J. R. Statist. Soc. B (2017)
79, *Part* 5, *pp.* 1–44

Sparse graphs using exchangeable random measures

François Caron

University of Oxford, UK

and Emily B. Fox

University of Washington, Seattle, USA

[*Read before* The Royal Statistical Society *at a meeting organized by the* Research Section *on Wednesday, May 10th, 2017*, Professor C. Leng *in the Chair*]

Summary. Statistical network modelling has focused on representing the graph as a discrete structure, namely the adjacency matrix. When assuming exchangeability of this array—which can aid in modelling, computations and theoretical analysis—the Aldous–Hoover theorem informs us that the graph is necessarily either dense or empty. We instead consider representing the graph as an exchangeable *random measure* and appeal to the Kallenberg representation theorem for this object. We explore using completely random measures (CRMs) to define the exchangeable random measure, and we show how our CRM construction enables us to achieve sparse graphs while maintaining the attractive properties of exchangeability. We relate the sparsity of the graph to the Lévy measure defining the CRM. For a specific choice of CRM, our graphs can be tuned from dense to sparse on the basis of a single parameter. We present a scalable Hamiltonian Monte Carlo algorithm for posterior inference, which we use to analyse network properties in a range of real data sets, including networks with hundreds of thousands of nodes and millions of edges.

Keywords: Exchangeability; Generalized gamma process; Lévy measure; Point process; Random graphs

1. Introduction

The rapid increase in the availability and importance of network data has been a driving force behind the significant recent attention on random-graph models. In devising such models, there are several competing forces:

(a) *flexibility* to capture network features like sparsity of connections between nodes, heavy-tailed node degree distributions, dense spots or clusters, block structure;
(b) *interpretability* of the network model and associated parameters;
(c) *theoretical tractability* of analysis of network properties;
(d) *computational tractability* of inference with the ability to scale analyses to large collections of nodes.

A plethora of network models have been proposed in recent decades, each with different trade-offs made between flexibility, interpretability and theoretical and computational tractability; we refer the interested reader to overviews of such models provided by Newman (2003, 2010),

Address for correspondence: François Caron, Department of Statistics, University of Oxford, 24–29 St Giles, Oxford, OX1 3LB, UK.
E-mail: caron@stats.ox.ac.uk

 1369–7412/17/79000

Bollobás (2001), Durrett (2007), Goldenberg *et al.* (2010) and Fienberg (2012). In this paper, our focus is on providing a new framework in which to make these trade-offs. We demonstrate the ability to make gains in multiple directions using this framework through a specific example where the goal is to capture

(i) *sparsity*—tunable from sparse to dense via interpretable parameters,
(ii) *heavy-tailed degree distributions*—again controlled via interpretable parameters—and
(iii) *computational tractability* of Bayesian inference, scaling to networks with hundreds of thousands of nodes and millions of edges.

Classically, the graph being modelled has been represented by a discrete structure, or *adjacency matrix*, Z where Z_{ij} is a binary variable with $Z_{ij} = 1$ indicating an edge from node i to node j. In the case of undirected graphs, we furthermore restrict $Z_{ij} = Z_{ji}$. Then, the statistical network model is devised with this structure representing the observable quantity.

From a modelling, computational and theoretical standpoint, making an assumption of *exchangeability* is attractive. Under the adjacency matrix graph representation, such a statement informally equates with an invariance in distribution to permutations of node orderings. One reason why this assumption is attractive can be seen from applying the celebrated Aldous–Hoover theorem (Aldous, 1981; Hoover, 1979) to the adjacency matrix: infinite exchangeability of the binary matrix implies a mixture model representation involving transformations of uniform random variables (see theorem 7 in Appendix A.1). For undirected graphs, this transformation is specified by the *graphon* (Borgs *et al.* 2008; Lovász, 2013), which was originally studied as the limit object of dense graph sequences (Lovász and Szegedy, 2006; Borgs *et al.* 2010). The connection with the Aldous–Hoover theorem was made by Diaconis and Janson (2008). The graphon provides an object by which to study the theoretical properties of the statistical network process and to devise new estimators, as has been studied extensively in recent years (Bickel and Chen, 2009; Bickel *et al.*, 2011; Rohe, *et al.*, 2011; Zhao *et al.*, 2012; Airoldi *et al.*, 2013; Choi and Wolfe, 2014). Furthermore, the mixture model is a cornerstone of Bayesian modelling and provides a framework in which computational strategies are straightforwardly devised. Indeed, the Aldous–Hoover constructive definition has motivated new models (Lloyd *et al.*, 2012) and many popular existing models can be recast in this framework, including the stochastic block model (Nowicki and Snijders, 2001; Airoldi *et al.*, 2008) and latent space model (Hoff *et al.*, 2002).

One consequence of the Aldous–Hoover theorem is that graphs that are represented by an exchangeable random array are either empty or dense, i.e. the number of edges grows quadratically with the number of nodes n (see Lovász (2013) and Orbanz and Roy (2015)). However, empirical analyses suggest that many real world networks are sparse (Newman, 2010). Formally, sparsity is an asymptotic property of a graph. Following Bollobás and Riordan (2009), we refer to graphs with $\Theta(n^2)$ edges as *dense* and graphs with $o(n^2)$ edges as *sparse* (for notation, see Appendix C). The conclusion appears to be that we cannot have both exchangeability, with the associated benefits described above, and sparse graphs. Although network models can often adapt parameters to fit finite graphs, it is appealing to have a modelling framework with theoretically provable properties that are consistent with observed network attributes.

There are a couple of approaches to handling this apparent issue. One is to give up on exchangeability to obtain sparse graphs, such as in the popular preferential attachment model (Barabási and Albert, 1999; Berger *et al.*, 2014) or configuration model (Bollobás, 1980; Newman, 2010). Indeed, the networks literature is dominated by sparse non-exchangeable models. Alternatively, there is a body of literature that examines rescaling graph properties with network size n, leading to sparse graph sequences where each graph is finitely exchangeable (Bollobás

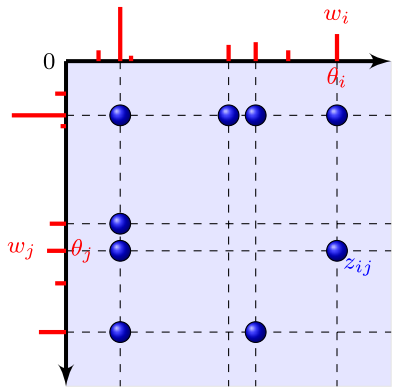


Fig. 1. Point process representation of a random graph: each node i is embedded in $\mathbb{R}_+$ at some location θ_i and is associated with a sociability parameter w_i; an edge between nodes θ_i and θ_j is represented by a point at locations (θ_i, θ_j) and (θ_j, θ_i) in $\mathbb{R}_+^2$

et al., 2007; Bollobás and Riordan, 2009; Olhede and Wolfe, 2012; Wolfe and Olhede, 2013; Borgs *et al.*, 2014a,b). Convergence of sparse graph sequences, analogous to the study of limiting objects for dense graph sequences, has also been studied (e.g. Borgs *et al.* (2016)). However, any method building on a rescaling approach provides a graph distribution π_n that lacks projectivity: marginalizing node n does not yield π_{n-1}, the distribution on graphs of size $n-1$.

We instead propose to set aside the discrete structure of the adjacency matrix and examine a different notion of exchangeability for a continuous space representation of networks. In particular, we consider a *point process* on $\mathbb{R}_+^2$:

$$Z = \sum_{i,j} z_{ij}\, \delta_{(\theta_i, \theta_j)}, \tag{1}$$

where z_{ij} is 1 if there is a link between node i and node j, and is 0 otherwise, and θ_i and θ_j are in $\mathbb{R}_+ = [0, +\infty)$ (Fig. 1). We can think of θ_i as a *time index* for node i. Exchangeability, as defined in Section 2, then equates with invariance to the time of arrival of the nodes. See Section 3.5 for a further interpretation of θ_i.

We note that exchangeability of the point process representation does not imply exchangeability of the associated adjacency matrix; however, the same modelling, computational and theoretical advantages remain. Interestingly, we arrive at a direct analogue to the constructive representation of the Aldous–Hoover theorem for exchangeable arrays and the associated graphon. Appealing to Kallenberg (1990, 2005), chapter 9 a point process on $\mathbb{R}_+^2$ is exchangeable if and only if it can be represented as a transformation of unit rate Poisson processes and uniform random variables (see theorem 1 in Section 2).

As a case-study in how this exchangeable random-measure framework can enable statistical network models with properties that are different from what can be achieved in the exchangeable array framework, we consider the following specification. To induce node heterogeneity in the link probabilities, we endow each node with a scalar *sociability parameter* $w_i > 0$. We then consider a straightforward link probability model. For any $i \neq j$,

$$\Pr(z_{ij} = 1 | w_i, w_j, \theta_i, \theta_j) = 1 - \exp(-2 w_i w_j). \tag{2}$$

This link function has been previously used by several others to build network models (Aldous, 1997; Norros and Reittu, 2006; Bollobás, *et al.*, 2007; van der Hofstad, 2014). Note that, under this specification, the 'time indices' θ_i and θ_j of nodes i and j do not influence the probability

of these two nodes to form a link. This is in contrast with, for example, standard latent space models (Hoff *et al.*, 2002). See Section 4 for further discussion.

To define the set of $(w_i, \theta_i)_{i=1,2,\ldots}$ underlying this statistical network model, we explore the use of *completely random measures* (CRMs) (Kingman, 1967). The $(w_i)_{i=1,2,\ldots}$ are the jumps of the CRM and the $(\theta_i)_{i=1,2,\ldots}$ the locations of the atoms. We show that, by carefully choosing the Lévy measure characterizing this CRM, we can construct graphs ranging from *sparse* to *dense*. In particular, any Lévy measure yielding an *infinite activity* CRM leads to sparse graphs; alternatively, *finite activity* CRMs yield dense graphs. For the class of infinite activity *regularly varying* CRMs, we can sharpen the results to obtain graphs where the number of edges increases at a rate below n^a, where $1 < a < 2$ depending on the Lévy measure. We focus on the flexible *generalized gamma process* CRM and show that one can tune the graph from dense to sparse via a single parameter.

Building on the framework of CRMs leads to other desirable properties as well. One is that our CRM-based exchangeable point process leads to an analytic representation for the graphon analogue in the Kallenberg framework (see Section 5.1). Another is that, by drawing on the considerable theory of CRMs that has been well studied in the Bayesian non-parametric community, we can derive network simulation techniques and develop a principled statistical estimation procedure. For the latter, in Section 7 we devise a scalable Hamiltonian Monte Carlo (HMC) sampler that can automatically handle a range of graphs from dense to sparse. We empirically show in Section 8 that our methods scale to graphs with hundreds of thousands of nodes and millions of edges. Importantly, exchangeability of the random measure underlies the efficiency of the sampler.

In summary, the CRM-based formulation combined with the specific link model of equation (2) serves as a proof of concept that moving to the point process representation of equation (1) can yield models with desirable attributes that are different from what can be obtained by using the discrete adjacency matrix representation. More generally, the notion of modelling the graph as an exchangeable random measure and appealing to a Kallenberg representation for such exchangeable random measures serves as an important new framework for devising and studying random-graph models, just as the original graphon concept stimulated considerable work in the network community in the past decade.

Our paper is organized as follows. In Section 2, we provide background on exchangeability and CRMs. Our statistical network models for directed multigraphs, undirected graphs and bipartite graphs are in Section 3. A discussion of our framework compared with related network models is provided in Section 4. Properties, such as exchangeability and sparsity, and methods for simulation are presented in Section 5. Specific cases of our formulation leading to dense and sparse graphs are considered in Section 6, including an empirical analysis of network properties of our proposed formulation. Our Markov chain Monte Carlo (MCMC) posterior computations are in Section 7. Finally, Section 8 provides a simulation study and an analysis of a variety of large, real world graphs.

The programs that were used to analyse the data can be obtained from

`http://www.rss.org.uk/preprints`

2. Background on exchangeability

Our focus is on exchangeable random structures that can represent networks. We first briefly review exchangeability for random sequences, continuous time processes and discrete network arrays. Thorough and accessible overviews of exchangeability of random structures have been presented in the surveys of Aldous (1985) and Orbanz and Roy (2015). Here, we simply abstract

Table 1. Overview of representation theorems

	Discrete structure	*Continuous time or space*
Exchangeability	de Finetti (1931)	Bühlmann (1960)
Joint or separate exchangeability	Aldous (1981) and Hoover (1979)	Kallenberg (1990)

away the notions that are relevant to placing our network formulation in context, as summarized in Table 1.

The classical representation theorem arising from a notion of exchangeability for discrete *sequences* of random variables is due to de Finetti (1931). The theorem states that a sequence $Z_1, Z_2, \ldots$ with $Z_i \in \mathbf{Z}$ is exchangeable if and only if there is a random probability measure $\mathbb{P}$ on $\mathbf{Z}$ with law ν such that the Z_i are conditionally independently and identically distributed (IID) given $\mathbb{P}$, i.e. all exchangeable infinite sequences can be represented as a mixture with mixing measure ν. If examining continuous time *processes* instead of sequences, the representation that is associated with exchangeable *increments* was given by Bühlmann (1960) (see also Freedman (1996) in terms of mixing Lévy processes.

The focus of our work, however, is on graph structures. For generic matrices Z in some space $\mathbf{Z}$, an (infinite) *exchangeable random array* (Diaconis and Janson, 2008; Lauritzen, 2008) is one such that

$$(Z_{ij}) \stackrel{\mathrm{d}}{=} (Z_{\pi(i)\tilde{\pi}(j)}) \qquad \text{for } (i, j) \in \mathbb{N}^2 \tag{3}$$

for any permutations $\pi, \tilde{\pi}$ of $\mathbb{N}$ (separate exchangeability), or for any permutation $\pi = \tilde{\pi}$ of $\mathbb{N}$ (joint exchangeability), where the notation '=d' stands for 'equal in distribution'. A representation theorem for exchangeability of the classical discrete adjacency *matrix* Z arises by considering a special case of the Aldous–Hoover theorem (Aldous, 1981; Hoover, 1979) to *2-arrays*. For undirected graphs where Z is a binary, symmetric adjacency matrix, the Aldous–Hoover representation can be expressed as the existence of a *graphon*. For completeness, the Aldous–Hoover theorem (specialized to 2-arrays under joint exchangeability) and further details on the graphon are provided in Appendix A.1.

Throughout this paper, we instead consider representing a graph as a point process $Z = \Sigma_{i,j} z_{ij} \delta_{(\theta_i, \theta_j)}$ with nodes θ_i embedded in $\mathbb{R}_+$, as in equation (1), and then examine notions of exchangeability in this context. Paralleling result (3), the point process Z on $\mathbb{R}_+^2$ is exchangeable if and only if (Kallenberg (2005), chapter 9)

$$(Z(A_i \times A_j)) \stackrel{\mathrm{d}}{=} (Z(A_{\pi(i)} \times A_{\tilde{\pi}(j)})) \qquad \text{for } (i, j) \in \mathbb{N}^2, \tag{4}$$

for any permutations $\pi, \tilde{\pi}$ of $\mathbb{N}$, with $\pi = \tilde{\pi}$ in the jointly exchangeable case, and any *intervals* $A_i = [h(i-1), hi]$ with $i \in \mathbb{N}$ and $h > 0$.

In words, result (4) states that the point process Z is exchangeable if, for any arbitrary regular square grid on the plane, the associated infinite array of increments (edge counts between nodes in a square) is exchangeable (Fig. 2). This provides a notion of exchangeability akin to that of the Aldous–Hoover theorem, but fundamentally different as the array being considered here is formed on the basis of an underlying continuous process. This array is not equivalent to an adjacency matrix, regardless of how fine a grid is considered.

Kallenberg (1990) derived de-Finetti-style representation theorems for separately and jointly exchangeable random measures on $\mathbb{R}_+^2$, which we present for the jointly exchangeable case in

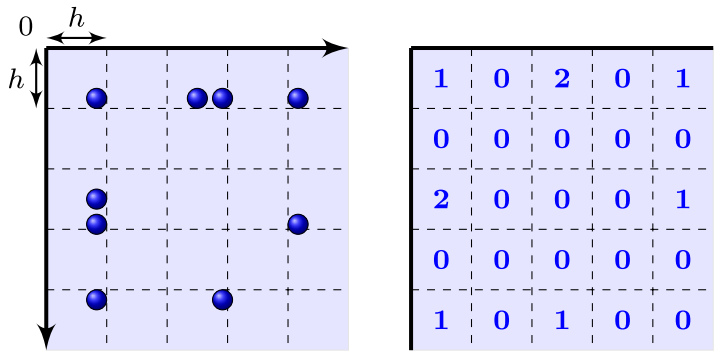


Fig. 2. Illustration of the notion of exchangeability for point processes on the plane: for any regular square grid on the plane (a), the associated infinite array counting the number of points in each square (b) is exchangeable in the sense of result (3)

theorem 1. In what follows λ denotes the Lebesgue measure on $\mathbb{R}_+$, λ_D the Lebesgue measure on the diagonal $D = \{(s,t) \in \mathbb{R}_+^2 | s = t\}$ and $\tilde{\mathbb{N}}_2 = \{\{i,j\} | (i,j) \in \mathbb{N}^2\}$. We also define a *U-array* to be an array of independent uniform random variables.

Theorem 1 (representation theorem for jointly exchangeable random measures on $\mathbb{R}_+^2$ (Kallenberg (1990) and Kallenberg (2005), theorem 9.24)). A random measure ξ on $\mathbb{R}_+^2$ is jointly exchangeable if and only if almost surely

$$\begin{aligned}\xi = &\sum_{i,j} f(\alpha_0, \vartheta_i, \vartheta_j, \zeta_{\{i,j\}})\delta_{\theta_i,\theta_j} + \beta_0 \lambda_D + \gamma_0(\lambda \times \lambda) \\ &+ \sum_{j,k}\{g(\alpha_0, \vartheta_j, \chi_{jk})\delta_{\theta_j,\sigma_{jk}} + g'(\alpha_0, \vartheta_j, \chi_{jk})\delta_{\sigma_{jk},\theta_j}\} \\ &+ \sum_{j}\{h(\alpha_0, \vartheta_j)(\delta_{\theta_j} \times \lambda) + h'(\alpha_0, \vartheta_j)(\lambda \times \delta_{\theta_j})\} \\ &+ \sum_{k}\{l(\alpha_0, \eta_k)\delta_{\rho_k,\rho'_k} + l'(\alpha_0, \eta_k)\delta_{\rho'_k,\rho_k}\} \end{aligned} \tag{5}$$

for some measurable functions $f : \mathbb{R}_+^4 \to \mathbb{R}_+$, $g, g' : \mathbb{R}_+^3 \to \mathbb{R}_+$ and $h, h', l, l' : \mathbb{R}_+^2 \to \mathbb{R}_+$. Here, $(\zeta_{\{i,j\}})$ with $\{i,j\} \in \tilde{\mathbb{N}}_2$ is a U-array. $\{(\theta_j, \vartheta_j)\}$ and $\{(\sigma_{ij}, \chi_{ij})\}$ on $\mathbb{R}_+^2$ and $\{(\rho_j, \rho'_j, \eta_j)\}$ on $\mathbb{R}_+^3$ are independent, unit rate Poisson processes. Furthermore, $\alpha_0, \beta_0, \gamma_0 \geqslant 0$ are an independent set of random variables.

We can think of the θ_i as random time indices, the sets $\{\theta_i\} \times \mathbb{R}_+$ and $\mathbb{R}_+ \times \{\theta_j\}$ forming Poisson processes of vertical and horizontal lines. The representation (1) is slightly more involved than the representation theorem for exchangeable arrays (see Appendix A.1). The first component of ξ is, however, similar to the representation for exchangeable arrays, the sequence of fixed indices $i = 1, 2, \ldots$ and uniform random variables $(U_i)_{i=1,2,\ldots}$ in equation (46) in Appendix A.1 being replaced by a unit rate Poisson process $\{(\theta_i, \vartheta_i)\}$ on $\mathbb{R}_+^2$. We place our proposed network model of Section 3 within this Kallenberg representation in Section 5.1, yielding direct analogues to the classical graphon representation of graphs based on exchangeability of the adjacency matrix.

3. Proposed statistical network model

Recall that we represent an undirected graph using an atomic measure

$$Z = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} z_{ij} \delta_{(\theta_i, \theta_j)},$$

with the convention $z_{ij} = z_{ji} \in \{0, 1\}$. Here, $z_{ij} = z_{ji} = 1$ indicates an undirected edge between nodes θ_i and θ_j. See Section 3.5 for the interpretation of θ_i.

There are many options for defining a statistical model for the point process graph representation Z. We consider one in particular in this paper based on a specific choice of

(a) link probability model and
(b) a prior on the model parameters.

Expanding on the discussion of Section 1, we introduce a collection of per-node *sociability* parameters $w = \{w_i\}$ and specify link probabilities via

$$\Pr(z_{ij} = 1 | w) = \begin{cases} 1 - \exp(-2 w_i w_j) & i \neq j, \\ 1 - \exp(-w_i^2) & i = j. \end{cases} \tag{6}$$

As mentioned in Section 1, this link probability model is not new to the statistical networks community and is a straightforward method for achieving node heterogeneity (see Aldous (1997) and Norros and Reittu (2006)).

3.1. Defining node parameters by using completely random measures

The model parameters consist of a collection of node-specific sociability parameters $w_i > 0$ and continuous-valued node indices $\theta_i \in \mathbb{R}_+$.

Our generative model jointly specifies $(w_i, \theta_i)_{i=1,2,\ldots}$ by first defining an atomic random measure

$$W = \sum_{i=1}^{\infty} w_i \delta_{\theta_i} \tag{7}$$

and then taking W to be distributed according to a homogeneous CRM (Kingman, 1967).

CRMs have been used extensively in the Bayesian non-parametric literature for proposing flexible classes of priors over functional space (Regazzini *et al.*, 2003; Lijoi and Prünster, 2010). We briefly review a few important properties of CRMs that are relevant to our construction; the reader can refer to Kingman (1993) or Daley and Vere-Jones (2008) for an exhaustive coverage.

A CRM W on $\mathbb{R}_+$ is a random measure such that, for all finite families of disjoint, bounded measurable sets $(A_1, \ldots, A_n)$ of $\mathbb{R}_+$, the random variables $W(A_1), \ldots, W(A_n)$ are mutually independent.

