arXiv: 1401.1137

Bayesian nonparametric models of sparse and exchangeable random graphs

François Caron* **and Emily B. Fox†**

Department of Statistics
University of Oxford
1 South Parks Road
Oxford, OX1 3TG United Kingdom
e-mail: caron@stats.ox.ac.uk

Department of Statistics
University of Washington
Box 354322
Seattle, WA 98195-4322
e-mail: ebfox@stat.washington.edu

Abstract: Statistical network modeling has focused on representing the graph as a discrete structure, namely the adjacency matrix, and considering the exchangeability of this array. In such cases, the Aldous-Hoover representation theorem (Aldous, 1981; Hoover, 1979) applies and informs us that the graph is necessarily either dense or empty. In this paper, we instead consider representing the graph as a measure on $\mathbb{R}_+^2$. For the associated definition of exchangeability in this continuous space, we rely on the Kallenberg representation theorem (Kallenberg, 2005). We show that for certain choices of the specified graph construction, our network process is both exchangeable and sparse with power-law degree distribution. In particular, we build on the framework of completely random measures (CRMs) and use the theory associated with such processes to derive important network properties, such as an urn representation for network simulation. The CRM framework also provides for interpretability of the network model in terms of node-specific sociability parameters, with properties such as sparsity and power-law behavior simply tuned by three hyperparameters. Our theoretical results are explored empirically and compared to common network models.

Primary 62F15, 05C80; secondary 60G09, 60G51, 60G55.
Keywords and phrases: random graphs, Lévy measure, point process, exchangeability, generalized gamma process.

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. This effort builds on a long history, with a popular early model being the Erdös Rényi random graph (Erdös and Rényi, 1959). However, the Erdös Rényi

*FC acknowledges the support of the European Commission under the Marie Curie Intra-European Fellowship Programme.

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

formulation has since been dismissed as overly simplistic since it fails to capture important real-world network properties such as *sparsity* with *power-law* degree distribution, as in preferential attachment (Price, 1976; Barabási and Albert, 1999), or community structure, as in stochastic block models (Nowicki and Snijders, 2001) and their mixed membership variants (Airoldi et al., 2008). A plethora of other network models, both generative and discriminative, have been proposed in recent years. There are many nice overviews of network models, including (Newman, 2003, 2009; Bollobás, 2001; Durrett, 2007; Goldenberg et al., 2010; Fienberg, 2012).

In many scenarios, it is appealing conceptually to assume that the order in which nodes are observed is of no importance (Bickel and Chen, 2009; Hoff, 2009). In statistical network models, this equates with the notion of exchangeability. Classically, the graph 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 $Z_{ij} = Z_{ji}$. For generic matrices Z in some space $\mathbf{Z}$, an (infinite) *exchangeable random array* (Diaconis and Janson, 2008; Lauritzen, 2008) is one such that

$$(Z_{ij}) \stackrel{d}{=} (Z_{\pi(i)\sigma(j)}) \text{ for } (i,j) \in \mathbb{N}^2 \tag{1}$$

for any permutation π, σ of $\mathbb{N}$, with $\pi = \sigma$ in the jointly exchangeable case.

The celebrated Aldous-Hoover theorem (Aldous, 1981; Hoover, 1979) shows that infinite exchangeability implies a mixture model representation for the matrix involving transformations of uniform random variables (see Theorem 1). For undirected graphs, this transformation is described via the *graphon*, ω (see (5)).

The Aldous-Hoover constructive definition has motivated the development of Bayesian statistical models for arrays, cf. (Lloyd et al., 2012). Likewise, many popular network models can be recast in the Aldous-Hoover framework (Hoff, Raftery and Handcock, 2002; Airoldi et al., 2008; Kim and Leskovec, 2012; Miller, Griffiths and Jordan, 2009). Estimators of models in this class and their associated properties have been studied extensively in recent years (Bickel and Chen, 2009; Bickel, Chen and Levina, 2011; Rohe, Chatterjee and Yu, 2011; Zhao, Levina and Zhu, 2012; Airoldi, Costa and Chan, 2014; Wolfe and Choi).

However, one unpleasing consequence of the Aldous-Hoover theorem is that graphs represented by an exchangeable random array are either trivially empty or dense, i.e. the number of edges grows quadratically with the number of nodes. To quote the survey of Orbanz and Roy (2013) *"the theory also clarifies the limitations of exchangeable models. It shows, for example, that most Bayesian models of network data are inherently misspecified."* The conclusion is that we cannot have both exchangeability of the nodes (in the sense of (1)), a cornerstone of Bayesian model construction, and sparse graphs, which is what we observe in the real world (Newman, 2009). Several models have been developed which give up this exchangeability assumption in order to obtain sparse graphs (Barabási and Albert, 1999; Bollobás, Janson and Riordan, 2007).

Alternatively, Wolfe and Olhede (2013) propose a sparse network model via a

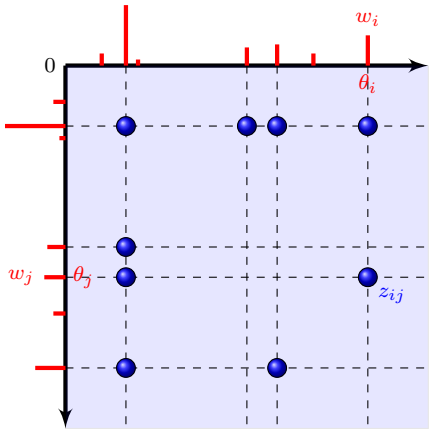


FIG 1. *Point process representation of a random graph. Each node i is embedded in $\mathbb{R}_+$ at some location θ_i and is associated with a sociability parameter w_i. An edge between nodes θ_i and θ_j is represented by a point at locations (θ_i, θ_j) and (θ_j, θ_i) in $\mathbb{R}_+^2$.*

scaling of the graphon with network size n, $\rho_n \omega$, leading to a sequence of *finitely exchangeable* graphs and a consistent estimator of the graphon, ω. However, such a formulation lacks a fully generative specification.

In this paper, we present the first class of generative models for graphs which is able to handle both sparsity and infinite exchangeability. To achieve this goal, instead of the discrete array structure of the adjacency matrix, we consider a continuous-space representation of networks based on a *point process* on $\mathbb{R}_+^2$ (see Figure 1)

$$Z = \sum_{i,j} z_{ij} \delta_{(\theta_i, \theta_j)}, \tag{2}$$

where $z_{ij} = 1$ if there is a link between nodes θ_i and θ_j in $\mathbb{R}_+$, and is 0 otherwise. Paralleling (1), the point process Z on $\mathbb{R}_+^2$ is exchangeable if and only if, for any permutations π, σ of $\mathbb{N}$,

$$(Z(A_i \times A_j)) \stackrel{d}{=} (Z(A_{\pi(i)} \times A_{\sigma(j)})) \text{ for } (i,j) \in \mathbb{N}^2, \tag{3}$$

where here we consider *intervals* $A_i = [h(i-1), hi]$, $h > 0$ and $i \in \mathbb{N}$.

In place of the Aldous-Hoover theorem, we now appeal to the continuous-space counterpart (Kallenberg, 2005, Chapter 9) which provides a representation theorem for exchangeable point processes on $\mathbb{R}_+^2$: a point process is exchangeable if and only if it can be represented as a transformation of unit-rate Poisson processes and uniform random variables (see Theorem 2). We show that by carefully choosing this transformation, or equivalently the corresponding Lévy measure on $\mathbb{R}_+^2$, we are able to construct sparse graphs while preserving exchangeability in this point process framework. In particular, when building on

a specific class of Lévy measures called *generalized gamma processes*, we can obtain graphs where the number of nodes increases at a rate below n^a for some constant $1 < a < 2$ that depends on the model parameters. In summary, our proposed framework captures a number of desirable properties:

- **Power Law.** Our formulation yields a power-law form, which is useful in practice since many real-world graphs exhibit power-law degree distributions (Newman, 2009).
- **Sparsity.** We can obtain graphs where the number of nodes increases sub-quadratically.
- **Exchangeability.** Our formulation provides an exchangeable process (in the sense of (3)).
- **Simplicity.** The resulting formulation is parsimonious: three hyperparameters tune the expected number of nodes, power-law properties, etc.
- **Interpretability** An interpretation of our generative model is in terms of *sociability* parameters $w_i > 0$, that influence the probability of node i forming an edge with any other node j as $1 - \exp(-2w_i w_j)$. This single node parameter leads to straightforward interpretability of the model.

