Completely random measures for modelling block-structured networks

Tue Herlau
Technical University of Denmark
Richard Petersens plads 31,
2800 Lyngby, Denmark

Mikkel N. Schmidt
Technical University of Denmark
Richard Petersens plads 31,
2800 Lyngby, Denmark

Morten Mørup
Technical University of Denmark
Richard Petersens plads 31,
2800 Lyngby, Denmark

Abstract

Many statistical methods for network data parameterize the edge-probability by attributing latent traits to the vertices such as block structure and assume exchangeability in the sense of the Aldous-Hoover representation theorem. Empirical studies of networks indicate that many real-world networks have a power-law distribution of the vertices which in turn implies the number of edges scale slower than quadratically in the number of vertices. These assumptions are fundamentally irreconcilable as the Aldous-Hoover theorem implies quadratic scaling of the number of edges. Recently Caron and Fox [2014] proposed the use of a different notion of exchangeability due to Kallenberg [2006] and obtained a network model which admits power-law behaviour while retaining desirable statistical properties, however this model does not capture latent vertex traits such as block-structure. In this work we re-introduce the use of block-structure for network models obeying Kallenberg's notion of exchangeability and thereby obtain a model which admits the inference of block-structure and edge inhomogeneity. We derive a simple expression for the likelihood and an efficient sampling method. The obtained model is not significantly more difficult to implement than existing approaches to block-modelling and performs well on real network datasets.

1 Introduction

Two phenomena are generally considered important for modelling complex network. The first is community or block structure where the vertices are partitioned into non-overlapping blocks (denoted by $\ell = 1, \ldots, K$ in the following) and the probability two vertices i, j are connected depend on their assignment to blocks:

$$P\big(\text{Edge between vertex } i \text{ and } j\big) = \xi_{\ell m}$$

where $\xi_{\ell m} \in [0, 1]$ is a number only depending on the blocks ℓ, m which i, j belongs to. *Stochastic block models* (SBMs) were first proposed by White et al. [1976] and today form the basic starting point for many important link-prediction methods such as the *infinite relational model* [Xu et al., 2006, Kemp et al., 2006].

While block-structure is important for link prediction the degree distribution of edges in complex networks has also attracted a deal of attention [Barabási and Albert, 1999, Newman et al., 2001, Strogatz, 2001]. In many large networks the degree distribution is often found to follow a power-law [Newman, 2010]

$$P\big(\text{Fraction of nodes with } k \text{ edges}\big) \sim k^{-\gamma} \quad (1)$$

where $\gamma \approx 2.5$. Explaining this scaling has lead to many important models of network *growth* where edges and vertices are typically added in a sequential fashion. A notably example is the preferential attachment (PA) model of Barabási and Albert [1999].

Models such as the IRM and the PA model have different goals. The PA model attempts to explain how network structure, such as the degree distribution, follows from simple rules of network growth and is not suitable for link prediction. The IRMs main goal is to discover latent block-structure and predict edges, tasks which the PA model is not suitable for. In the following we will use network model to refer to a model with the same aims as the IRM, most notably predicting missing edges.
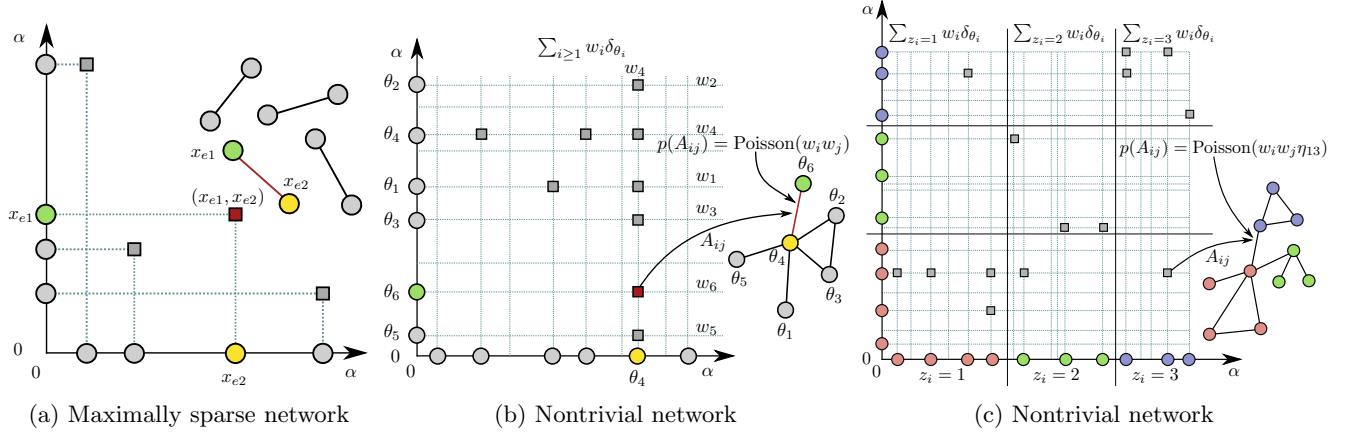
Figure 1: (Left:) A network is generated by randomly selecting points from $[0, \alpha[^2 \subset \mathbb{R}_+^2$ (squares) and identifying the unique coordinates with a vertex (circles) giving the maximally disconnected graph. (Middle:) The edges are restricted to lie at the intersection of randomly generated gray lines at θ_i each with a mass/sociability parameter w_i such that the probability of selecting an intersection is proportional to $w_i w_j$ giving a non-trivial network structure. (Right:) Using three independent measures $\sum_{i \geq 1} w_{\ell i} \delta_{\theta_{\ell i}}$ and modulating the interacting with a parameter $\eta_{\ell m} \geq 0$ allows block-structured networks where each random measure corresponds to a block.

1.1 Exchangeability

Invariances is an important theme in Baysian approaches to network modelling [Kallenberg, 2006]. For network data the criteria which has received most attention is infinite exchangeability of random arrays. Suppose we represent the network as a subset of an infinite matrix $A = (A_{ij})_{ij \geq 1}$ such that A_{ij} is the number of edges between vertex i and j (we will allow multi and self-edges in the following). Infinite exchangeability of the random array $(A_{ij})_{ij \geq 1}$ is the requirement that [Hoover, 1979, Aldous, 1981]

$$(A_{ij})_{ij \geq 1} \stackrel{d}{=} (A_{\sigma(i)\sigma(j)})_{ij \geq 1} \tag{2}$$

for all finite permutations σ of $\mathbb{N}$. The distribution of a finite network is then obtained by marginalization. An important consequence of infinite exchangeability is given by the Aldous-Hoover theorem [Hoover, 1979, Aldous, 1981] according to which a random network obeying infinite exchangeability eqn. (2) has a representation in terms of a random function and furthermore the number of edges in the network must either scale as the square of the number of vertices or (with probability 1) be zero [Orbanz and Roy, 2015]. None of these options are compatible with a power-law degree distribution and one is faced with the dilemma of either giving up the power-law distribution or exchangeability. It is the first horn of this dilemma which have been pursued by much work on Bayesian network modelling [Orbanz and Roy, 2015].

It is however possible to substitute the notation of infinite exchangeability in the sense of eqn. (2) with a different definition due to Kallenberg [2006, chapter 9]. The new notion retains many important characteristics of the original including a powerful representation theorem parallelling the Aldous-Hoover theorem but expressed in terms of a random set. Important progress has recently been made in exploring network models based on this representation by Caron and Fox [2014] who demonstrate the ability to model power-law behaviour of the degree distribution and construct an efficient sampler for parameter inference. The reader is encouraged to consult this reference for more details.

