}$ be a sequence of simple random graphs of size n defined as follows. The probability that two nodes i and j are connected in the graph G_n is given by

$$p_{ij}^{(n)} = 1 - \exp\left(-\frac{w_i w_j}{s^{(n)}}\right) \tag{4}$$

where $s^{(n)} = \sum_{i=1}^n w_i$ and the positive weights $(w_1, w_2, \ldots)$ are independently and identically distributed (iid) from some distribution F with $\mathbb{E}(w_1) < \infty$. The model (4) is known as the Norros-Reittu (NR) inhomogeneous random graph model (Norros and Reittu, 2006). This model has been the subject of a lot of interest in the applied probability and graph theory literature (Bollobás et al., 2007; Bhamidi et al., 2012; van der Hofstad, 2013, 2016; Broutin et al., 2018). The parameter $w_i > 0$ accounts for degree heterogeneity in the graph and can be interpreted as a sociability parameter of node i. The larger this parameter, the more likely node i is to connect to other nodes.

3.2 Sparsity and scale-free properties

The random graph sequence defined by Equation (4) satisfies a number of remarkable asymptotic properties. The first result, which follows from Bollobás et al. (2007) (see details in the supplementary material), shows that the resulting graphs are sparse.

Theorem 3.1 (Bollobás et al. (2007))**.** *Let $|E_n|$ denote the number of edges in the graph G_n. Then*

$$\frac{\mathbb{E}(|E_n|)}{n} \to \frac{\mathbb{E}(w_1)}{2} \quad \text{and} \quad \frac{|E_n|}{n} \overset{p}{\to} \frac{\mathbb{E}(w_1)}{2}. \tag{5}$$

The following result is a corollary of Theorem 3.13, remark 2.4 and the discussion in Section 16.4 in (Bollobás et al., 2007). It states that the asymptotic degree distribution is a mixture of Poisson distributions, with mixing distribution F.

Theorem 3.2 (Bollobás et al. (2007)). *Let $N_k^{(n)}$ be the number of vertices of degree k in the graph G_n of size n and link probability $p_{ij}^{(n)}$ given by Equation (4). Then, for each $k = 1, 2, \ldots$, $N_k^{(n)}/n \xrightarrow{p} \pi_k$ as n tends to infinity, where*

$$\pi_k := \int_0^\infty \frac{x^k}{k!} e^{-x} dF(x). \tag{6}$$

Our analysis on the asymptotic degree distribution is based on the following theorem for the asymptotic behavior of mixed Poisson distributions.

Theorem 3.3. *(Willmot, 1990) Suppose that*

$$f(x) \sim \ell(x) x^\eta e^{-\zeta x}, \quad x \to \infty, \tag{7}$$

where $\ell(x)$ is a locally bounded function on $(0, \infty)$ which varies slowly at infinity, $\zeta \geq 0$, and $-\infty < \eta < \infty$ (with $\eta < -1$ when $\zeta = 0$). For $\lambda > 0$, define the probabilities of the mixed Poisson distribution as

$$\pi_k = \int_0^\infty \frac{(\lambda x)^k e^{-\lambda x}}{k!} f(x) dx; \quad k = 0, 1, 2, \ldots. \tag{8}$$

Then,

$$\pi_k \sim \frac{\ell(k)}{(\lambda + \zeta)^{\eta+1}} \left(\frac{\lambda}{\lambda + \zeta} \right)^k k^\eta, \quad k \to \infty. \tag{9}$$

The following result is a corollary of Theorem 3.2 and Theorem 3.3. It states that if the random variables w_i are regularly varying (see definition in the supplementary material), then the sequence of random graphs is scale-free.

Corollary 3.1. *Let $N_k^{(n)}$ be the number of vertices of degree k in the graph G_n of size n and link probability $p_{ij}^{(n)}$ given by Equation (4). Assume that the distribution F is absolutely continuous with pdf f verifying $f(w) \sim \ell(w) w^{-\eta}$ as w tends to infinity, for some locally bounded slowly varying function ℓ and $\eta > 1$. Then, for each $k = 1, 2, \ldots$, $N_k^{(n)}/n \xrightarrow{p} \pi_k$ as n tends to infinity, where*

$$\pi_k \sim \ell(k) k^{-\eta}, \quad k \to \infty. \tag{10}$$

3.3 Particular examples

We now consider two special cases. The first case yields scale-free graphs with asymptotic power-law degree distributions with exponent $\eta > 2$. The second yields non-scale-free graphs, where the asymptotic degree distribution is power-law with exponential cut-off.