By building on the framework of completely random measures (CRMs) (Kingman, 1967, 1993; Lijoi and Prünster, 2010), we are able to harness the considerable theory of such processes to derive important properties of our proposed model. For example, our sparsity result utilizes an urn representation provided by a reformulation of our model in terms of normalized CRMs. Thus, our generative specification enjoys both an analytic representation in the Kallenberg framework and a formulation in terms of CRMs. The former allows us to nicely connect with existing random graph models whereas the latter provides interpretability and theoretical analysis of the formulation beyond what is leant by exchangeability alone.

Our paper is organized as follows. In Section 2, we provide background on exchangeability for sequences, arrays, and random measures on $\mathbb{R}_+^2$. The latter provides an important theoretical foundation for the graph structures we propose. We also present background on CRMs, which form the key building block within our graph construction. The generic formulation for directed multigraphs, undirected graphs, and bipartite graphs is presented in Section 3 with specific cases considered in Section 5. The associated generic graph properties and those tailored to the special cases, including methods for simulation, are presented in Section 4 and 5, respectively. Section 6 provides an empirical analysis of network properties of our proposed formulation relative to common network models.

2. Background

2.1. Exchangeability and de Finetti-type representation theorems

Our focus is on exchangeable random structures that can represent networks. To build to such constructs, we first present a brief review of exchangeability

for random sequences, continuous-time processes, and discrete network arrays. Thorough and accessible overviews of exchangeability of random structures are presented in the surveys of Aldous (1985) and Orbanz and Roy (2013). Here, we simply abstract away the notions relevant to placing our network formulation in context, as summarized in Table 1.

TABLE 1
Overview of representation theorems

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

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 exists a random probability measure Θ on $\mathbf{Z}$ with law ν such that the Z_i are conditionally i.i.d. given Θ. That is, all exchangeable infinite sequences can be represented as a mixture with directing measure Θ and mixing measure ν. If examining continuous-time *processes* instead of sequences, the representation associated with exchangeable *increments* is given by Bühlmann (1960) (see also Freedman (1996)) in terms of mixing Lévy processes.

The focus of our work, however, is on graph structures. Recall the definition of exchangeability of arrays in (1). A representation theorem for exchangeability of the classical discrete adjacency *matrix*, Z, follows in Theorem 1 by considering a special case of the Aldous-Hoover theorem to *2-arrays*. We additionally focus here on *joint exchangeability*—that is, symmetric permutations of rows and columns—because of the applicability to our graph structures 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 we consider in Section 3.3. Extensions of Theorem 1 to higher dimensional arrays are likewise straightforward (Orbanz and Roy, 2013).

Theorem 1 (Aldous-Hoover representation of jointly exchangeable matrices (Aldous, 1981; Hoover, 1979)). *A random 2-array $(Z_{ij})_{i,j\in\mathbb{N}}$ is jointly exchangeable if and only if there exists a random measurable function $f : [0,1]^3 \to \mathbf{Z}$ such that*

$$(Z_{ij}) \stackrel{d}{=} (f(U_i, U_j, U_{ij})), \tag{4}$$

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 i.i.d. 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* $\omega : [0,1]^2 \to [0,1]$, symmetric in its arguments, where

$$f(U_i, U_j, U_{ij}) = \begin{cases} 1 & U_{ij} < \omega(U_i, U_j) \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Exchangeability is a fundamentally important concept in modeling. For example, an assumption of joint exchangeability in network models implies that the probability of a given graph depends on certain structural features, such as number of edges, triangle, and five-stars, but not on where these features occur in the network. Likewise, for separate exchangeability, the probability of the matrix is invariant to reordering of the rows and columns, e.g., users and items. However, based on the Aldous-Hoover representation theorem, one can derive the important consequence that *if a random graph is exchangeable, it is either dense or empty.* Note, crucially, that this result assumes the graph is modeled via a *discrete* adjacency matrix structure.

We instead consider here that a graph will be represented as a point process $Z = \sum_{i,j} z_{ij}\delta_{(\theta_i,\theta_j)}$ with nodes θ_i embedded in $\mathbb{R}_+$, as in (2). 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 Theorem 2. Recall the definition of joint exchangeability of a random measure on $\mathbb{R}_+$ in (3). In the following, 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 $\widetilde{\mathbb{N}}_2 = \{\{i,j\} | (i,j) \in \mathbb{N}^2\}$. We also define a *U-array* to be an array of independent uniform random variables.

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

$$\begin{aligned}\xi = &\sum_{i,j} f(\alpha,\vartheta_i,\vartheta_j,\zeta_{\{i,j\}})\delta_{\theta_i,\theta_j} + \beta\lambda_D + \gamma(\lambda\times\lambda)\\ &+ \sum_{j,k}\left(g(\alpha,\vartheta_j,\chi_{jk})\delta_{\theta_j,\sigma_{jk}} + g'(\alpha,\vartheta_j,\chi_{jk})\delta_{\sigma_{jk},\theta_j}\right)\\ &+ \sum_{j}\left(h(\alpha,\vartheta_j)(\delta_{\theta_j}\times\lambda) + h'(\alpha,\vartheta_j)(\lambda\times\delta_{\theta_j})\right)\\ &+ \sum_{k}\left(l(\alpha,\eta_k)\delta_{\rho_k,\rho'_k} + l'(\alpha,\eta_k)\delta_{\rho'_k,\rho_k}\right)\end{aligned} \tag{6}$$

for some measurable functions $f : \mathbb{R}_+^4 \to \mathbb{R}_+$, $g : \mathbb{R}_+^4 \to \mathbb{R}_+$ and h, h', l, l' from $\mathbb{R}_+^2$ to $\mathbb{R}_+$. Here, $(\zeta_{\{i,j\}})$ with $\{i,j\} \in \widetilde{\mathbb{N}}_2$ 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, $\alpha,\beta,\gamma \geq 0$ are an independent set of random variables.

We place our proposed network model of Section 3 within this Kallenberg representation in Section 4.1.

2.2. Completely Random Measures

Our models for graphs build on the completely random measure (CRM) (Kingman, 1967) framework. CRMs have been used extensively in the Bayesian nonparametric literature for proposing flexible classes of priors over functional

spaces, cf. (Regazzini, Lijoi and Prünster, 2003; Lijoi and Prünster, 2010). We recall in this section basic properties of CRMs and describe some examples that will be of particular interest later on. The reader can refer to the monograph of Kingman (1993) for an exhaustive coverage.

A CRM W on $\mathbb{R}_+$ is a random measure such that for any countable number of disjoint measurable sets $A_1, A_2, \ldots$ of $\mathbb{R}_+$, the random variables $W(A_1), W(A_2), \ldots$ are independent and

$$W(\cup_j A_j) = \sum_j W(A_j). \tag{7}$$

If one additionally assumes that the distribution of $W([t,s])$ only depends on $t-s$, (i.e. we have i.i.d. increments of fixed size) then the CRM takes the following form

$$W = \sum_{i=1}^{\infty} w_i \delta_{\theta_i}, \tag{8}$$

where $(w_i)_{i\in\mathbb{N}}, (\theta_i)_{i\in\mathbb{N}}$ are independent sequences of random variables on $\mathbb{R}_+$; moreover, the Laplace transform of $W(A)$ for any measurable set A admits the following representation. For any $t > 0$,

$$\mathbb{E}[\exp(-tW(A))] = \exp\left(-\int_{\mathbb{R}_+\times A} [1-\exp(-tw)]\,\rho(dw)\lambda(d\theta)\right), \tag{9}$$

where ρ is a measure on $\mathbb{R}_+$ such that

$$\int_0^{\infty} (1-e^{-w})\rho(dw) < \infty. \tag{10}$$

The measure ρ is referred to as the jump part of the Lévy measure of the CRM W. For a CRM W with i.i.d. increments, which are intimately connected to subordinators (Kingman, 1993, Chapter 8), ρ characterizes these increments. We denote this process as $W \sim \mathrm{CRM}(\rho, \lambda)$. Note that $W([0,T]) < \infty$ for any $T < \infty$, while $W(\mathbb{R}_+) = \infty$ if ρ is not degenerate at 0.