In this paper, we will apply the ideas of Caron and Fox [2014] to block-structured network data, thereby obtaining a model based on the same structural invariance but able to capture both block-structure and degree heterogeneity.

The remainder of the paper is structured as follows. In section 2.1 and 2.2 we will provide a basic description of the construction of Caron and Fox [2014] and a sketch of our proposed model, dubbed the completely random measure stochastic block model (CRMSBM). Section 2.3 contains a brief review of random measure theory which will be used to define the proposed model more explicitly. In section 2.5 we will introduce a simple sampling scheme and section 3 will investigate the model on 11 network datasets.

2 Methods

Before introducing the full method we will describe the construction informally omitting details relating to completely random measures.

2.1 A simple approach to sparse networks

Suppose the vertices in the graph are labelled by real numbers in $\mathbb{R}_+$. An edge e (edges are considered directed and we allow for self-edges) then consists of two numbers $(x_{e1}, x_{e2}) \in \mathbb{R}_+^2$ denoted the *edge endpoint* and a network X of L edges (possibly $L = \infty$) is simply the collection of points $X = ((x_{e1}, x_{e2}))_{e=1}^L \subset \mathbb{R}_+^2$. We adopt the convention that multi-edges implies duplicates in the list of edges.

Suppose X is generated by a Poisson process with base measure ξ on $\mathbb{R}_+^2$:

$$X \sim \mathrm{PP}(\xi). \tag{3}$$

a *finite* network X_α can then be obtained by considering the restriction of X to $[0, \alpha[^2$: $X_\alpha = X \cap [0, \alpha[^2$. This representation easily admits sparse networks as illustrated in the left pane of figure 1a. Suppose ξ is the Borel measure, the number of edges is then $L \sim \mathrm{Poisson}(\alpha^2)$ and the edge-endpoints x_{e1}, x_{e2} are i.i.d. on $[0, \alpha[$. The edges are indicated by the gray squares in figure 1a and the vertices as circles. Notice the vertices will be distinct with probability 1 and the procedure therefore gives rise to the de-generated but *sparse* network of $2L$ vertices and L edges shown in figure 1a.

To generate non-trivial networks the edge-endpoints must co-inside with nonzero probability. Similar to Caron and Fox [2014], suppose the coordinates are restricted to only take a countable number of potential values, $\theta_1, \theta_2, \cdots \in \mathbb{R}_+$ and each value has an associated *sociability* (or *mass*) parameter $w_1, w_2, \cdots \in [0, \infty[$ (we use the shorthand $(\theta_i)_i = (\theta_i)_{i=1}^\infty$ for series). If we define the measure $\mu = \sum_{i \geq 1} w_i \delta_{\theta_i}$ and let $\xi = \mu \times \mu$, then generating X_α according to the procedure of eqn. (3) the position of the edges is still selected i.i.d., but with probability proportional to $w_i w_j$ of selecting coordinate (θ_i, θ_j). Since the edge-endpoints coincide with non-zero probability this procedure allows the generation of non-trivial associative network structure, see figure 1b. With proper choice of $(w_i, \theta_i)_{i \geq 1}$ these networks exhibit many desirable properties such as a power-law degree distribution and sparsity [Caron and Fox, 2014].

This process can be intuitively extended to block-structured networks as illustrated in figure 1c: Each vertice is assigned a *sociability parameter* (here indicated by the colors and the symbol $z_i \in \{1, \ldots, K\}$) indicating the assignment of vertex i to one of K blocks. We can then consider a measure of the form

$$\xi = \sum_{i,j \geq 1} \eta_{z_i z_j} w_i w_j \delta_{(\theta_i, \theta_j)} = \sum_{\ell, m=1}^K \eta_{\ell m} \mu_\ell \times \mu_m \tag{4}$$

Where we have introduced $\mu_\ell = \sum_{i:z_i = \ell} w_i \delta_{\theta_i}$. ξ is then a measure on $[0, \alpha[^2$ and $\eta_{\ell m}$ parameterize the interaction strength between community ℓ and m. Notice the number of edges $L_{\ell m}$ between block ℓ and m is, by basic properties of the Poisson process, distributed as $L_{\ell m} \sim \mathrm{Poisson}(\eta_{\ell m} T_\ell T_m)$ where $T_\ell = \mu_\ell([0, \alpha[)$ and in figure 1c the location θ_i of the vertices have been artificially ordered according to color for easy visualization. The following section will show the connection between the above construction of eq. (4) and the exchangeable representation due to Kallenberg [2006], however for greater generality we will let the latent trait be a general continuous parameter $u_i \in [0, 1]$ and later show block-structured models can be obtained as a special case.

2.2 Exchangeability and point-process network models

Since the networks in the point-set representation are determined by the properties of the measure ξ, invariance (i.e. exchangeability) of random point-set networks is defined as invariance of this random measure. Recall infinite exchangeability for infinite matrices eqn. (2) was the requirement the distribution of the random matrix did not change by permutation of rows/columns in the network. For a random measure on $\mathbb{R}_+^2$ the corresponding requirement is that it should be possibly to partition $\mathbb{R}_+$ into intervals $I_1, I_2, I_3, \ldots$, permute the intervals, and the random measure should be invariant to this permutation. Formally, a random measure ξ on $\mathbb{R}_+^2$ is then said to be *jointly exchangeable* if $\xi \circ (\varphi \otimes \varphi)^{-1} \stackrel{d}{=} \xi$ for all measure-preserving transformation φ of $\mathbb{R}_+$. According to Kallenberg [2006, theorem 9.24] this is ensured provided the measure has a representation of the form:

$$\xi = \sum_{i,j \geq 1} h(\zeta, u_i, u_j) \delta_{(\theta_i, \theta_j)} \tag{5}$$

where h is a measurable function, ζ is a random variable and $\{(x_i, \theta_i)\}_{i \geq 1}$ is a unit rate Poisson process on $\mathbb{R}_+^2$ (the converse involves five additional terms [Kallenberg, 2006]). In this representation, the locations $(\theta_i)_i$ and the parameters $(x_i)_i$ are de-coupled, however we are free to select the random parameters $(x_i)_{i \geq 1}$ to lie in a more general space than $\mathbb{R}_+$. Specifically, define

$$x_i = (u_i, v_i) \in [0, 1] \times \mathbb{R}_+$$

with the interpretation that each v_i corresponds to a random mass w_i through a transformation $w_i = g(v_i)$ and each $u_i \in [0, 1]$ is a general *latent trait* of the vertex (In figure 1 this parameter corresponded to the assignment to blocks). We then consider the following

choice:

$$h(\zeta, x_i, x_j) = f(u_i, u_j)g_{z_i}(v_i)g_{z_j}(v_j) \qquad (6)$$

where $f : [0,1]^2 \to \mathbb{R}_+$ is a measurable function (compare to the graphon in the Aldous-Hoover representation) and $\{(u_i, v_i, \theta_i)\}_{i \geq 1}$ follows a unit-rate Poisson process on $[0,1] \times \mathbb{R}_+^2$.

To see the connection with the block-structured model, suppose the function f as the piece-wise constant function

$$f(u, u') = \sum_{\ell, m=1}^{K} \eta_{\ell m} 1_{J_\ell}(u) 1_{J_m}(u')$$

where

$$J_\ell = \left[\sum_{m=1}^{\ell-1} \beta_m, \sum_{m=1}^{\ell} \beta_m \right[, \quad \sum_{\ell=1}^{K} \beta_\ell = 1$$

and $\beta_\ell > 0$ and $z_i = \ell$ denotes the event $1_{J_\ell}(u_i) = 1$. Notice this choice for f is exactly equivalent to the graphon for the block-structured network model in the Aldous-Hoover representation [Orbanz and Roy, 2015]. The procedure is illustrated in figure 2.