3.3.1 Scale-free graph with power-law degree distribution

For $i = 1, 2, \ldots$, let $w_i \overset{\text{i.i.d.}}{\sim} \text{invgamma}(\alpha, \beta)$ where $\text{invgamma}(\alpha, \beta)$ denotes the inverse gamma distribution with parameters $\alpha > 1$ and $\beta > 0$, whose probability density function (pdf) is given by

$$f(w) = \frac{\beta^\alpha}{\Gamma(\alpha)} w^{-\alpha-1} e^{-\beta/w}.$$

Here, the constraint $\alpha > 1$ is required for the condition $\mathbb{E}[w_1] < \infty$. By Theorem 3.2, the asymptotic degree distribution is a mixed Poisson-inverse-gamma distribution with probability mass function

$$\pi_k = \frac{2\beta^{\frac{k+\alpha}{2}}}{k!\Gamma(\alpha)} K_{k-\alpha}(2\sqrt{\beta}), \tag{11}$$

where K is the modified Bessel function of the second kind. Using Corollary 3.1, we obtain

$$\pi_k \sim \frac{\beta^\alpha}{\Gamma(\alpha)} k^{-\alpha-1} \quad \text{as } k \to \infty. \tag{12}$$

The resulting asymptotic degree distribution is a power-law and the graph is scale-free with arbitrary index $\alpha + 1 > 2$. The two hyperparameters of the inverse gamma prior play an important role to decide the asymptotic properties of graphs. The shape parameter α tunes the index of power-law, and is also related to the sparsity of graphs. The scale parameter β is also related to the sparsity of graphs. Fig. 1 shows the empirical degree distributions and number of edges of graphs generated from inverse gamma NR model.

3.3.2 Non scale-free graph with power-law degree distribution with exponential cut-off

Now we consider another model with generalized inverse Gaussian (GIG) prior. Let $w_i \overset{\text{i.i.d.}}{\sim} \text{GIG}(\nu, a, b)$ where the density of the GIG distribution with parameter ν, $a > 0$ and $b > 0$ is given by

$$f(w) = \frac{(a/b)^{\nu/2}}{2K_\nu(\sqrt{ab})} w^{\nu-1} \exp\left\{ -\frac{1}{2}\left(aw + \frac{b}{w} \right) \right\}. \tag{13}$$

Note that by taking $a \to 0$, one obtains the pdf of an inverse gamma distribution as a limiting case. By Theorem 3.2, the asymptotic degree distribution is

$$\pi_k = \frac{(a/b)^{\nu/2}}{k!\{(a+2)/b\}^{(k+\nu)/2}} \frac{K_{k+\nu}(\sqrt{(a+2)b})}{K_\nu(\sqrt{ab})}. \tag{14}$$

This distribution is sometimes called the Sichel distribution, after Herbert Sichel (Sichel, 1974). Note that $f(w) \sim (a/b)^{\nu/2}/2/K_\nu(\sqrt{ab}) w^{\nu-1} \exp(-aw/2)$ as $w \to \infty$ hence, by Theorem 3.3,

$$\pi_k \sim \frac{(a/b)^{\nu/2} k^{\nu-1} e^{-\log(1+a/2)k}}{2(1+a/2)^\nu K_\nu(\sqrt{ab})} \quad \text{as } k \to \infty. \tag{15}$$