We shall focus here on CRMs with no deterministic component and stationary increments (i.e. the distribution of $W([t, s])$ depends only on $t - s$). In this case, the CRM takes the form (7), with $(w_i, \theta_i)_{i \in \mathbb{N}}$ the points of a Poisson point process on $(0, \infty) \times \mathbb{R}_+$ defined by a mean measure $\nu(\mathrm{d}w, \mathrm{d}\theta) = \rho(\mathrm{d}w)\lambda(\mathrm{d}\theta)$, where λ is the Lebesgue measure and ρ is a Lévy measure on $(0, \infty)$.

We denote this by

$$W \sim \mathrm{CRM}(\rho, \lambda). \tag{8}$$

Note that $W([0, T]) < \infty$ almost surely for any $T < \infty$, whereas $W(\mathbb{R}_+) = \infty$ almost surely if $\int_0^{\infty} \rho(\mathrm{d}w) > 0$.

The jump part ρ of the mean measure—which characterizes the increments of W—is of particular interest in our graph construction, as explored in Section 5. If ρ satisfies the condition

$$\int_0^\infty \rho(\mathrm{d}w) = \infty, \tag{9}$$

then there will be an infinite number of jumps in any interval $[0, T]$, and we refer to the CRM as *infinite activity*. Otherwise, the number of jumps will be finite almost surely. In our model, the jumps correspond to potentially connected nodes, i.e. these nodes need not be connected to any other node within a bounded interval and instead represent an upper bound on the set of connected nodes. See Section 3.5 for further discussion.

In Section 6, we consider special cases including the (compound) Poisson process and generalized gamma process (GGP) (Brix, 1999; Lijoi *et al.*, 2007).

3.2. Directed multigraphs

Formally, our undirected graph model is viewed as a transformation of a directed integer-weighted graph, or *multigraph*, as detailed in Section 3.3. We now specify this directed multigraph. Although our primary focus is on undirected network models, in some applications the directed multigraph might actually be the direct quantity of interest. For example, in social networks, interactions are often not only directed ('person i messages person j'), but also have an associated count. Additionally, interactions might be typed ('message', 'SMS', 'like', 'tag'). Our proposed framework could be directly extended to model such data.

Let $V = (\theta_1, \theta_2, \ldots)$ be a countably infinite set of node indices with $\theta_i \in \mathbb{R}_+$. We represent the directed multigraph of interest with an atomic measure on $\mathbb{R}_+^2$

$$D = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} n_{ij}\, \delta_{(\theta_i, \theta_j)}, \tag{10}$$

where n_{ij} counts the number of directed edges from node i to node j, with time indices θ_i and θ_j. See Fig. 3 for an illustration.

Given W as defined in expressions (7) and (8), D is simply generated from a Poisson process with intensity given by the product measure $\tilde{W} = W \times W$ on $\mathbb{R}_+^2$:

$$D | W \sim PP(\tilde{W}), \tag{11}$$

i.e., informally, the individual counts n_{ij} are generated as Poisson$(w_i w_j)$. (We consider a generalized definition of a Poisson process, where the mean measure is allowed to have atoms (Daley and Vere-Jones (2003), section 2.4).) By construction, for any $A, B \subset \mathbb{R}$, we have $\tilde{W}(A \times B) = W(A)W(B)$. On any bounded interval A of $\mathbb{R}_+$, $W(A) < \infty$, implying that $\tilde{W}(A \times A)$ has finite mass.

3.3. Undirected graphs via transformations of directed graphs

We arrive at the undirected graph via a simple transformation of the directed graph: set $z_{ij} = z_{ji} = 1$ if $n_{ij} + n_{ji} > 0$ and $z_{ij} = z_{ji} = 0$ otherwise, i.e. place an undirected edge between nodes θ_i and θ_j if and only if there is at least one directed interaction between the nodes. In this definition of an undirected graph, we allow self-edges. This could represent, for example, a person posting a message on his or her own profile page. The resulting hierarchical model is

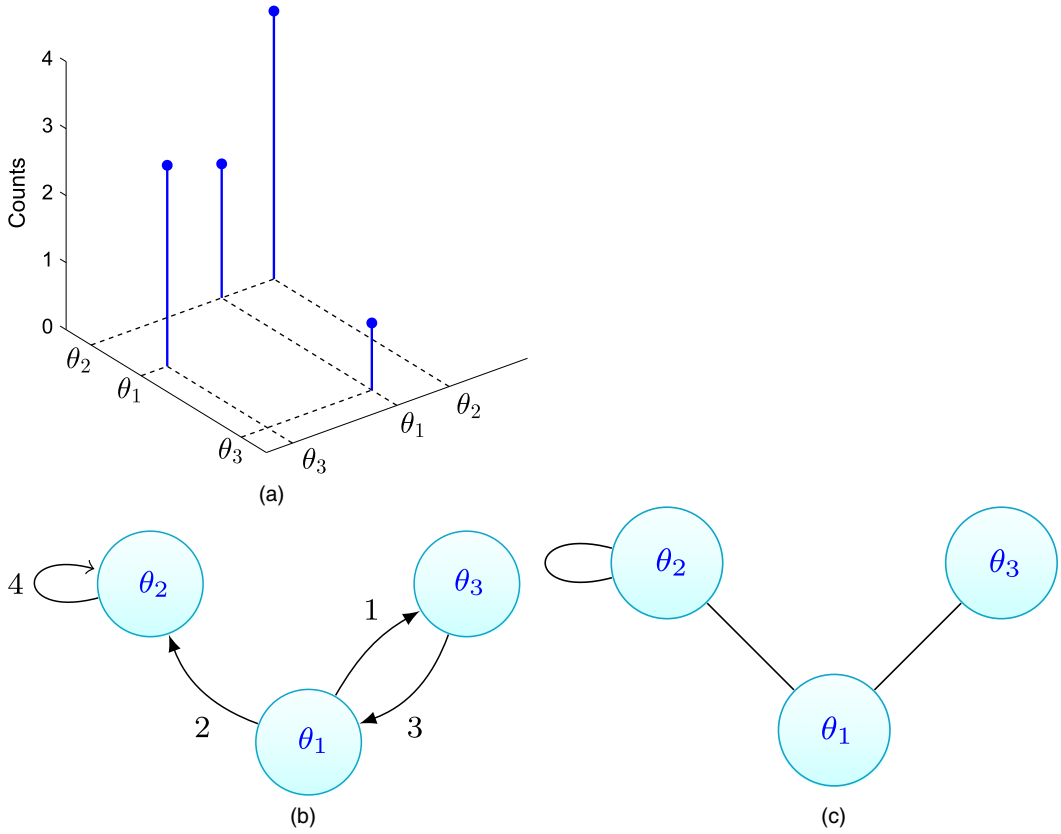


Fig. 3. Example of (a) an atomic measure D as in equation (10) restricted to $[0,1]^2$, (b) the corresponding directed multigraph and (c) the corresponding undirected graph

$$\left.\begin{aligned} W &= \sum_{i=1}^{\infty} w_i \delta_{\theta_i} && W \sim \mathrm{CRM}(\rho, \lambda), \\ D &= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} n_{ij} \delta_{(\theta_i, \theta_j)} && D|W \sim \mathrm{PP}(W \times W), \\ Z &= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \min(n_{ij} + n_{ji}, 1) \delta_{(\theta_i, \theta_j)}. \end{aligned}\right\} \tag{12}$$

This process is depicted graphically in Fig. 4.

To see the equivalence between this formulation and that specified in equation (6), note that, for $i \neq j$, $\Pr(z_{ij} = 1|w) = \Pr(n_{ij} + n_{ji} > 0|w)$. By properties of the Poisson process, n_{ij} and n_{ji} are independent random variables conditioned on W. The sum of two Poisson random variables, each with rate $w_i w_j$, is again Poisson with rate $2w_i w_j$. Result (6) arises from the fact that $\Pr(n_{ij} + n_{ji} > 0|w) = 1 - \Pr(n_{ij} + n_{ji} = 0|w)$. Likewise, the $i = j$ case arises by using a similar reasoning for $\Pr(z_{ii} = 1|w) = \Pr(n_{ii} > 0|w)$.

We note that our computational strategy of Section 7 relies on this interpretation of our model for undirected graphs as a transformation of a directed multigraph. In particular, we introduce the directed edge counts as latent variables and do inference over these counts.

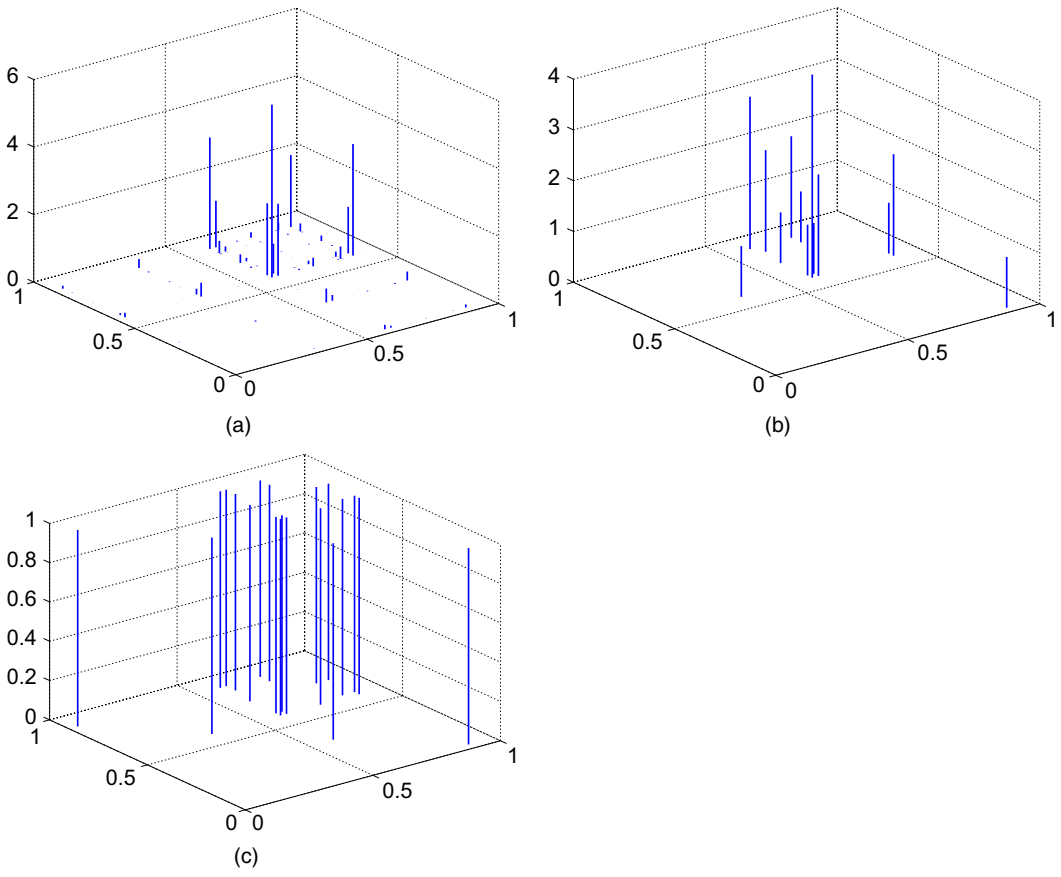


Fig. 4. Example of (a) the product measure $\tilde{W} = W \times W$ for CRM W, (b) a draw of the directed multigraph measure $D|W \sim PP(W \times W)$ and (c) the corresponding undirected measure $Z = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \min(n_{ij} + n_{ji}, 1)\delta_{(\theta_i,\theta_j)}$

3.4. Bipartite graphs

The above construction can also be extended to bipartite graphs. Let $V = (\theta_1, \theta_2, \ldots)$ and $V' = (\theta'_1, \theta'_2, \ldots)$ be two countably infinite sets of nodes with $\theta_i, \theta'_i \in \mathbb{R}_+$. We assume that only connections between nodes of different sets are allowed.

We represent the *directed bipartite multigraph* of interest by using an atomic measure on $\mathbb{R}_+^2$:

$$D = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} n_{ij}\delta_{(\theta_i,\theta'_j)}, \tag{13}$$

where n_{ij} counts the number of directed edges from node θ_i to node θ'_j. Similarly, the *bipartite graph* is represented by an atomic measure

$$Z = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} z_{ij}\delta_{(\theta_i,\theta'_j)}.$$

Our bipartite graph formulation introduces two independent CRMs, $W \sim CRM(\rho, \lambda)$ and $W' \sim CRM(\rho', \lambda)$, whose jumps correspond to sociability parameters for nodes in sets V and V' respectively. The generative model for the bipartite graph mimics that of the non-bipartite graph:

$$\left.\begin{aligned} W &= \sum_{i=1}^{\infty} w_i \delta_{\theta_i} && W \sim \mathrm{CRM}(\rho, \lambda), \\ W' &= \sum_{j=1}^{\infty} w'_j \delta_{\theta'_j} && W' \sim \mathrm{CRM}(\rho', \lambda), \\ D &= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} n_{ij} \delta_{(\theta_i, \theta'_j)} && D|W, W' \sim \mathrm{PP}(W \times W'), \\ Z &= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \min(n_{ij}, 1) \delta_{(\theta_i, \theta'_j)}. \end{aligned}\right\} \tag{14}$$

Model (14) has been proposed by Caron (2012) in a slightly different formulation. In this paper, we recast this model within our general framework, enabling new theoretical and practical insights.

3.5. *Interpretation of* θ_i

We can think of the positive, continuous valued node index θ_i as representing the time at which a *potential node* enters the network and has the opportunity to link with other existing nodes $\theta_j < \theta_i$. We use the terminology 'potential node' here to clarify that this node need not form any observed connections with other nodes existing before time θ_i. We emphasize that an observed link between θ_i and some other node $\theta_k > \theta_i$ will eventually occur almost surely as time progresses. This could represent, for example, signing on to a social networking service before your friends do and only forming a link once they join. On the basis of our CRM specification, we have almost surely an infinite number of potential nodes as time goes to ∞. For infinite activity CRMs, we have almost surely an infinite set of potential nodes even at any finite time.

In Section 5, we examine properties of the network process across time, and we describe methods for simulating networks at any finite time. There, our focus is on the observed link process from this set of potential nodes. For example, sparsity is examined with respect to the set of nodes with degree at least 1, not with respect to the set of potential nodes. Since we need not think of θ_i as a time index, but rather just a general construct of our formulation, we also generically refer to θ_i as the node *location* in the remainder of the paper.

4. Related work

There has been extensive work over recent years on flexible Bayesian non-parametric models for networks, allowing complex latent structures of unknown dimension to be uncovered from real world networks (Kemp *et al.*, 2006; Miller *et al.*, 2009; Lloyd *et al.*, 2012; Palla *et al.*, 2012; Herlau *et al.*, 2014). However, as mentioned in the unifying overview of Orbanz and Roy (2015), these methods all fit in the Aldous–Hoover framework and as such produce dense graphs.

Norros and Reittu (2006) proposed a conditionally Poissonian multigraph process with similarities to be drawn to our multigraph process. In their formulation, each node has a given sociability parameter and the number of edges between two nodes i and j is drawn from a Poisson distribution with rate the product of the sociability parameters, normalized by the sum of the sociability parameters of all the nodes. The normalization makes this model similar to models based on rescaling of the graphon and, as such, does not define a projective model, as explained in Section 1. See van der Hofstad (2014) for a review of this model and Britton *et al.*, (2006) for a similar model.

As pointed out by Jacobs and Clauset (2014) in their discussion of an earlier version of this paper, another related model is the degree-corrected random-graph model (Karrer and

Newman, 2011), where edges of the multigraph are drawn from a Poisson distribution whose rate is the product of node-specific sociability parameters and a parameter tuning the interaction between the latent communities to which these nodes belong. When the sociability parameters are assumed to be IID from some distribution, this model yields an exchangeable adjacency matrix and thus a dense graph.

Additionally, there are similarities to be drawn with the extensive literature on latent space modelling (e.g. Hoff *et al.* (2002), Penrose (2003) and Hoff (2009)). In such models, nodes are embedded in a low dimensional, continuous latent space and the probability of an edge is determined by a distance or similarity metric of the node-specific latent factors. In our case, the continuous node index θ_i is of no importance in forming edge probabilities. It would, however, be possible to extend our approach to time- or location-dependent connections by considering inhomogenous CRMs.

Finally, as we shall detail in Section 5.5, our model admits a construction with connections to the configuration model (Bollobás, 1980; Newman, 2010), which is a popular model for generating simple graphs with a given degree sequence.

The connections with this broad set of past work place our proposed network model within the context of existing literature. Importantly, however, to the best of our knowledge this work represents the first fully generative and projective approach to sparse graph modelling (see Section 5), and with a notion of exchangeability that is essential for devising our scalable statistical estimation procedure, as shown in Section 7.

5. General properties and simulation

We provide general properties of our network model depending on the properties of the Lévy measure ρ.

5.1. Exchangeability under the Kallenberg framework

Proposition 1 (joint exchangeability of undirected graph measure). For any CRM $W \sim \text{CRM} \times (\rho, \lambda)$, the point process Z defined by equation (12), or equivalently by equation (6), is jointly exchangeable.

The proof is given in Appendix B. In the adjacency matrix representation, we think of exchangeability as invariance to node orderings. Here, we have invariance to the time of arrival of the nodes, thinking of θ_i as a time index.

We now reformulate our network process in the Kallenberg representation (5). Because of exchangeability, we know that such a representation exists. What we show here is that our CRM-based formulation has an analytic and interpretable representation. In particular, the CRM W can be constructed from a two-dimensional unit rate Poisson process on $\mathbb{R}_+^2$ by using the inverse Lévy method (Khintchine, 1937; Ferguson and Klass, 1972). Let (θ_i, ϑ_i) be a unit rate Poisson process on $\mathbb{R}_+^2$. Let $\bar{\rho}(x)$ be the *tail Lévy intensity*

$$\bar{\rho}(x) = \int_x^\infty \rho(\mathrm{d}w). \tag{15}$$

Then the CRM $W = \Sigma_i w_i \delta_{\theta_i}$ with Lévy measure $\rho(\mathrm{d}w)\mathrm{d}\theta$ can be constructed from the bidimensional point process by taking $w_i = \bar{\rho}^{-1}(\vartheta_i)$. Note that the *inverse Lévy intensity* $\bar{\rho}^{-1}$ is a monotone function. It follows that our undirected graph model can be formulated under representation (5) by selecting any α_0, $\beta_0 = \gamma_0 = 0$, $g = g' = 0$, $h = h' = l = l' = 0$ and

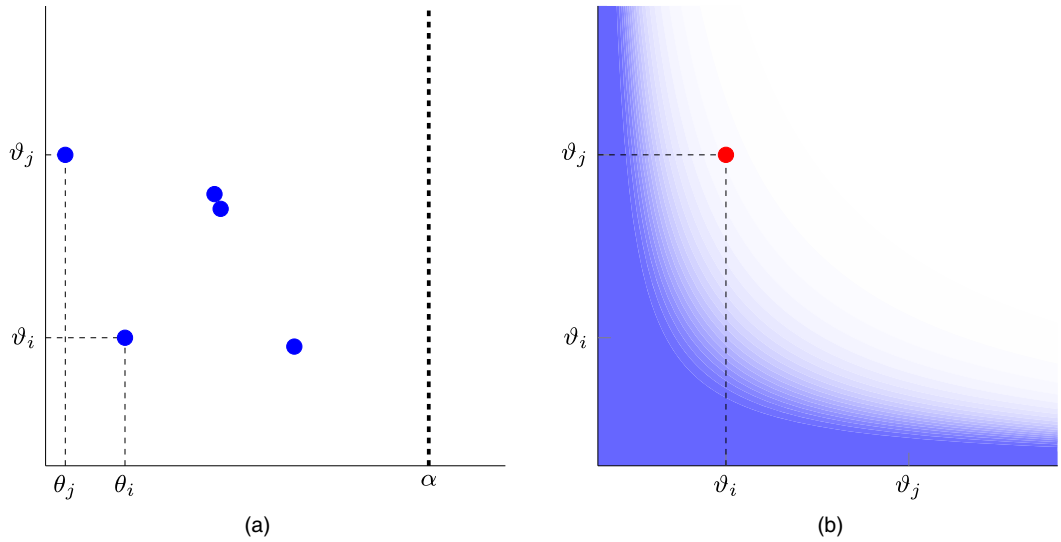


Fig. 5. Illustration of the model construction based on the Kallenberg representation: (a) a unit rate Poisson process (θ_i, ϑ_i), $i \in \mathbb{N}$, on $[0, \alpha] \times \mathbb{R}_+$; (b) for each pair $\{i, j\} \in \mathbb{N}^2$, set $z_{ij} = z_{ji} = 1$ with probability $M(\vartheta_i, \vartheta_j)$ (here, M is indicated by the blue shading (darker shading indicates higher value) for a stable process (GGP with $\tau = 0$); in this case there is an analytic expression for $\bar{\rho}^{-1}$ and therefore M)

$$f(\alpha_0, \vartheta_i, \vartheta_j, \zeta_{\{i,j\}}) = \begin{cases} 1 & \zeta_{\{i,j\}} \leqslant M(\vartheta_i, \vartheta_j), \\ 0 & \text{otherwise} \end{cases} \tag{16}$$

where $M : \mathbb{R}_+^2 \to [0, 1]$ is defined by

$$M(\vartheta_i, \vartheta_j) = \begin{cases} 1 - \exp\{-2\,\bar{\rho}^{-1}(\vartheta_i)\,\bar{\rho}^{-1}(\vartheta_j)\} & \text{if } \vartheta_i \neq \vartheta_j, \\ 1 - \exp\{-\bar{\rho}^{-1}(\vartheta_i)^2\} & \text{if } \vartheta_i = \vartheta_j. \end{cases}$$

In Section 6, we provide explicit forms for $\bar{\rho}$ depending on our choice of Lévy measure ρ. Expression (16) represents a direct analogue to that arising from the Aldous–Hoover framework. In particular, M here is akin to the graphon ω of expression (47) in Appendix A.1, and thus allows us to connect our CRM-based formulation with the extensive literature on graphons. An illustration of the network construction from the Kallenberg representation, including the function M, is in Fig. 5. Note that, if we had started from the Kallenberg representation and selected an f (or M) arbitrarily, we would probably not have obtained a network model with the normalized CRM interpretation that enables both interpretability and analysis of network properties.

For the bipartite graph, Kallenberg's representation theorem for *separate* exchangeability (Kallenberg (1990) and Kallenberg (2005), theorem 9.23) can likewise be applied.