Realizations of this point-process is shown in figure 3 for various values of K (top-row) and the choice of random measure introduced in section 2.3 and using the simulation method of Caron and Fox [2014]. For comparison the bottom row indicates the corresponding standard stochastic block-model representation where the edges are distributed uniformly within each tile. Notice the $K = 1, \eta_{11} = 1$ case corresponds to the method of Caron and Fox [2014].

To fully define the method we must first introduce the relevant prior for the measure $\mu = \sum_{i \geq 1} w_i \delta_{(\theta_i, u_i)}$. As a prior we will use the Generalized Gamma-process (GGP) [Hougaard, 1986]. In the following section we will briefly review properties of completely random measures and use these to derive a simple expression of the posterior.

2.3 Random measures

As a prior for μ we will use completely random measures (CRMs) and the reader is referred to [Kallenberg, 2006, Kingman, 1967] for a comprehensive account. Recall first the definition of a CRM. Assume $\mathbb{S}$ is a separable complete metric space with the Borel σ-field $\mathcal{B}(\mathbb{S})$ (comparing to the preceding discussion $\mathbb{S} = [0, \alpha[$). A *random measure* μ is a random variable whose values are random measure on $\mathbb{S}$. For each measurable set $A \in \mathcal{B}(\mathbb{S})$, the random measure induces a random variable $\mu(A)$, and the random measure μ will be said to be *completely random* if for any finite collection $A_1, \ldots, A_n$ of disjoint measurable sets the random
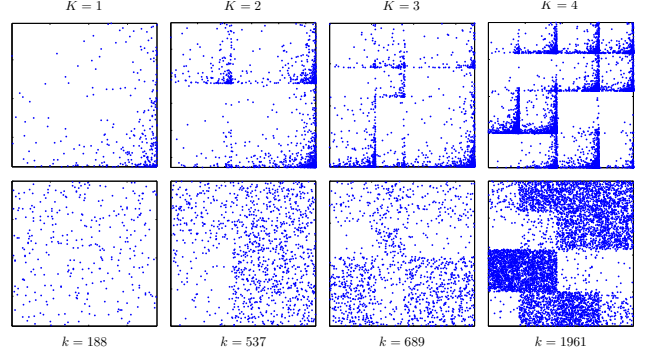

Figure 3: (Top row:) Example of four randomly generated graphs for $K = 1, 2, 3$ and 4. The other parameters were fixed at $\alpha = 20, \tau = 1, \sigma = 0.5$ and $\lambda_a = \lambda_b = 1$. Vertices has been sorted according to their assignment to blocks and sociability parameters. (Bottom row:) The same networks as above but applying a random permutation to the edges within each tile. A standard SBM assumes a network structure of this form.

variables $\mu(A_1), \ldots, \mu(A_n)$ are independent. It was shown by Kingman [1967] any such random measure μ is discrete almost certainly with a representation

$$\mu = \sum_{i=1}^{\infty} w_i \delta_{\theta_i} \qquad (7)$$

where the sequence of *masses* and *locations* $(w_i, \theta_i)_i$ (also known as the *atoms*) is a Poisson random measure on $\mathbb{R}^+ \times \mathbb{S}$ with mean measure ν. The mean measure ν thus fully characterizes the CRM and is known as the Lévy intensity measure. We will consider *homogeneous* CRMs where locations are independent, $\nu(dw, d\theta) = \rho(dw)\kappa_\alpha(d\theta)$, and assume κ_α is the standard Borel measure on $[0, \alpha[$.

Since the construction as outlined in figure 1c depends on sampling the edge start and end-points at random from the locations $(\theta_i)_i$ with probability proportional to w_i of particular interest will be the normalized form of eqn. (7). Specifically, the chance of selecting a particular location from a random draw is governed by

$$P = \frac{\mu}{T} = \sum_{i=1}^{\infty} p_i \delta_{\theta_i}, \ p_i = \frac{w_i}{T}, \ T = \mu(\mathbb{S}) = \sum_{i=1}^{\infty} w_i \ (8)$$

which is known as the *normalized random measure* (NRM) and T is known as the *total mass* of the CRM μ. A random draw from a Poisson process based on the CRM can thus be realized by first sampling the number of generated points, $L \sim \text{Poisson}(T)$, and then drawing their locations iid. from the NRM of eqn. (8).

Notice for this definition to make sense the Lévy intensity must satisfy the requirements: $\int_0^\infty \rho(dw) = \infty$
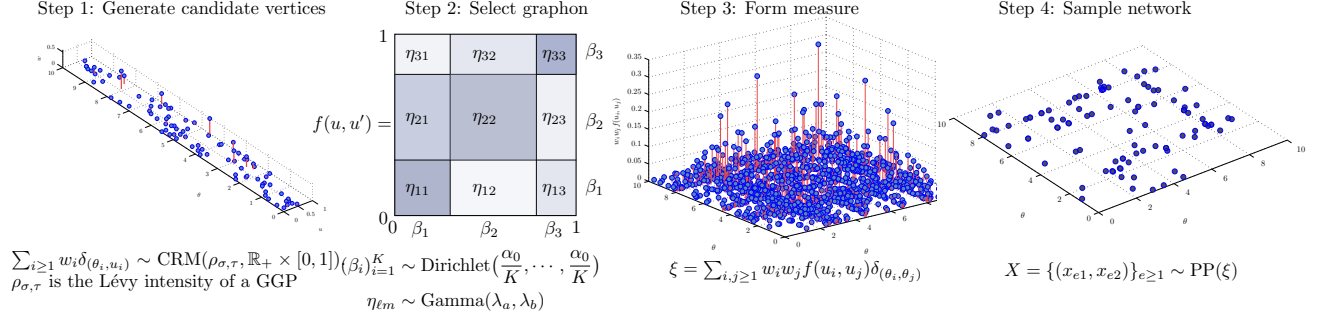
Step 1: Generate candidate vertices

Step 2: Select graphon

Step 3: Form measure

Step 4: Sample network

$$\sum_{i \geq 1} w_i \delta_{(\theta_i, u_i)} \sim \mathrm{CRM}(\rho_{\sigma,\tau}, \mathbb{R}_+ \times [0,1])$$
$\rho_{\sigma,\tau}$ is the Lévy intensity of a GGP

$$(\beta_i)_{i=1}^K \sim \mathrm{Dirichlet}(\tfrac{\alpha_0}{K}, \cdots, \tfrac{\alpha_0}{K})$$

$$\eta_{\ell m} \sim \mathrm{Gamma}(\lambda_a, \lambda_b)$$

$$\xi = \sum_{i,j \geq 1} w_i w_j f(u_i, u_j) \delta_{(\theta_i, \theta_j)}$$

$$X = \{(x_{e1}, x_{e2})\}_{e \geq 1} \sim \mathrm{PP}(\xi)$$

Figure 2: (Step 1:) The potential vertex locations, θ_i, latent traits u_i and sociability parameters w_i are generated using a *generalized gamma process* (Step 2:) The interaction of the latent traits $f : [0,1]^2 \to \mathbb{R}_+$, *the graphon*, is chosen to be a piece-wise constant function (Step 3-4:) Together, these determine the random measure ξ which is used to generate the network from a Poisson process

and $\int_0^\infty (1 - \exp(-w))\rho(dw) < \infty$, guaranteeing respectively that T is positive and finite, thereby ensuring eqn. (8) is well-defined. The reader is referred to James [2002] for a comprehensive treatment. For many important Lévy intensities the density of the total mass does not have a simple analytical expression, however the Laplace transform of the random variable S corresponding to the total mass T has the general form [James, 2002]

$$\mathcal{L}[S](u) = \Psi(u) = \kappa(\mathbb{S}) \int \rho(dw)(1 - e^{-uw}). \quad (9)$$

Consider a sequence of random draws made from a NRM of the form eqn. (8). Since the NRM is discrete almost surely there is a positive probability any two draws will select the same atom θ_i. This induces a random partition Π of $\mathbb{N}$ known as the *exchangeable partition probability function* where $i, j \in \mathbb{N}$ are in the same block iff. they select the same atom. This partition is exchangeable and it's properties can be derived from the choice of Lévy intensity, see [Kingman, 1978, Pitman, 2006] for additional details.