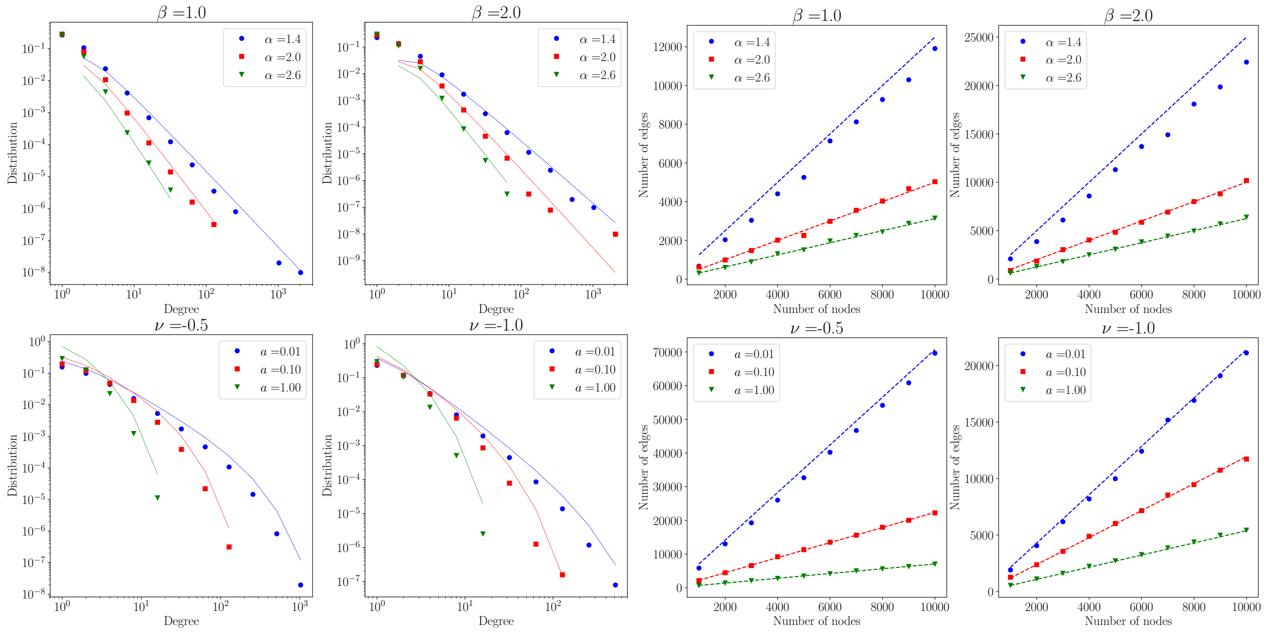


Figure 1: First row, first and second boxes: empirical degree distributions (dashed lines) of graphs with 10,000 nodes sampled from IG-NR compared to the theoretically expected asymptotic degree distribution (dotted lines), with various values of α and β. First row, third and fourth boxes: empirical number of edges (dashed lines) of graphs sampled from IG-NR 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 GIG-NR with various values of ν and a with fixed $b = 2.0$. Best viewed magnified in color.

In this case, the asymptotic degree distribution is not of the form of Equation (2), and the graph sequence is therefore not scale-free. However, the asymptotic degree distribution has the form $k^{\nu-1}e^{-\tau k}$ of a power-law distribution with exponential cut-off. This class of probability distributions has been shown to provide a good fit to the degree distributions of a wide range of real-world networks (Clauset et al., 2009). As for the inverse gamma NR model, the hyperparameters (ν, a, b) tunes the asymptotic properties. ν determines the power-law index of degree distribution, a is related to the exponential cutoff and sparsity, and b is related to the sparsity. Fig. 1 shows the empirical degree distributions and the number of edges of graphs generated from GIG NR model.

4 Extension to Latent Overlapping Communities

4.1 Definition

The inhomogeneous random graphs considered so far only account for degree heterogeneity. However, the connections in real-world networks are often due to some latent interactions between the vertices. Recently, several models that combine a degree correction together with a latent structure to define edge probabilities were proposed (Zhou, 2015; Todeschini et al., 2016; Herlau et al., 2016; Lee et al., 2017). In this section, we propose an extension of the NR model that includes some latent overlapping structure, and study the sparsity, scale-freeness properties and asymptotic degree distribution of this model. Let the edge probability between the vertex i and j be given by

$$p_{ij}^{(n)} = 1 - \exp\left(- \frac{w_i w_j}{s^{(n)}} \sum_{q=1}^{c} \frac{v_{iq} v_{jq}}{r_q^{(n)}/n} \right). \tag{16}$$

where $(w_i)_{i=1,2,\dots}$ are iid random variables with distribution F with $\mathbb{E}(w_1) < \infty$ and $(v_{i1}, \dots, v_{ic})_{i=1,2,\dots}$ are i.i.d. with $\mathbb{E}(v_{1q}) < \infty$ for all q and $r_q^{(n)} = \sum_{i=1}^{n} v_{iq}$. We call this model with c communities the *rank-c model.* As in the rank-1 model, the parameter w_i can be interpreted as an overall sociability parameter of node i, or degree-correction. The parameter v_{iq} can be interpreted as the level of affiliation of individual of i to community q. Similar models, in a different asymptotic framework have been used in (Yang and Leskovec, 2013; Zhou, 2015; Todeschini et al., 2016).

Theorem 4.1. *Let $|E_n|$ denote the number of edges in* *the graph G_n defined with link probability* (16). *Then,*

$$\frac{\mathbb{E}(|E_n|)}{n} \to \frac{\mathbb{E}(w_1)\sum_{q=1}^c \mathbb{E}(v_{1q})}{2} \tag{17}$$

$$\frac{|E_n|}{n} \overset{p}{\to} \frac{\mathbb{E}(w_1)\sum_{q=1}^c \mathbb{E}(v_{1q})}{2}. \tag{18}$$

Recall that $N_k^{(n)}$ is the number of vertices of degree k in the graph G_n of size n. Then, for each $k = 1, 2, \ldots$, $N_k^{(n)}/n \overset{p}{\to} \pi_k$ as n tends to infinity, where

$$\pi_k = \int_0^\infty \int_0^\infty \frac{(uw)^k}{k!} e^{-uw} dF(w) dH(u) \tag{19}$$

where H is the distribution of the random variable $U = \sum_{q=1}^c v_{1q}$. If additionally F is absolutely continuous with pdf f verifying $f(w) \sim \ell(w) w^{-\eta}$ as $w \to \infty$ for some locally bounded slowly varying function ℓ and $\eta > 1$ and $\mathbb{E}(U^{\eta-1+\epsilon}) < \infty$ for some $\epsilon > 0$, then

$$\pi_k \sim \mathbb{E}(U^\eta)\ell(k)k^{-\eta} \quad \text{as } k \to \infty.$$

The proof of Theorem 4.1 is given in the supplementary material. In this paper, we consider in particular

$$(v_{i1}, \ldots, v_{iq}) \sim \mathrm{Dir}(\gamma), \tag{20}$$

where $\mathrm{Dir}(\gamma)$ denotes the standard Dirichlet distribution with parameter $\gamma = (\gamma_1, \ldots, \gamma_c)$, where $\gamma_q > 0$ for $q = 1, \ldots, c$.