5.2. Interactions between groups

For any disjoint set of nodes $A, B \subset \mathbb{R}_+$, $A \cap B = \emptyset$, the probability that there is at least one connection between a node in A and a node in B is given by

$$\Pr\{Z(A \times B) > 0 | W\} = 1 - \exp\{-2W(A)\,W(B)\},$$

i.e. the probability of a between-group edge depends on the sum of the sociabilities in each group, $W(A)$ and $W(B)$.

5.3. Graph restrictions

Let us consider the restriction of our process to the square $[0, \alpha]^2$. For finite activity CRMs, there will be a finite number of potential nodes (jumps) in the interval $[0, \alpha]$. For infinite activity CRMs, we shall have an infinite number of potential nodes. We are interested in the properties of the process as α grows, where we can think of α as representing *time* and observing the process as new potential nodes and any resulting edges enter the network. We note that, in the limit of $\alpha \to \infty$, the number of edges approaches ∞ since $W(\mathbb{R}_+) = \infty$ almost surely.

Let D_α and Z_α be the restrictions of D and Z-respectively to the square $[0, \alpha]^2$. Then, $(D_\alpha)_{\alpha \geqslant 0}$ and $(Z_\alpha)_{\alpha \geqslant 0}$ are measure-valued stochastic processes, indexed by α. We also denote by W_α and λ_α the corresponding CRM and Lebesgue measure on $[0, \alpha]$. In what follows, our interests are in studying how the following quantities vary with α:

(a) N_α, the number of nodes with degree at least one in the network, and
(b) $N_\alpha^{(e)}$, the number of edges in the undirected network.

We refer to N_α as the number of *observed nodes*. In our construction, recall that $(N_\alpha)_{\alpha \geqslant 0}$ and $(N_\alpha^{(e)})_{\alpha \geqslant 0}$ are non-decreasing, integer-valued stochastic processes corresponding to the number of nodes with at least one connection in Z_α and the number of edges in Z_α respectively. Formally,

$$N_\alpha = \text{card}(\{\theta_i \in [0, \alpha] | Z(\{\theta_i\} \times [0, \alpha]) > 0\}), \tag{17}$$

$$N_\alpha^{(e)} = Z[\{(x, y) \in \mathbb{R}_+^2 | 0 \leqslant x \leqslant y \leqslant \alpha\}]. \tag{18}$$

The two processes have the same jump times, which correspond to the addition of one or more new nodes with at least one connection in the graph. An example of these processes is represented in Fig. 6. In later sections we use $Z_\alpha^* = Z_\alpha([0, \alpha]^2)$ to denote the total mass on $[0, \alpha]^2$, and similarly for D_α^* and W_α^*.

5.4. Sparsity

In this section we state the sparsity properties of our graph model, which relate to the properties of the Lévy measure ρ. In particular, we are interested in the relative asymptotic behaviour of the number of edges $N_\alpha^{(e)}$ with respect to the number of observed nodes N_α as $\alpha \to \infty$. Henceforth, we consider $\int_0^\infty \rho(\mathrm{d}w) > 0$, since the case of $\int_0^\infty \rho(\mathrm{d}w) = 0$ trivially gives $N_\alpha^{(e)} = N_\alpha = 0$ almost surely.

In theorem 2 we characterize the sparsity of the graph with respect to the properties of its Lévy measure: graphs obtained from infinite activity CRMs are sparse, whereas graphs obtained from finite activity CRMs are dense. The rate of growth can be further specified when ρ is a *regularly varying* Lévy measure (Feller, 1971; Karlin, 1967; Gnedin *et al.*, 2006, 2007), as defined in Appendix A.2. We follow the notation of Janson (2011) for probability asymptotics (see Appendix C.1 for details).

Theorem 2. Consider a point process Z representing an undirected graph. Let $N_\alpha^{(e)}$ be the number of edges and N_α be the number of observed nodes in the point process restriction Z_α (see equations (17) and (18)). Assume that the defining Lévy measure is such that $\int_0^\infty w\, \rho(\mathrm{d}w) < \infty$. If the CRM W is *finite activity*, i.e.

$$\int_0^\infty \rho(\mathrm{d}w) < \infty,$$

then the number of edges scales quadratically with the number of observed nodes

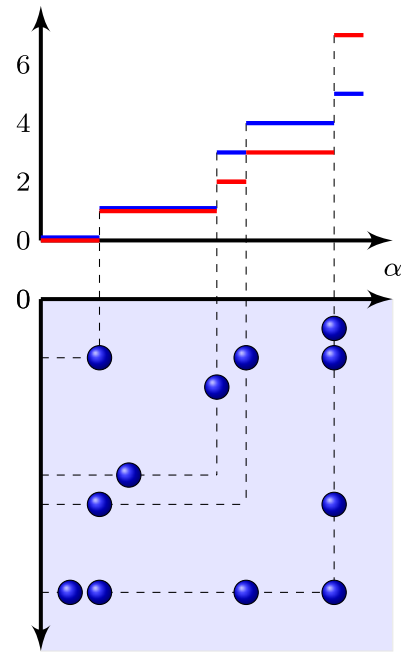


Fig. 6. Example of point process Z and above it the associated integer-valued stochastic processes for the number of observed nodes $(N_\alpha)_{\alpha \geqslant 0}$ (—) and edges $(N_\alpha^{(e)})_{\alpha \geqslant 0}$ (—)

$$N_\alpha^{(e)} = \Theta(N_\alpha^2) \tag{19}$$

almost surely as $\alpha \to \infty$, implying that the graph is *dense*.

If the CRM is *infinite activity*, i.e.

$$\int_0^\infty \rho(\mathrm{d}w) = \infty,$$

then the number of edges scales subquadratically with the number of observed nodes

$$N_\alpha^{(e)} = o(N_\alpha^2) \tag{20}$$

almost surely as $\alpha \to \infty$, implying that the graph is *sparse*.

Furthermore, if the Lévy measure ρ is *regularly varying* (see definition 1 in Appendix A.2), with exponent $\sigma \in (0, 1)$ and slowly varying function l satisfying $\liminf_{t \to \infty} l(t) > 0$, then

$$N_\alpha^{(e)} = O(N_\alpha^{2/1+\sigma}) \qquad \text{almost surley as } \alpha \to \infty. \tag{21}$$

Theorem 2 is a direct consequence of two theorems that we state now and prove in Appendix C. The first theorem states that the number of edges grows quadratically with α, whereas the second states that the number of nodes scales superlinearly with α for infinite activity CRMs, and linearly otherwise.

Theorem 3. Consider the point process Z. If $\int_0^\infty w\,\rho(dw) < \infty$, then the number of edges in Z_α grows quadratically with α:

$$N_\alpha^{(e)} = \Theta(\alpha^2) \tag{22}$$

almost surely. Otherwise, $N_\alpha^{(e)} = \Omega(\alpha^2)$ almost surely.

Theorem 4. Consider the point process Z. Then

$$N_\alpha = \begin{cases} \Theta(\alpha) & \text{if } W \text{ is a finite activity CRM,} \\ \omega(\alpha) & \text{if } W \text{ is an infinite activity CRM} \end{cases} \tag{23}$$

almost surely as $\alpha \to \infty$. In words, the number of nodes with degree at least 1 in Z_α scales linearly with α for finite activity CRMs and superlinearly with α for infinite activity CRMs. Furthermore, for a regularly varying Lévy measure with slowly varying function l such that $\liminf_{t\to\infty} l(t) > 0$, we have

$$N_\alpha = \Omega(\alpha^{\sigma+1}) \qquad \text{almost surely as } \alpha \to \infty. \tag{24}$$

We finally give the expressions of the expectations for the number of edges and nodes in the model. The proof is given in Appendix C.4. (Equations (26) and (27) could alternatively be derived as particular cases of the results in Veitch and Roy (2015).)

Theorem 5. The expected number of edges D_α^* in the multigraph, edges $N_\alpha^{(e)}$ in the undirected graph and observed nodes N_α are given as follows:

$$\mathbb{E}[D_\alpha^*] = \alpha^2 \left\{ \int_0^\infty w\, \rho(\mathrm{d}w) \right\}^2 + \alpha \int_0^\infty w^2 \rho(\mathrm{d}w), \tag{25}$$

$$\mathbb{E}[N_\alpha^{(e)}] = \alpha^2 \int_0^\infty \psi(w)\, \rho(\mathrm{d}w) + \alpha \int_0^\infty \{1 - \exp(-w^2)\} \rho(\mathrm{d}w), \tag{26}$$

$$\mathbb{E}[N_\alpha] = \alpha \int_0^\infty [1 - \exp\{-w^2 - \alpha\psi(2w)\}] \rho(\mathrm{d}w), \tag{27}$$

where $\psi(t) = \int_0^\infty \{1 - \exp(-wt)\} \rho(\mathrm{d}w)$ is the Laplace exponent. Additionally, if ρ is a regularly varying Lévy measure with exponent $\sigma \in [0, 1)$ and slowly varying function l, and $\int_0^\infty w\rho(dw) < \infty$ then

$$\mathbb{E}[N_\alpha] \overset{\alpha\uparrow\infty}{\sim} \alpha^{1+\sigma} l(\alpha)\, \Gamma(1-\sigma) \left\{ 2 \int_0^\infty w\, \rho(\mathrm{d}w) \right\}^\sigma. \tag{28}$$

5.5. Simulation

5.5.1. *Direct simulation of graph restrictions*

By definition, the directed multigraph restriction D_α is drawn from a Poisson process with finite mean measure $W_\alpha \times W_\alpha$, where $W_\alpha \sim \mathrm{CRM}(\rho, \lambda_\alpha)$. Leveraging standard properties of the CRM and Poisson process, we can first simulate the total number of directed edges D_α^* based on the total mass W_α^*:

$$D_\alpha^* | W_\alpha^* \sim \mathrm{Poisson}(W_\alpha^{*\,2}). \tag{29}$$

For $k = 1, \ldots, D_\alpha^*$ a particular edge is drawn by sampling a pair of nodes

$$U_{kj} | W_\alpha \overset{\mathrm{IID}}{\sim} \frac{W_\alpha}{W_\alpha^*} \qquad j = 1, 2, \tag{30}$$

where W_α / W_α^* is called a *normalized CRM*. We form directed edges (U_{k1}, U_{k2}), resulting in

$$D_\alpha = \sum_{k=1}^{D_\alpha^*} \delta_{(U_{k1}, U_{k2})}. \tag{31}$$

Because of the discreteness of W_α, there will be ties between the (U_{k1}, U_{k2}), and the number of such ties corresponds to the multiplicity of that edge. In particular, a total of $2D^*_\alpha$ nodes U_{kj} are drawn but result in some $N_\alpha \leqslant 2D^*_\alpha$ distinct values. We overload the notation N_α here because this quantity also corresponds to the number of nodes with degree at least 1 in the resulting undirected network. Recall that the undirected network construction simply forms an undirected edge between a set of nodes if there is at least one directed edge between them. If we consider unordered pairs $\{U_{k1}, U_{k2}\}$, the number of such unique pairs takes a number $N^{(e)}_\alpha \leqslant D^*_\alpha$ of distinct values, where $N^{(e)}_\alpha$ corresponds to the number of edges in the undirected network.

The construction above enables us to re-express our Cox process model in terms of normalized CRMs (Regazzini *et al.*, 2003). This is very attractive both practically and theoretically. As we show in Section 6 for special cases of CRMs, one can use the results surrounding normalized CRMs to derive an exact simulation technique for our directed and undirected graphs.

Remark 1. The construction above enables us to draw connections with the configuration model (Bollobás, 1980; Newman, 2010), which proceeds as follows. First, the degree k_i of each node $i = 1, \ldots, n$ is specified such that the sum of k_i is an odd number. Each node i is given a total of k_i stubs, or *demi edges*. Then, we repeatedly choose pairs of stubs uniformly at random, without replacement, and connect the selected pairs to form an edge. The simple graph is obtained either by discarding the multiple edges and self-loops (an *erased* configuration model), or by repeating the above sampling until obtaining a simple graph. In our case, we have an infinite set of (potential) nodes and do not prespecify the node degrees. Furthermore, each node in the pair (U_{k1}, U_{k2}) is drawn from a normalized CRM rather than the pair being selected uniformly at random. However, at a high level, there is a similar flavour to our construction.

5.5.2. *Urn-based simulation of graph restrictions*

We now describe an urn formulation that allows us to obtain a finite dimensional generative process. Recall that, in practice, we cannot sample $W_\alpha \sim CRM(\rho, \lambda_\alpha)$ if the CRM is infinite activity since there will be an infinite number of jumps.

Let $(U'_1, \ldots, U'_{2D^*_\alpha}) = (U_{11}, U_{12}, \ldots, U_{D^*_\alpha 1}, U_{D^*_\alpha 2})$. For some classes of Lévy measure ρ, it is possible to integrate out the normalized CRM $\mu_\alpha = W_\alpha / W^*_\alpha$ in expression (30) and derive the conditional distribution of U'_{n+1} given $(W^*_\alpha, U'_1, \ldots, U'_n)$. We first recall some background on random partitions. As μ_α is discrete with probability 1, variables $U'_1, \ldots, U'_n$ take $k \leqslant n$ distinct values $\tilde{\theta}_j$, with multiplicities $1 \leqslant \tilde{m}_j \leqslant n$. The distribution on the underlying partition is usually defined in terms of an exchangeable partition probability function (EPPF) (Pitman, 1995) $\Pi^{(k)}_n(\tilde{m}_1, \ldots, \tilde{m}_k | W^*_\alpha)$ which is symmetric in its arguments. The predictive distribution of U'_{n+1} given $(W^*_\alpha, U'_1, \ldots, U'_n)$ is then given in terms of the EPPF:

$$U'_{n+1} | (W^*_\alpha, U'_1, \ldots, U'_n) \sim \frac{\Pi^{(k+1)}_{n+1}(\tilde{m}_1, \ldots, \tilde{m}_k, 1 | W^*_\alpha)}{\Pi^{(k)}_n(\tilde{m}_1, \ldots, \tilde{m}_k | W^*_\alpha)} \frac{1}{\alpha} \lambda_\alpha + \sum_{j=1}^{k} \frac{\Pi^{(k)}_{n+1}(\tilde{m}_1, \ldots, \tilde{m}_j + 1, \ldots, \tilde{m}_k | W^*_\alpha)}{\Pi^{(k)}_n(\tilde{m}_1, \ldots, \tilde{m}_k | W^*_\alpha)} \delta_{\tilde{\theta}_j}. \tag{32}$$

Using this urn representation, we can rewrite our generative process as

$$\begin{aligned} W^*_\alpha &\sim P_{W^*_\alpha}, \\ D^*_\alpha | W^*_\alpha &\sim \text{Poisson}(W^{*\,2}_\alpha), \\ (U_{kj})_{k=1,\ldots,D^*_\alpha; j=1,2} | W^*_\alpha &\sim \text{urn process (32)}, \end{aligned}$$

$$D_\alpha = \sum_{k=1}^{D_\alpha^*} \delta_{(U_{k1}, U_{k2})}, \tag{33}$$

where $P_{W_\alpha^*}$ is the distribution of the CRM total mass W_α^*. Representation (33) can be used to sample *exactly* from our graph model, assuming that we can sample from $P_{W_\alpha^*}$ and evaluate the EPPF. In Section 6 we show that this is indeed possible for specific CRMs of interest.

5.5.3. *Approximate simulation of graph restrictions*

If we cannot sample from $P_{W_\alpha^*}$ in expression (33) and evaluate the EPPF in expression (32), we resort to approximate simulation methods. In particular, we harness the directed multigraph representation and approximate the draw of W_α. For our undirected graphs, we simply transform the (approximate) draw of a directed multigraph as described in Section 3.3.

One approach to approximate simulation of W_α, which is possible for some Lévy measures ρ, is to resort to adaptive thinning (Lewis and Shedler, 1979; Ogata, 1981; Favaro and Teh, 2013). A related alternative approximate approach, but applicable to any Lévy measure ρ satisfying condition (9), is the inverse Lévy method. This method first defines a threshold ε and then samples the weights $\Omega = \{w_i | w_i > \varepsilon\}$ by using a Poisson measure on $[\varepsilon, \infty]$. One then simulates D_α using these truncated weights Ω.

A naive application of this truncated method that considers sampling directed or undirected edges as in expression (12) or expression (6), respectively can prove computationally problematic since a large number of possible edges must be considered (one Poisson or Bernoulli draw for each (θ_i, θ_j) pair for the directed or undirected case). Instead, we can harness the Cox process representation and resulting sampling procedure of expression (29)–(30) to sample first the total number of directed edges and then their specific instantiations. More specifically, to simulate approximately a point process on $[0, \alpha]^2$, we use the inverse Lévy method to sample

$$\Pi_{\alpha,\varepsilon} = \{(w, \theta) \in \Pi, 0 < \theta \leqslant \alpha, w > \varepsilon\}. \tag{34}$$

Let $W_{\alpha,\varepsilon} = \Sigma_{i=1}^K w_i \delta_{\theta_i}$ be the associated truncated CRM and $W_{\alpha,\varepsilon}^* = W_{\alpha,\varepsilon}([0, \alpha])$ its total mass. We then sample $D_{\alpha,\varepsilon}^*$ and U_{kj} as in expression (29)–(30), and set $D_{\alpha,\varepsilon} = \Sigma_{k=1}^{D_{\alpha,\varepsilon}^*} \delta_{(U_{k1}, U_{k2})}$. The undirected graph measure $Z_{\alpha,\varepsilon}$ is set to the manipulation of $D_{\alpha,\varepsilon}$ as in expression (12).

6. Special cases

In this section, we examine the properties of various models and their link to classical random-graph models depending on the Lévy measure ρ. We show that, in the GGP case, the resulting graph can be either dense or sparse, with the sparsity tuned by a single hyperparameter. Furthermore, exact simulation is possible via expression (33). We focus on the undirected graph case, but similar results can be obtained for directed multigraphs and bipartite graphs.

6.1. Poisson process

Consider a Poisson process with fixed increments $w_0 > 0$:

$$\rho(\mathrm{d}w) = \delta_{w_0}(\mathrm{d}w).$$

This measure ρ defines a finite activity CRM. Recalling the definition $\bar{\rho}(x) = \int_x^\infty \rho(\mathrm{d}w)$, in this case, we have

$$\bar{\rho}(x) = \begin{cases} 1 & \text{if } x < w_0, \\ 0 & \text{otherwise.} \end{cases}$$

Ignoring self-edges, the graph construction can be described as follows. To sample $W_\alpha \sim \text{CRM}(\rho, \lambda_\alpha)$, we generate $n \sim \text{Poisson}(\alpha)$ and then sample $\theta_i \sim \text{Unif}([0, \alpha])$ for $i = 1, \ldots, n$. We then sample edges according to expression (6): for $0 < i < j < n$, set $z_{ij} = z_{ji} = 1$ with probability $1 - \exp(-2w_0^2)$ and 0 otherwise. The model is therefore equivalent to the Erdös–Rényi random-graph model $G(n, p)$ with $n \sim \text{Poisson}(\alpha)$ and $p = 1 - \exp(-2w_0^2)$. Therefore, this choice of ρ leads to a *dense* graph, as our theory suggests, where the number of edges grows quadratically with the number of nodes n.

6.2. Compound Poisson process

A compound Poisson process is a process where

$$\rho(\mathrm{d}w) = h(w)\,\mathrm{d}w$$

and $h : \mathbb{R}_+ \to \mathbb{R}_+$ is such that $\int_0^\infty h(w)\,\mathrm{d}w = 1$ and defines a finite activity CRM. In this case, we have $\bar{\rho}(x) = 1 - H(x)$ where H is the distribution function that is associated with h. Here, we arrive at a framework that is similar to the standard graphon. Leveraging the Kallenberg representation (16), we first sample $n \sim \mathit{Poisson}(\alpha)$. Then, for $i = 1, \ldots, n$ we set $z_{ij} = z_{ji} = 1$ with probability $M(U_i, U_j)$ where U_i are uniform $[0, 1]$ variables and M is defined by

$$M(U_i, U_j) = 1 - \exp\{-2H^{-1}(U_i)H^{-1}(U_j)\}.$$

This representation is the same as with the Aldous–Hoover theorem, except that the number of nodes is random and follows a Poisson distribution. As such, the resulting random graph is either trivially empty or *dense*, again agreeing with our theory.

6.3. Generalized gamma process

The GGP (Hougaard, 1986; Aalen, 1992; Lee and Whitmore, 1993; Brix, 1999) is a flexible two-parameter CRM with interpretable parameters and remarkable conjugacy properties (James, 2002; Lijoi and Prünster, 2003; Lijoi *et al.*, 2007; Caron *et al.*, 2014). The process is also known as the Hougaard process (Hougaard, 1986) when λ is the Lebesgue measure, as in this paper, but we shall use the more standard term GGP in the rest of this paper. The Lévy measure of the GGP is given by

$$\rho(\mathrm{d}w) = \frac{1}{\Gamma(1-\sigma)} w^{-1-\sigma} \exp(-\tau w)\mathrm{d}w, \tag{35}$$

where the two parameters (σ, τ) satisfy

$$(\sigma, \tau) \in (-\infty, 0] \times (0, +\infty) \quad \text{or} \quad (\sigma, \tau) \in (0, 1) \times [0, +\infty). \tag{36}$$

The GGP has different properties if $\sigma \geqslant 0$ or $\sigma < 0$. When $\sigma < 0$, the GGP is a finite activity CRM (i.e. a compound Poisson process); more precisely, the number of jumps in $[0, \alpha]$ is finite with probability 1 and drawn from a Poisson distribution with rate $-(\alpha/\sigma)\tau^\sigma$ whereas the jumps w_i are IID gamma$(-\sigma, \tau)$.

When $\sigma \geqslant 0$, the GGP has an infinite number of jumps over any interval $[s, t]$. It includes as special cases the gamma process ($\sigma = 0$, $\tau > 0$), the stable process ($\sigma \in (0, 1)$, $\tau = 0$) and the inverse Gaussian process ($\sigma = \frac{1}{2}$, $\tau > 0$).