The jump part ρ of the Lévy measure is of particular interest for our construction for graphs. If it satisfies the condition

$$\int_0^{\infty} \rho(dw) = \infty \tag{11}$$

then there will be an infinite number of jumps in any interval $[0,T]$. Otherwise, the number of jumps will be finite almost surely. In our case, these jumps will map directly to the nodes in the graph. We will in particular consider the following cases

1. Poisson process with fixed increments a and

$$\rho(dw) = \delta_{w_0}(dw),$$

where δ_{w_0} is the dirac delta mass at $w_0 > 0$.

2. Compound Poisson process, where
$$\rho(dw) = h(w)dw$$
and $h : \mathbb{R}_+ \to \mathbb{R}_+$ is such that $\int_0^\infty h(w)dw = 1$.
3. Gamma process, where
$$\rho(dw) = w^{-1}\exp(-\tau w)dw.$$
with $\tau > 0$. In this case, the CRM has an infinite number of jumps over any interval $[s, t]$.
4. Generalized gamma process (Brix, 1999; Lijoi, Mena and Prünster, 2007), where
$$\rho(dw) = \frac{1}{\Gamma(1-\sigma)} w^{-1-\sigma}\exp(-\tau w)dw, \tag{12}$$
with $\sigma \in [0, 1)$ and $\tau \geq 0$. It includes as a special case the gamma process ($\sigma = 0$), the stable process ($\tau = 0$) and the inverse-Gaussian process ($\sigma = \frac{1}{2}$). The CRM has an infinite number of jumps over any interval $[s, t]$.

3. Statistical network models

Our primary focus is on undirected network models, but implicit in our construction is the definition of a directed integer-weighted, or *multigraph*, which in some applications might 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.

Our undirected graph simply transforms the directed multigraph by forming an undirected edge if there is any directed edge between two nodes. Due to the straightforward relationship between the two graphs, much of the intuition gained from the directed case carries over to the undirected scenario.

3.1. Directed multigraphs

Let $V = (\theta_1, \theta_2, ...)$ be a countably infinite set of nodes with $\theta_i \in \mathbb{R}_+$.

We represent the directed multigraph of interest 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. See Figure 2 for an illustration of the restriction of D to $[0, 1]^2$ and the corresponding directed graph.

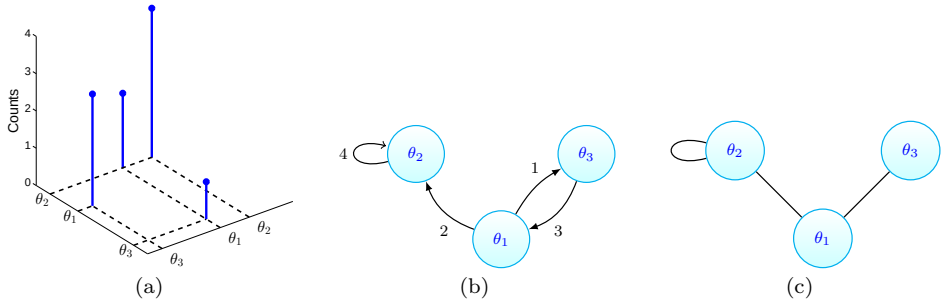


FIG 2. *An example of (a) the restriction on* $[0,1]^2$ *of an atomic measure* D*, (b) the corresponding directed multigraph, and (c) corresponding undirected graph.*

Our generative approach for modeling D associates with each node θ_i a *sociability* parameter $w_i > 0$ and employs a homogeneous CRM to define the atomic measure

$$W = \sum_{i=1}^{\infty} w_i \delta_{\theta_i} \quad W \sim \mathrm{CRM}(\rho, \lambda). \tag{14}$$

Given W, D is simply generated from a Poisson process (PP) with intensity given by the product measure $\widetilde{W} = W \times W$ on $\mathbb{R}_+^2$:

$$D \mid W \sim \mathrm{PP}(W \times W).$$

That is, informally, the individual counts n_{ij} are generated as $\mathrm{Poisson}(w_i w_j)$. Note that by construction, for any $A, B \subset \mathbb{R}$, $\widetilde{W}(A \times B) = W(A)W(B)$. Therefore, on any bounded interval A of $\mathbb{R}_+$, since $W(A) < \infty$, $\widetilde{W}(A \times A)$ has finite mass as well.

The full generative model for the directed multigraph is therefore simply

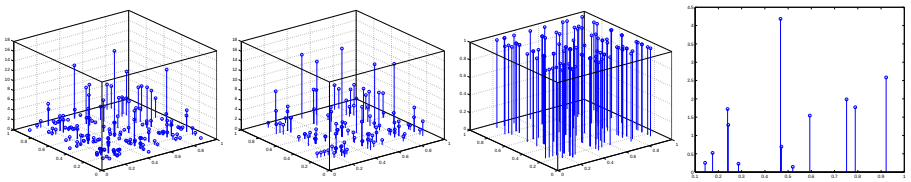
$$W \sim \mathrm{CRM}(\rho, \lambda)$$
$$D|W \sim \mathrm{PP}(W \times W).$$

3.2. Undirected graphs

We now turn to the primary focus of modeling undirected graphs. Similarly to the directed case of Section 3.1, 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. We arise at the undirected graph via a simple transformation of the directed graph: set $z_{ij} = z_{ji} = 1$ if $n_{ij} + n_{ji} > 0$



(a) Intensity measure $\widetilde{W}$ (b) Integer point process D (c) Point process Z (d) CRM W

FIG 3. *An example of (a) the product measure $\widetilde{W} = W \times W$, (b) a draw of the directed multigraph measure $D \mid W \sim PP(W \times W)$, (c) corresponding undirected measure $Z = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \min(n_{ij}, 1)\delta_{(\theta_i,\theta_j)}$ and (d) the sociability measure W used above.*

and $z_{ij} = z_{ji} = 0$ otherwise. That is, 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. Note that 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 as follows:

$$\begin{array}{ll} W = \sum_{i=1}^{\infty} w_i \delta_{\theta_i} & W \sim \mathrm{CRM}(\rho, \lambda) \\ D = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} n_{ij}\delta_{(\theta_i,\theta_j)} & D \mid W \sim \mathrm{PP}\left(W \times W\right) \\ Z = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \min(n_{ij} + n_{ji}, 1)\delta_{(\theta_i,\theta_j)}. & \end{array} \tag{15}$$

This process is depicted graphically in Figure 3.

Equivalently, given the sociability parameters $w = \{w_i\}$, we can directly specify the undirected graph model as

$$\Pr(z_{ij} = 1 \mid w) = \begin{cases} 1 - \exp(-2w_i w_j) & i \neq j \\ 1 - \exp(-w_i^2) & i = j. \end{cases} \tag{16}$$

To see the equivalence between this formulation and the one obtained from manipulating the directed multigraph, note that for $i \neq j$, $\Pr(z_{ij} = 1 \mid w) = \Pr(n_{ij} + n_{ji} > 0 \mid w)$. 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 $2w_i w_j$. The result (16) arises from the fact that $\Pr(n_{ij} + n_{ji} > 0 \mid w) = 1 - \Pr(n_{ij} + n_{ji} = 0 \mid w)$. Likewise, the $i = j$ case arises using a similar reasoning for $\Pr(z_{ii} = 1 \mid w) = \Pr(n_{ii} > 0 \mid w)$.