5 Posterior inference

5.1 Posterior inference for the rank-1 NR

Let $Y = \{y_{ij}\}_{1 \le i < j \le n}$ be an (upper triangular part of) adjacency matrix of a graph G_n and $w = (w_1, \ldots, w_n)$. The joint density is written as

$$p(Y, w) = \prod_{i=1}^n f(w_i) \prod_{i<j} \left(1 - e^{-\frac{w_i w_j}{s^{(n)}}}\right)^{y_{ij}} e^{(y_{ij}-1)\frac{w_i w_j}{s^{(n)}}} \tag{21}$$

Following Caron and Fox (2017) and Zhou (2015), we introduce a set of auxiliary truncated Poisson random variables m_{ij} for the pairs with $y_{ij} = 1$.

$$p(m_{ij}|w) = \frac{(\frac{w_i w_j}{s^{(n)}})^{m_{ij}} \exp(-\frac{w_i w_j}{s^{(n)}}) \mathbb{1}_{\{m_{ij}>0\}}}{m_{ij}!(1 - \exp(-\frac{w_i w_j}{s^{(n)}}))}. \tag{22}$$

The log joint density is then given as

$$\log p(Y, M, w) = \sum_{(i,j)\in E_n} \left(m_{ij} \log \frac{w_i w_j}{s^{(n)}} - \log m_{ij}!\right) + \frac{1}{2}\left(\sum_{i=1}^n \frac{w_i^2}{s^{(n)}} - s^{(n)}\right) + \sum_{i=1}^n \log f(w_i). \tag{23}$$

Note that the terms for the pairs without edges ($y_{ij} = 0$) are collapsed into a single summation, and hence the overall computations of the log joint density and its gradient take $O(n + |E_n|)$ time. This is a huge advantage of the link function of NR model, while other link functions for rank-1 inhomogeneous random graphs (Britton et al., 2006; Chung and Lu, 2002b,a, 2003) suffer from $O(n^2)$ computing times.

For the posterior inference, we use a Markov chain Monte Carlo (MCMC) algorithm. At each step, given the gradient of the log joint density, we update w via Hamiltonian Monte Carlo (HMC, (Duane et al., 1987; Neal, 2011)). Then we resample the auxiliary variables m from truncated Poisson, and update hyperparameters for $f(w)$ using a Metropolis-Hastings step. Details can be found in the supplementary material.

5.2 Posterior inference for the rank-c NR

The posterior inference for the rank-c model is similar to that of the rank-1 model. Following Todeschini et al. (2016), for tractable inference, we introduce a set of multivariate truncated Poisson random variables $M = ((m_{ijq})_{q=1}^c)_{(i,j)\in E_n}$,

$$p(M|w, V) = \prod_{(i,j)\in E_n} \prod_{q=1}^c \frac{\lambda_{ijq}^{m_{ijq}} e^{-\lambda_{ijq}} \mathbb{1}_{\{\sum_{q'=1}^c m_{ijq'}>0\}}}{1 - \exp(-\sum_{q'=1}^c \lambda_{ijq'})}. \tag{24}$$

where $\lambda_{ijq} = \frac{w_i w_j}{s^{(n)}} \frac{v_{iq} v_{jq}}{r_q^{(n)}/n}$ and $V = (v_{iq})_{i=1,\ldots,n, q=1,\ldots,c}$. The log joint density is

$$\log p(Y, M, w, V) = \prod_{(i,j)\in E_n} \sum_{q=1}^c (m_{ijq} \log \lambda_{ijq} - \log m_{ijq}!) - \sum_{i<j}\sum_{q=1}^c \lambda_{ijq} + \sum_{i=1}^n \log f(w_i) + \sum_{i=1}^n \log g(v_{i1}, \ldots, v_{ic}; \gamma), \tag{25}$$

where $g(\cdot; \gamma)$ is the density for Dirichlet distribution with parameters γ. As for the rank-1 model, we can efficiently compute this log joint density and its gradient w.r.t. w and V with $O(cn + c|E_n|)$ time. At each step of MCMC, we first sample w and V via HMC, resample M from multivariate truncated Poisson, and update hyperparameters via Metropolis-Hastings. The detailed procedure can be found in the supplementary material.