In the following we will focus exclusively on the *generalized gamma process* (GGP) as the choice of intensity measure[James, 2002]. The GGP is parameterized with two parameters σ, τ and has the functional form

$$\rho_{\sigma,\tau}(dw) = \frac{1}{\Gamma(1-\sigma)} w^{-1-\sigma} e^{-\tau w} dw$$

The parameters (σ, τ) may lie in either the regions $]-\infty, 0] \times]0, \infty]$ or $]0, 1[\times [0, \infty[$. We will focus exclusively on the later region as the first corresponds to a *finite* number of atoms. Notice this is the same choice as in [Caron and Fox, 2014]. In conjunction with κ_α we thus obtain three parameters (α, σ, τ) which fully describes the CRM and induced partition structure. To signify this dependence we will use $\mu_{\alpha,\sigma,\tau}$ for the CRM μ. For this particular choice the eqn. (9) becomes [Pitman,

2006]

$$\psi_{\alpha,\sigma,\tau}(u) = \int \rho_{\alpha,\sigma,\tau}(dw)(1 - e^{-uw})$$
$$= \frac{\alpha}{\sigma} \left[(u + \tau)^\sigma - \tau^\sigma \right] \quad (10)$$

For $\tau = 0, \alpha = 1$ this is the Laplace transform of a σ-stable random variable S_σ [Pitman, 2006, Devroye and James, 2014] which according to Zolotarev's integral representation has the following form [Zolotarev, 1964]

$$f_\sigma(x) = \frac{\sigma x^{\frac{-1}{1-\sigma}}}{\pi(1-\sigma)} \int_0^\pi du \; A(\sigma, u) e^{-A(\sigma,u)/x^{\sigma/(1-\sigma)}}$$
$$A(\sigma, u) = \left[\frac{\sin((1-\sigma)u)^{1-\sigma} \sin(\sigma u)^\sigma}{\sin(u)} \right]^{\frac{1}{1-\sigma}}. \quad (11)$$

Thus introducing a new random variable $T_{\alpha,\sigma,\tau}$ with density

$$g_{\alpha,\sigma,\tau}(t) = \theta^{-\frac{1}{\sigma}} f_\sigma(t\theta^{-\frac{1}{\sigma}}) \phi_\lambda(t\theta^{-\frac{1}{\sigma}}) \quad (12)$$

where $\phi_\lambda(t) = e^{\lambda^\sigma - \lambda t}$, $\lambda = \tau \theta^{\frac{1}{\sigma}}$, $\theta = \frac{\alpha}{\sigma}$. Standard properties of the Laplace transform reveals $\mathcal{L}[g_{\alpha,\sigma,\tau}] = \Psi$ as defined in eqn. (10). Thus, eqn. (12) is the density of total mass T for the GGP. Additional details, including efficient sampling methods for this distribution, is discussed by Devroye and James [2014].

With the notation in place we provide the final form of the generative process. Suppose the random measure μ (restricted to the region $[0, \alpha[$) have been generated from a GGP. Assume $z_i = \ell$ iff. $u_i \in J_\ell$ and define the K sub-measures on $[0, \alpha[$:

$$\mu_\ell = \sum_{i: z_i = \ell} w_i \delta_{\theta_i}$$

each with total mass $T_\ell = \mu_\ell([0, \alpha[)$. The number of points in each tile $L_{\ell m}$ is then Poisson($\eta_{\ell m} T_\ell T_m$) distributed, and given $L_{\ell m}$ the edge-endpoints

$(x_{e1\ell}, x_{e2m})$ between atoms in measure ℓ and m can then be drawn from the corresponding NRM. The generative process is then simply:

$$(\beta_\ell)_{\ell=1}^K \sim \text{Dirichlet}(\frac{\beta_0}{K}, \ldots, \frac{\beta_0}{K})$$

$$\mu \overset{\text{iid}}{\sim} \text{CRM}(\rho_{\sigma,\tau}, U_{[0,1]} \times U_{\mathbb{R}_+})$$

$$\eta_{\ell k} \overset{\text{iid}}{\sim} \text{Gamma}(\lambda_a, \lambda_b)$$

$$L_{\ell m} \overset{\text{iid}}{\sim} \text{Poisson}(\eta_{\ell m} T_\ell T_m)$$

$$\text{for } e=1,\ldots,L_{\ell m}: x_{e1\ell} \overset{\text{iid}}{\sim} \text{Categorical}\big((w_i/T_\ell)_{z_i=\ell}\big)$$

$$\text{and } x_{e2m} \overset{\text{iid}}{\sim} \text{Categorical}\big((w_j/T_m)_{z_j=m}\big)$$

For sampling random graphs we used the same adaptive thinning strategy as Caron and Fox [2014][1].

2.4 Posterior Distribution

In order to define a sampling procedure of the CRMSBM we must first characterize the posterior distribution. In Caron and Fox [2014] this was calculated using a specially tailored version of Palms formula. In this work we will use a counting argument inspired by Pitman [2003, eqn. (32)].

To this end, consider first the case where the interaction strengths $(\eta_{\ell m})_{\ell m}$ and block sizes $(\beta_\ell)_\ell$ has a fixed value and that number of edges $L_{\ell m}$ within each tile (ℓ, m) is given. Since not all potential vertices (i.e. terms $w_i \delta_{\theta_i}$ in μ) will have edges attached to them it is useful to introduce a variable which encapsulates this distinction. We therefore define the variable $\tilde{z}_i = 0, 1, \ldots, K$ with the definition:

$$\tilde{z}_i = \begin{cases} z_i & \text{if there exist } (x,y) \in X_\alpha \text{ st. } \theta_i \in \{x,y\}, \\ 0 & \text{otherwise.} \end{cases}$$

Suppose in addition for each measure μ_ℓ, the endpoints of the edges associated with this measure selects

$$k_\ell = |\{i : \tilde{z}_i = \ell\}|$$

unique atoms and that the number of edge-endpoints selecting any particular atom w_i is n_i. This naturally divides the edge-endpoints associated with a particular measure ℓ into a partition, $\{B_1, \ldots, B_{k_\ell}\}$ [Pitman, 2003], and we denote by $\Pi_{\ell,2L}$ this random partition for measure ℓ. For a particular measure the joint distribution

$$P(\Pi_{\ell,2L} = \{B_1, \ldots, B_{k_\ell}\}, w_i \in dw_i, T_\ell \in dT_\ell) \quad (13)$$

is obtained from three contributions (with $\alpha_\ell \equiv \beta_\ell \alpha$):

- The mass parameter $T_{\alpha\ell}$ is distributed as $g_{\alpha_\ell, \sigma, \tau}$

- For each $\ell = 1, \ldots, K$, there must be a Poisson atom in dw_i for each i such that $\tilde{z}_i = \ell$

- For each ℓ, we know there are Poisson atoms in $(dw_i)_{\tilde{z}_i=\ell}$, however since the measure of these intervals is infinitesimal, the remaining mass $T_\ell - \sum_{\tilde{z}_i=\ell} w_i$ is still distributed as $g_{\alpha_\ell, \sigma, \tau}$.