Our general network process is defined on $\mathbb{R}_+^2$ and, due to the fact that $W(\mathbb{R}_+) = \infty$, yields an infinite number of edges. In applications, we are typically interested in considering graphs with a finite number of edges. We therefore consider restrictions D_α and Z_α of D and Z, respectively, to the box $[0, \alpha]^2$. We also denote by W_α and λ_α the corresponding CRM and Lebesgue measure on $[0, \alpha]$. We write $Z_\alpha^* = Z_\alpha([0, \alpha]^2)$, the total mass on $[0, \alpha]^2$, and similarly for D_α^* and W_α^*. By definition, D_α is drawn from a Poisson process with finite mean measure $W_\alpha \times W_\alpha$, so we have the following generative model for simulating D_α

and Z_α directly:

$$W_\alpha \sim \text{CRM}(\rho, \lambda_\alpha)$$
$$D_\alpha^* | W_\alpha^* \sim \text{Poisson}(W_\alpha^{*\,2}).$$

For $k = 1, \ldots, D_\alpha^*$ and $j = 1, 2$

$$U_{kj} | W_\alpha \overset{iid}{\sim} \frac{W_\alpha}{W_\alpha^*}$$
$$D_\alpha = \sum_{k=1}^{D_\alpha^*} \delta_{(U_{k1}, U_{k2})}. \tag{17}$$

Here, the variables $U_{kj} \in \mathbb{R}_+$ correspond to nodes in the graph, and pairs of variables (U_{k1}, U_{k2}) correspond to a directed edge from node U_{k1} to node U_{k2}. The number of such directed edges, D_α^*, depends on the total mass of the CRM, W_α^*. For each such directed edge, the defining nodes U_{kj} are drawn from a normalized CRM, $\frac{W_\alpha}{W_\alpha^*}$. As the normalized CRM $\frac{W_\alpha}{W_\alpha^*}$ is discrete with probability 1, the U_{kj} take a number $N_\alpha \leq 2D_\alpha^*$ of distinct values. Here, N_α corresponds to the number of nodes with degree at least one in the network. Recall that the undirected network construction simply forms an undirected edge between a set of nodes if there exists 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 $N_\alpha^{(e)} \leq D_\alpha^*$ of distinct values, where $N_\alpha^{(e)}$ corresponds to the number of edges in the undirected network.

The construction (17), derived from the construction of the directed graph as a Cox process, enables us to re-express the model in terms of normalized CRMs (Regazzini, Lijoi and Prünster, 2003). This is very attractive both practically and theoretically; as we show in Section 5, one can use this framework to build on the various results on urn processes and power-law properties of normalized CRMs in order to get exact samplers for our graph models as well as to show its sparsity.

Finite-dimensional generative process We now describe this urn formulation, that allows us to obtain a finite dimensional generative process. 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 random measure $\mu_\alpha = \frac{W_\alpha}{W_\alpha^*}$ in (17) and to 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 a number $k \leq n$ of distinct values $\widetilde{U}'_j$, with multiplicities $1 \leq m_j \leq k$. The distribution on the underlying partition is usually defined in terms of an exchangeable partition probability function (EPPF) (Pitman, 1995) $\Pi_n^{(k)}(m_1, \ldots, m_k | W_\alpha^*)$ which is symmetric in its arguments. Note that the EPPF associated to the normalized CRM depends in general on the total mass of the CRM. The predictive distribution of U'_{n+1} given

$(W_\alpha^*, U_1', \ldots, U_n')$ can be given in terms of the EPPF:

$$U_{n+1}'|(W_\alpha^*, U_1', \ldots, U_n') \sim \frac{\Pi_{n+1}^{(k+1)}(m_1, \ldots, m_k, 1|W_\alpha^*)}{\Pi_n^{(k)}(m_1, \ldots, m_k|W_\alpha^*)} \frac{1}{\alpha}\lambda_\alpha + \sum_{j=1}^{k} \frac{\Pi_{n+1}^{(k)}(m_1, \ldots, m_j+1, \ldots, m_k|W_\alpha^*)}{\Pi_n^{(k)}(m_1, \ldots, m_k|W_\alpha^*)}\delta_{\widetilde{U}_j'}. \quad (18)$$

We now summarize the whole generative process, where $P_{W_\alpha^*}$ is the distribution of W_α^*:

$$\begin{aligned} W_\alpha^* &\sim P_{W_\alpha^*} \\ D_\alpha^*|W_\alpha^* &\sim \text{Poisson}({W_\alpha^*}^2). \\ (U_{kj})_{k=1,\ldots,D_\alpha^*;j=1,2}|W_\alpha^* &\sim \text{Urn process (18)} \\ D_\alpha &= \sum_{k=1}^{D_\alpha^*} \delta_{(U_{k1},U_{k2})}. \end{aligned} \quad (19)$$

The above alternative representation of the model can be used to sample exactly from our graph model; it requires us to be able to sample from the distribution $P_{W_\alpha^*}$ as well as to evaluate the EPPF. In Section 5 we show that this is indeed possible for the generalized gamma process class of CRMs. If this is not possible, in Section 4.3 we present alternative, though potentially more computationally complex, methods for simulation.

Remark 3 *The urn construction highlights a connection with the configuration model (Bollobás, 1980; Newman, 2009), a popular model for generating simple graphs with a given degree sequence. The configuration model 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.*

3.3. 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 set 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 using an atomic measure on $\mathbb{R}_+^2$

$$D = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} n_{ij}\delta_{(\theta_i,\theta_j')}, \quad (20)$$

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 CRMs $W \sim \text{CRM}(\rho,\lambda)$ and $W' \sim \text{CRM}(\rho',\lambda)$, 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 one:

$$\begin{array}{ll} W = \sum_{i=1}^{\infty} w_i\delta_{\theta_i} & W \sim \text{CRM}(\rho,\lambda) \\ W' = \sum_{j=1}^{\infty} w'_j\delta_{\theta'_j} & W' \sim \text{CRM}(\rho',\lambda) \\ D = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} n_{ij}\delta_{(\theta_i,\theta'_j)} & D \mid W,W' \sim \text{PP}\left(W \times W'\right) \\ Z = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \min(n_{ij},1)\delta_{(\theta_i,\theta'_j)}. & \end{array} \tag{21}$$

The model (21) has been proposed by Caron (2012) in a slightly different formulation. Here we cast the model within our general framework for sparse and exchangeable networks, garnering the properties derived in this paper.

4. Simulation and general properties

We provide here general properties of our network model that apply for any choice of Lévy intensity ρ. In the next section, we provide more refined properties, depending on specific choices of ρ.

4.1. Exchangeability under the Kallenberg framework

Proposition 4 (Joint exchangeability of the undirected graph measure).
For any CRM $W \sim CRM(\rho,\lambda)$, the point process Z defined by (15), or equivalently by (16), is jointly exchangeable.

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

$$(W(A_i)) \overset{d}{=} (W(A_{\pi(i)})) \tag{22}$$

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

$$(D(A_i \times A_j)) \overset{d}{=} (D(A_{\pi(i)} \times A_{\pi(j)})) \tag{23}$$

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

We now reformulate our network process in the Kallenberg representation of (6). Due to exchangeability, we know that such a representation exits. 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 $\mathbb{R}_+^2$ using the inverse Lévy method (Khintchine, 1937; Ferguson and Klass, 1972). Let (θ_i, ϑ_i) be a unit-rate Poisson process on $\mathbb{R}_+^2$. Define $L(x) = \int_x^\infty \rho(dw)$. Then the CRM $W = \sum w_i \delta_{\theta_i}$ with Lévy measure $\rho(dw)d\theta$ can be constructed from the bi-dimensional point process by taking $w_i = L^{-1}(\vartheta_i)$. L^{-1} is a monotone function, known as the inverse Lévy intensity. It follows that our undirected graph model can be formulated under the representation of (6) by selecting

$$f(\alpha, \vartheta_i, \vartheta_j, \zeta_{\{i,j\}}) = \begin{cases} 1 & \zeta_{\{i,j\}} \leq M(\vartheta_i, \vartheta_j) \\ 0 & \text{otherwise} \end{cases} \tag{24}$$

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

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

In Section 5, we provide explicit forms for L depending upon our choice of Lévy intensity ρ. The expression (24) represents a direct analog to that of (5) arising from the Aldous-Hoover framework. In particular, M here is akin to the graphon ω, 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 provided in Figure 4. Note that had we started from the Kallenberg representation and selected an f (or M) arbitrarily, we would likely not have yielded a network model with the normalized CRM interpretation that enables both interpretability and analysis of network properties, such as those presented in Sections 5.3 and 5.4.

