

---

# Random clique covers for graphs with local density and global sparsity

---

Sinead A. Williamson<sup>1,2,3</sup> and Mauricio Tec<sup>1</sup>

<sup>1</sup>Department of Statistics and Data Science, University of Texas at Austin

<sup>2</sup>Department of Information, Risk and Operations Management, University of Texas at Austin

<sup>3</sup>CognitiveScale

## Abstract

Large real-world graphs tend to be sparse, but they often contain many densely connected subgraphs and exhibit high clustering coefficients. While recent random graph models can capture this sparsity, they ignore the local density, or vice versa. We develop a Bayesian non-parametric graph model based on random edge clique covers, and show that this model can capture power law degree distribution, global sparsity and non-vanishing local clustering coefficient. This distribution can be used directly as a prior on observed graphs, or as part of a hierarchical Bayesian model for inferring latent graph structures.

## 1 INTRODUCTION

Random graph models provide statistical tools for network analysis and can often be used as prior distributions in a Bayesian framework. Such models aim to capture various properties of real-world graphs, such as power-law degree distributions [Albert and Barabási, 2002, Dangalchev, 2004, BloemReddy and Orbanz, 2018], small-world properties [Watts and Strogatz, 1998], or latent community structure [Holland et al., 1983, Airolidi et al., 2008, Karrer and Newman, 2011].

One statistic of interest is the density of a graph, defined as the number of edges over the number of *possible* edges for a binary, undirected graph. This can be extended to distributions over graphs: we think of a random graph as being dense if the number of edges grows quadratically with the number of vertices, and sparse if it grows sub-quadratically with the number of vertices [Neetil and Osona de Mendez, 2012]. Many commonly-used random graphs, such as Erdős-Rényi graphs [Gilbert, 1959, Erdős and Rényi, 1960] or stochastic blockmodels and

their variants [Holland et al., 1983, Airolidi et al., 2008, Karrer and Newman, 2011], concentrate on dense graphs [see Lloyd et al., 2012, for a discussion]. Dense behavior is commonly seen in small, closed communities, where every vertex has the option to interact with all other vertices and a fully connected graph is at least conceivable. However it is not generally seen in larger graphs, where a given vertex will not have the opportunity to interact with more than a small subset of the other vertices—for example international online social networks.

Recently, a number of Bayesian nonparametric models have been proposed that can generate sparse graphs. The sparse exchangeable graph framework, developed by Caron and Fox [2017] and further explored by Veitch and Roy [2015] and Borgs et al. [2018], constructs sparse graphs or multigraphs based on a random atomic measure on the space  $\mathcal{V}$  of potential vertices. The edge exchangeable graph framework [Cai et al., 2016, Crane and Dempsey, 2018] constructs sparse multigraphs as a sequence of edges, whose distribution is parametrized by a random atomic measure on  $\mathcal{V}$ . In both classes of model, appropriate design choices yield power-law degree distribution, another statistical property often seen in real-world graphs.

While distributions such as these offer the ability to model sparse graphs, they do not capture certain other important graph behaviors. While large-scale graphs are usually sparse, locally they exhibit transitivity: if two vertices  $j$  and  $k$  are both connected to vertex  $i$ , a connection between vertices  $j$  and  $k$  is more likely than a connection between  $j$  and a randomly selected vertex. This means that the graph will tend to exhibit densely connected subgraphs, even if the overall graph is sparse. Such graphs will typically exhibit a high average local clustering coefficient, indicating that most vertices' neighborhoods are close to being cliques and that the graph has a relatively high number of triangles. This type of behavior is a key component of “small world” graphs [Watts and Strogatz, 1998]. While extensions to the

sparse exchangeable graphs [Herlau et al., 2016, Todeschini et al., 2016] and the edge-exchangeable graphs [Williamson, 2016] incorporate community-type structure, they do not yield this form of transitivity.