- Each edge-endpoint selects the atom independently with probability given by the NRM of eqn. (8), w_i/T_ℓ.

The probability eqn. (13) can then be obtained from these three contributions as (with $k = \sum_{\ell=1}^K k_\ell$ being the total number of vertices in the network):

$$\left\{ \prod_{i=1}^k \alpha \rho_{\sigma,\tau}(dw_i) \right\} \quad (14)$$

$$\times \prod_\ell^K \left\{ g_{\alpha_\ell, \sigma, \tau}(T_\ell - \sum_{i:\tilde{z}_i=\ell} w_i) \right\} \left\{ \prod_{i:\tilde{z}_i=\ell} \left(\frac{w_i}{T_\ell} \right)^{n_i} \right\} \quad (15)$$

where n_i is the total number of times a particular atom i of μ_ℓ is selected in the process. To connect these definitions to actual network data, i.e. an array $(A_{ij})_{i,j=1}^k$, notice if the atom (w_i, θ_i) corresponds to a particular vertex i in the network then $n_i = \sum_j (A_{ij} + A_{ji})$.

Returning to eqn. (15) for a particular ℓ, the expression can be integrated by introducing the variables $s_\ell = \sum_{i:\tilde{z}_i=\ell} w_i$ corresponding to the sum of the *selected* atoms, introducing the parameters $x_i = w_i/s_{z_i}$, and integrating [Pitman, 2003, Lijoi and Walker, 2008, Favaro et al., 2013]. With $n_\ell = \sum_{i:\tilde{z}_i=\ell} n_i$ eqn. (15) can be written as a product over K factors:

$$\prod_{i:\tilde{z}_i=\ell}^k (1-\sigma)_{n_i} \int_0^{T_\ell} ds_\ell \frac{s_\ell^{n_\ell - k_\ell \sigma - 1} g_{\alpha, \sigma, \tau}(T_\ell - s_\ell)}{\Gamma(n_\ell - k_\ell \sigma) T_\ell^{n_\ell} \alpha_\ell^{-k_\ell} e^{\tau s_\ell}}. \quad (16)$$

Recall the number of edges within each tile $L_{\ell m}$ is Poisson with rate $\eta_{\ell m} T_\ell T_m$. In addition, when considering a concrete observed data matrix the edges does not have a particular labelling which is otherwise introduced in the proceeding counting argument. Thus, if we observe a number A_{ij} of edges between vertices i, j in a particular tile, we must consider all ways a network with this number of edges can be obtained by our generative process. This is equivalent to the number of ways of selecting the particular edge-counts of the total edge-counts within each tile. The multiplicity becomes the multinomial coefficient:

$$\binom{L_{\ell m}}{(A_{ij})_{\tilde{z}_i=\ell, \tilde{z}_j=m}} = \frac{L_{\ell m}!}{\prod_{\tilde{z}_i=\ell, \tilde{z}_j=m} A_{ij}!}. \quad (17)$$

The probability of obtaining a particular observed network A_{ij} can be obtained by combining eqs. (16), (17) and the Poisson rates for the edge-counts within each tile to obtain:

$$P(A, (z_i)_i | (\eta_{\ell m})_{\ell m}, (\beta_\ell)_\ell) = \left\{ \prod_{\ell=1}^{K} \int_0^\infty dT_\ell \{\text{eqn. (16)}\} \right\}$$

$$\times \prod_{\ell m} \left\{ \text{Poisson}(L_\ell | \eta_{\ell m} T_\ell T_m) \frac{L_{\ell m}!}{\prod_{\substack{\tilde{z}_i=\ell, \\ \tilde{z}_j=m}} A_{ij}!} \right\}. \quad (18)$$

Defining $n_{\ell m} = \sum_{\tilde{z}_i=\ell, \tilde{z}_m=j} A_{ij}$ and simplifying

$$P(A, (z_i)_i | (\eta_{\ell m}), (\beta_\ell)_\ell) = \frac{1}{\prod_{ij} A_{ij}!} \prod_\ell \left[\int_0^\infty \int_0^{T_\ell} dT_\ell ds_\ell \right]$$

$$\times \left[\prod_{\ell m} \eta_{\ell m}^{n_{\ell m}} e^{-\sum_{\ell m} \eta_{\ell m} T_\ell T_m} \right] \left\{ \prod_\ell E_\ell \right\}$$

where we have defined

$$E_\ell = \frac{\alpha^{k_\ell} s_\ell^{n_\ell - k_\ell \sigma - 1}}{\Gamma(n_\ell - k_\ell \sigma) e^{\tau s_\ell}} g_{\alpha_\ell, \tau, \sigma}(T_\ell - s_\ell) \prod_{\tilde{z}_i=\ell} (1-\sigma)_{n_i}.$$

Similar to Lijoi and Walker [2008] we will use the simple change-of-variable from T to $t = T - s$ and a change in the order of integration to obtain:

$$\int_{\mathbb{R}^+} \int_0^s dT ds \ h(T, s) = \iint_{\mathbb{R}_+^2} ds dt \ h(t+s, s). \quad (19)$$

Then introducing the Gamma-priors for $\eta_{\ell m}$, Dirichlet prior for $(\beta_\ell)_\ell$ and integrating over $\eta_{\ell m}$ we obtain the final expression:

$$P(A, (z_i)_i, \sigma, \tau, (\alpha_\ell, s_\ell, t_\ell)_\ell) = \frac{\Gamma(\beta_0) \prod_{\ell=1}^K \alpha_\ell^{\frac{\beta_0}{K} - 1}}{\Gamma(\frac{\beta_0}{K})^K \alpha^{\beta_0}}$$

$$\times \frac{\prod_\ell E_\ell}{\prod_{ij} A_{ij}!} \prod_{\ell m} \frac{G(\lambda_a + n_{\ell m}, \lambda_b + (s_\ell + t_\ell)(s_m + t_m))}{G(\lambda_a, \lambda_b)}$$

$$(20)$$

where $G(a, b) = \Gamma(a) b^{-a}$ is the normalization factor of the Gamma distribution. Finally notice the $\eta = 1$ case, corresponding to Caron and Fox [2014], can be obtained by taking the limit $\lambda_a = \lambda_b \to \infty$ in which case $\frac{G(\lambda_a + n, \lambda_b + T)}{G(\lambda_a, \lambda_b)} \to e^{-T}$. When discussing the $K = 1$ case we will assume this limit has been taken.