Such a statement of joint exchangeability and the associated Kallenberg representation is straightforward to derive for the directed multigraph measure D as well. For the bipartite graph, an application of Kallenberg's representation theorem for *separate* exchangeability can likewise be made.

4.2. 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

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

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

4.3. Simulation

To simulate an undirected graph, we harness the directed multigraph representation. That is, we first sample a directed graph and then transform it to an

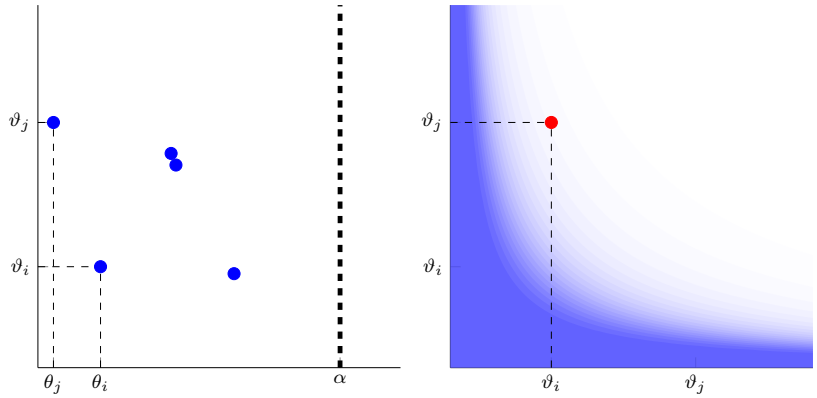


FIG 4. *Illustration of the model construction based on the Kallenberg representation. (left) A unit-rate Poisson process* (θ_i, ϑ_i), $i \in \mathbb{N}$ *on* $[0, \alpha] \times \mathbb{R}_+$. *(right) For each pair* $\{i, j\} \in \widetilde{\mathbb{N}}^2$, *set* $z_{ij} = z_{ji} = 1$ *with probability* $M(\vartheta_i, \vartheta_j)$. *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 are analytic expressions of* L^{-1} *and therefore* M.

undirected graph as described in Section 3.2. One might imagine simulating a directed network by first sampling W_α and then sampling D_α given W_α. However, recall that W_α may have an infinite number of jumps. One approach to coping with this issue, which is possible for some Lévy intensities ρ, is to resort to adaptive thinning (Lewis and Shedler, 1979; Ogata, 1981; Favaro and Teh, 2012). A related alternative approach, but applicable to any Lévy intensity ρ satisfying (11), is the inverse Lévy method. This method first defines a threshold ε and then samples the weights $\Omega = \{w_i | w_i > \varepsilon\}$ 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 (15) or (16), respectively, can prove computationally problematic since a large number of possible edges must be considered (one Poisson/Bernoulli draw for each θ_i, θ_j pair for the directed/undirected case). Instead, we can harness the Cox process representation and resulting sampling procedure of (17) that first samples the total number of directed edges and then their specific instantiations. More specifically, to simulate a point process on $[0, \alpha]^2$, we use the inverse Lévy method to sample

$$\Pi_{\alpha,\varepsilon} = \{(w, \theta) \in \Pi, 0 < \theta \leq \alpha, w > \varepsilon\}. \tag{25}$$

Let $W_{\alpha,\varepsilon} = \sum_{i=1}^{K} w_i \delta_{\theta_i}$ be the associated truncated CRM and $W^*_{\alpha,\varepsilon} = W_{\alpha,\varepsilon}([0, \alpha])$. We then sample $D^*_{\alpha,\varepsilon}$ and $U_{k,j}$ as in (17) and set $D_{\alpha,\varepsilon} = \sum_{k=1}^{D^*_{\alpha,\varepsilon}} \delta_{U_{k1}, U_{k2}}$. The undirected graph measure $Z_{\alpha,\varepsilon}$ is set to the manipulation of $D_{\alpha,\varepsilon}$ as in (15).

In the next section, we show that it is possible to sample a graph exactly when considering the classes of (generalized) gamma processes by resorting to an urn scheme.

5. Special cases

In this section, we review the properties of various models depending on the Lévy measure ρ, and their link to classical random graph models. We show that in some cases, our proposed formulation can lead to sparse random graphs in which the number of edges increases at rate $o(n^a)$, where n is the number of nodes and $1 < a < 2$. We focus on the undirected graph case, but similar results can be obtained for directed multigraphs and bipartite graphs.

5.1. Poisson process

In this case, we have

$$L(x) = \begin{cases} 1 & \text{if } x < w_0 \\ 0 & \text{otherwise} \end{cases}$$

and the model can be defined as follows, ignoring self-edges. Sample $n \sim \text{Poisson}(\alpha)$. For $i = 1, \dots n$, sample $\theta_i \sim \text{Uniform}([0, \alpha])$. 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 where the number of edges grows quadratically with the number of nodes n.

5.2. Compound Poisson process

In this case, we have

$$L(x) = 1 - H(x)$$

where H is the distribution function associated with h.

Here, we arrive at a framework that is similar to the standard graphon. Specifically, sample $n \sim \text{Poisson}(\alpha)$. Then, for $i = 1, \dots n$ set $z_{ij} = z_{ji} = 1$ with probability $M(U_i, U_j)$ where U_i are uniform 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, where the number of nodes is random and follows a Poisson distribution. As such, the resulting random graph is either trivially empty or dense.

5.3. Gamma process

In this case, we have

$$L(x) = E_1(\tau x)$$

where $E_1(z) = \int_1^\infty t^{-1} \exp(-zt)dt$ is the exponential integral.

Exact sampling via an urn approach In the case of the gamma process, the distribution of the total mass W_α^* is Gamma(α, τ). The urn process is known as the Blackwell-MacQueen urn process (Blackwell and MacQueen, 1973; Pitman, 1996)

$$U'_{n+1}|(W_\alpha^*, U'_1, \ldots, U'_n) \sim \frac{1}{\alpha+n}\lambda_\alpha + \sum_{j=1}^{k} \frac{m_j}{\alpha+n}\delta_{\widetilde{U}'_j}. \tag{26}$$

One can therefore use the generative process (19) in order to sample exactly from the model.

Expected number of nodes and edges The urn process described above is useful in deriving approximations for the mean and variance of the number of nodes with degree at least one and of the number of edges in the multigraph, as given in Proposition 5. The proof is in the Appendix.