In this paper we explore distributions over graphs that, while sparse, exhibit local density in the form of arbitrarily-large fully-connected subgraphs and non-vanishing average local clustering coefficient. We achieve this by explicitly modeling graphs as collections of cliques, or fully connected subsets of vertices. Our construction takes the form of a random clique cover, with cliques selected using a nonparametric feature selection model (in this paper, the stable beta Indian buffet process [Teh and Görür, 2009], but other choices are possible). The resulting graphs exhibit local density and non-zero local clustering coefficients, with clique sizes controlled primarily by a single parameter. On a global scale, the graphs are sparse, with the degree of sparsity primarily controlled by a separate parameter. We observe power-law distributions over both the number of cliques a vertex belongs to, and its degree.

When used directly as a distribution over graphs, we show that this model’s statistical properties mimic the statistics of real-world graphs (Section 5.1). When used in a hierarchical modeling context, the cliques can be thought of as latent communities (Section 5.2).

## 2 RANDOM CLIQUE COVERS

A graph  $G$  is most commonly described using an ordered pair  $(V, E)$  of vertices  $V$  and edges  $E$ .  $G$  can alternatively be described in terms of its edge clique cover [Orlin, 1977]. An edge clique cover (or intersection graph) is a set of cliques—i.e. fully connected subgraphs—such that two vertices share an edge iff they have at least one clique in common.

This formulation suggests a method for generating random graphs by placing a distribution over cliques. Concretely, let a random clique be a random, finite subset  $\mathcal{C} \subset \mathcal{V}$  of a (possibly uncountable) set  $\mathcal{V}$  of potential vertices. Given a distribution over sequences of such subsets, we can generate a sequence  $\mathcal{C}_1, \dots, \mathcal{C}_N$  of cliques, that can be translated into a graph  $G = (V, E)$  by adding undirected edges between all vertices with at least one clique in common, so that

$$\begin{aligned} (i, j) \in E & \text{ iff } \sum_{n=1}^N I(i, j \in \mathcal{C}_n \text{ and } i \neq j) > 0 \\ i \in V & \text{ iff } \sum_{n=1}^N I(i \in \mathcal{C}_n) > 0. \end{aligned}$$

If we number the vertices in our graph, and represent our cliques in terms of a binary matrix  $Z$  where  $Z_{ni} = 1$  iff clique  $n$  includes vertex  $i$ , then we can represent the

graph’s adjacency matrix as

$$\min (Z^T Z - \text{diag}(Z^T Z), 1)$$

where the min is taken as an element-wise operation.

We can modify this construction to give a distribution over random multigraphs, by letting the number of edges between vertices  $i$  and  $j$  equal the number of cliques they have in common. We discuss this multigraph formulation in Appendix A.

### 2.1 RANDOM CLIQUES GENERATED USING EXCHANGEABLE RANDOM MEASURES

We use the idea of random clique covers to construct a distribution over graphs of fixed (but random) dimensionality. To do so, we must specify both a distribution over the number of cliques, and distributions over the vertices appearing in those cliques. We let the number of cliques  $N \sim \text{Poisson}(\tau)$ , although other choices could easily be made.

Reasonable desiderata for the random clique selection mechanism might be that the total number of vertices (and edges) is unbounded; that the size of each clique is random; and that cliques overlap with finite probability (avoiding the trivial solution where our graph is made up of a series of disconnected subgraphs).

One choice that meets these criteria is the stable beta process [Teh and Görür, 2009]. The stable beta process is a completely random measure [Kingman, 1967] whose atoms lie in  $(0, 1]$ , with Lévy measure

$$\nu(d\mu) = \alpha \frac{\Gamma(1+c)}{\Gamma(1-\sigma)\Gamma(c+\sigma)} \mu^{-\sigma-1} (1-\mu)^{c+\sigma-1} d\mu \quad (1)$$

where  $\sigma \in (0, 1]$ ,  $c \geq -\sigma$ , and  $\alpha > 0$ . If  $\sigma = 0$ , this reduces to the (homogeneous) beta process [Hjort, 1990, Thibaux and Jordan, 2007]. If  $c = 1 - \sigma$ , the stable beta process corresponds to a stable process with all atoms of size larger than one removed.