2.5 Inference

Sampling the expression eqn. (20) requires three types of sampling updates: For A_{ij} we must apply a sampling procedure to impute missing values, the sequence of block-assignments $(z_i)_i$ must be updated, the parameters associated with the random measure σ, τ

must be updated and finally the remaining variables $(\alpha_\ell, s_\ell, t_\ell)$ associated with each expression E_ℓ must be updated. We will first consider the later problem:

Update of variables associated with each E_ℓ: All terms except the densities $g_{\alpha, \sigma, \tau}$ are amenable to standard sampling techniques. In [Caron and Fox, 2014] this expression was sampled by employing a proposal distribution proportional to the density, thus allowing their value to cancel. In our work we opted for the approach of Lomelí et al. [2014] in which u in Zolotarev's integral representation eqn. (11) is considered an auxiliary parameter. Thus, introducing $u_\ell \in]0, \pi[$ for each RPM gives the full set of variables $\Phi_\ell = (\alpha_\ell, s_\ell, t_\ell, u_\ell)$ for each RPM. For convenience, the domain of the variables are in turn transformed to $\mathbb{R}$ using the standard change-of-variables $x \mapsto e^x$ for α, t and s and the logistic mappings $x \mapsto (1 + e^{-x})^{-1}$, $x \mapsto \pi(1 + e^{-x})^{-1}$ for σ and u. We found a simple random-walk Metropolis-Hastings sampling with a $\mathcal{N}(0, \sigma = 0.1)$ kernel (50 steps per iteration) was robust and efficient compared to the other updates.

Update of z_i: These variables can be updated directly from the likelihood eqn. (20), however we opted to re-impute the weights $(w_i)_{\tilde{z}_i=\ell}$ by inverting the integration step from eqn. (15) to eqn. (16) to obtain

$$(w_i/s_\ell)_{i:\tilde{z}_i=\ell} \sim \text{Dirichlet}\left((n_i - \sigma)_{i:\tilde{z}_i=\ell}\right) \quad (21)$$

doing this for each $\ell = 1, \dots, K$ allows all variables z_i to be updated in a regular Gibbs sweep.

Update of A_{ij}: Most networks are binary whereas the model assumes count-data. Furthermore to test the model it is useful to predict the presence of unobserved edges. Both of these difficulties are resolved by imputation. Suppose we are given a matrix W such that $W_{ij} = 1$ iff. the edge-count A_{ij} is unobserved. Furthermore assume A_{ij} is binary and must be imputed. Edges can then in principle be imputed directly by performing MCMC updates of A_{ij} and accepting/rejecting according to the likelihood eqn. (20), however the coupling between different counts through the gamma functions in E_ℓ would make such a sampling procedure prohibitively expensive. This difficulty is not present in Caron and Fox [2014] where the sociability-vector $(w_i)_i$ are retained and updates using Hamiltonian Monte-Carlo, however we can re-sample $(w_i)_i$ and $(\eta_{\ell m})_{\ell m}$ from their marginal distributions and use the re-sampled values of $(w_i)_i$ to impute the corresponding values of (A_{ij}). Thus for each plate (ℓ, m) we sample $(w_i)_{\tilde{z}_i=\ell}$ from (21) and $\eta_{\ell m}$ from

$$\eta_{\ell m} \sim \text{Gamma}\left(n_{\ell m} + \lambda_a, (t_\ell + s_\ell)(t_m + s_m) + \lambda_b\right) \quad (22)$$

the distribution of each unobserved A_{ij} is then simply $\text{Poisson}(\eta_\ell w_i w_j)$, $z_i = \ell, z_j = m$.
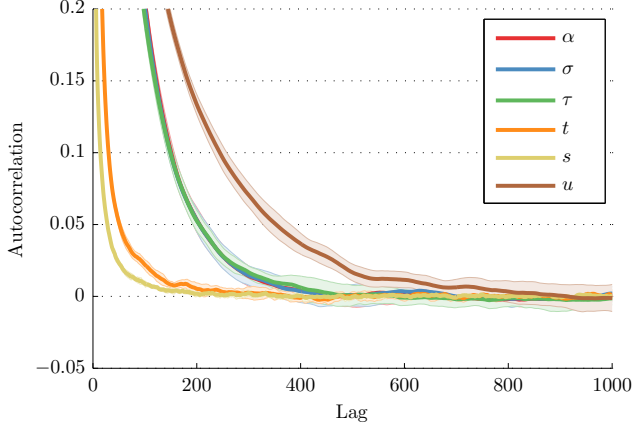
Figure 4: Autocorrelation plots for lags up to $1\,000$ of the parameters α, σ, τ (top row) and s, t, u (bottom row) for a $K = 1$ network drawn from the prior distribution using $\alpha = 25$, $\sigma = 0.5$ and $\tau = 2$. The plots were obtained by evaluating the proposed sampling procedure for 10^6 iterations and the shaded region indicates standard deviation obtained over 12 re-runs. The simulation indicates reasonable mixing for all parameters with u being the most affected by excursions.

2.5.1 Full inference procedure

Collecting the preceding steps we obtain

- For $\ell = 1, \ldots, K$, update the four variables in Φ_ℓ and σ, τ using random-walk metropolis hastings

- Impute $(w_i)_{\bar{z}_i = \ell}$ using eqs. (22) once for each ℓ and then iterate over i and update each z_i using a Gibbs sweep from the likelihood.

- Impute $(\eta_{\ell m})_{\ell m}$ and $(w_i)_i$ using eqs. (22) and (21) and for each (ij) such that the edge is either unobserved ($W_{ij} = 1$) or must be imputed ($A_{ij} \geq 1$) generate a candidate $a \sim \text{Poisson}(\eta_{\ell m} w_i w_j)$. Then, if $W_{ij} = 1$ simply set $A_{ij} = a$, otherwise if $W_{ij} = 0$ and $a = 0$ reject the update.

The parameters Φ_ℓ and σ, τ are important for determining the sparsity and power-law properties of the network model [Caron and Fox, 2014] and to investigate the sampling of these parameters we generated a single network problem using $\alpha = 25, \sigma = 0.5, \tau = 2$ and evaluated 12 samplers with $K = 1$ on the problem. Autocorrelation plots (mean and standard deviation computed over 12 restarts) can be seen in figure 4. All parameters mix, however the different parameters have different mixing times with in particular u being affected by excursions. This indicates many slice-sampling update of Φ_ℓ are required to explore the state space appreciably and we therefore applied 150 slice-sampling updates of Φ_ℓ for each update of $(z_i)_i$ and
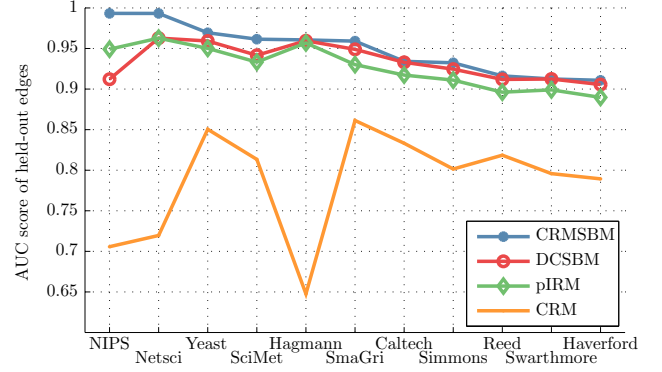

Figure 5: AUC score on held-out edges for the selected methods (averaged over 4 restarts) on 11 network datasets. For the same number of blocks, the CRMSBM offers good link-prediction performance compared to the method of Caron and Fox [2014] (CRM), a SBM with Poisson observations (pIRM) and the degree-corrected SBM (DCSBM) [Herlau et al., 2014]. Additional information is found in the supplementary material.

A_{ij}. Basic validation of the sampling procedure can be found in the supplementary material.

3 Experiments

The proposed method was evaluated on 11 network datasets (a description of how the datasets were obtained and prepared can be found in the supplementary material). As a criteria of evaluation we decided for AUC score on held-out edges, i.e. predicting the presence or absence of unobserved edges using the imputation method described in the previous section. All networks were initially processed by thresholds at 0 and vertices with zero edges were removed. A fraction of 5% of the edges were removed and considered as held-out data.