Proposition 5 *Let N_α be the number of nodes with at least one edge. Then for $\alpha \gg \tau^2 > 0$*

$$\mathbb{E}[N_\alpha] \simeq \alpha\left(\log\left(\frac{2\alpha}{\tau^2}\right)\right) \quad Var[N_\alpha] \simeq \alpha\left(\log\left(\frac{2\alpha}{\tau^2}\right) + 4\alpha\psi(1,\alpha)\right).$$

where $\psi(1,\alpha)$ is the first derivative of the digamma function evaluated at α. Let D_α^* be the number of edges in the directed multigraph. Then

$$\mathbb{E}[D_\alpha^*] = \frac{\alpha(\alpha+1)}{\tau^2} \quad Var[D_\alpha^*] = \frac{\alpha(\alpha+1)}{\tau^2}\left(1 + \frac{4\alpha+6}{\tau^2}\right).$$

5.4. Generalized gamma process

In this case, we have

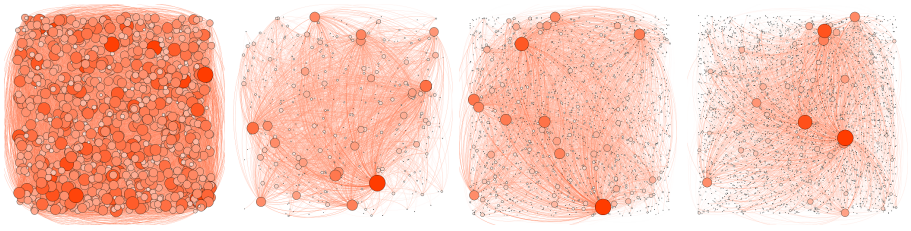
$$L(x) = \int_x^\infty \frac{1}{\Gamma(1-\sigma)} w^{-1-\sigma}\exp(-\tau w)dw = \begin{cases} \frac{\tau^\sigma \Gamma(-\sigma,\tau x)}{\Gamma(1-\sigma)} & \text{if } \tau > 0 \\ \frac{x^{-\sigma}}{\Gamma(1-\sigma)\sigma} & \text{if } \tau = 0 \end{cases}$$

where $\Gamma(a,x)$ is the incomplete gamma function.

Example realizations of the process for various values of σ are displayed in Figure 5 alongside a realization of an Erdös Rényi graph.

Exact sampling via an urn approach In the generalized gamma process case, W_α^* is an exponentially tilted stable random variable, for which exact samplers exist (Devroye, 2009). As shown by Pitman (2003) (see also (Lijoi, Prünster and Walker, 2008)), the EPPF conditional on the total mass $W_\alpha^* = t$ only depends on the parameter σ (and not τ, α) and is given by

$$\Pi_k^{(n)}(m_1,\ldots,m_k|t) = \frac{\sigma^k t^{-n}}{\Gamma(n-k\sigma)g_\sigma(t)}\int_0^t s^{n-k\sigma-1}g_\sigma(t-s)ds\left(\prod_{i=1}^k \frac{\Gamma(m_i-\sigma)}{\Gamma(1-\sigma)}\right) \tag{27}$$



(a) $G(1000, 0.05)$ (b) $GGP(100, 2, 0)$ (c) $GGP(100, 2, 0.5)$ (d) $GGP(100, 2, 0.8)$

FIG 5. *Sample graphs: (a) Erdös-Rényi graph $G(n, p)$ with $n = 1000$ and $p = 0.05$ (b-c) Generalized gamma process graph $GGP(\alpha, \tau, \sigma)$ with $\alpha = 100$, $\tau = 2$ and (b) $\sigma = 0$, (c) $\sigma = 0.5$, (d) $\sigma = 0.8$. The size of a node is proportional to its degree. Graphs have been generated with the software Gephi.*

where g_σ is the pdf of the positive stable distribution. Plugging the EPPF of (27) in to (18) yields the urn process for sampling in the generalized gamma process case.

Power-law properties

Theorem 6 *Let $N_{\alpha,j}$, $1 \leq j$ be the number of nodes in the directed graph D with j outgoing or ingoing edges (a self edge counts twice for a given node). Then we have the following asymptotic results*

$$\frac{N_{\alpha,j}}{N_\alpha} \xrightarrow{\alpha \to \infty} p_{\sigma,j} = \frac{\sigma\Gamma(j-\sigma)}{\Gamma(1-\sigma)\Gamma(j+1)}, \tag{28}$$

almost surely. In particular, for large j, we have the following tail behavior

$$p_{\sigma,j} \sim \frac{\sigma}{\Gamma(1-\sigma)} j^{-1-\sigma} \tag{29}$$

corresponding to a power-law behavior.

The proof, which builds on the asymptotic properties of normalized generalized gamma processes (Lijoi, Mena and Prünster, 2007), is given in the Appendix.

Sparsity of the graph The following theorem states that the number of edges grows sub-quadratically in the number of nodes in the graph. There is some discrepancy in the literature as to the definition of network "sparsity", and some may simply refer to our network growth as "not dense".

Theorem 7 *Let N_α be the number of nodes and $N_\alpha^{(e)}$ be the number of edges in the undirected graph. Assume $\sigma > 0$ and let $0 < \varepsilon < \sigma$ be some small constant. Then*

$$N_\alpha^{(e)} = O(N_\alpha^{2-\sigma+\varepsilon})$$

almost surely as $\alpha \to \infty$, i.e. the underlying graph is sparse.

The proof, which builds on Theorem 6, is given in the Appendix.

6. Empirical analysis of graph properties

For the GGP-based formulation, we provide an empirical analysis of our network properties in Figure 6 by simulating undirected graphs using the approach described in Section 4.3 for various values of σ, τ. We compare to an Erdös Rényi random graph, preferential attachment (Barabási and Albert, 1999), and the Bayesian nonparametric network model of (Lloyd et al., 2012). The particular features we explore are

- **Degree distribution** Figure 6(a) demonstrates that the model can exhibit power-law behavior providing a heavy-tailed degree distribution. As shown in Figure 6(b), the model can also handle an exponential cut-off in the tails of the degree distribution, which is an attractive property (Newman, 2009).
- **Number of degree 1 nodes** Figure 6(c) examines the fraction of degree 1 nodes versus number of nodes.
- **Sparsity** Figure 6(d) plots the number of edges versus the number of nodes. The larger σ, the sparser the graph. 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 $O(n^2)$.

Interpretation of hyperparameters Based on the properties derived and explored empirically in this section, we see that our hyperparameters have the following interpretations:

- $\boldsymbol{\sigma}$— From Figure 6(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.
- $\boldsymbol{\alpha}$— From Proposition 5, α provides an overall scale that affects the number of nodes and directed interactions, with larger α leading to larger networks.
- $\boldsymbol{\tau}$— From Figure 6(b), τ determines the exponential decay of the tails of the power-law degree distribution, with τ small looking like pure power-law. This is intuitive from the form of $\rho(dw)$ in (12), where we see that τ affects large weights more than small ones.

Appendix A: Proof of Proposition 5

We derive $E[N_\alpha]$ using iterated expectation. First, as N_α corresponds to the number of different values when sampling $2D_\alpha^*$ data from a Dirichlet process,

$$\mathbb{E}[N_\alpha | D_\alpha^*] \simeq \alpha \log\left(1 + \frac{2D_\alpha^*}{\alpha}\right),$$

and for $\alpha >> \tau^2 > 0$, we have $\frac{2D_\alpha^*}{\alpha} >> 1$ such that

$$\mathbb{E}[N_\alpha | D_\alpha^*] \simeq \alpha \log\left(\frac{2D_\alpha^*}{\alpha}\right).$$