The tail Lévy intensity of the GGP is given by

$$\bar{\rho}(x) = \int_x^\infty \frac{1}{\Gamma(1-\sigma)} w^{-1-\sigma} \exp(-\tau w)\,\mathrm{d}w = \begin{cases} \dfrac{\tau^\sigma \Gamma(-\sigma, \tau x)}{\Gamma(1-\sigma)} & \text{if } \tau > 0, \\ \dfrac{x^{-\sigma}}{\Gamma(1-\sigma)\sigma} & \text{if } \tau = 0, \end{cases}$$

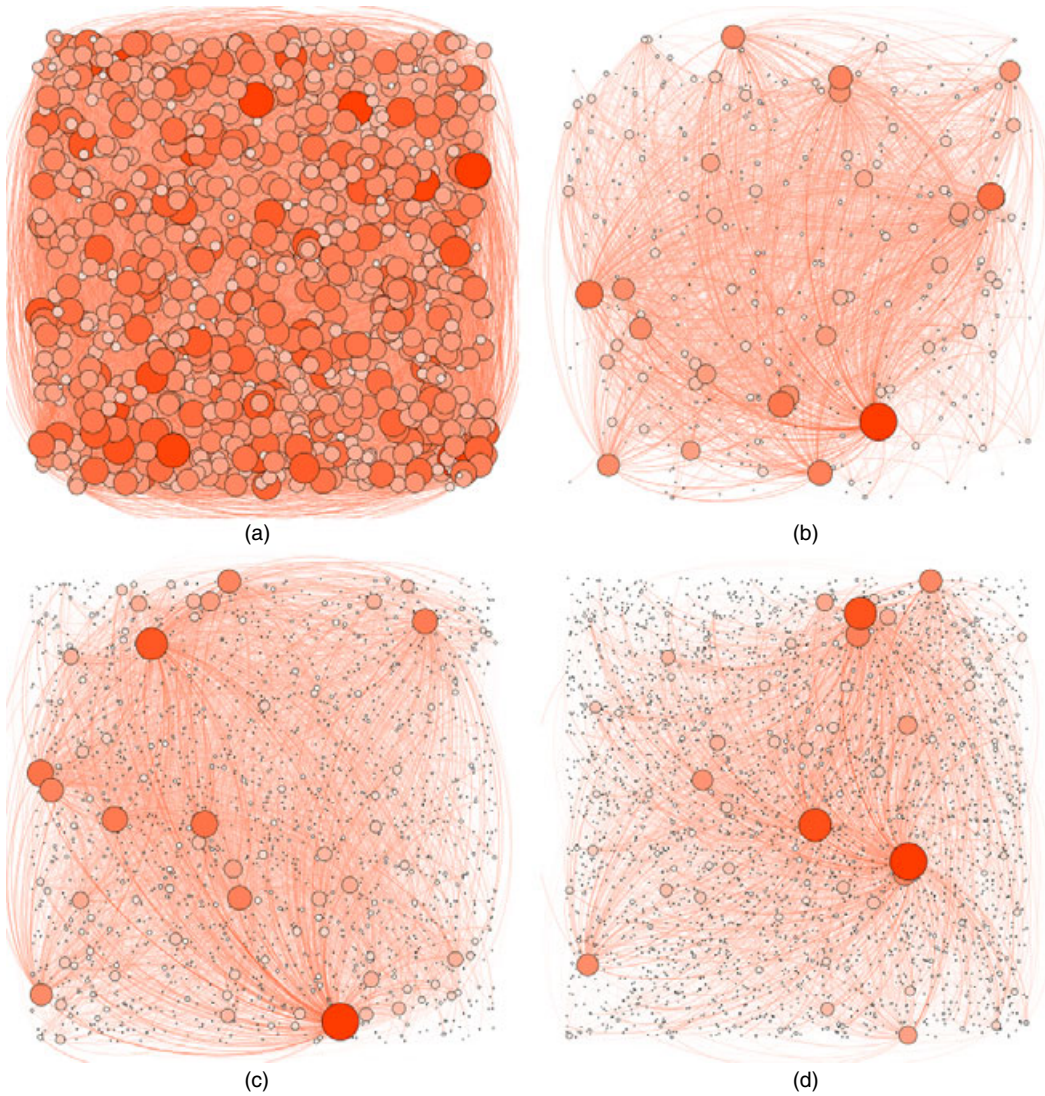


Fig. 7. Sample graphs: (a) Erdös–Rényi graph $G(n,p)$ with $n=1000$ and $p=0.05$, and GGP graphs GGP(α,τ,σ) with (b)-(d) $\alpha=100$, $\tau=2$ and (b) $\sigma=0$, (c) $\sigma=0.5$ and (d) $\sigma=0.8$ (the size of a node is proportional to its degree; the graphs were generated with the software Gephi (Bastian *et al.*, 2009))

where $\Gamma(a,x)$ is the incomplete gamma function. Example realizations of the process for various values of $\sigma \geqslant 0$ are displayed in Fig. 7 alongside a realization of an Erdös–Rényi graph.

6.3.1. Exact sampling via an urn approach

In the case $\sigma > 0$, W_α^* is an exponentially tilted stable random variable, for which exact samplers exist (Devroye, 2009). As shown by Pitman (2003) (see also Lijoi *et al.* (2008)), the EPPF conditional on the total mass $W_\alpha^* = t$ depends only on the parameter σ (and not τ and α) and is given by

$$\Pi_k^{(n)}(\tilde{m}_1,\ldots,\tilde{m}_k|t) = \frac{\sigma^k t^{-n}}{\Gamma(n-k\sigma)g_\sigma(t)} \int_0^t s^{n-k\sigma-1} g_\sigma(t-s)\,\mathrm{d}s \prod_{i=1}^k \frac{\Gamma(\tilde{m}_i-\sigma)}{\Gamma(1-\sigma)}, \tag{37}$$

where g_σ is the probability density function of the positive stable distribution. Plugging the EPPF (37) into expression (32) yields the urn process for sampling in the GGP case. In particular, we can use the generative process (33) to sample exactly from the model.

In the special case of the gamma process ($\sigma = 0$), W_α^* is a gamma(α, τ) random variable and the resulting urn process is given by (Blackwell and MacQueen, 1973; Pitman, 1996):

$$U'_{n+1}|(W_\alpha^*, U'_1, \ldots, U'_n) \sim \frac{\alpha}{\alpha+n}\frac{\lambda_\alpha}{\alpha} + \sum_{j=1}^{k}\frac{\tilde{m}_j}{\alpha+n}\delta_{\tilde{\theta}_j}. \tag{38}$$

When $\sigma < 0$, the GGP is a compound Poisson process and can thus be sampled exactly.

6.3.2. *Sparsity*

Appealing to theorem 2, we use the following facts about the GGP to characterize the sparsity properties of this special case.

(a) For $\sigma < 0$, the CRM is *finite activity* with $\int_0^\infty w\rho(\mathrm{d}w) < \infty$; thus theorem 2 implies that the graph is *dense*.
(b) When $\sigma \geqslant 0$ the CRM is *infinite activity*; moreover, for $\tau > 0$, $\int_0^\infty w\rho(\mathrm{d}w) < \infty$, and thus theorem 2 implies that the graph is *sparse*.
(c) For $\sigma > 0$, the tail Lévy intensity has the asymptotic behaviour

$$\bar{\rho}(x) \overset{x\downarrow 0}{\sim} \frac{1}{\sigma\,\Gamma(1-\sigma)}x^{-\sigma}$$

and, as such, is *regularly varying* with exponent σ and constant slowly varying function.

We thus conclude that

$$N_\alpha^{(e)} = \begin{cases} \Theta(N_\alpha^2) & \text{if } \sigma < 0, \\ o(N_\alpha^2) & \text{if } \sigma = 0, \tau > 0, \\ O(N_\alpha^{2/(1+\sigma)}) & \text{if } \sigma \in (0,1), \tau > 0, \end{cases} \tag{39}$$

almost surely as $\alpha \to \infty$, i.e. the GGP parameter σ tunes the sparsity of the graph: The underlying graph is sparse if $\sigma \geqslant 0$ and dense otherwise.

Remark 2. The proof technique of theorem 2 requires $\int_0^\infty w\,\rho(\mathrm{d}w) < \infty$ and thus excludes the stable process ($\tau = 0, \sigma \in (0,1)$), although we conjecture that the graph is also sparse in that case.

Additionally, applying theorem 5, we obtain

$$\mathbb{E}[N_\alpha] \overset{\alpha\uparrow\infty}{\sim} \begin{cases} \alpha\dfrac{-\tau^\sigma}{\sigma} & \text{if } \sigma < 0, \\ \alpha\log(\alpha) & \text{if } \sigma = 0, \\ \alpha^{1+\sigma}\dfrac{2^\sigma\tau^{\sigma(\sigma-1)}}{\sigma} & \text{if } \sigma > 0, \tau > 0. \end{cases}$$

6.3.3. *Empirical analysis of graph properties*

For the GGP-based formulation, we provide an empirical analysis of our network properties in Fig. 8 by simulating undirected graphs by using the approach that was described in Section 5.5 for various values of σ and τ. We compare with an Erdös–Rényi random graph, preferential attachment (Barabási and Albert, 1999) and the Bayesian non-parametric network model of Lloyd *et al.* (2012). The particular features that we explore are as follows.

(a) *Degree distribution*: Fig. 8(a) suggests empirically that the model can exhibit power law

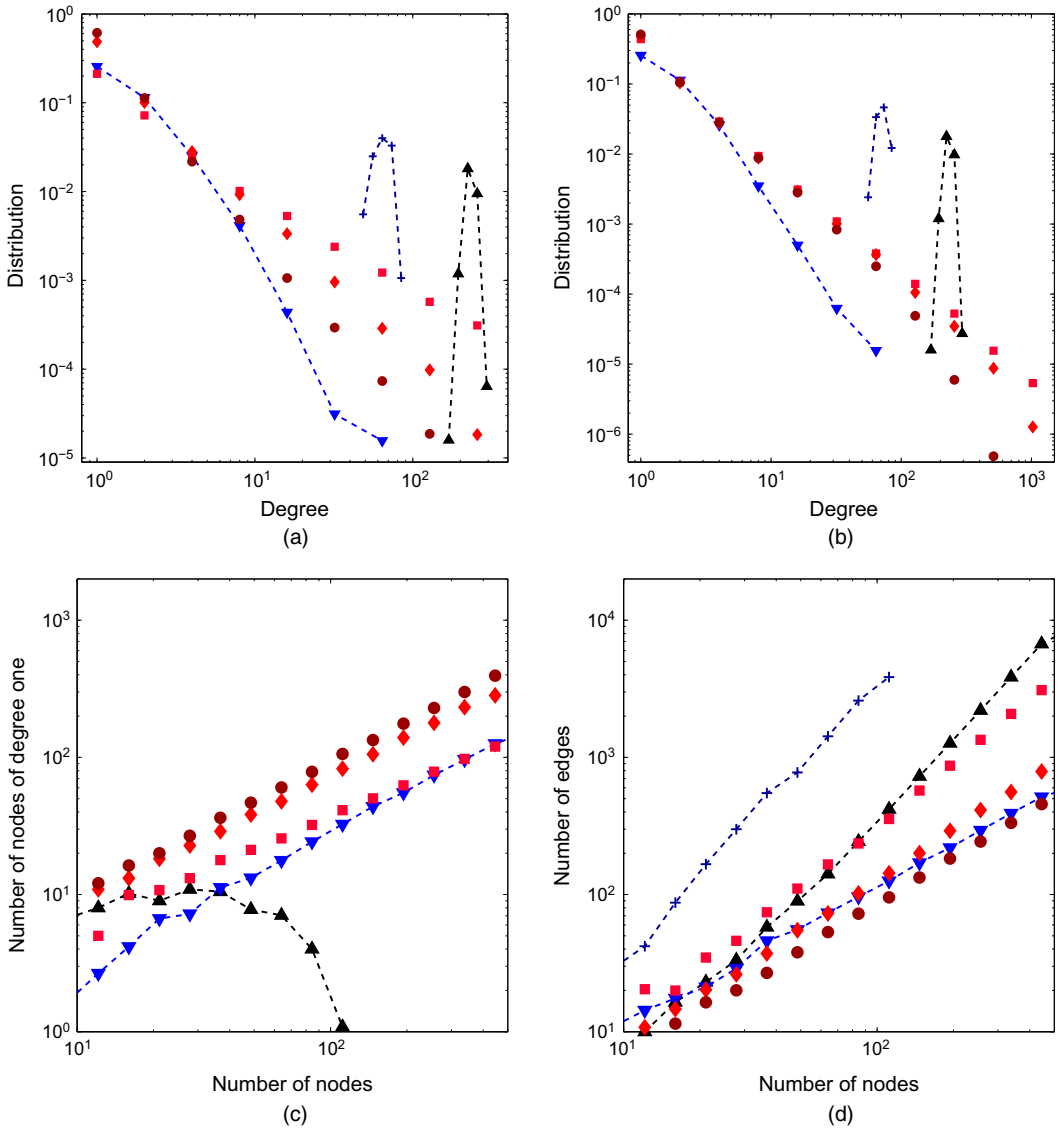


Fig. 8. Examination of the GGP undirected network properties (averaging over graphs with various α) in comparison with an Erdös–Rényi $G(n,p)$ model with $p=0.05$ (▲), the preferential attachment model of Barabási and Albert (1999) (▼) and the non-parametric formulation of Lloyd *et al.* (2012) (✚): (a) degree distribution on a log–log-scale for (a) various values of σ (■, $\sigma=0.2$; ♦, $\sigma=0.5$; ●, $\sigma=0.8$) ($\tau=10^{-2}$) and (b) various values of τ (■, $\tau=10^{-1}$; ♦, $\tau=1$; ●, $\tau=5$) ($\sigma=0.5$) for the GGP; (c) number of nodes with degree 1 *versus* number of nodes on a log–log-scale (■, $\sigma=0.2$; ♦, $\sigma=0.5$; ●, $\sigma=0.8$) (note that the Lloyd method leads to dense graphs such that no node has only degree 1); (d) number of edges *versus* number of nodes (■, $\sigma=0.2$; ♦, $\sigma=0.5$; ●, $\sigma=0.8$) (here we note growth at a rate $o(n^2)$ for our GGP graph models, and $\Theta(n^2)$ for the Erdös-Rényi and Lloyd models (dense graphs))

behaviour providing a heavy-tailed degree distribution. As shown in Fig. 8(b), the model can also handle an exponential cut-off in the tails of the degree distribution, which is an attractive property (Clauset *et al.*, 2009; Olhede and Wolfe, 2012).

(b) *Number of degree 1 nodes*: Fig. 8(c) examines the fraction of degree 1 nodes *versus* the number of nodes.

(c) *Sparsity*: Fig. 8(d) plots number of edges *versus* number of nodes. The larger σ, the sparser the graph is. In particular, for the GGP random-graph model, we have network growth at a rate $O(n^a)$ for $1 < a < 2$ whereas the Erdös–Rényi (dense) graph grows as $\Theta(n^2)$.

6.3.4. Interpretation of hyperparameters

On the basis of the properties derived and illustrated empirically in this section, we see that our hyperparameters have the following interpretations.

(a) $\boldsymbol{\sigma}$—from Figs 8(a) and (d), σ relates to the slope of the degree distribution in its power law regime and the overall network sparsity. Increasing σ leads to higher power law exponent and sparser networks.

(b) $\boldsymbol{\alpha}$—from theorem 5, α provides an overall scale that affects the number of nodes and directed interactions, with larger α leading to larger networks.

(c) $\boldsymbol{\tau}$—from Fig. 8(b), τ determines the exponential decay of the tails of the degree distribution, with τ small looking like pure power law. This is intuitive from the form of $\rho(\mathrm{d}w)$ in equation (35), where we see that τ affects large weights more than small weights.

7. Posterior characterization and inference

In this section, we consider the posterior characterization and MCMC inference of parameters and hyperparameters in our statistical network models.

Assume that we have observed a set of undirected connections $(z_{ij})_{1\leqslant i,j\leqslant N_\alpha}$ or directed connections $(n_{ij})_{1\leqslant i,j\leqslant N_\alpha}$ where N_α is the observed number of nodes with at least one connection. Without loss of generality, we assume that the locations of these nodes $0 < \theta_1 < \ldots < \theta_{N_\alpha} < \alpha$ are ordered, and we write $w_i = W(\{\theta_i\})$ as their associated sociability parameters. For simplicity, we are overloading notation here with the *unordered* nodes in $W = \sum_i w_i \delta_{\theta_i}$ of equation (7).

We aim to infer the sociability parameters w_i, $i = 1, \ldots, N_\alpha$, for each of the observed nodes. We also aim to infer the sociability parameters of the nodes with no connections (the difference between the set of potential nodes and those with observed interactions). We refer to these as *unobserved nodes*. Under our framework, the number of such nodes is either finite but unknown or infinite. The observed connections, however, provide information about only the sum of their sociabilities, denoted w_*. The node locations θ_i of both observed and unobserved nodes are also not likelihood identifiable and are thus ignored. We additionally aim to estimate α and the hyperparameters of the Lévy intensity ρ of the CRM; we write ϕ for the set of hyperparameters. We therefore aim to approximate the posterior $p(w_1, \ldots, w_{N_\alpha}, w_*, \phi | (z_{ij})_{1<i,j<N_\alpha})$ for an observed undirected graph and $p(w_1, \ldots, w_{N_\alpha}, w_*, \phi | (n_{ij})_{1<i,j<N_\alpha})$ for an observed directed graph. (Formally, this density is with respect to a product measure that has a Dirac mass at 0 for w_*, as detailed in Appendix F.)

7.1. Directed multigraph posterior

In theorem 6, we characterize the posterior in the directed multigraph case. This plays a key role in the undirected case that is explored in Section 7.2 as well.

Theorem 6. For $N_\alpha \geqslant 1$, let $\theta_1 < \ldots < \theta_{N_\alpha}$ be the set of support points of the measure D_α such that $D_\alpha = \Sigma_{1 \leqslant i,\, j \leqslant N_\alpha} n_{ij} \delta_{(\theta_i, \theta_j)}$. Let $w_i = W_\alpha(\{\theta_i\})$ and $w_* = W_\alpha^* - \Sigma_{i=1}^{N_\alpha} w_i$. We have

$$\begin{aligned} &\mathbb{P}\{(w_i \in dw_i)_{1 \leqslant i \leqslant N_\alpha}, w_* \in \mathrm{d}w_* | (n_{ij})_{1 \leqslant i,j \leqslant N_\alpha}, \phi) \\ &\quad \propto \exp\left\{ - \left(\sum_{i=1}^{N_\alpha} w_i + w_* \right)^2 \right\} \left\{ \prod_{i=1}^{N_\alpha} w_i^{m_i} \rho(dw_i) \right\} G_\alpha^*(dw_*) \end{aligned} \tag{40}$$

where $m_i = \Sigma_{j=1}^{N_\alpha} (n_{ij} + n_{ji}) > 0$ for $i = 1, \ldots, N_\alpha$ are the node degrees of the multigraph and G_α^* is the probability distribution of the random variable W_α^*, with Laplace transform

$$\mathbb{E}[\exp(-tW_\alpha^*)] = \exp\{-\alpha\psi(t)\}. \tag{41}$$

Additionally, conditionally on observing an empty graph, i.e. $N_\alpha = 0$, we have

$$\mathbb{P}(w_* \in dw_* | N_\alpha = 0, \phi) \propto \exp(-w_*^2)\, G_\alpha^*(dw_*). \tag{42}$$

The proof builds on posterior characterizations for normalized CRMs (James, 2002, 2005; Prünster, 2002; Pitman, 2003; James *et al.*, 2009) using the hierarchical construction of expression (29)–(30). See Appendix E.

The conditional distribution of $(w_1, \ldots, w_{N_\alpha}, w_*)$ given $(n_{ij})_{1 \leqslant i,j \leqslant N_\alpha}$ does not depend on the locations $(\theta_1, \ldots, \theta_{N_\alpha})$ because we considered a homogeneous CRM. This fact is important since the locations $(\theta_1, \ldots, \theta_{N_\alpha})$ are typically not observed, and our algorithm outlined below will not consider these terms in the inference.

7.2. Markov chain Monte Carlo Sampling for generalized gamma process based directed and undirected graphs

We now specialize to the case of the GGP, for which we derive an MCMC sampler for posterior inference. Let $\phi = (\alpha, \sigma, \tau)$ be the set of hyperparameters that we also want to estimate. We assume improper priors on the hyperparameters:

$$\left.\begin{array}{ll} p(\alpha) & \propto 1/\alpha, \\ p(\sigma) & \propto 1/(1-\sigma), \\ p(\tau) & \propto 1/\tau. \end{array}\right\} \tag{43}$$

To emphasize the dependence on the hyperparamters of the Lévy measure and distribution of the total mass w_*, we write $\rho(w|\sigma, \tau)$ and $G_{\alpha,\sigma,\tau}^*(\mathrm{d}w_*)$.

In the case of an undirected graph, we simply impute the missing directed edges in the graph. For each $i \leqslant j$ such that $z_{ij} = 1$, we introduce latent variables $\bar{n}_{ij} = n_{ij} + n_{ji}$ with conditional distribution

$$\bar{n}_{ij} | z, w \sim \left\{ \begin{array}{ll} \delta_0 & \text{if } z_{ij} = 0, \\ \mathrm{tPoisson}(2w_i w_j) & \text{if } z_{ij} = 1, i \neq j, \end{array} \right. \tag{44}$$

and $n_{ii} | z_{ii} = 1, w_i \sim \mathrm{tPoisson}(w_i^2)$, where $\mathrm{tPoisson}(\lambda)$ is the zero-truncated Poisson distribution with probability mass function

$$\frac{\lambda^k \exp(-\lambda)}{\{1 - \exp(-\lambda)\} k!}, \qquad \text{for } k = 1, 2, \ldots.$$

By convention, we set $\bar{n}_{ij} = \bar{n}_{ji}$ for $j < i$ and $m_i = \Sigma_{j=1}^{N_\alpha} \bar{n}_{ij}$.

For scalable exploration of the target posterior, we propose to use HMC (Duane *et al.*, 1987; Neal, 2011) within Gibbs sampling to update the weights $(w_1, \ldots, w_{N_\alpha})$. The HMC step requires

computation of the gradient of the log-posterior, which in our case, letting $\omega_i = \log(w_i)$, is given by

$$[\nabla_{\omega_{1:N_\alpha}} \log\{p(\omega_{1:N_\alpha}, w_*|D_\alpha)\}]_i = m_i - \sigma - w_i\left(\tau + 2\sum_{j=1}^{N_\alpha} w_j + 2w_*\right). \tag{45}$$

For the update of the total mass w_* and hyperparameters ϕ, we use a Metropolis–Hastings step. Unless $\sigma = 0$ or $\sigma = \frac{1}{2}$, $G^*_{\alpha,\sigma,\tau}(\mathrm{d}w_*)$ does not admit any tractable analytical expression. We therefore use a specific proposal for w_* based on exponential tilting of $G^*_{\alpha,\sigma,\tau}$ that alleviates the need to evaluate this probability density function in the Metropolis–Hasting ratio (see the details in Appendix F). To summarize, the MCMC sampler is defined as follows.