To examine the effect of using blocks, we compared the method against the method of Caron and Fox [2014] (CRM) (corresponding to $\eta_{\ell m} = 1$ and $K = 1$), a standard block-structured model with Poisson observations (pIRM) [Kemp et al., 2006] and the degree-corrected stochastic block model (DCSBM) Herlau et al. [2014], a simple model which allows both block-structure and degree-heterogeneity but is not exchangeable. More details on the simulations and methods are found in the supplementary material.

The pIRM was selected since it is the closest block-structured model to the CRMSBM without degree-correction. This allows us to determine the relative benefit of inferring the degree-distribution compared to only the block-structure.

For priors we selected uniform priors for σ, τ, α and a Gamma$(2,1)$ prior for $\beta_0, \lambda_a, \lambda_b$. For consistency similar choices were made for the alternative models.

All methods were evaluated for $T = 2\,000$ iterations and the later half of the chains was used for link prediction. We used 4 random selections of held-out edges per network to obtain the results seen in figure 5 (same sets of held-out edges were used for all methods). It is evident the use of block-structure is crucial to obtain good link prediction performance. For the block-structured methods, the results indicate additional benefits from using models which allows the modelling of degree-heterogenity for most networks except the Hagmann brain connectivity graph. This result is possibly explained by the Hagmann graph being constructed by fixing the number of outgoing edges for each vertex. Comparing the CRMSBM and the DCSBM, these models perform either on par or with a slight advantage to the CRMSBM. Models of networks based on the CRM representation of Kallenberg [2006] offers one of the most important new ideas in statistical modelling of networks in recent years and to our knowledge Caron and Fox [2014] were the first to realize the benefits of this modelling approach as well as describing it's statistical properties and provide an efficient sampling procedure.

The degree distribution of a network is only one of several important characteristics of a complex network. In this work we have examined how the ideas presented in Caron and Fox [2014] can be applied for a simple block-structured network model to obtain a model which admits block structure and degree correction. Our approach is a fairly straightforward generalization of the methods of Caron and Fox [2014]. However, we have opted to explicitly represent the density of the total mass $g_{\alpha_\ell, \sigma, \tau}$ using Zolotarev's integral representation and integrate out the sociability parameters $(w_i)_i$ reducing the number of parameters associated with the CRM from the order of vertices to the order of blocks.

The resulting model has the increased flexibility of being able to control the degree distribution within each block. In practice results of the model on 11 real-world datasets indicates this flexibility offers benefits over a purely block-structured approaches to link prediction for most networks as well as some potential benefit over an alternative approach to modelling block-structure and degree-heterogeneity. The results heavily indicates structural assumptions (such as block-structure) is important to obtain reasonable link prediction.

Block-structured network modelling is in turn the simplest structural assumption for block-modelling. The extension of the method of Caron and Fox [2014] to overlapping blocks, possibly using the correlated random measures of Chen et al. [2013], appears fairly straight-forward and should potentially offer a generalization of overlapping block models.
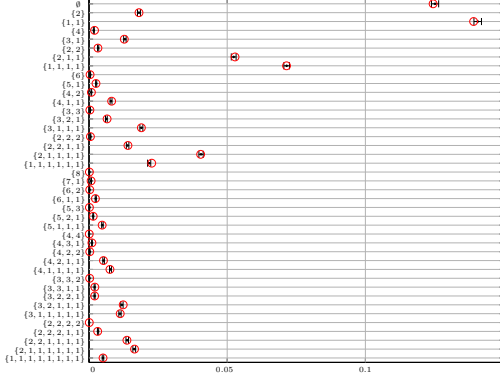
Acknowledgments

Figure 6: The estimated frequency of all unique networks binned according to their unique edge-endpoint counts $(n_1, \ldots, n_k)$. (ordered decreasingly) for $L = 0, \ldots, 4$ edges (41 in total, red circles), as well as the frequency obtained by computing the probability (see text for details).

A Supplementary Material

Validation of the sampler

To investigate the validity of the sampling procedure, we considered the $K = 1, \lambda_a = \lambda_b \to \infty$ case and used the sampling procedure of [Caron and Fox, 2014] with $(\alpha = 2, \sigma = 0.5, \tau = 1)$ to generate 250 000 random networks. As described in the previous section the probability of any given network is fully determined by the edge-endpoint counts $(n_1, \ldots, n_k)$ and the probability of a particular sequence of counts is permutation invariant. If ordered decreasingly this gives 41 unique vectors of edge-endpoint counts $(n_1, \ldots, n_k)$ for $L = 0, 1, 2, 3, 4$ (see vertical axis on figure 6) and the generated networks were binned according to their edge-endpoint count signature (networks with more than 4 edges were discarded). In this manner we obtained an estimate of the true frequency of a particular network signature. This estimate of the frequency was compared against the probability of a given network as computed by eqn. (17). Notice that due to permutation invariance the probability of each network signature must be corrected by multiplying eqn. (17) with a factor obtained by a combinatorial argument (see for instance Pitman [2006, eqn. (2.2)])

$$\frac{n!}{\sum_{i=1}^{m_i}(i!)^{m_i}m_i!}, \quad \text{where} \quad m_i = \sum_{i=1}^{k} 1(n_i = 1).$$

We thus obtain two estimates of the probability of a particular network signature shown in figure 6, both in close agreement. In figure 7 is shown the estimated density of the total mass T obtained by numerically
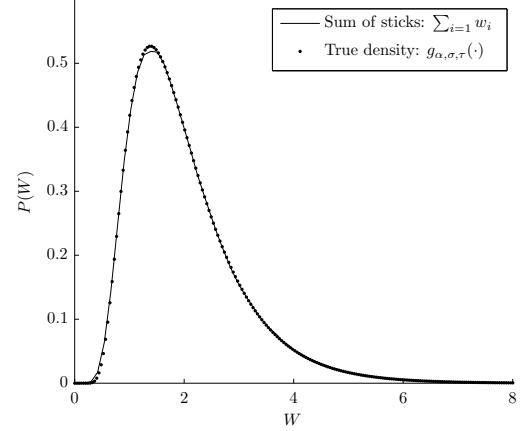

Figure 7: The density of the stick length $\sum_i w_i$ for the randomly generated networks as well as the true density eqn. (24) obtained by numerical integration of eqn. (23)

integrating Zolotarev's integral representation of f_σ

$$f_\sigma(x) = \frac{\sigma x^{\frac{-1}{1-\sigma}}}{\pi(1-\sigma)} \int_0^\pi du \ A(\sigma, u) e^{-A(\sigma,u)/x^{\sigma/(1-\sigma)}},$$

$$A(\sigma, u) = \left[\frac{\sin((1-\sigma)u)^{1-\sigma}\sin(\sigma u)^\sigma}{\sin(u)} \right]^{\frac{1}{1-\sigma}}, \quad (23)$$

$$g_{\alpha,\sigma,\tau}(t) = \theta^{-\frac{1}{\sigma}} f_\sigma(t\theta^{-\frac{1}{\sigma}})\phi_\lambda(t\theta^{-\frac{1}{\sigma}}) \quad (24)$$

and the estimated density of the total mass T obtained by summing the generated sticks (w_i). Both the estimates of the networks signatures and the density of T are in close agreement.

Datasets and preparation To test the methods we selected 11 publicly available datasets describing social networks, co-authorship networks and biological networks.

Yeast: Interaction network of 2361 proteins in yeast [Bu et al., 2003].