6 Discussion

The models described in this paper can capture sparsity, scale-freeness with exponent $\eta > 2$ and latent

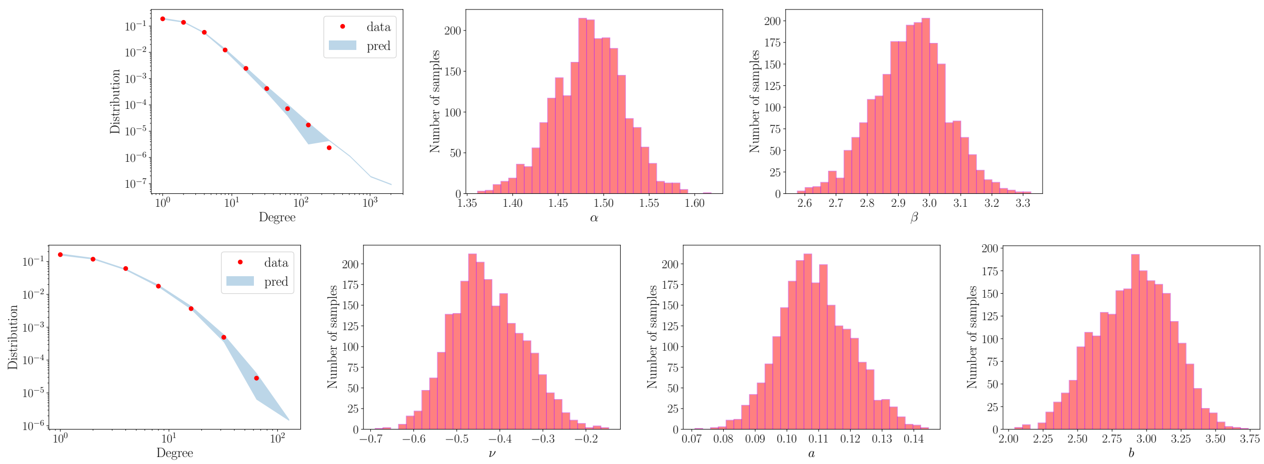


Figure 2: (Top row): IG-NR. 95% credible intervals of the predictive degree distribution, posterior samples of the hyperparameters α and β. The true values are $\alpha = 1.5$ and $\beta = 3.0$. (Bottom row): GIG-NR. 95% credible intervals of the predictive degree distribution and posterior samples of the hyperparameters ν, a and b. The true values are $\nu = -0.5, a = 0.1$ and $b = 3.0$.

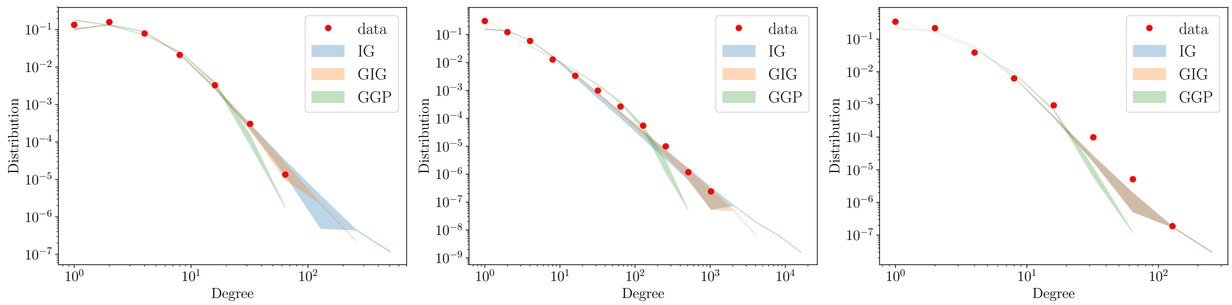
community structure. One drawback of the construction is that the model lacks projectivity, due to normalisation by s_n in the link probability (4). While this is an undesirable feature of the approach, we stress that there does not exist any projective class of random graphs that can capture all those properties, as we explain below. A popular class of models is the graphon-based or vertex-exchangeable graphs, which include as special cases stochastic blockmodels, latent factor models and their extensions, see (Orbanz and Roy, 2015) for a review. While these models have been successfully applied in a wide range of application, they produce dense graphs with probability one, as stressed by Orbanz and Roy (2015). Alternative models have been proposed, either based on exchangeable point processes (Caron and Fox, 2017; Veitch and Roy, 2015; Borgs et al., 2016), or on the notion of edge-exchangeability (Crane and Dempsey, 2015, 2017; Cai et al., 2016). Caron and Rousseau (2017) showed that using exchangeable point processes, one can obtain scale-free graphs with exponent $\eta \in (1, 2]$, but not above. While no results exist for the scale-freeness of edge-exchangeable random graphs in the sense of Definition 2.1 (see (Janson, 2017, Problem 9.8)), it is likely that a similar range is achieved for this class of models. Another family of models are non-exchangeable models based on preferential attachment (Barabási and Albert, 1999). The generated graphs are scale-free with exponent $\eta > 2$. However, the generative process makes it difficult to consider more general constructions that take into account community structure. Additionally, the non-exchangeability implies that the ordering of nodes must be known or need to be inferred for inference, which limits its applicability. By contrast, our model is finitely exchangeable for each n, and so the ordering of the nodes needs not to be known in order to make inference. As a consequence, no other projective class of model can give scale-free networks with exponent $\eta > 2$, interpretable parameters capturing community structure, and scalable inference, as described in this paper. While the model has a number of attractive properties, it also has some limitations. The mean number of triangles in inhomogeneous random graphs converges to a constant as n tends to infinity (van der Hofstad, 2018). Although the latent community structure introduced may mitigate this effect for reasonable n, this property appears undesirable for real-world network.