Step 1: update the weights $(w_1, \ldots, w_{N_\alpha})$ given the rest by using an HMC update.
Step 2: update the total mass w_* and hyperparameters $\phi = (\alpha, \sigma, \tau)$ given the rest by using a Metropolis–Hastings update.
Step 3: (undirected graph) update the latent counts $(\bar{n}_{ij})$ given the rest by using the conditional distribution (44) or a Metropolis–Hastings update.

The computational bottlenecks lie in steps 1 and 3, which roughly scale linearly in the number of nodes and edges respectively, although one can parallelize step 3 over edges. If L is the number of leapfrog steps in the HMC algorithm and n_{iter} the number of MCMC iterations, the overall complexity is in $O\{n_{\text{iter}}(LN_\alpha + N^{(e)}_\alpha)\}$. We show in Section8 that the algorithm scales well to large networks with hundreds of thousands of nodes and edges. To scale the HMC algorithm to even larger collections of nodes of edges, one could explore the methods of Chen *et al.* (2014).

8. Experiments

8.1. Simulated data

We first study the convergence of the MCMC algorithm on simulated data where the graph is simulated from our model. We simulated a GGP undirected graph with parameters $\alpha = 300$, $\sigma = 0.5$ and $\tau = 1$, which places us in the sparse regime. The sampled graph resulted in 13 995 nodes and 76 605 edges. We ran three MCMC chains each with 40 000 iterations and with different initial values and $L = 10$ leapfrog steps; the step size of the leapfrog algorithm was adapted during the first 10 000 iterations to obtain an acceptance rate of 0.6. Standard deviations of the random-walk Metropolis–Hastings steps for $\log(\tau)$ and $\log(1 - \sigma)$ were set to 0.02. The computing time for running the three chains successively was 10 min using a MATLAB implementation on a standard computer (central processor unit at 3.10 GHz; four cores). Trace plots of the parameters α, σ, τ and w_* are given in Fig. 9. We computed the potential scale factor reduction (Brooks and Gelman, 1998; Gelman *et al.*, 2014) for all 13 999 parameters ($w_{1:N_\alpha}, w_*, \alpha, \sigma$ and τ) and found a maximum value of 1.01, suggesting convergence of the algorithm. This is quite remarkable as the MCMC sampler actually samples from a target distribution of dimension $13\,995 + 76\,605 + 4 = 90\,604$. Posterior credible intervals of the sociability parameters w_i of the nodes with highest degrees and log-sociability parameters $\log(w_i)$ of the nodes with lowest degrees are displayed in Figs 10(a) and 10(b) respectively, showing the ability of the method to recover sociability parameters of both low and high degree nodes accurately.

To show the versatility of the GGP graph model, we now examine our approach when the observed graph is actually generated from an Erdös–Rényi model with $n = 1000$ and $p = 0.01$. The generated graph had 1000 nodes and 5058 edges. We ran three MCMC chains with the same specifications as above. In this dense graph regime, the following transformation of our parameters α, σ and τ is more informative: $\varsigma_1 = -(\alpha/\sigma)\tau^\sigma$, $\varsigma_2 = -\sigma/\tau$ and $\varsigma_3 = -\sigma/\tau^2$. When

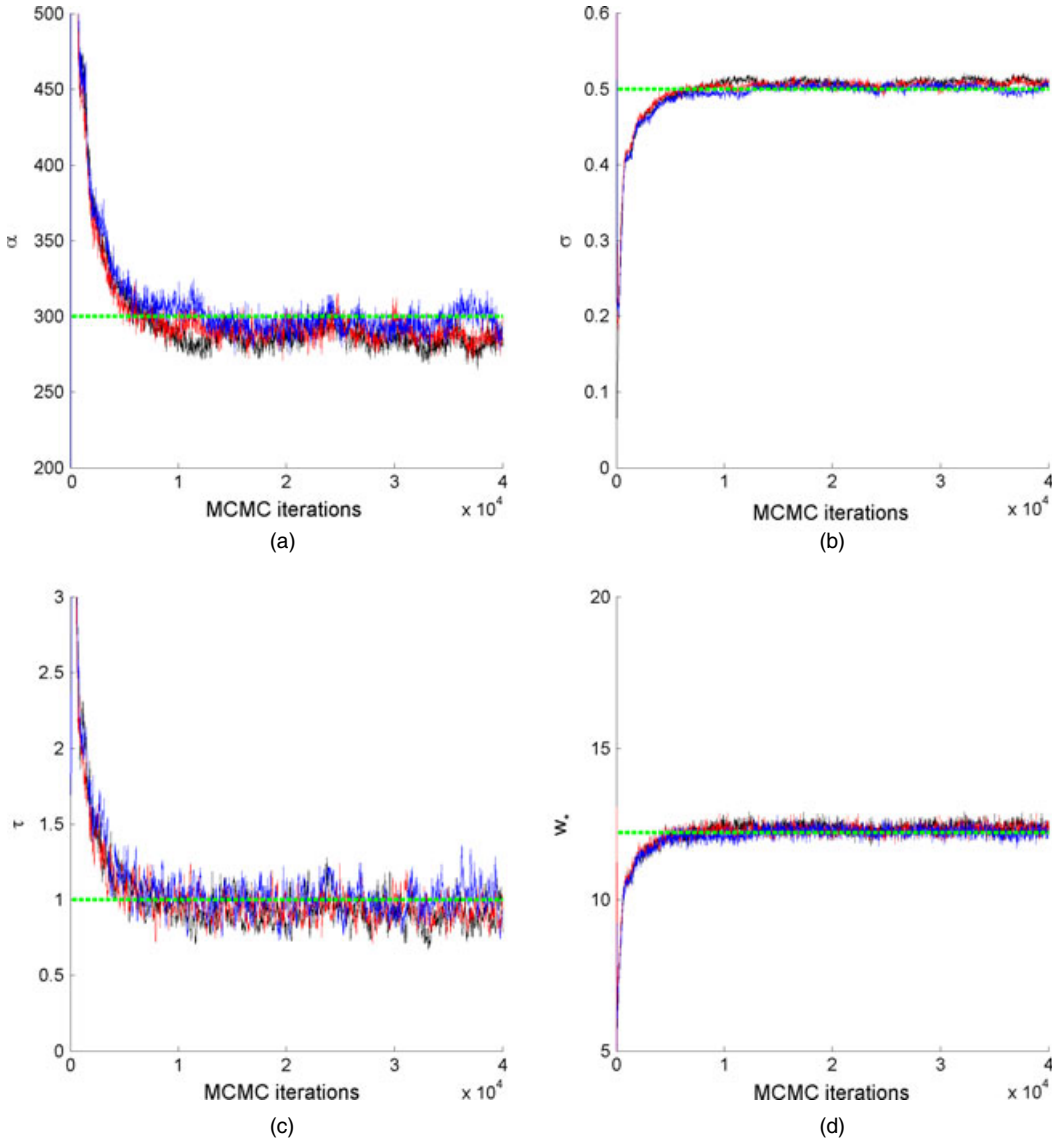
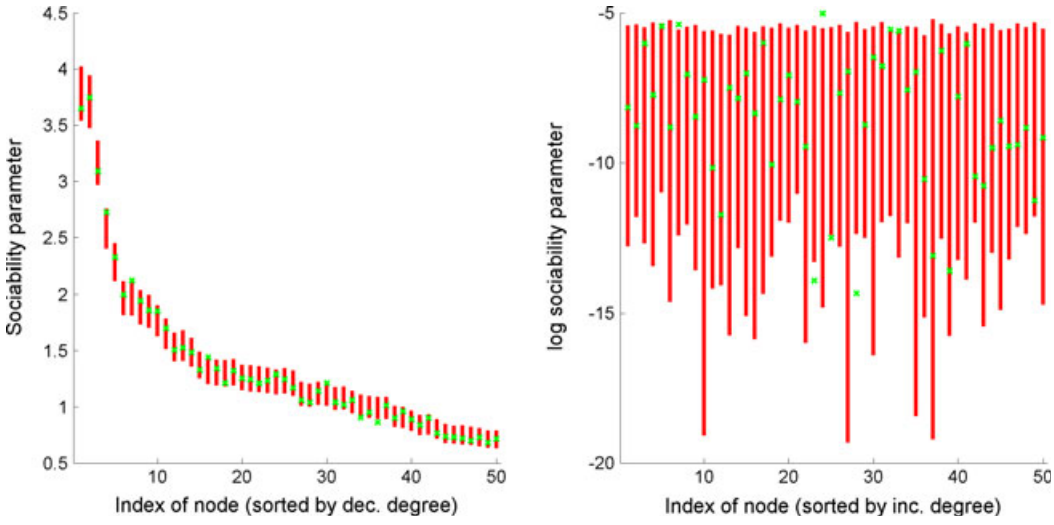


Fig. 9. MCMC trace plots of parameters (a) α, (b) σ, (c) τ and (d) w_* for a graph generated from a GGP model with parameters $\alpha = 300, \sigma = 0.5$ and $\tau = 1$: ——, chain 1; ——, chain 2; ——, chain 3; - - - - -, true

$\sigma < 0$, ς_1 corresponds to the expected number of nodes, ς_2 to the mean of the sociability parameters and ς_3 to their variance (see Section 6.3). In contrast, the parameters σ and τ are only weakly identifiable in this case. The potential scale reduction factor is computed on $(w_{1:N_\alpha}, w_*, \varsigma_1, \varsigma_2, \varsigma_3)$, and its maximum value was 1.01, suggesting convergence.

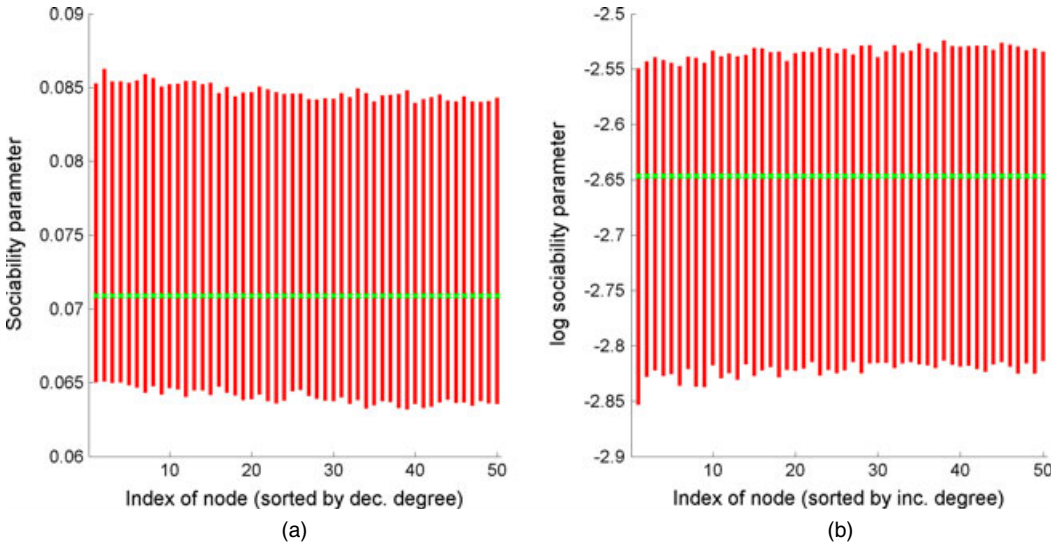
The value of ς_1 converges around the true number of nodes and ς_2 to the true sociability parameter $\sqrt{\{-\frac{1}{2}\log(1-p)\}}$ (constant across nodes for the Erdös–Rényi model), whereas ς_3 is close to 0 as the variance over the sociability parameters is very small. The total mass is also very close to 0, indicating that there are no nodes with degree 0.

Posterior credible intervals for the nodes with highest and lowest degrees are in Fig. 11,



(a) (b)

Fig. 10. 95% posterior intervals (I) of (a) the sociability parameters w_i of the 50 nodes with highest degree and (b) the log-sociability parameter $\log(w_i)$ of the 50 nodes with lowest degree, for a graph generated from a GGP model with parameters $\alpha = 300, \sigma = 0.5$ and $\tau = 1$: ×, true values



(a) (b)

Fig. 11. 95% posterior intervals (I) of (a) sociability parameters w_i of the 50 nodes with highest degree and (b) log-sociability parameters $\log(w_i)$ of the 50 nodes with lowest degree, for a graph generated from an Erdös-Rényi model with parameters $n = 1000$ and $p = 0.01$: in this case, all nodes have the same true sociability parameter $\sqrt{\{-\frac{1}{2}\log(1-p)\}}$ (×)

showing that the model can accurately recover sociability parameters of both low and high degree nodes in the dense regime as well.

8.2. Assessing properties of real world graphs

We now turn to using our methods to assess properties of a collection of real world graphs, including their degree distributions and aspects of sparsity. For the latter, evaluation based on

a single finite graph is notoriously challenging as sparsity relates to the asymptotic behaviour of the graph. Measures of sparsity from finite graphs exist but can be costly to implement (Nešetřil and Ossona de Mendez, 2012). On the basis of our GGP-based formulation and associated theoretical results described in Section 6, we consider $\Pr(\sigma \geqslant 0|z)$ as informative of the connectivity structure of the graph since the GGP graph model yields dense graphs for $\sigma < 0$, and sparse graphs for $\sigma \in [0, 1)$ (see equation (39)). For our analyses, we consider improper priors on the unknown parameters (α, σ, τ). We report $\Pr(\sigma \geqslant 0|z)$ based on a set of observed connections $(z_{ij})_{1<i,j<N_\alpha}$, which can be directly approximated from the MCMC output.We consider 12 different data sets:

(a) *facebook107*—social circles in Facebook (https://snap.stanford.edu/data/egonets-Facebook.html) (McAuley and Leskovec, 2012);
(b) *polblogs*—political blogosphere (February 2005) (http://www.cise.ufl.edu/research/sparse/matrices/Newman/polblogs) (Adamic and Glance, 2005);
(c) *USairport*—US airport connection network in 2010 (http://toreopsahl.com/datasets/) (Colizza *et al.*, 2007);
(d) *UCirvine*—social network of students at the University of California, Irvine (http://toreopsahl.com/datasets/) (Opsahl and Panzarasa, 2009);
(e) *yeast*—yeast protein interaction network (http://www.cise.ufl.edu/research/sparse/matrices/Pajek/yeast.html) (Bu *et al.*, 2003);
(f) *USpower*—network of the high-voltage power grid in the western states of the USA (https://snap.stanford.edu/data/email-Enron.html) (Watts and Strogatz, 1998);
(g) *IMDB*—actor collaboration network based on acting in the same movie (http://www.cise.ufl.edu/research/sparse/matrices/Pajek/IMDB.html);
(h) *cond-mat1*—co-authorship network (https://snap.stanford.edu/data/email-Enron.html) (Newman, 2001), based on preprints posted to condensed matter of arXiv between 1995 and 1999, obtained from the bipartite preprints–authors network using a one-mode projection;
(i) *cond-mat2*—as in cond-mat1, but using Newman's projection method;
(j) *Enron*—Enron collaboration network from a multigraph e-mail network (https://snap.stanford.edu/data/email-Enron.html)
(k) *internet*—connectivity of internet routers (http://www.cise.ufl.edu/research/sparse/matrices/Pajek/internet.html);
(l) *www*—linked World Wide Web pages in the nd.edu domain (http://lisgi1.engr.ccny.cuny.edu/∼makse/soft_data.html).

The sizes of the various data sets are given in Table 2 and range from a few hundred nodes or edges to a million. The adjacency matrices for these networks are plotted in Fig. 12 and empirical degree distributions in Fig. 13 (red).

We ran three MCMC chains for 40 000 iterations with the same specifications as above and report the estimate of $\Pr(\sigma \geqslant 0|z)$ and 99% posterior credible intervals of σ in Table 2; we additionally provide run times. MCMC trace plots suggested rapid convergence of the sampler. Since sparsity is an asymptotic property of a graph, and we are analysing finite graphs, our inference of σ here simply provides insight into some structure of the graph and is not formally a test of sparsity. From Table 2, we note that we infer negative σ-values for many of the smaller networks. This might indicate that these graphs have dense connectivity; for example, our facebook107 data set represents a small social circle that is probably highly interconnected and the polblogs data set represents two tightly connected political parties. We infer positive σ-values for three of the data sets (USairport, Enron and www); note that two of these data sets are in the top three

Table 2. Size of real world data sets and posterior probability of sparsity

Data set	*Number of nodes*	*Number of edges*	*Time (min)*	$Pr(\sigma \geqslant 0 \mid z)$	*99% credible interval* σ
facebook107	1034	26749	1	0.00	$[-1.06, -0.82]$
polblogs	1224	16715	1	0.00	$[-0.35, -0.20]$
USairport	1574	17215	1	1.00	$[0.10, 0.18]$
UCirvine	1899	13838	1	0.00	$[-0.14, -0.02]$
yeast	2284	6646	1	0.28	$[-0.09, 0.05]$
USpower	4941	6594	1	0.00	$[-4.84, -3.19]$
IMDB	14752	38369	2	0.00	$[-0.24, -0.17]$
cond-mat1	16264	47594	2	0.00	$[-0.95, -0.84]$
cond-mat2	7883	8586	1	0.00	$[-0.18, -0.02]$
Enron	36692	183831	7	1.00	$[0.20, 0.22]$
internet	124651	193620	15	0.00	$[-0.20, -0.17]$
www	325729	1090108	132	1.00	$[0.26, 0.30]$

largest networks considered, where sparse connectivity is more commonplace. In the remaining large network, internet, a question is why the inferred σ is negative. This may be due to dense subgraphs or spots (for example, spatially proximate routers may be highly interconnected, but sparsely connected outside the group) (Borgs *et al.*, 2014b). This relates to the idea of *community structure*, though not every node need be associated with a community. As in many sparse network models that assume no dense spots (Bollobás and Riordan, 2009; Wolfe and Olhede, 2013), our approach does not explicitly model such effects. Capturing such structure remains a direction of future research that is likely to be feasible within our generative framework. However, our current method has the benefit of simplicity with three hyperparameters tuning the network properties. Finally, we note in Table 2 that our analyses finish in a remarkably short time although the code base was implemented in MATLAB on a standard desktop machine, without leveraging possible opportunities for parallelizing and other mechanisms for scaling the sampler (see Section 7 for a discussion).

To assess our fit to the empirical degree distributions, we used the methods that were described in Section 5.5 to simulate 5000 graphs from the posterior predictive distribution in each scenario. Fig. 13 provides a comparison between the empirical degree distributions and those based on the simulated graphs. In all cases, we see a reasonably good fit. For the largest networks, Figs 13(j)–13(l), we see a slight underestimate of the tail of the distribution, i.e. we do not capture as many high degree nodes as truly present. This may be because these graphs exhibit a power law behaviour, but only after a certain node degree (Clauset *et al.*, 2009), which is not an effect that is explicitly modelled by our framework. Instead, our model averages the error in the low and high degree nodes. Another reason for underestimating the tails might be dense spots, which we also do not explicitly model. However, our model does capture power law behaviour with possible exponential cut off in the tail. We see a similar trend for cond-mat1, but not cond-mat2. Based on the bipartite articles–authors graph, cond-mat1 uses the standard one-mode projection and sets a connection between two authors who have co-authored a paper; this projection clearly creates dense spots in the graph. In contrast, cond-mat2 uses Newman's projection method (Newman *et al.*, 2001). This method constructs a weighted undirected graph by counting the number of papers that were co-authored by two scientists, where each count is normalized by the number of authors on the paper. To construct the undirected graph, an edge is created if the weight is equal to or greater than 1; cond-mat1 and cond-mat2 thus have a different number of edges

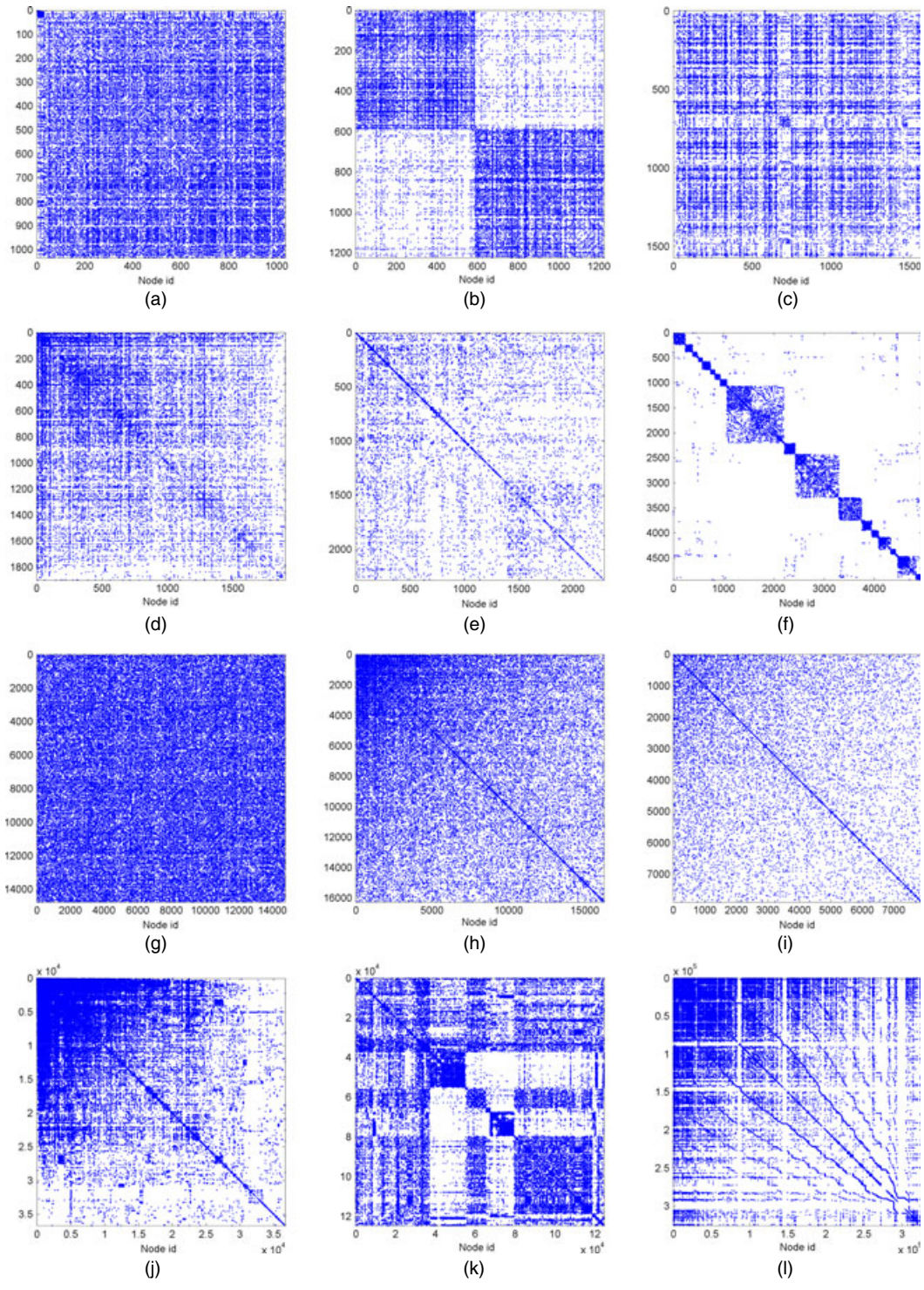


Fig. 12. Adjacency matrices for various real world networks: (a) facebook107; (b) polblogs; (c) USairport; (d) UCirvine; (e) yeast; (f) USpower; (g) IMDB; (h) cond-mat1; (i) cond-mat2; (j) Enron; (k) internet; (l) www