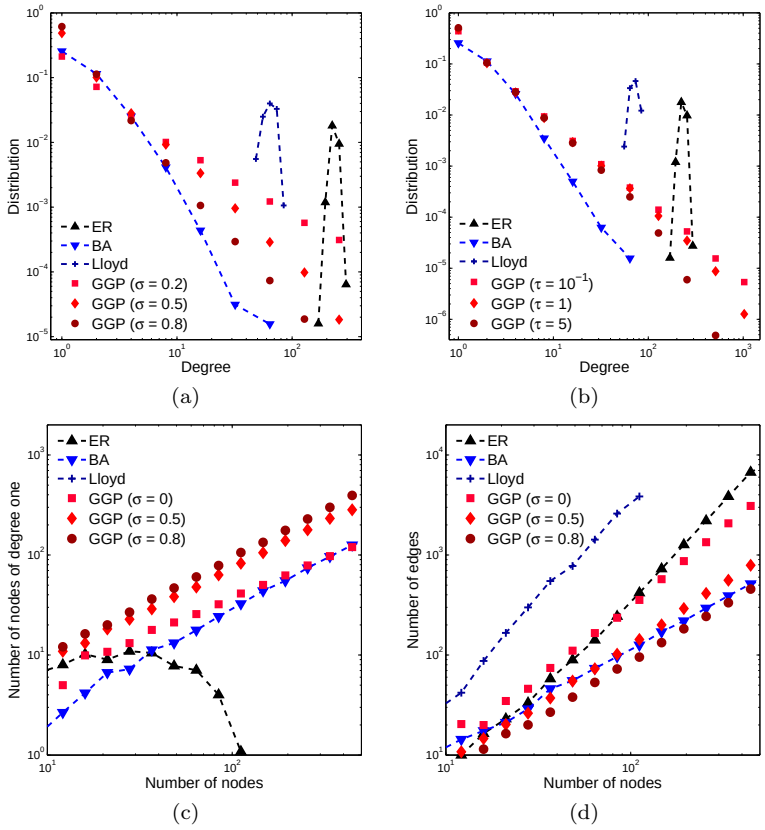


FIG 6. *Examination of the undirected network properties with generalized gamma process (averaging over graphs with various α) in comparison to an Erdös Rényi $G(n,p)$ model with $p = 0.05$ (ER), the preferential attachment model of (Barabási and Albert, 1999) (BA), and the nonparametric formulation of (Lloyd et al., 2012) (Lloyd). (a-b) Degree distribution on a log-log scale for (a) various values of σ ($\tau = 10^{-2}$) and (b) various values of τ ($\sigma = 0.5$) for the GGP. (c) Number of nodes with degree one versus the number of nodes on a log-log scale. Note that the Lloyd method leads to dense graphs such that no node has only degree 1. (d) Number of edges versus the number of nodes. In (d) we note growth at a rate $O(n^a)$ for all models, with $a = 2$ for the Erdös Rényi model (dense graphs) and $1 < a < 2$ for the GGP model with $\sigma > 0$. Simulations seem to indicate a subquadratic rate for the gamma process ($\sigma = 0$), although our proof does not apply in this case.*

Conditioning on W_α^* alone,

$$\mathbb{E}[N_\alpha|W_\alpha^*] = \mathbb{E}[\mathbb{E}[N_\alpha|D_\alpha^*]|W_\alpha^*] = \alpha \log\left(\frac{2}{\alpha}\right) + \alpha\mathbb{E}\left[\log\left(D_\alpha^*\right)|W_\alpha^*\right]$$
$$\simeq \alpha \log\left(\frac{2}{\alpha}\right) + \alpha \log\left({W_\alpha^*}^2\right).$$

Finally, using the fact that in the gamma process case, we have $W_\alpha^* \sim \text{Gamma}(\alpha, \tau)$,

$$\begin{aligned}\mathbb{E}[N_\alpha] &= \alpha \log\left(\frac{2}{\alpha}\right) + 2\alpha\mathbb{E}\left[\log\left(W_\alpha^*\right)\right] \\ &= \alpha \log\left(\frac{2}{\alpha}\right) + 2\alpha(\psi(\alpha) - \log(\tau)) \\ &\simeq \alpha \log\left(\frac{2\alpha}{\tau^2}\right)\end{aligned}$$

as $\psi(\alpha) \simeq \log\alpha$ for large α.

Similarly, for the variance, we have

$$\mathbb{E}[N_\alpha^2|D_\alpha^*] \simeq \alpha \log\left(\frac{2D_\alpha^*}{\alpha}\right) + \left(\alpha \log\left(\frac{2D_\alpha^*}{\alpha}\right)\right)^2$$
$$\mathbb{E}[N_\alpha^2|W_\alpha^*] \simeq \alpha \log\left(\frac{2{W_\alpha^*}^2}{\alpha}\right) + \left(\alpha \log\left(\frac{2{W_\alpha^*}^2}{\alpha}\right)\right)^2.$$

Using $\mathbb{E}\left[(\log\frac{2}{\alpha}{W_\alpha^*}^2)^2\right] = 4(\log\frac{\sqrt{\alpha}\tau}{\sqrt{2}} - \psi(\alpha))^2 + 4\psi(1,\alpha) \simeq \log^2\frac{\tau^2}{2\alpha} + 4\psi(1,\alpha)$, we obtain

$$\begin{aligned}Var(N_\alpha) &= \alpha \log\left(\frac{2\alpha}{\tau^2}\right) + \alpha^2\left[\log^2\frac{\tau^2}{2\alpha} + 4\psi(1,\alpha)\right] - \alpha^2\log^2\left(\frac{2\alpha}{\tau^2}\right) \\ &= \alpha \log\left(\frac{2\alpha}{\tau^2}\right) + 4\alpha^2\psi(1,\alpha).\end{aligned}$$

For the number of edges in the directed multigraph, we have

$$\mathbb{E}[D_\alpha^*] = \mathbb{E}[{W_\alpha^*}^2] = \frac{\alpha(\alpha+1)}{\tau^2}$$

$$\mathbb{E}[{D_\alpha^*}^2] = \mathbb{E}[{W^*}^2] + \mathbb{E}[{W^*}^4] = \frac{\alpha(\alpha+1)}{\tau^2} + \frac{\alpha(\alpha+1)(\alpha+2)(\alpha+3)}{\tau^4}.$$

Hence

$$\begin{aligned}Var(D_\alpha^*) &= \frac{\alpha(\alpha+1)}{\tau^2}(1 - \frac{\alpha(\alpha+1)}{\tau^2}) + \frac{\alpha(\alpha+1)(\alpha+2)(\alpha+3)}{\tau^4} \\ &= \frac{\alpha(\alpha+1)}{\tau^2}\left(1 - \frac{\alpha(\alpha+1)}{\tau^2} + \frac{(\alpha+2)(\alpha+3)}{\tau^2}\right) = \frac{\alpha(\alpha+1)}{\tau^2}\left(1 + \frac{4\alpha+6}{\tau^2}\right).\end{aligned}$$

Appendix B: Proof of Theorem 6

Consider the conditionally Poisson construction

$$D_\alpha^* \sim \text{Poisson}({W_\alpha^*}^2)$$
$$(U_1', \ldots, U_{2D_\alpha^*}')|D_\alpha^*, W_\alpha \sim \frac{W_\alpha}{W_\alpha^*}.$$

From this construction, it is clear that the number of U_j' in any interval $[a, b]$ goes to infinity as α goes to infinity. We can therefore invoke asymptotic results on iid sampling from a normalized generalized gamma process.

Let $N_{\alpha,j}$ be the number of clusters of size j in $(U_1', \ldots, U_{2D_\alpha^*}')$. In the directed graph model, $N_{\alpha,j}$ corresponds to the number of nodes with j incoming/outgoing edges (self-edges count twice for a given node).

As the U_j' are drawn from a normalized generalized gamma process of parameters (α, σ, τ), we have the following asymptotic result (Pitman, 2006; Lijoi, Mena and Prünster, 2007, Corollary 1)

$$\frac{N_{\alpha,j}}{N_\alpha} \xrightarrow{\alpha \to \infty} p_{\sigma,j} = \frac{\sigma\Gamma(j-\sigma)}{\Gamma(1-\sigma)\Gamma(j+1)}.$$

almost surely, for $j = 1, 2, \ldots$.