7 Experiments

7.1 Experiments with the rank-1 models

In this section, we test our inverse-gamma NR model (IG-NR) and generalized inverse Gaussian NR model (GIG-NR) on synthetic and real world graphs. For all experiments, we ran three MCMC chains for 10,000 iterations for our algorithms, and collected every 10th samples after 5,000 burn-in samples. The prior distributions for the hyperparameters of the different models are given in the supplementary material. The codes to replicate our experiments are available at `https://github.com/OxCSML-BayesNP/BNRG`.

Experiments with synthetic graphs. We first fitted the basic models with Inverse-gamma prior (IG) and generalized inverse Gaussian prior (GIG) on synthetic graphs generated from IG-NR model and GIG-



(a) `cond-mat` (b) `Enron` (c) `internet`

Figure 3: 95% credible intervals of predictive degree distributions of IG, GIG, GGP prior random graph models on (a) `cond-mat`, (b) `Enron` and (c) `internet` (right) graphs.

Table 1: Average reweighted KS statistic of predictive degree distributions and 95% credible intervals of estimated hyperparameters for IG, GIG and GGP models.

	`cond-mat`		`Enron`		`internet`	
	D	hyperparams	D	hyperparams	D	hyperparams
IG	**0.07**±0.01	$\alpha \in (2.55, 2.72)$ $\beta \in (9.20, 9.95)$	0.13±0.05	$\alpha \in (1.29, 1.34)$ $\beta \in (3.23, 3.41)$	**0.19**±0.00	$\alpha \in (3.20, 3.28)$ $\beta \in (6.51, 6.72)$
GIG	**0.07**±0.01	$\nu \in (-2.61, -2.37)$ $a \in (0.01, 0.02)$ $b \in (17.41, 19.14)$	**0.12**±0.01	$\nu \in (-1.33, -1.28)$ $a \in (0.00, 0.00)$ $b \in (6.42, 6.75)$	**0.19**±0.00	$\nu \in (-3.25, -3.18)$ $a \in (0.00, 0.00)$ $b \in (12.93, 13.30)$
GGP	0.15±0.06	$\sigma \in (-0.93, -0.80)$ $\tau \in (75.81, 85.52)$	0.18±0.02	$\sigma \in (0.19, 0.22)$ $\tau \in (11.53, 12.98)$	0.40±0.10	$\sigma \in (-0.18, -0.04)$ $\tau \in (92.05, 196.17)$

NR model. For IG, we generated a graph with $n = 5,000$ nodes with parameters $\alpha = 1.5$ and $\beta = 3.0$. For GIG, we generated a graph with 5,000 nodes with parameters $\nu = 0.5, a = 0.1, b = 3.0$. As summarized in Fig. 2, the posterior distribution recovers the hyperparameter values used to generated the graphs, and the posterior predictive distribution provides a good fit to the empirical degree distribution.

Experiments with real-world graphs. Now we evaluate our models on three real-world networks:
• `cond-mat`[1]: co-authorship network based on arXiv preprints for condensed matter, 16,264 nodes and 47,594 edges.
• `Enron`[2]: Enron collaboration e-mail network, 36,692 nodes and 183,831 edges.
• `internet`[3]: Network of internet routers, 124,651 nodes and 193,620 edges.
To evaluate the goodness-of-fit in terms of degree distributions, as suggested in Clauset et al. (2009), we sample graphs from the posterior predictive distribution based on the posterior samples, and computed the reweighted Kolmogorov-Sminorov (KS) statistic:

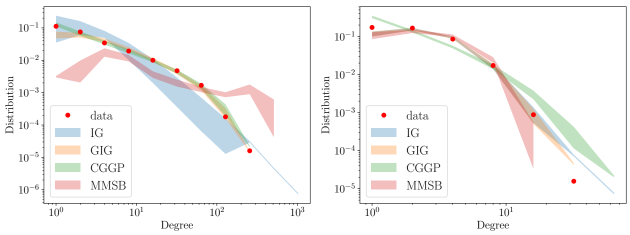
$$D = \max_{x \geq x_{\min}} \frac{|S(x) - P(x)|}{\sqrt{P(x)(1 - P(x))}}, \tag{26}$$

where $S(x)$ is the CDF of observed degrees, $P(x)$ is the CDF of degrees of graphs sampled from the predictive distribution, and $x_{\min}$ is the minimum x values among the observed degree and predictive degree. We compare our model to the random graph model with generalized gamma process prior (GGP, (Caron and Fox, 2017)), whose asymptotic degree distribution is a power-law with exponent in $(1, 2)$. We ran MCMC for the GGP model with 40,000 iterations and three chains. Posterior predictive degree distribution are reported in Fig. 3. Credible intervals of the hyperparameters and KS statistics for the different models are given in Table 1. Both IG and GIG provide a good to the degree distribution, with an exponent greater than 2, while the GGP model fails to capture the shape of the degree distribution.