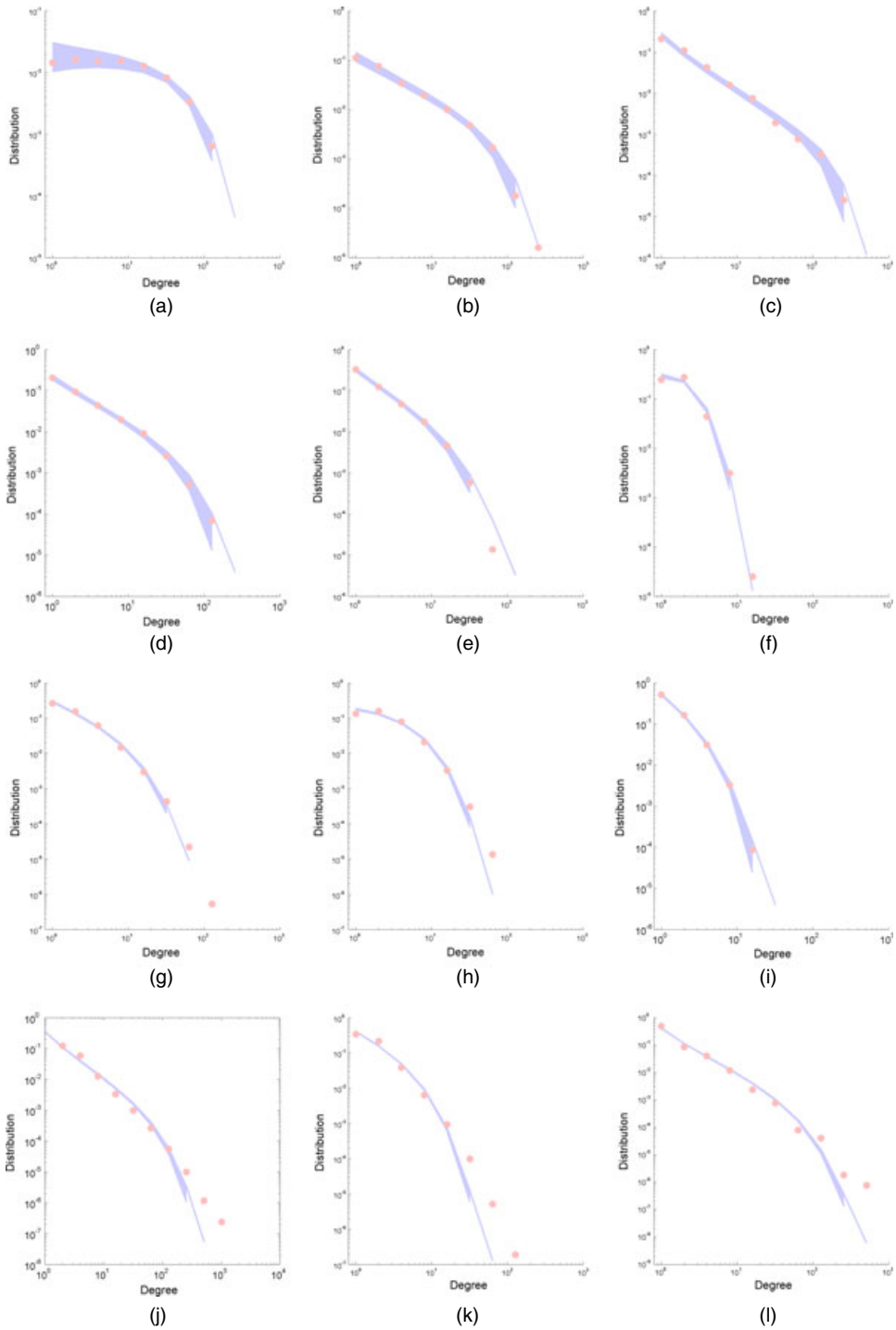


Fig. 13. Empirical degree distribution (——) and posterior predictive (——) for various real world networks (●, data): (a) facebook107; (b) polblogs; (c) USairport; (d) UCirvine; (e) yeast; (f) USpower; (g) IMDB; (h) cond-mat1; (i) cond-mat2; (j) Enron; (k) internet; (l) www

and nodes, as only nodes with at least one connection are considered. It is interesting that the projection method that was used for the cond-mat data set has a clear influence on the sparsity of the resulting graph, cond-mat2 being less dense than cond-mat1 (see Figs 13(h) and 13(i)). The degree distribution for cond-mat1 is similar to that of internet, thus inheriting the same issues as previously discussed. Overall, it appears that our model better captures homogeneous power law behaviour with possible exponential cut-off in the tails than it does a graph with perhaps structured dense spots or power law after a point behaviour.

9. Conclusion

We proposed a class of statistical network models building on exchangeable random measures. Using this representation, we showed how it is possible to specify models with properties that are different from those of models based on exchangeable adjacency matrices. As an example, we considered a model building on the framework of CRMs that can yield sparse graphs while maintaining attractive exchangeability properties. For a choice of CRMs, our fully generative formulation can yield networks ranging from dense to sparse, as tuned by a single hyperparameter.

In this paper, exchangeability is in the context of random measures for which we appealed to the Kallenberg representation in place of the Aldous–Hoover theorem for exchangeable arrays. Using this framework, we arrived at a structure that is analogous to the graphon, which opens up new modelling and theoretical analysis possibilities beyond those of the special case that is considered herein. Importantly, through the exchangeability of the underlying random measures and leveraging HMC sampling, we devised a scalable algorithm for posterior computations. This scheme enables inference of the graph parameters, including the parameter determining the sparsity of the graph. We examined our methods on a range of real world networks, demonstrating that our model yields a practically useful statistical tool for network analysis.

We believe that the foundational modelling tools and theoretical results that we presented represent an important building block for future developments. Such developments can be divided along two dimensions:

(a) modelling advances, such as incorporating notions of community structure and node attributes, within this framework and
(b) theoretical analyses looking at the properties of the corresponding class of networks.

For the latter, we considered just one simplified version of the Kallenberg representation; examining a more general form could yield graphs with additional structure. Building on an initial version of this paper (Caron and Fox, 2014), initial forays into advances on the modelling side can be found in Herlau *et al.* (2016) and Todeschini and Caron (2016) and theoretical analyses in Veitch and Roy (2015) and Borgs *et al.* (2016).

Acknowledgements

The authors thank Bernard Bercu, Arnaud Doucet, Yee Whye Teh, Stefano Favaro, Dan Roy, Lancelot James, Jennifer Chayes, Christian Borgs, Judith Rousseau, George Deligiannidis and Konstantina Palla for helpful discussions and feedback on earlier versions of this paper.

FC acknowledges the support of the European Commission under the Marie Curie intra-European fellowship programme. This work was partially supported by the Alan Turing Institute under Engineering and Physical Sciences Research Council grant EP/N510129/1 and by the BNPSI ANR project ANR-13-BS-03-0006-01.

EBF was supported in part by DARPA grant FA9550-12-1-0406 negotiated by AFOSR and AFOSR grant FA9550-12-1-0453.

Appendix A: Further background

A.1. Aldous–Hoover theorem and graphons

In theorem 7, we present the Aldous–Hoover theorem for the case of *joint exchangeability*—i.e. symmetric permutations of rows and columns—which is applicable to matrices Z where both rows and columns index the same set of nodes. *Separate exchangeability* allows for different row and column permutations, making it applicable to scenarios where one has distinct node identities on rows and columns, such as in the bipartite graphs that we considered in Section 3.4. Extensions to higher dimensional arrays are likewise straightforward. For a more general statement of the Aldous–Hoover theorem, which holds for separate exchangeability and higher dimensional arrays, see Orbanz and Roy (2015).

Theorem 7 (Aldous-Hoover representation of jointly exchangeable matrices (Aldous, 1981; Hoover, 1979)). A random matrix $(Z_{ij})_{i,\, j\in\mathbb{N}}$ is jointly exchangeable if and only if there is a random measurable function $f:[0,1]^3 \to \mathbf{Z}$ such that

$$(Z_{ij}) \stackrel{d}{=} (f(U_i, U_j, U_{ij})), \tag{46}$$

where $(U_i)_{i\in\mathbb{N}}$ and $(U_{ij})_{i,\, j>i\in\mathbb{N}}$ with $U_{ij} = U_{ji}$ are a sequence and matrix respectively of IID uniform[0, 1] random variables.

For undirected graphs where Z is a binary, symmetric adjacency matrix, the Aldous–Hoover representation can be expressed as the existence of a *graphon* $M:[0,1]^2 \to [0,1]$, symmetric in its arguments, where

$$f(U_i, U_j, U_{ij}) = \begin{cases} 1 & U_{ij} < M(U_i, U_j), \\ 0 & \text{otherwise.} \end{cases} \tag{47}$$

A.2. Regularly varying Lévy measures

Here we provide a formal definition of a regularly varying Lévy measure that is ρ referred to in theorems 2 and 5.

Definition 1 (regular variation). A Lévy measure ρ on $(0,\infty)$ is said to be *regularly varying* if its tail Lévy intensity $\bar{\rho}(x) = \int_x^\infty \rho(dw)$ is a regularly varying function (Feller, 1971), i.e. it satisfies

$$\bar{\rho}(x) \stackrel{x\downarrow 0}{\sim} l(1/x)x^{-\sigma} \tag{48}$$

for $\sigma \in [0,1)$ and l a slowly varying function satisfying $\lim_{t\to\infty} l(at)/l(t) = 1$ for any $a > 0$.

For example, constant and logarithmic functions are slowly varying.

Appendix B: Proof of proposition 1

The proof of proposition 1 follows from the properties of $W \sim \mathrm{CRM}(\rho, \lambda)$. Let $A_i = [h(i-1), hi]$ for $h > 0$ and $i \in \mathbb{N}$. We have

$$(W(A_i)) \stackrel{\mathrm{d}}{=} (W(A_{\pi(i)})) \tag{49}$$

for any permutation π of $\mathbb{N}$. As $D(A_i \times A_j) \sim \mathrm{Poisson}\{W(A_i)W(A_j)\}$, it follows that

$$(D(A_i \times A_j)) \stackrel{\mathrm{d}}{=} (D(A_{\pi(i)} \times A_{\pi(j)})) \tag{50}$$

for any permutation π of $\mathbb{N}$. Joint exchangeability of Z follows directly.

Appendix C: Proofs of results on the sparsity

C.1. Probability asymptotics notation

We first describe the asymptotic notation that is used in the remainder of this section, which follows the notation of Janson (2011). All unspecified limits are as $\alpha \to \infty$.

Let $(X_\alpha)_{\alpha\geqslant 0}$ and $(Y_\alpha)_{\alpha\geqslant 0}$ be two $[0,\infty)$-valued stochastic processes defined on the same probability space and such that $\lim_{\alpha\to\infty} X_\alpha = \lim_{\alpha\to\infty} Y_\alpha = \infty$ almost surely. We have

$$\begin{aligned}
X_\alpha = O(Y_\alpha) \text{ almost surely} &\Leftrightarrow \limsup_{\alpha\to\infty} X_\alpha/Y_\alpha < \infty \text{ almost surely},\\
X_\alpha = o(Y_\alpha) \text{ almost surely} &\Leftrightarrow \lim_{\alpha\to\infty} X_\alpha/Y_\alpha = 0 \text{ almost surely},\\
X_\alpha = \Omega(Y_\alpha) \text{ almost surely} &\Leftrightarrow Y_\alpha = O(X_\alpha) \text{ almost surely},\\
X_\alpha = \omega(Y_\alpha) \text{ almost surely} &\Leftrightarrow Y_\alpha = o(X_\alpha) \text{ almost surely},\\
X_\alpha = \Theta(Y_\alpha) \text{ almost surely} &\Leftrightarrow X_\alpha = O(Y_\alpha) \text{ and } X_\alpha = \Omega(Y_\alpha) \text{ almost surely}.
\end{aligned}$$

The equivalence notation $f(x) \sim^{x\downarrow 0} g(x)$ is used for $\lim_{x\to 0} f(x)/g(x) = 1$ (not to be confused with the notation '$\sim$' alone for 'distributed from').

C.2. Proof of theorem 3

Assume the moment condition $0 < \int_0^\infty w\,\rho(dw) < \infty$. Let

$$\tilde{Z}_{ij} = \begin{cases} 1 & \text{if } Z([i-1,i]\times[j-1,j]) > 0,\\ 0 & \text{otherwise} \end{cases} \tag{51}$$

and $\tilde{D}_{ij} = D([i-1,i]\times[j-1,j])$. Then, almost surely for any $k\in\mathbb{N}$,

$$\sum_{1\leqslant i<j\leqslant k} \tilde{Z}_{ij} \leqslant N_k^{(e)} \leqslant \sum_{1\leqslant i,j\leqslant k} \tilde{D}_{ij}. \tag{52}$$

As Z is a jointly exchangeable point process, $(\tilde{Z}_{ij})_{i,j\in\mathbb{N}}$ is a jointly exchangeable binary matrix, and so

$$\sum_{1\leqslant i<j\leqslant k} \tilde{Z}_{ij} = \Theta(k^2) \qquad \text{almost surely as } k\to\infty. \tag{53}$$

Moreover, we have

$$\tilde{D}_{ij}|W \overset{\text{ind}}{\sim} \text{Poisson}\{W([i-1,i])\,W([j-1,j])\} \tag{54}$$

So lemma 1 in Appendix D and the strong law of large numbers for V-statistics (Arcones and Giné, 1992; Giné and Zinn, 1992) imply that

$$\sum_{1\leqslant i,j\leqslant k} \tilde{D}_{ij} = \Theta(k^2) \qquad \text{almost surely as } k\to\infty. \tag{55}$$

We therefore conclude that $N_k^{(e)} = \Theta(k^2)$ almost surely as $k\to\infty$. Finally, for any $k\leqslant\alpha\leqslant k+1$,

$$\frac{k^2}{(k+1)^2}\frac{N_k^{(e)}}{k^2} \leqslant \frac{N_\alpha^{(e)}}{\alpha^2} \leqslant \frac{(k+1)^2}{k^2}\frac{N_{k+1}^{(e)}}{(k+1)^2}$$

and as $(k+1)/k\to 1$ we conclude that

$$N_\alpha^{(e)} = \Theta(\alpha^2) \qquad \text{almost surely as } \alpha\to\infty. \tag{56}$$

C.3. Proof of theorem 4

C.3.1. Finite activity case

We first consider the case of a finite activity CRM. In this case, the number of nodes is bounded below by the square root of the number of edges, and bounded above by the (finite) number of jumps of the CRM. Let $T = \int_0^\infty \rho(\mathrm{d}w)$, $0<T<\infty$. Let J_α denote the number of points (w_i,θ_i) such that $\theta_i<\alpha$. $(J_\alpha)_{\alpha\geqslant 0}$ is a homogeneous Poisson process of rate T, and thus $J_\alpha/\alpha\to T$ almost surely. Since $\sqrt{N_\alpha^{(e)}} \leqslant N_\alpha \leqslant J_\alpha$ almost surely, it follows from result (56) that $N_\alpha = \Theta(\alpha)$ almost surely as $\alpha\to\infty$.

C.3.2. Infinite activity case

We now consider the infinite activity case where $\int_0^\infty \rho(\mathrm{d}w) = \infty$. First note that, by monotone convergence, $\lim_{t\to\infty}\psi(t) = \infty$.

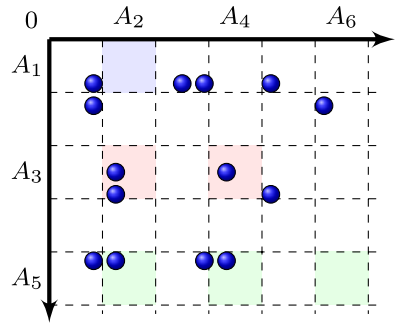


Fig. 14. Illustration of the proof of theorem 4 in the infinite activity case (●, support points of the measure D): in this example, the coloured regions indicate the intersection of A_{2n-1} with $\mathcal{S}_n^{(2)}$ for $n=1$ (●), $n=2$ (●), and $n=3$ (●); from this, we see that $X_1=0$ (the count in blue), $X_2=2$ (the count in red) and $X_3=1$ (the count in green)

Consider sets $A_k=[(k-1)/2,k/2)$ for $k=1,2,\ldots$, and let $\mathcal{S}_n^{(1)}=\cup_{k=1}^n A_{2k-1}$ and $\mathcal{S}_n^{(2)}=\cup_{k=1}^n A_{2k}$. $\mathcal{S}_n^{(1)}$ and $\mathcal{S}_n^{(2)}$ define a partition of $[0,n]$. For $n\in\mathbb{N}$, define the random variable X_n as

$$X_n=\#\{\theta_i\in A_{2n-1}|D(\{\theta_i\}\times\mathcal{S}_n^{(2)})>0\} \tag{57}$$

and let

$$\tilde{N}_\alpha=\sum_{n=1}^{\lfloor\alpha\rfloor}X_n.$$

See Fig. 14 for an illustration. Clearly, $\tilde{N}_\alpha$ is a lower bound for the number of nodes:

$$\tilde{N}_\alpha\leqslant N_\alpha \qquad \text{almost surely.} \tag{58}$$

Using the notation $\mathcal{S}^{(1)}=\cup_{k=1}^\infty A_{2k-1}$ and $\mathcal{S}^{(2)}=\cup_{k=1}^\infty A_{2k}$, let $W^{(1)}$ and $W^{(2)}$ be respectively the restriction of W to the set $\mathcal{S}^{(1)}$ and $\mathcal{S}^{(2)}$. As $\mathcal{S}^{(1)}$ and $\mathcal{S}^{(2)}$ are non-overlapping and W is a CRM, $W^{(1)}$ and $W^{(2)}$ are independent. Integrating over $W^{(1)}$ and using the marking theorem for Poisson processes (see below for more details), we obtain for $n\geqslant 1$

$$X_n|W^{(2)}\overset{\text{ind}}{\sim}\text{Poisson}[\tfrac{1}{2}\psi\{W(\mathcal{S}_n^{(2)})\}]. \tag{59}$$

Lemma 1 thus implies that

$$\frac{\sum_{k=1}^n X_k}{\frac{1}{2}\sum_{k=1}^n\psi\{W(\mathcal{S}_k^{(2)})\}}\to 1 \qquad \text{almost surely.} \tag{60}$$

We have $\lambda(\mathcal{S}_n^{(2)})=n/2$ and, using the law of large numbers,

$$\frac{W(\mathcal{S}_n^{(2)})}{n/2}\to\int_0^\infty w\,\rho(\mathrm{d}w) \qquad \text{almost surely.} \tag{61}$$

Therefore $\psi\{W(\mathcal{S}_n^{(2)})\}\to\infty$ almost surely. Its Cesàro mean also diverges and

$$\frac{\sum_{k=1}^n\psi\{W(\mathcal{S}_k^{(2)})\}}{n}\to\infty \qquad \text{almost surely,} \tag{62}$$

which, together with result (60), implies that $1/n\Sigma_{k=1}^n X_k\to\infty$ almost surely. We conclude that $\tilde{N}_\alpha/\alpha\to\infty$ almost surely and, using inequality (58), $N_\alpha/\alpha\to\infty$ almost surely.

Consider now the case where $\bar{\rho}(x)\sim^{x\downarrow 0}l(1/x)x^{-\sigma}$ where $\sigma\in(0,1)$ and $l(t)$ is a slowly varying function such that $\liminf_{t\to\infty}l(t)>0$. Then lemma 2 in Appendix D implies that $\liminf_{t\to\infty}\psi(t)/t^\sigma>0$ and thus, using result (61),

$$\liminf_{n\to\infty} \frac{\psi\{W(\mathcal{S}_n^{(2)})\}}{n^\sigma} > 0 \qquad \text{almost surely.}$$

Riemann integration and the Stolz–Cesàro theorem then imply that

$$\liminf_{n\to\infty} \frac{\sum_{k=1}^n \psi\{W(\mathcal{S}_k^{(2)})\}}{n^{\sigma+1}} = \liminf_{n\to\infty} \frac{\sum_{k=1}^n \psi\{W(\mathcal{S}_k^{(2)})\}}{(\sigma+1)\sum_{k=1}^n k^\sigma} > 0 \qquad \text{almost surely.}$$

and finally $N_\alpha = \Omega(\alpha^{\sigma+1})$ almost surely as $\alpha \to \infty$.

C.3.2.1. Proof of result (59). Let $n \geqslant 1$. For any $\theta_i \in A_{2n-1}$, let u_i be a binary mark such that

$$\Pr(u_i = 1 | W) = 1 - \exp\{-w_i W(\mathcal{S}_n^{(2)})\}$$

Conditionally on $W_n^{(2)}$, the marking theorem for Poisson processes implies that the set of points $\{(w_i)|\theta_i \in A_{2n-1}, u_i = 1\}$ is drawn from a Poisson process of intensity $\frac{1}{2}\rho(\mathrm{d}w)[1-\exp\{-wW(\mathcal{S}_n^{(2)})\}]$ and the number X_n of those points is Poisson distributed with rate

$$\int_0^\infty \tfrac{1}{2}\rho(\mathrm{d}w)[1-\exp\{-w\,W(\mathcal{S}_n^{(2)})\}] = \tfrac{1}{2}\psi\{W(\mathcal{S}_n^{(2)})\}.$$

Finally, independence of the X_ns follows from the complete randomness of $W_n^{(1)}$.