Appendix C: Proof of Theorem 7

The sparsity of the directed and undirected graphs is a consequence of the power-law properties of the directed multigraph. Let η_j be the average number of undirected edges for nodes with j directed edges. Then we have

$$N_\alpha^{(e)} \leq \sum_{j=1}^{\infty} \eta_j N_{\alpha,j}.$$

Using Theorem 6, we have

$$\sum_{j=1}^{\infty} \eta_j N_{\alpha,j} \to N_\alpha \sum_{j=1}^{\infty} \eta_j p_{\sigma,j}$$

almost surely, as $\alpha \to \infty$. Using the fact that $\eta_j \leq \min(j, N_\alpha)$, we can further upper bound the quantity defined above

$$\begin{aligned} N_\alpha \sum_{j=1}^{\infty} \eta_j p_{\sigma,j} &\leq N_\alpha \sum_{j=1}^{\infty} \min(j, N_\alpha) p_{\sigma,j} \\ &= N_\alpha \sum_{j=1}^{N_\alpha - 1} j \cdot p_{\sigma,j} + N_\alpha^2 \sum_{j=N_\alpha}^{\infty} p_{\sigma,j} \\ &= N_\alpha \sum_{j=1}^{N_\alpha} j^{1-\sigma+\varepsilon} \cdot j^{-\varepsilon+\sigma} p_{\sigma,j} + N_\alpha^2 \frac{\Gamma(N_\alpha - \sigma)}{\Gamma(N_\alpha)\Gamma(1-\sigma)} \\ &\leq N_\alpha^{2-\sigma+\varepsilon} \sum_{j=1}^{N_\alpha} j^{-\varepsilon+\sigma} p_{\sigma,j} + N_\alpha^2 \frac{\Gamma(N_\alpha - \sigma)}{\Gamma(N_\alpha)\Gamma(1-\sigma)} \end{aligned}$$

where $0 < \epsilon < \sigma$. As j is large, $\frac{\Gamma(j+a)}{\Gamma(j+b)} \sim j^{a-b}$. So

$$N_\alpha^2 \frac{\Gamma(N_\alpha - \sigma)}{\Gamma(N_\alpha)} \sim N_\alpha^{2-\sigma} \text{ and } \sum_{j=1}^{\infty} j^{\sigma-\varepsilon} p_{\sigma,j} < \infty$$

for any $0 < \epsilon$. Hence

$$N_\alpha^{(e)} \leq C N_\alpha^{2+\varepsilon-\sigma} \tag{30}$$

almost surely, as $\alpha \to \infty$, for some constant $C > 0$.

References

Airoldi, E. M., Costa, T. B. and Chan, S. H. (2014). Stochastic blockmodel approximation of a graphon: Theory and consistent estimation. In *Advances in Neural Information Processing Systems* **26**.

Airoldi, E. M., Blei, D., Fienberg, S. E. and Xing, E. (2008). Mixed membership stochastic blockmodels. *The Journal of Machine Learning Research* **9** 1981–2014.

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

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

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

Bickel, P. J. and Chen, A. (2009). A nonparametric view of network models and Newman–Girvan and other modularities. *Proceedings of the National Academy of Sciences* **106** 21068–21073.

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

Blackwell, D. and MacQueen, J. B. (1973). Ferguson distributions via Pólya urn schemes. *The annals of statistics* 353–355.

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

Bollobás, B. (2001). *Random graphs* **73**. Cambridge university press.

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

Brix, A. (1999). Generalized gamma measures and shot-noise Cox processes. *Advances in Applied Probability* **31** 929–953.

Bühlmann, H. (1960). Austauschbare stochastische Variablen und ihre Grenzwertsätze PhD thesis, University of California, Berkeley.

Caron, F. (2012). Bayesian nonparametric models for bipartite graphs. In *NIPS*.

De Finetti, B. (1931). Funzione caratteristica di un fenomeno aleatorio. *Atti della R. Academia Nazionale dei Lincei, Serie 6. Memorie, Classe di Scienze Fisiche, Mathematice e Naturale* **4** 251-299.

Devroye, L. (2009). Random variate generation for exponentially and polynomially tilted stable distributions. *ACM Transactions on Modeling and Computer Simulation (TOMACS)* **19** 18.

Diaconis, P. and Janson, S. (2008). Graph limits and exchangeable random graphs. *Rendiconti di Matematica e delle sue Applicazioni. Serie VII* 33–61.

Durrett, R. (2007). *Random graph dynamics*. Cambridge university press.

Erdös, P. and Rényi, A. (1959). On random graphs. *Publicationes Mathematicae* **6** 290–297.

Favaro, S. and Teh, Y. W. (2012). MCMC for normalized random measure mixture models Technical Report, University of Turin.

Ferguson, T. S. and Klass, M. J. (1972). A representation of independent increment processes without Gaussian components. *The Annals of Mathematical Statistics* **43** 1634–1643.

Fienberg, S. E. (2012). A brief history of statistical models for network analysis and open challenges. *Journal of Computational and Graphical Statistics* **21** 825–839.

Freedman, D. A. (1996). De Finetti's theorem in continuous time. *Lecture Notes-Monograph Series* 83–98.

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

Hoff, P. D. (2009). Multiplicative latent factor models for description and prediction of social networks. *Computational and Mathematical Organization Theory* **15** 261–272.

Hoff, P. D., Raftery, A. E. and Handcock, M. S. (2002). Latent space approaches to social network analysis. *Journal of the american Statistical association* **97** 1090–1098.

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

Kallenberg, O. (1990). Exchangeable random measures in the plane. *Journal of Theoretical Probability* **3** 81–136.

Kallenberg, O. (2005). *Probabilistic symmetries and invariance principles.* Springer.

Khintchine, A. (1937). Zur Theorie der unbeschrankt teilbaren Verteilungsgesetze. *Mat. Sbornik* **2** 79–119.

Kim, M. and Leskovec, J. (2012). Multiplicative attribute graph model of real-world networks. *Internet Mathematics* **8** 113–160.

Kingman, J. F. C. (1967). Completely random measures. *Pacific Journal of Mathematics* **21** 59–78.

Kingman, J. F. C. (1993). *Poisson processes* **3**. Oxford University Press, USA.

Lauritzen, S. (2008). Exchangeable Rasch matrices. *Rendiconti di Matematica, Serie VII* **28** 83–95.

Lewis, P. A. and Shedler, G. S. (1979). Simulation of nonhomogeneous Poisson processes by thinning. *Naval Research Logistics Quarterly* **26** 403–413.

Lijoi, A., Mena, R. H. and Prünster, I. (2007). Controlling the reinforcement in Bayesian non-parametric mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **69** 715–740.

Lijoi, A., Prünster, I. and Walker, S. G. (2008). Investigating nonparametric priors with Gibbs structure. *Statistica Sinica* **18** 1653.

Lijoi, A. and Prünster, I. (2010). Models beyond the Dirichlet process. In *Bayesian Nonparametrics* (P. M. S. G. W. N. L. Hjort C. Holmes, ed.) Cambridge University Press.

Lloyd, J., Orbanz, P., Ghahramani, Z. and Roy, D. (2012). Random function priors for exchangeable arrays with applications to graphs and relational data. In *NIPS* **25** 1007–1015.

Miller, K., Griffiths, T. and Jordan, M. (2009). Nonparametric latent feature models for link prediction. In *NIPS*.

Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM review* 167–256.

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

Nowicki, K. and Snijders, T. (2001). Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association* **96** 1077–1087.

Ogata, Y. (1981). On Lewis' simulation method for point processes. *IEEE Transactions on Information Theory* **27** 23–31.

Orbanz, P. and Roy, D. M. (2013). Bayesian Models of Graphs, Arrays and Other Exchangeable Random Structures Technical Report, University of Cambridge.

Pitman, J. (1995). Exchangeable and partially exchangeable random partitions. *Probability Theory and Related Fields* **102** 145–158.

Pitman, J. (1996). Some developments of the Blackwell-MacQueen urn scheme. *Lecture Notes-Monograph Series* 245–267.

Pitman, J. (2003). Poisson-Kingman partitions. *Lecture Notes-Monograph Series* 1–34.

Pitman, J. (2006). Combinatorial Stochastic Processes. In *Ecole d'Eté de Probabilités de Saint-Flour XXXII–2002. Lecture Notes in Mathematics.* Springer.

Price, D. (1976). A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science* **27** 292–306.

Regazzini, E., Lijoi, A. and Prünster, I. (2003). Distributional results for means of normalized random measures with independent increments. *The Annals of Statistics* **31** 560–585.

Rohe, K., Chatterjee, S. and Yu, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *The Annals of Statistics* **39** 1878–1915.

Wolfe, P. and Choi, D. S. Co-clustering separately exchangeable network data. *To appear in Annals of Statistics.*

Wolfe, P. J. and Olhede, S. C. (2013). Nonparametric graphon estimation. *arXiv preprint arXiv:1309.5936.*

Zhao, Y., Levina, E. and Zhu, J. (2012). Consistency of community detection in networks under degree-corrected stochastic block models. *The Annals of Statistics* **40** 2266–2292.