7.2 Experiments with latent overlapping communities

Finally, we tested our models with latent overlapping communities on two real-world graphs with ground-truth communities.

[1]`https://toreopsahl.com/datasets/#newman2001`
[2]`https://snap.stanford.edu/data/email-Enron.html`
[3]`https://www.cise.ufl.edu/research/sparse/matrices/Pajek/internet.html`



(a) `polblogs` (b) `DBLP`

Figure 4: 95% credible intervals of predictive degree distributions on (a) `polblogs` and (b) `DBLP`.

Table 2: Average reweighted KS statistics and clustering accuracies.

	`polblogs`		`DBLP`	
	D	Acc (%)	D	Acc (%)
IG	0.71±0.50	**94.28**	**0.08**±0.03	72.46
GIG	0.14 ± 0.03	93.79	0.09±0.03	**76.58**
CGGP	**0.12**±0.03	94.12	0.33±0.02	57.49
MMSB	3.74±1.18	52.12	0.37±0.07	39.94

• `polblogs`[4]: the network of Americal political blogs. 1,224 nodes and 16.715 edges, two true communities (left or right).
• `DBLP`[5]: Co-authorship network of DBLP computer science bibliography. The original network has 317,080 nodes. Based on the ground-truth communities extracted in Yang and Leskovec (2012), we took three largest communities and subsampled 1,990 nodes among them. The subsampled graph contains 4,413 edges.

We compared our two models IG-NR and GIG-NR models to the random graph model based on compound generalized gamma process (CGGP, (Todeschini et al., 2016)), and mixed membership stochastic blockmodel (MMSB, Airoldi et al. (2009)). CGGP can capture the latent overlapping communities and has asymptotic power-law degree distsribution of exponent in $(1, 2)$. MMSB can capture the latent communities, but does not include a degree correction term. For all three models, we set the number of communities to be equal to two for `polblogs`, and three for `DBLP`. The CGGP was ran for 200,000 iterations after 10,000 initial iterations where w was initialized by running the model without communities (GGP). Each iteration of the sampler for MMSB scales quadratically with the number of nodes, and the sampler was therefore ran for a smaller number of iterations (5,000) for fair comparison. We found that longer iterations did not lead to improved performances. All methods were ran with three MCMC chains. For CGGP and MMSB methods, point estimates of the parameters measuring the level of affiliation of each individual were obtained using the Bayesian estimator described in Todeschini et al. (2016). For IG-NR and GIG-IR, we simply took the maximum a posteriori estimate of V. To compare to the ground truth communities, nodes are then assigned to the community where they have the strongest affiliation. The learned communities are shown in the supplementary material. Posterior predictive of the degree distributions for the different models are given in Fig. 4, and the KS statistic in Table 2. Both GIG-NR and CGGP exhibit a good fit to the `polblogs` dataset, where there does not seem to be evidence for a power-law exponent greater than 2. For the DBLP, both IG-NR and GIG-NR provide a good fit, while CGGP fails to capture adequately the degree distribution. The classification accuracy is also reported in Table 2. The classification accuracy is similar for IG-NR, GIG-NR and CGGP on `polblogs`. IG-NR and GIG-NR outperform other methods on the `DBLP` network. MMSB failed to capture both degree distributions and community structures, due to the large degree heterogeneity, a limitation already reported in previous articles (Karrer and Newman, 2011; Gopalan et al., 2013).

[4]`http://www.cise.ufl.edu/research/sparse/matrices/Newman/polblogs`
[5]`https://snap.stanford.edu/data/com-DBLP.html`

8 Conclusion

In this paper, we proposed a non-projective inhomogeneous random graph models with theoretically attractive asymptotic properties and overlapping community structures. We presented two parametrizations; 1) IG-NR, which is sparse and scale-free with power-law exponent greater than 2, and 2) GIG-NR, which is also sparse but has asymptotically power-law with exponential cutoff degree distribution. We described efficient MCMC algorithms that scale to large graphs, and demonstrated the benefits of the proposed models on various real-world graphs.

Acknowledgement JL and FC's research leading to these results has received funding from European Research Council under the European Unions Seventh Framework Programme (FP7/2007-2013) ERC grant agreement no. 617071 and from EPSRC under grant EP/P026753/1. FC acknowledges support from the Alan Turing Institute under EPSRC grant EP/N510129/1. LFJ was supported in part by the grant RGC-GRF 16300217 of the HKSAR.