C.4. Proof of theorem 5

$$\begin{aligned}\mathbb{E}[D_\alpha^*] &= \mathbb{E}[\mathbb{E}[D_\alpha^*|W]] = \mathbb{E}[W_\alpha^{*2}] = \mathbb{E}[W_\alpha^*]^2 + \mathrm{var}[W_\alpha^*] \\ &= \alpha^2 \left\{\int_0^\infty w\,\rho(\mathrm{d}w)\right\}^2 + \alpha \int_0^\infty w^2 \rho(\mathrm{d}w)\end{aligned}$$

where the last line follows from Campbell's theorem.

$$\begin{aligned}\mathbb{E}[N_\alpha^{(e)}] &= \mathbb{E}[\mathbb{E}[N_\alpha^{(e)}|W]] = \mathbb{E}[\textstyle\sum_i [\{1-\exp(-w_i^2)\} + \sum_{j\neq i}\{1-\exp(-2w_iw_j)\}\mathbf{1}_{\theta_j\leqslant\alpha}]\mathbf{1}_{\theta_i\leqslant\alpha}] \\ &= \mathbb{E}[\textstyle\sum_i [\{1-\exp(-w_i^2)\} - (1-\exp(-2w_i^2))]\mathbf{1}_{\theta_i\leqslant\alpha}] + \mathbb{E}[\sum_i [\sum_j [1-\exp(-2w_iw_j)]\mathbf{1}_{\theta_j\leqslant\alpha}]\mathbf{1}_{\theta_i\leqslant\alpha}].\end{aligned}$$

Using the Palm formula for Poisson point processes (Bertoin, 2006; Daley and Vere-Jones 2008)

$$\mathbb{E}[\textstyle\sum_i [\sum_j \{1-\exp(-2w_iw_j)\}\mathbf{1}_{\theta_j\leqslant\alpha}]\mathbf{1}_{\theta_i\leqslant\alpha}] = \alpha \int_0^\infty \mathbb{E}[\{1-\exp(-2w^2)\} + \textstyle\sum_j \{1-\exp(-2ww_j)\}\mathbf{1}_{\theta_j\leqslant\alpha}]\rho(\mathrm{d}w)$$

and the final expression is obtained by applying Campbell's theorem. Finally, the expected number of nodes

$$\begin{aligned}\mathbb{E}[N_\alpha] &= \mathbb{E}[\mathbb{E}[N_\alpha|W]] = \mathbb{E}[\textstyle\sum_i \{1-\exp(-2w_i \sum_j w_j\mathbf{1}_{\theta_j\leqslant\alpha} + w_i^2)\}\mathbf{1}_{\theta_i\leqslant\alpha}] \\ &= \alpha \int_0^\infty \mathbb{E}[\{1-\exp(-2w\textstyle\sum_j w_j\mathbf{1}_{\theta_j\leqslant\alpha} - w^2)\}]\rho(\mathrm{d}w) \\ &= \alpha \int_0^\infty (1-\mathbb{E}[\exp(-2w\textstyle\sum_j w_j\mathbf{1}_{\theta_j\leqslant\alpha} - w^2)])\rho(\mathrm{d}w) \\ &= \alpha \int_0^\infty [1-\exp\{-w^2 - \alpha\psi(2w)\}]\,\rho(\mathrm{d}w)\end{aligned}$$

where we successively used the Palm formula and Campbell's theorem. By dominated convergence, $\mathbb{E}[N_\alpha] \sim^{\alpha\uparrow\infty} \alpha\int_0^\infty \rho(\mathrm{d}w)$ if the CRM is finite activity. Consider now that the CRM is infinite activity. Using integration by parts, we have

$$\mathbb{E}[N_\alpha] = \alpha \int_0^\infty \{2w + 2\alpha\psi'(2w)\} \exp\{-w^2 - \alpha\psi(2w)\}\bar{\rho}(w)\mathrm{d}w. \tag{63}$$

The Lévy exponent is a strictly increasing function with $\psi(0) = 0$ and $\lim_{t\to\infty} \psi(t) = \infty$ and therefore admits a well-defined inverse, denoted $\psi^{-1} : [0, \infty) \to [0, \infty)$. Using the change of variable $u = \psi(2w)$, we obtain

$$\int_0^\infty 2\psi'(2w) \exp\{-w^2 - \alpha\psi(2w)\}\bar{\rho}(w)\,\mathrm{d}w = \int_0^\infty \exp[-\{\psi^{-1}(u)/2\}^2 - \alpha u]\bar{\rho}\{\psi^{-1}(u)/2\}\,\mathrm{d}u.$$

Assume that $\int_0^\infty w\,\rho(\mathrm{d}w) < \infty$. Now note that $\psi(t) \sim^{t\downarrow 0} t\int_0^\infty w\,\rho(\mathrm{d}w)$ and therefore $\psi^{-1}(t) \sim^{t\downarrow 0} t/\int_0^\infty w\rho(\mathrm{d}w)$. If ρ is a regularly varying Lévy measure, then

$$\bar{\rho}(x) \sim^{x\downarrow 0} l(1/x)x^{-\sigma}$$

where $\sigma \in [0, 1)$ and l is a slowly varying function, and it therefore follows from lemma 3 in Appendix D and $\psi^{-1}(0) = 0$ that

$$g(u) := \exp[-\{\psi^{-1}(u)/2\}^2]\bar{\rho}\{\psi^{-1}(u)/2\} \overset{u\downarrow 0}{\sim} l(1/u)u^{-\sigma}\left\{2\int_0^\infty w\,\rho(\mathrm{d}w)\right\}^\sigma$$

where g is a monotone decreasing function. Applying the Tauberian theorem of proposition 2 in Appendix D, we therefore have

$$\int_0^\infty \exp(-\alpha u)g(u)\mathrm{d}u \overset{\alpha\uparrow\infty}{\sim} \alpha^{\sigma-1}l(\alpha)\,\Gamma(1-\sigma)\left\{2\int_0^\infty w\,\rho(\mathrm{d}w)\right\}^\sigma.$$

Finally, combining the above asymptotics with equation (63), and noting that

$$\int_0^\infty w \exp\{-w^2 - \alpha\psi(2w)\}\bar{\rho}(w)\mathrm{d}w = o(1)$$

by dominated convergence, and $\lim_{\alpha\to\infty} \alpha^\sigma l(\alpha) > 0$ for $\sigma \in [0, 1)$, we obtain

$$\mathbb{E}[N_\alpha] \overset{\alpha\uparrow\infty}{\sim} \alpha^{1+\sigma}l(\alpha)\Gamma(1-\sigma)\left\{2\int_0^\infty w\rho(\mathrm{d}w)\right\}^\sigma.$$

Appendix D: Technical lemmas

The following lemma is a corollary of theorem 3, page 239, in Feller (1971).

Lemma 1. Let $(X_n)_{n=1,2,\ldots}$ be a sequence of mutually independent random variables with arbitrary distribution and such that $\mathrm{var}(X_n) \leqslant \mathbb{E}[X_n] < \infty$. Let $S_n = \Sigma_{k=1}^n X_k$. If

$$\lim_{n\to\infty} \mathbb{E}[S_n] = \infty$$

then

$$S_n/\mathbb{E}[S_n] \to 1 \qquad \text{almost surely as } n \to \infty.$$

Proof. Assume for simplicity that $\mathbb{E}[X_n] > 0$ for all n (otherwise, consider the subsequence of random variables with strictly positive mean). We have

$$\sum_{k=1}^n \frac{\mathrm{var}(X_k)}{1 + \mathbb{E}[S_k]^2} \leqslant \sum_{k=1}^n \frac{\mathbb{E}[X_k]}{1 + \mathbb{E}[S_k]^2} \leqslant \int_0^\infty \frac{1}{1+x^2}\mathrm{d}x < \infty$$

by Riemann integration. The result then follows from theorem 3, page 239, in Feller (1971) with $b_n = \mathbb{E}[S_n]$.

Lemma 2 (relating tail Lévy intensity and Laplace exponent). (Gnedin *et al.* (2007), propositions 17 and 19) Let ρ be a Lévy measure, $\bar{\rho}(x) = \int_x^\infty \rho(\mathrm{d}w)$ be the tail Lévy intensity and $\psi(t) = \int_0^\infty \{1 - \exp(-wt)\}\rho(dw)$ its Laplace exponent. The following conditions are equivalent:

$$\bar{\rho}(x) \overset{x\downarrow 0}{\sim} l(1/x)x^{-\sigma}, \tag{64}$$

$$\psi(t) \overset{t\uparrow\infty}{\sim} \Gamma(1-\sigma)t^{\sigma}l(t) \tag{65}$$

where $0 \leqslant \sigma < 1$ and l is a function slowly varying at ∞, i.e. satisfying $l(cy)/l(y) \to 1$ as $y \to \infty$ for every $c > 0$.

Lemma 3 (Resnick (1987), chapter 0, proposition 0.8). If U is a regularly varying function at 0 with exponent $\sigma \in \mathbb{R}$, and f is a positive function such that $f(t) \sim^{t\downarrow 0} tc$, for some constant $0 < c < \infty$, then $U\{f(t)\} \sim^{t\downarrow 0} c^{\sigma}U(t)$.

Proposition 2 (Tauberian theorem). (Feller (1971), chapter XIII, section 5, theorems 3 and 4). Let $U(dw)$ be a measure on $(0, \infty)$ with ultimately monotone density u, i.e. monotone in some interval (x_0, ∞). Assume that

$$\mathcal{L}(t) = \int_0^{\infty} \exp(-tw)u(w)\mathrm{d}w$$

exists for $t > 0$. If l is slowly varying at ∞ and $0 \leqslant a < \infty$, then the following two relationships are equivalent:

$$\mathcal{L}(t) \overset{t\uparrow\infty}{\sim} t^{-a}l(t), \tag{66}$$

$$u(x) \overset{x\downarrow 0}{\sim} \frac{1}{\Gamma(a)}x^{a-1}l\left(\frac{1}{x}\right). \tag{67}$$

Appendix E: Proof of theorem 6

The proof of theorem 6 relies on results on posterior characterization with models involving normalized CRMs. We first state a corollary of lemma 5 by Pitman (2003) and Theorem 8.1 by James (2002). Similar results appear in (Prünster (2002), James (2005) and James *et al.* (2009). The corollary involves the introduction of a discrete random variable R, conditional on which the CRM has strictly positive mass.

Corollary 1. Let W_α be a (finite or infinite) CRM on $[0, \alpha]$ without fixed atoms nor deterministic component, with mean measure $\rho(\mathrm{d}w)d\theta$. Denote $W_\alpha^* = W_\alpha([0, \alpha])$, with probability distribution G_α^*. Let $R \in \{0, 1, 2, \ldots\}$ be a discrete random variable such that, for $r \geqslant 0$,

$$\zeta_r(t) := \Pr(R = r | W_\alpha^* = t)$$

with $\zeta_0(0) = 1$. The condition $\Pr(R = 0 | W_\alpha^* = 0) = 1$ ensures that, conditionally on $R > 0$, $W_\alpha^* > 0$ almost surely, and the normalized CRM below is properly defined.

Conditionally on $R = r > 0$, let $X_1, \ldots, X_r \sim \mathrm{IID} W_\alpha/W_\alpha^*$. Let $\tilde{\theta}_1, \ldots, \tilde{\theta}_k$, $k \leqslant r$, be the unique values in $(X_1, \ldots, X_r)$, in order of appearance, with multiplicities $1 \leqslant \tilde{m}_j \leqslant r$, $\tilde{w}_i = W_\alpha(\{\tilde{\theta}_i\})$ the associated weights and $\Pi_r = \{A_1, \ldots, A_k\}$ with $A_i = \{j | X_j = \tilde{\theta}_i\}$ be the associated random partition of $\{1, \ldots, r\}$. Let $w_* = W_\alpha^* - \Sigma_i \tilde{w}_i$. For $r > 0$, we have

$$\begin{aligned}\Pr[R = r, \Pi_R = \{A_1, \ldots, A_k\}, (\tilde{w}_i \in \mathrm{d}\tilde{w}_i)_{i=1,\ldots k}, w_* \in \mathrm{d}w_*) \\ = \left(w_* + \sum_{i=1}^{k} \tilde{w}_i\right)^{-r} \zeta_r\left(w_* + \sum_{i=1}^{k} \tilde{w}_i\right) G_\alpha^*(\mathrm{d}w_*)\alpha^k \prod_{i=1}^{k} \tilde{w}_i^{\tilde{m}_i} \rho(\mathrm{d}\tilde{w}_i)\end{aligned} \tag{68}$$

and

$$\Pr(R = 0, w_* \in \mathrm{d}w_*) = \zeta_0(w_*)G_\alpha^*(\mathrm{d}w_*).$$

We now prove theorem 6. Consider the conditionally Poisson construction that was described in Section 5.5:

$$D_\alpha^* | W_\alpha^* \sim \mathrm{Poisson}(W_\alpha^{*\,2}),$$

$$(U_1', \ldots, U_{2D_\alpha^*}') | D_\alpha^*, W_\alpha \overset{\mathrm{IID}}{\sim} W_\alpha/W_\alpha^*.$$

First, define, for $r \in \{0, 2, 4, \ldots\}$ and $t \geqslant 0$,

$$\zeta_r(t) := \Pr(2D_\alpha^* = r | W_\alpha^* = t) = \frac{t^r \exp(-t^2)}{(r/2)!}$$

with $\zeta_0(0) = 1$. Conditionally on $2D_\alpha^* = r > 0$, $U_1', \ldots, U_r'$ are IID from W_α / W_α^*. The variables $U_1', \ldots, U_r'$ take $N_\alpha \leqslant r$ distinct values $\tilde{\theta}_j$, with multiplicities $1 \leqslant \tilde{m}_j \leqslant r$. Let $\Pi_r = \{A_1, \ldots, A_{N_\alpha}\}$ be the associated partition of $\{1, \ldots, r\}$. From corollary 1 we have, for $r \in \{2, 4, 6, \ldots\}$,

$$\begin{aligned}\Pr[(D_\alpha^* = r/2, \Pi_{2D_\alpha^*} = \{A_1, \ldots, A_{N_\alpha}\}, (\tilde{w}_i \in \mathrm{d}\tilde{w}_i)_{i=1,\ldots N_\alpha}, w_* \in \mathrm{d}w_*)] \\ = \frac{1}{(r/2)!} \exp\left\{ -\left(w_* + \sum_{i=1}^{N_\alpha} \tilde{w}_i \right)^2 \right\} G_\alpha^*(\mathrm{d}w_*) \alpha^{N_\alpha} \prod_{i=1}^{N_\alpha} \tilde{w}_i^{\tilde{m}_i} \rho(\mathrm{d}\tilde{w}_i)\end{aligned} \tag{69}$$

and

$$\Pr(D_\alpha^* = 0, w_* \in \mathrm{d}w_*) = \exp(-w_*^2) G_\alpha^*(\mathrm{d}w_*). \tag{70}$$

Finally, let π be the permutation of $\{1, \ldots, N_\alpha\}$ such that $\tilde{\theta}_{\pi(1)} < \tilde{\theta}_{\pi(2)} < \ldots < \tilde{\theta}_{\pi(N_\alpha)}$ and let $w_i = \tilde{w}_{\pi(i)}$ and $m_i = \tilde{m}_{\pi(i)}$. As the $\tilde{\theta}_i$ are IID and independent of $\tilde{w}_i$, π is uniformly distributed over the set of permutations of $\{1, \ldots, N_\alpha\}$. The vector $(m_1, \ldots, m_{N_\alpha})$ corresponds to the sizes of the partition $\Pi_{2D_\alpha^*}$ in exchangeable random order (Pitman (2006), equation (2.7), page 39), and, for $r \in \{2, 4, 6, \ldots\}$,

$$\begin{aligned}\Pr\{D_\alpha^* = r/2, (m_1, \ldots, m_{N_\alpha}), (w_i \in dw_i)_{i=1,\ldots N_\alpha}, w_* \in dw_*\} \\ = \frac{r!}{(r/2)!\, N_\alpha! \prod_{i=1}^{N_\alpha} m_i!} \exp\left\{ -\left(w_* + \sum_{i=1}^{N_\alpha} w_i \right)^2 \right\} G_\alpha^*(\mathrm{d}w_*) \alpha^{N_\alpha} \prod_{i=1}^{N_\alpha} w_i^{m_i} \rho(\mathrm{d}w_i).\end{aligned} \tag{71}$$

Appendix F: Details on the Markov chain Monte Carlo algorithms

The undirected graph sampler that was outlined in Section 7.2 iterates as follows.

Step 1: update $w_{1:N_\alpha}$ given the rest with HMC sampling.
Step 2: update $(\alpha, \sigma, \tau, w_*)$ given the rest by using a Metropolis–Hastings step.
Step 3: update the latent counts $\bar{n}_{ij}$ given the rest by using either the full conditional or a Metropolis–Hastings step.

F.1. Step 1: update of $w_{1:N_\alpha}$

We use an HMC update for $w_{1:N_\alpha}$ via an augmented system with momentum variables p. See Neal (2011) for on overview. Let $L \geqslant 1$ be the number of leapfrog steps and $\varepsilon > 0$ the step size. For conciseness, we write

$$U'(w_{1:N_\alpha}, w_*, \phi) = \nabla_{\omega_{1:N_\alpha}} \log\{p(\omega_{1:N_\alpha}, w_*, \phi | D_\alpha)\}|_{w_{1:N_\alpha}, w_*, \phi}$$

the gradient of the log-posterior in equation (45). The algorithm proceeds by first sampling momentum variables as

$$p \sim \mathcal{N}(0, I_{N_\alpha}). \tag{72}$$

The Hamiltonian proposal $q(\tilde{w}_{1:N_\alpha}, \tilde{p} | w_{1:N_\alpha}, p)$ is obtained by the following leapfrog algorithm (for simplicity of exposition, we omit indices $1{:}N_\alpha$). Simulate L steps of the discretized Hamiltonian via

$$\begin{aligned}\tilde{p}^{(0)} &= p + \frac{\varepsilon}{2} U'(w, w_*, \phi), \\ \tilde{w}^{(0)} &= w\end{aligned}$$

and, for $l = 1, \ldots, L-1$,

$$\begin{aligned}\log(\tilde{w}^{(l)}) &= \log(\tilde{w}^{(l-1)}) + \varepsilon \tilde{p}^{(l-1)}, \\ \tilde{p}^{(l)} &= \tilde{p}^{(l-1)} + \varepsilon U'(\tilde{w}^{(l)}, w_*, \phi)\end{aligned}$$

and finally set

$$\log(\tilde{w}) = \log(\tilde{w}^{(L-1)}) + \varepsilon \tilde{p}^{(L-1)},$$
$$\tilde{p} = -\left\{ \tilde{p}^{(L-1)} + \frac{\varepsilon}{2} U'(\tilde{w}, w_*, \phi) \right\}$$
$$\tilde{w} = \tilde{w}^{(L)}.$$

Accept the proposal $(\tilde{w}, \tilde{p})$ with probability $\min(1, r)$ with

$$r = \frac{\left(\prod_{i=1}^{N_\alpha} \tilde{w}_i^{m_i}\right) \exp\left\{ -\left(\sum_{i=1}^{N_\alpha} \tilde{w}_i + w_*\right)^2 \right\} \prod_{i=1}^{N_\alpha} \tilde{w}_i \, \rho(\tilde{w}_i)}{\left(\prod_{i=1}^{N_\alpha} w_i^{m_i}\right) \exp\left\{ -\left(\sum_{i=1}^{N_\alpha} w_i + w_*\right)^2 \right\} \prod_{i=1}^{N_\alpha} w_i \, \rho(w_i)} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{N_\alpha} (\tilde{p}_i^2 - p_i^2) \right\}$$
$$= \left\{ \prod_{i=1}^{N_\alpha} \left(\frac{\tilde{w}_i}{w_i}\right)^{m_i - \sigma} \right\} \exp\left\{ -\left(\sum_{i=1}^{N_\alpha} \tilde{w}_i + w_*\right)^2 + \left(\sum_{i=1}^{N_\alpha} w_i + w_*\right)^2 - \tau \left(\sum_{i=1}^{N_\alpha} \tilde{w}_i - \sum_{i=1}^{N_\alpha} w_i \right) \right\} \times \exp\left\{ -\frac{1}{2} \sum_{i=1}^{N_\alpha} (\tilde{p}_i^2 - p_i^2) \right\}.$$

F.2. Step 2: update of w_*, α, σ and τ

Let $g^*_{\alpha,\sigma,\tau}$ denote the density of $G^*_{\alpha,\sigma,\tau}$ with respect to the reference measure $\lambda_\alpha + \delta_0$, where we recall that λ_α denotes the Lebesgue measure over $[0, \alpha]$. For our Metropolis–Hastings step, we propose $(\tilde{\alpha}, \tilde{\sigma}, \tilde{\tau}, \tilde{w}_*)$ from $q(\tilde{\alpha}, \tilde{\sigma}, \tilde{\tau}, \tilde{w}_* | \alpha, \sigma, \tau, w_*)$ and accept with probability $\min(1, r)$ where

$$r = \frac{\exp\left\{ -\left(\sum_{i=1}^{N_\alpha} w_i + \tilde{w}_*\right)^2 \right\}}{\exp\left\{ -\left(\sum_{i=1}^{N_\alpha} w_i + w_*\right)^2 \right\}} \left\{ \prod_{i=1}^{N_\alpha} \frac{\rho(w_i | \tilde{\sigma}, \tilde{\tau})}{\rho(w_i | \sigma, \tau)} \right\} \times \frac{g^*_{\tilde{\alpha},\tilde{\sigma},\tilde{\tau}}(\tilde{w}_*)}{g^*_{\alpha,\sigma,\tau}(w_*)} \frac{p(\tilde{\alpha}, \tilde{\sigma}, \tilde{\tau})}{p(\alpha, \sigma, \tau)} \frac{q(\alpha, \sigma, \tau, w_* | \tilde{\alpha}, \tilde{\sigma}, \tilde{\tau}, \tilde{w}_*)}{q(\tilde{\alpha}, \tilde{\sigma}, \tilde{\tau}, \tilde{w}_* | \alpha, \sigma, \tau, w_*)}. \quad (73)$$