SmaGri: Coauthorship network of 1059 authors from the Garfield's collection of citation networks [Batagelj and Mrvar, 2014].

SciMet: Coauthorship network of 3084 authors from the Scientometrics journal, 1978-2000 [Batagelj and Mrvar, 2014].

Netscience: Coauthorship network of 1589 authors working in network theory as compiled by Newman [2006].

Hagman: Structural brain networks where edges correspond to the number of fiber tracts between 998 brain regions. All five networks in the dataset

were simply averaged to produce a single network [Hagmann et al., 2008].

NIPS: Consisting of the 2865 authors who have coauthored papers together at the 1-12'th NIPS conference [Roweis, 2009].

Caltech, Simmons, Reed, Haverford, Swarthmore: Five social networks of $769, 1446, 962, 1518, 1659$ students respectively obtained from the Facebook100 dataset [Traud et al., 2011].

The datasets were processes similarly by first removing any vertices without edges, i.e. where $n_i = 0$, and thresholding at 0 to produce binary networks. Selection of the missing edges for link prediction was done by first removing a fraction of 5% of all potential edges at random and then, if this procedure left any vertices without attached edges, re-introducing one of the edges attached to each such vertex and removing (at random from all other potential edges) a single edge. This procedure was repeated until 5% of the potential edges were removed and all vertices had at least one edge attached.

Models considered In addition to non-parametric extensions of the Poisson SBM we compared the CRMSBM against a degree corrected block model, the *degree-corrected stochastic block model* (DCSBM) of Herlau et al. [2014]. This model is not exchangeable but does model block structure and sociability.

Specifically the DCSBM assumes a generative process of the form:

$$(z_1, \ldots, z_n) \sim \mathrm{CRP}(\alpha)$$
$$\eta_{\ell m} \sim \mathrm{Gamma}(\lambda_a, \lambda_b)$$
$$(\theta_{i\ell}^{(1)}), (\theta_{i\ell}^{(2)}) \sim \mathrm{Dirichlet}((\gamma)_{i=1}^{k_\ell})$$
$$A_{ij} \sim \mathrm{Poisson}(k_{z_i} k_{z_j} \theta_{iz_i}^{(1)} \theta_{jz_j}^{(2)} \eta_{z_i z_j}).$$

To be consistent with the CRMSBM we selected a prior of the form Gamma$(2, 1)$ for α, λ_a and λ_b. The model is somewhat sensitive to the choice of prior for γ however we found a prior of the form Gamma$(2, 1)$ to perform reasonably well. The DCSBM reduces to a model without degree-correction, the pIRM [Kemp et al., 2006], by the choice $\gamma_{i\ell} = \frac{1}{n_\ell}$.

References

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

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

Vladimir Batagelj and Andrej Mrvar. Pajek. In *Encyclopedia of Social Network Analysis and Mining*, pages 1245–1256. Springer, 2014.

Dongbo Bu, Yi Zhao, Lun Cai, Hong Xue, Xiaopeng Zhu, Hongchao Lu, Jingfen Zhang, Shiwei Sun, Lunjiang Ling, Nan Zhang, et al. Topological structure analysis of the protein–protein interaction network in budding yeast. *Nucleic acids research*, 31 (9):2443–2450, 2003.

Francois Caron and Emily B Fox. Bayesian nonparametric models of sparse and exchangeable random graphs. *arXiv preprint arXiv:1401.1137*, 2014.

Changyou Chen, Vinayak Rao, Wray Buntine, and Yee Whye Teh. Dependent normalized random measures. In *Proceedings of The 30th International Conference on Machine Learning*, pages 969–977, 2013.

Luc Devroye and Lancelot James. On simulation and properties of the stable law. *Statistical methods & applications*, 23(3):307–343, 2014.

Stefano Favaro, Yee Whye Teh, et al. Mcmc for normalized random measure mixture models. *Statistical Science*, 28(3):335–359, 2013.

P Hagmann, L Cammoun, X Gigandet, R Meuli, C J Honey, V J Wedeen, and O Sporns. Mapping the structural core of human cerebral cortex. *PLoS biology*, 6(7):e159, 2008.

Tue Herlau, Mikkel N. Schmidt, and Morten Mørup. Infinite-degree-corrected stochastic block model. *Phys. Rev. E*, 90:032819, Sep 2014. doi: 10.1103/PhysRevE.90.032819. URL http://link.aps.org/doi/10.1103/PhysRevE.90.032819.

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

Philip Hougaard. Survival models for heterogeneous populations derived from stable distributions. *Biometrika*, 73(2):387–396, 1986.

Lancelot F James. Poisson process partition calculus with applications to exchangeable models and bayesian nonparametrics. *arXiv preprint math/0205093*, 2002.

Olav Kallenberg. *Probabilistic symmetries and invariance principles*. Springer Science & Business Media, 2006.

Charles Kemp, Joshua B Tenenbaum, Thomas L Griffiths, Takeshi Yamada, and Naonori Ueda. Learning systems of concepts with an infinite relational model. In *AAAI*, volume 3, page 5, 2006.

JFC Kingman. The representation of partition structures. *Journal of the London Mathematical Society*, 2(2):374–380, 1978.

John Kingman. Completely random measures. *Pacific Journal of Mathematics*, 21(1):59–78, 1967.

Antonio Lijoi and Stephen G Walker. Investigating nonparametric priors with gibbs structure. *Statistica Sinica*, 18(4):1653, 2008.

María Lomelí, Stefano Favaro, and Yee Whye Teh. A marginal sampler for σ-stable poisson-kingman mixture models. *arXiv preprint arXiv:1407.4211*, 2014.

M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Random graphs with arbitrary degree distributions and their applications. *Physical Review E*, 64(2), July 2001. ISSN 1063-651X. URL `http://pre.aps.org/abstract/PRE/v64/i2/e026118`.

Mark Newman. *Networks: an introduction.* Oxford University Press, 2010.

Mark EJ Newman. Finding community structure in networks using the eigenvectors of matrices. *Physical review E*, 74(3):036104, 2006.

Peter Orbanz and Daniel M Roy. Bayesian models of graphs, arrays and other exchangeable random structures. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 37(2):437–461, 2015.

Jim Pitman. Poisson-kingman partitions. *Lecture Notes-Monograph Series*, pages 1–34, 2003.

Jim Pitman. *Combinatorial Stochastic Processes: Ecole D'Eté de Probabilités de Saint-Flour XXXII-2002.* Springer, 2006.

Sam Roweis. Data for MATLAB hackers. `http://www.cs.nyu.edu/ roweis/data.html`, 2009. [Online; accessed 24-April-2015].

Steven H Strogatz. Exploring complex networks. *Nature*, 410(6825):268–276, 2001.

Amanda L. Traud, Peter J. Mucha, and Mason A. Porter. Facebook data scrape related to paper "The Social Structure of Facebook Networks". `https://archive.org/details/oxford-2005-facebook-matrix`, 2011. [Online; accessed 24-April-2015].

Harrison C White, Scott A Boorman, and Ronald L Breiger. Social structure from multiple networks. i. blockmodels of roles and positions. *American journal of sociology*, pages 730–780, 1976.

Z Xu, V Tresp, K Yu, and H P Kriegel. Infinite hidden relational models. In *Proceedings of the 22nd International Conference on Uncertainty in Artificial Intelligence (UAI 2006)*, 2006.

Vladimir Mikhailovich Zolotarev. On the representation of stable laws by integrals. *Trudy Matematicheskogo Instituta im. VA Steklova*, 71:46–50, 1964.