References

Airoldi, E. M., Blei, D. M., Fienberg, S. E., and Xing, E. P. (2009). Mixed membership stochastic blockmodels. In *Advances in Neural Information Processing Systems 22*.

Aldous, D. (1997). Brownian excursions, critical random graphs and the multiplicative coalescent. *The Annals of Probability*, pages 812–854.

Barabási, A.-L. (2016). *Network science*, chapter 4. Cambridge university press.

Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439):509–512.

Bhamidi, S., van Der Hofstad, R., and Van Leeuwaarden, J. (2012). Novel scaling limits for critical inhomogeneous random graphs. *The Annals of Probability*, 40(6):2299–2361.

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

Borgs, C., Chayes, J. T., Cohn, H., and Holden, N. (2016). Sparse exchangeable graphs and their limits via graphon processes. *ArXiv preprint arXiv:1601.07134*.

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*.

Broutin, N., Duquesne, T., and Wang, M. (2018). Limits of multiplicative inhomogeneous random graphs and Lévy trees. *arXiv preprint arXiv:1804.05871*.

Cai, D., Campbell, T., and Broderick, T. (2016). Edge-exchangeable graphs and sparsity. In Lee, D. D., Sugiyama, M., Luxburg, U. V., Guyon, I., and Garnett, R., editors, *Advances in Neural Information Processing Systems 29*, pages 4249–4257. Curran Associates, Inc.

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.

Caron, F. and Rousseau, J. (2017). On sparsity and power-law properties of graphs based on exchangeable point processes. *arXiv preprint arXiv:1708.03120*.

Chung, F. and Lu, L. (2002a). The average distances in random graphs with given expected degrees. *Proceedings of the National Academy of Sciences of the Unites States of America*, 99(25):15879–15882.

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

Chung, F. and Lu, L. (2003). The average distance in a random graph with given expected degrees. *Internet Mathematics*, 1(1):91–113.

Clauset, A., Shalizi, C. R., and Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Reivew*, 51:661–703.

Crane, H. and Dempsey, W. (2015). A framework for statistical network modeling. *arXiv preprint arXiv:1509.08185*.

Crane, H. and Dempsey, W. (2017). Edge exchangeable models for interaction networks. *Journal of the American Statistical Association*.

Duane, S., Kennedy, A. D., Pendleton, B. J., and Roweth, D. (1987). Hybrid Monte Carlo. *Physics Letters B*, 195(2):216–222.

Goldenberg, A., Zheng, A. X., Fienberg, S. E., and Airoldi, E. M. (2010). A survey of statistical network models. *Foundations and Trends in Machine Learning*, 2(2):129–233.

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

Herlau, T., Schmidt, M. N., and Mørup, M. (2016). Completely random measures for modelling block-structured sparse networks. In *Advances in Neural Information Processing Systems 29*.

Janson, S. (2017). On edge exchangeable random graphs. *Journal of Statistical Physics*, pages 1–37.

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

Kolaczyk, E. D. (2009). *Statistical analysis of network data: methods and models.* Springer Science & Business Media.

Lee, J., Heakulani, C., Ghahramani, Z., James, L. F., and Choi, S. (2017). Bayesian inference on random simple graphs with power-law degree distributions. In *Proceedings of the 34th International Conference on Machine Learning*.

Neal, R. M. (2011). *MCMC using Hamiltonian Monte Carlo*, volume 2. Chapman & Hall / CRC Press.

Newman, M. (2009). *Networks: an introduction.* OUP Oxford.

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

Orbanz, P. and Roy, D. M. (2015). Bayesian models of graphs, arrays and other exchangeable random structures. *IEEE transactions on pattern analysis and machine intelligence*, 37(2):437–461.

Sichel, H. S. (1974). On a distribution representing sentence-length in written prose. *Journal of the Royal Statistical Society. Series A (General)*, pages 25–34.

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

van der Hofstad, R. (2013). Critical behavior in inhomogeneous random graphs. *Random Structures & Algorithms*, 42(4):480–508.

van der Hofstad, R. (2016). *Random graphs and complex networks: volume 1*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

van der Hofstad, R. (2018). *Random graphs and complex networks: volume 2*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Veitch, V. and Roy, D. M. (2015). The class of random graphs arising from exchangeable random measures. *arXiv preprint arXiv:1512.03099*.

Willmot, G. E. (1990). Asymptotic tail behaviour of Poisson mixtures with applications. *Advances in Applied Probability*, 22(1):147–159.

Yang, J. and Leskovec, J. (2012). Defining and evaluating network communities based on ground-truth. In *IEEE 12th International Conference on Data Mining*.

Yang, J. and Leskovec, J. (2013). Overlapping community detection at scale: a nonnegative matrix factorization approach. In *Proceedings of the sixth ACM international conference on Web search and data mining*, pages 587–596. ACM.

Zhou, M. (2015). Infinite edge partition models for overlapping community detection and link prediction. In *Proceedings of the 18th International Conference on Artificial Intelligence and Statistics*, pages 1135–1143.