We shall use the proposal

$$q(\tilde{\alpha}, \tilde{\sigma}, \tilde{\tau}, \tilde{w}_* | \alpha, \sigma, \tau, w_*) = q(\tilde{\tau}|\tau)\, q(\tilde{\sigma}|\sigma)\, q(\tilde{\alpha}|\tilde{\sigma}, \tilde{\tau}, w_*) q(\tilde{w}_* | \tilde{\alpha}, \tilde{\sigma}, \tilde{\tau}, w_*)$$

where

$$q(\tilde{\tau}|\tau) = \text{lognormal}\{\tilde{\tau}; \log(\tau), \sigma_\tau^2\},$$
$$q(\tilde{\sigma}|\sigma) = \text{lognormal}\{1 - \tilde{\sigma}; \log(1 - \sigma), \sigma_\tau^2\}, \quad (74)$$
$$q(\tilde{\alpha}|\tilde{\sigma}, \tilde{\tau}, w_*) = \text{gamma}\{\tilde{\alpha}; N_\alpha, \psi_{\tilde{\sigma},\tilde{\tau}}(2\textstyle\sum_i w_i + 2w_*)\}$$
$$q(\tilde{w}_* | \tilde{\alpha}, \tilde{\sigma}, \tilde{\tau}, w_*) = g^*_{\tilde{\alpha},\tilde{\sigma},\tilde{\tau}+2\Sigma_i w_i + 2w_*}(\tilde{w}_*). \quad (75)$$

The choice of the proposal for w_* and α is motivated as follows. From equations (71) and (43), the conditional density of (α, w_*) given the rest is given by

$$p(\alpha, w_* | \text{rest}) \propto \alpha^{N_\alpha - 1} g^*_{\alpha,\sigma,\tau}(w_*) \exp\{-(w_* + \textstyle\sum_i w_i)^2\}$$

which is not of a standard form because of the square in the exponential. Motivated by a first-order Taylor approximation around the current MCMC value w_*,

$$(\tilde{w}_* + \textstyle\sum_i w_i)^2 \simeq (w_* + \sum_i w_i)^2 + 2(\tilde{w}_* - w_*)(w_* + \sum_i w_i),$$

we use a proposal

$$q(\tilde{\alpha}, \tilde{w}_* | \tilde{\sigma}, \tilde{\tau}, w_*) \propto \tilde{\alpha}^{N_\alpha - 1} g^*_{\tilde{\alpha},\tilde{\sigma},\tilde{\tau}}(\tilde{w}_*) \exp\{-2\tilde{w}_*(w_* + \textstyle\sum_i w_i)\}$$

which corresponds to the product of the proposals (74) and (75). The proposal for w_* can be written as an exponential tilting of the probability density function $g^*_{\tilde{\alpha},\tilde{\sigma},\tilde{\tau}}(\tilde{w}_*)$:

$$g^*_{\tilde{\alpha},\tilde{\sigma},\tilde{\tau}+2\Sigma w_i + 2w_*}(\tilde{w}_*) = \frac{\exp\{-2\tilde{w}_*(\sum w_i + w_*)\} g^*_{\tilde{\alpha},\tilde{\sigma},\tilde{\tau}}(\tilde{w}_*)}{\exp\{-\tilde{\alpha}\psi_{\tilde{\sigma},\tilde{\tau}}(2\sum_i w_i + 2w_*)\}}$$

which will allow the terms involving the intractable probability density function g^* to cancel in the Metropolis–Hastings ratio (73). g^* is either a gamma density ($\sigma = 0$), a Poisson mixture of gamma densities ($\sigma < 0$) or an exponentially tilted stable density ($\sigma > 0$) for which efficient samplers exist (Devroye, 2009; Hofert, 2011).

Under the improper priors (43), the acceptance probability reduces to having

$$r = \exp\left\{ -\left(\sum_{i=1}^{N_\alpha} w_i + \tilde{w}_*\right)^2 + \left(\sum_{i=1}^{N_\alpha} w_i + w_*\right)^2 \right\} \exp\left\{ -(\tilde{\tau} - \tau + 2w_* - 2\tilde{w}_*) \sum_{i=1}^{N_\alpha} w_i \right\}$$
$$\times \left(\prod_{i=1}^{N_\alpha} w_i\right)^{-\tilde{\sigma}+\sigma} \left\{ \frac{\Gamma(1-\sigma)\psi_{\sigma,\tau}(2\tilde{w}_* + 2\sum_i w_i)}{\Gamma(1-\tilde{\sigma})\psi_{\tilde{\sigma},\tilde{\tau}}(2w_* + 2\sum_i w_i)} \right\}^{N_\alpha}.$$

F.3. Step 3: update of the latent variables $\bar{n}_{ij}$

Concerning the latent $\bar{n}_{ij}$, the conditional distribution is a truncated Poisson distribution (44) from which we can sample directly. An alternative strategy, which may be more efficient for a large number of edges, is to use a Metropolis–Hastings random-walk proposal.

References

Aalen, O. (1992) Modelling heterogeneity in survival analysis by the compound Poisson distribution. *Ann. Appl. Probab.*, 951–972.

Adamic, L. A. and Glance, N. (2005) The political blogosphere and the 2004 US election: divided they blog. In *Proc. 3rd Int. Wrkshp Link Discovery*, pp. 36–43. New York: Association for Computing Machinery.

Airoldi, E. M., Blei, D., Fienberg, S. E. and Xing, E. (2008) Mixed membership stochastic blockmodels. *J. Mach. Learn. Res.*, **9**, 1981–2014.

Airoldi, E. M., Costa, T. B. and Chan, S. H. (2013) Stochastic blockmodel approximation of a graphon: theory and consistent estimation. In *Advances in Neural Information Processing Systems*, vol. 26.

Aldous, D. (1985) Exchangeability and related topics. In *Ecole d'Été de Probabilités de Saint-Flour XIII–1983*, pp. 1–198. Springer.

Aldous, D. (1997) Brownian excursions, critical random graphs and the multiplicative coalescent. *Ann. Probab.*, 812–854.

Aldous, D. J. (1981) Representations for partially exchangeable arrays of random variables. *J. Multiv. Anal.*, **11**, 581–598.

Arcones, M. and Giné, E. (1992) On the bootstrap of U and V statistics. *Ann. Statist.*, 655–674.

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

Bastian, M., Heymann, S. and Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. *ICWSM*, **8**, 361–362.

Berger, N., Borgs, C., Chayes, J. T. and Saberi, A. (2014) Asymptotic behavior and distributional limits of preferential attachment graphs. *Ann. Probab.*, **42**, 1–40.

Bertoin, J. (2006) *Random Fragmentation and Coagulation Processes*. Cambridge: Cambridge University Press.

Bickel, P. J. and Chen, A. (2009) A nonparametric view of network models and Newman–Girvan and other modularities. *Proc. Natn. Acad. Sci.*, **106**, 21068–21073.

Bickel, P. J., Chen, A. and Levina, E. (2011) The method of moments and degree distributions for network models. *Ann. Statist.*, **39**, 2280–2301.

Blackwell, D. and MacQueen, J. B. (1973) Ferguson distributions via Pólya urn schemes. *Ann. Statist.*, 353–355.

Bollobás, B. (1980) A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *Eur. J. Combin.*, **1**, 311–316.

Bollobás, B. (2001) *Random Graphs*. Cambridge: Cambridge University Press.

Bollobás, B., Janson, S. and Riordan, O. (2007) The phase transition in inhomogeneous random graphs. *Rand. Struct. Algs*, **31**, 3–122.

Bollobás, B. and Riordan, O. (2009). Metrics for sparse graphs. In *Surveys in Combinatorics* (eds. S. Huczynska, J. Mitchell and C. Roney-Dougal), pp. 211–287. Cambridge: Cambridge University Press.

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

Borgs, C., Chayes, J., Cohn, H. and Zhao, Y. (2014a) An L^p theory of sparse graph convergence I: Limits, sparse random graph models, and power law distributions. *Preprint arXiv:1401.2906*.

Borgs, C., Chayes, J., Cohn, H. and Zhao, Y. (2014b) An L^p theory of sparse graph convergence II: LD convergence, quotients, and right convergence. *Preprint arXiv:1408.0744*.

Borgs, C., Chayes, J. T. and Gamarnik, D. (2016) Convergent sequences of sparse graphs: a large deviations approach. *Rand. Struct. Algs*, to be published.
Borgs, C., Chayes, J. T. and Lovász, L. (2010) Moments of two-variable functions and the uniqueness of graph limits. *Geometr. Functnl Anal.*, **19**, 1597–1619.
Borgs, C., Chayes, J. T., Lovász, L., Sós, V. T. and Vesztergombi, K. (2008) Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing. *Adv. Math.*, **219**, 1801–1851.
Britton, T., Deijfen, M. and Martin-Löf, A. (2006) Generating simple random graphs with prescribed degree distribution. *J. Statist. Phys.*, **124**, 1377–1397.
Brix, A. (1999) Generalized gamma measures and shot-noise Cox processes. *Adv. Appl. Probab.*, **31**, 929–953.
Brooks, S. P. and Gelman, A. (1998) General methods for monitoring convergence of iterative simulations. *J. Computnl Graph. Statist.*, **7**, 434–455.
Bu, D., Zhao, Y., Cai, L., Xue, H., Zhu, X., Lu, H., Zhang, J., Sun, S., Ling, L. and Zhang, N. (2003) Topological structure analysis of the protein–protein interaction network in budding yeast. *Nucleic Acids Res.*, **31**, 2443–2450.
Bühlmann, H. (1960) *Austauschbare stochastische Variablen und ihre Grenzwertsätze*. *PhD Thesis*. University of California at Berkeley, Berkeley.
Caron, F. (2012) Bayesian nonparametric models for bipartite graphs. In *Advances in Neural Information Processing Systems*, vol. 25.
Caron, F. and Fox, E. B. (2014) Bayesian nonparametric models of sparse and exchangeable random graphs. *Preprint ArXiv 1401.1137*.
Caron, F., Teh, Y. W. and Murphy, T. B. (2014) Bayesian nonparametric Plackett-Luce models for the analysis of preferences for college degree programmes. *Ann. Appl. Statist.*, **8**, 1145–1181.
Chen, T., Fox, E. and Guestrin, C. (2014). Stochastic gradient Hamiltonian Monte Carlo. In *Proc. Int. Conf. Machine Learning*, pp. 1683–1691.
Choi, D. and Wolfe, P. J. (2014) Co-clustering separately exchangeable network data. *Ann. Statist.*, **42**, 29–63.
Clauset, A., Shalizi, C. R. and Newman, M. E. J. (2009) Power-law distributions in empirical data. *SIAM Rev.*, **51**, 661–703.
Colizza, V., Pastor-Satorras, R. and Vespignani, A. (2007) Reaction–diffusion processes and metapopulation models in heterogeneous networks. *Nat. Phys.*, **3**, 276–282.
Daley, D. J. and Vere-Jones, D. (2003) *An Introduction to the Theory of Point Processes, vol. I, Elementary Theory and Methods*, 2nd edn. New York: Springer.
Daley, D. J. and Vere-Jones, D. (2008) *An Introduction to the Theory of Point Processes, vol. II, General Theory and Structure* 2nd edn. New York: Springer.
Devroye, L. (2009) Random variate generation for exponentially and polynomially tilted stable distributions. *ACM Trans. Modlng Comput. Simuln*, **19**, 18.
Diaconis, P. and Janson, S. (2008) Graph limits and exchangeable random graphs. *Rend. Mat. Applic. Ser.*, VII, 33–61.
Duane, S., Kennedy, A. D., Pendleton, B. J. and Roweth, D. (1987) Hybrid Monte Carlo. *Phys. Lett.* B, **195**, 216–222.
Durrett, R. (2007). *Random Graph Dynamics*. New York: Cambridge University Press.
Favaro, S. and Teh, Y. (2013) MCMC for normalized random measure mixture models. *Statist. Sci.*, **28**, 335–359.
Feller, W. (1971) *An Introduction to Probability Theory and its Applications*, vol. II, 2nd edn. New York: Wiley.
Ferguson, T. and Klass, M. (1972) A representation of independent increment processes without gaussian components. *Ann. Math. Statist.*, **43**, 1634–1643.
Fienberg, S. E. (2012) A brief history of statistical models for network analysis and open challenges. *J. Computnl Graph. Statist.*, **21**, 825–839.
de Finetti, B. (1931) Funzione caratteristica di un fenomeno aleatorio. *Atti R. Acad. Nazn. Linc. Ser. 6*, **4**, 251–299.
Freedman, D. A. (1996) De Finetti's theorem in continuous time. *Lect. Notes Monogr. Ser.*, 83–98.
Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D., Vehtari, A. and Rubin, D. (2014) *Bayesian Data Analysis*. Boca Raton: Chapman and Hall–CRC.
Giné, E. and Zinn, J. (1992). Marcinkiewicz type laws of large numbers and convergence of moments for U-statistics. In *8th Proc. Int. Conf. Probability in Banach Spaces*, pp. 273–291.
Gnedin, A., Hansen, B. and Pitman, J. (2007) Notes on the occupancy problem with infinitely many boxes: general asymptotics and power laws. *Probab. Surv.*, **4** (146–171), 88.
Gnedin, A., Pitman, J. and Yor, M. (2006) Asymptotic laws for compositions derived from transformed subordinators. *Ann. Probab.*, **34**, 468–492.
Goldenberg, A., Zheng, A., Fienberg, S. and Airoldi, E. (2010) A survey of statistical network models. *Foundns Trends Mach. Learn.*, **2**, 129–233.
Herlau, T., Schmidt, M. N. and Mørup, M. (2014) Infinite-degree-corrected stochastic block model. *Phys. Rev.* E, **90**, 032819.
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*, vol. 29.
Hofert, M. (2011) Sampling exponentially tilted stable distributions. *ACM Trans. Modlng Comput. Simuln*, **22**, 3.

Hoff, P. D. (2009) Multiplicative latent factor models for description and prediction of social networks. *Computnl Math. Organizn Theory*, **15**, 261–272.
Hoff, P. D., Raftery, A. E. and Handcock, M. S. (2002) Latent space approaches to social network analysis. *J. Am. Statist. Ass.*, **97**, 1090–1098.
van der Hofstad, R. (2014) *Random Graphs and Complex Networks*, vol. I, *Technical Report*. Eindhoven: Eindhoven University of Technology.
Hoover, D. N. (1979) Relations on probability spaces and arrays of random variables. *Preprint*. Institute for Advanced Study, Princeton.
Hougaard, P. (1986) Survival models for heterogeneous populations derived from stable distributions. *Biometrika*, **73**, 387–396.
Jacobs, A. Z. and Clauset, A. (2014) A unified view of generative models for networks: models, methods, opportunities and challenges. *Preprint arXiv:1411.4070*.
James, L. F. (2002) Poisson process partition calculus with applications to exchangeable models and bayesian nonparametrics. *Preprint arXiv math/0205093*.
James, L. F. (2005) Bayesian Poisson process partition calculus with an application to Bayesian Lévy moving averages. *Ann. Statist.*, 1771–1799.
James, L. F., Lijoi, A. and Prünster, I. (2009) Posterior analysis for normalized random measures with independent increments. *Scand. J. Statist.*, **36**, 76–97.
Janson, S. (2011) Probability asymptotics: notes on notation. *Preprint arXiv:1108.3924*.
Kallenberg, O. (1990) Exchangeable random measures in the plane. *J. Theoret. Probab.*, **3**, 81–136.
Kallenberg, O. (2005) *Probabilistic Symmetries and Invariance Principles*. New York: Springer.
Karlin, S. (1967) Central limit theorems for certain infinite urn schemes. *J. Math. Mech.*, **17**, 373–401.
Karrer, B. and Newman, M. E. (2011) Stochastic blockmodels and community structure in networks. *Phys. Rev.* E, **83**, 016107.
Kemp, C., Tenenbaum, J. B., Griffiths, T. L., Yamada, T. and Ueda, N. (2006) Learning systems of concepts with an infinite relational model. In *AAAI*, vol. 21, pp. 381.
Khintchine, A. (1937) Zur theorie der unbeschrankt teilbaren Verteilungsgesetze. *Mat. Sborn.*, **2**, 79–119.
Kingman, J. F. C. (1967) Completely random measures. *Pacif. J. Math.*, **21**, 59–78.
Kingman, J. F. C. (1993) *Poisson Processes*, Vol. 3. New York: Oxford University Press.
Lauritzen, S. (2008). Exchangeable Rasch matrices. *Rend. Mat. Ser. VII*, **28**, 83–95.
Lee, M.-L. T. and Whitmore, G. A. (1993) Stochastic processes directed by randomized time. *J. Appl. Probab.*, 302–314.
Lewis, P. A. and Shedler, G. S. (1979) Simulation of nonhomogeneous Poisson processes by thinning. *Navl Res. Logist. Q.*, **26**, 403–413.
Lijoi, A., Mena, R. H. and Prünster, I. (2007) Controlling the reinforcement in Bayesian non-parametric mixture models. *J. R. Statist. Soc.* B, **69**, 715–740.
Lijoi, A. and Prünster, I. (2003) On a normalized random measure with independent increments relevant to Bayesian nonparametric inference. In *Proc. 13th Eur. Young Statisticians Meet.*, pp. 123–134. Bernoulli Society.
Lijoi, A. and Prünster, I. (2010) Models beyond the Dirichlet process. In *Bayesian Nonparametrics* (eds. N. L. Hjort, C. Holmes, P. Müller and S. G. Walker). Cambridge: Cambridge University Press.
Lijoi, A., Prünster, I. and Walker, S. G. (2008) Investigating nonparametric priors with Gibbs structure. *Statist. Sin.*, **18**, 1653.
Lloyd, J., Orbanz, P., Ghahramani, Z. and Roy, D. (2012) Random function priors for exchangeable arrays with applications to graphs and relational data. In *Advances in Neural Information Processing Systems*, vol. 25.
Lovász, L. (2013) *Large Networks and Graph Limits*, vol. 60. American Mathematical Society.
Lovász, L. and Szegedy, B. (2006) Limits of dense graph sequences. *J. Combin. Theory* B, **96**, 933–957.
McAuley, J. and Leskovec, J. (2012) Learning to discover social circles in ego networks. In *Advances in Neural Information Processing Systems*, vol. 25, pp. 539–547.
Miller, K., Griffiths, T. and Jordan, M. (2009) Nonparametric latent feature models for link prediction. In *Advances in Neural Information Processing Systems*, vol. 22.
Neal, R. M. (2011) MCMC using Hamiltonian dynamics. In *Handbook of Markov Chain Monte Carlo* (eds. S. Brooks, A. Gelman, G. Jones and X.-L. Meng), vol. 2. Boca Raton: Chapman and Hall–CRC.
Nešetřil, J. and Ossona de Mendez, P. (2012) *Sparsity (Graphs, Structures, and Algorithms)*. Berlin: Springer.
Newman, M. E. J. (2001) The structure of scientific collaboration networks. *Proc. Natn. Acad. Sci. USA*, **98**, 404–409.
Newman, M. E. J. (2003) The structure and function of complex networks. *SIAM Rev.*, 167–256.
Newman, M. E. J. (2010) *Networks: an Introduction*. New York: Oxford University Press.
Newman, M. E. J., Strogatz, S. H. and Watts, D. J. (2001) Random graphs with arbitrary degree distributions and their applications. *Phys. Rev.* E, **64**, 26118.
Norros, I. and Reittu, H. (2006) On a conditionally Poissonian graph process. *Adv. Appl. Probab.*, **38**, 59–75.
Nowicki, K. and Snijders, T. (2001) Estimation and prediction for stochastic blockstructures. *J. Am. Statist. Ass.*, **96**, 1077–1087.
Ogata, Y. (1981) On Lewis' simulation method for point processes. *IEEE Trans. Inform. Theory*, **27**, 23–31.

Olhede, S. C. and Wolfe, P. J. (2012) Degree-based network models. *Preprint arXiv:1211.6537*. University College London: London.
Opsahl, T. and Panzarasa, P. (2009) Clustering in weighted networks. *Socl Netwrks*, **31**, 155–163.
Orbanz, P. and Roy, D. M. (2015) Bayesian models of graphs, arrays and other exchangeable random structures. *IEEE Trans. Pattn Anal. Mach. Intell.*, **37**, 437–461.
Palla, K., Knowles, D. A. and Ghahramani, Z. (2012) An infinite latent attribute model for network data. In *Proc. Int. Conf. Machine Learning*.
Penrose, M. (2003) *Random Geometric Graphs*, vol. 5. New York: Oxford University Press.
Pitman, J. (1995) Exchangeable and partially exchangeable random partitions. *Probab. Theory Reltd Flds*, **102**, 145–158.
Pitman, J. (1996) Some developments of the Blackwell-MacQueen urn scheme. *Lect. Notes Monogr. Ser.*, 245–267.
Pitman, J. (2003) Poisson-Kingman partitions. *Lect. Notes Monogr. Ser.*, 1–34.
Pitman, J. (2006) Combinatorial stochastic processes. In *Ecole d'Eté de Probabilités de Saint-Flour XXXII–2002*. New York: Springer.
Prünster, I. (2002) Random probability measures derived from increasing additive processes and their application to Bayesian statistics. *PhD Thesis*. University of Pavia, Pavia.
Regazzini, E., Lijoi, A. and Prünster, I. (2003) Distributional results for means of normalized random measures with independent increments. *Ann. Statist.*, **31**, 560–585.
Resnick, S. (1987) *Extreme Values, Point Processes and Regular Variation*. New York: Springer.
Rohe, K., Chatterjee, S. and Yu, B. (2011) Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Statist.*, **39**, 1878–1915.
Todeschini, A. and Caron, F. (2016) Exchangeable random measures for sparse and modular graphs with overlapping communities. *Preprint arXiv:1602.02114*.
Veitch, V. and Roy, D. M. (2015) The class of random graphs arising from exchangeable random measures. *Preprint arXiv:1512.03099*.
Watts, D. J. and Strogatz, S. H. (1998) Collective dynamics of 'small-world' networks. *Nature*, **393**, 440–442.
Wolfe, P. J. and Olhede, S. C. (2013) Nonparametric graphon estimation. *Preprint arXiv:1309.5936*. University College London, London.
Zhao, Y., Levina, E. and Zhu, J. (2012) Consistency of community detection in networks under degree-corrected stochastic block models. *Ann. Statist.*, **40**, 2266–2292.