We can use the stable beta process to construct a distribution over cliques, which we can represent in terms of a binary matrix. If  $\mu = \sum_{i=1}^{\infty} \mu_i \delta_{\theta_i} \sim \text{SB}(c, \sigma, \alpha)$  where each location  $\theta_i$  corresponds to a vertex, we can sample an exchangeable sequence of  $N$  cliques by including the  $i$ th vertex in the  $n$ th clique with probability  $\mu_i$ . We can represent the resulting clique allocation as a binary matrix  $Z$  with  $N$  rows and infinitely many columns, where  $Z_{ni} = 1$  iff vertex  $i$  is in clique  $n$ .

If we marginalize out  $\mu$ , we obtain a predictive distribution for  $Z_n | Z_1, \dots, Z_{n-1}$  that is known as the stable-beta Indian buffet process [SB-IBP, Teh and Görür,

2009]. If  $m_i = \sum_{j=1}^{n-1} Z_{ji}$  is the number of times vertex  $i$  has been previously selected, then the  $n$ th clique will include that vertex (i.e.  $Z_{ni} = 1$ ) with probability

$$\mathbf{P}(Z_{ni} = 1 | Z_1, \dots, Z_{n-1}) = \frac{m_i - \sigma}{n + c - 1}.$$

In addition to vertices that have previously appeared,  $Z_n$  will also select Poisson  $\left(\alpha \frac{\Gamma(1+c)\Gamma(n+c+\sigma-1)}{\Gamma(n+c)\Gamma(c+\sigma)}\right)$  new vertices.

The SB-IBP exhibits a number of power-law behaviors [Teh and Görür, 2009, Broderick et al., 2012, Heaukulani and Roy, 2018]. First, the total number of non-zero columns of  $Z$  (in our case, observed vertices) in  $N$  rows (cliques) follows a power law, growing, as  $N \rightarrow \infty$ , as  $\frac{\alpha}{\sigma} \frac{\Gamma(c+1)}{\Gamma(c+\sigma)} N^\sigma$ .

Second, the number of non-zero entries per column (in our case, the number of cliques each vertex appears in) follows a Zipf’s law. Let  $K_{N,j}$  be the number of features that appear exactly  $j$  times in  $N$  rows. Then, following Broderick et al. [2012],

$$K_{N,j} \sim \frac{\alpha \Gamma(j - \sigma) \Gamma(1 + c)}{j! \Gamma(1 - \sigma) \Gamma(c + \sigma)} N^\sigma \sim j^{-\sigma-1}. \quad (2)$$

As we will see in Section 3, when applied to edge clique covers, the SB-IBP yields many interesting graph-specific properties, many of which follow from these two observations.

## 2.2 A MODEL FOR PARTIALLY OBSERVED CLIQUES

In a modeling context, we may wish to interpret cliques as latent communities [Holland et al., 1983, Airolidi et al., 2008, Karrer and Newman, 2011]. For example in a social network, a clique might represent a shared interest or hobby. In this context, we would not necessarily expect all vertices in the community to be connected. Instead, we can think of the community as being a noisy instantiation of a fully connected latent clique.

To model this in the current context, start with the distribution over edge clique covers described in Section 2.1. If we associate each latent clique  $\mathcal{C}_n$  with a probability  $\pi_n$ , we can form a graph by including an edge between vertices  $i$  and  $j$  with probability

$$P((i, j) \in E) = 1 - \prod_{n: i, j \in \mathcal{C}_n} (1 - \pi_n).$$

If we use a single probability  $\pi$  across all cliques, this corresponds to including edges according to a noisy-OR likelihood: if two vertices have  $m$  latent cliques in common, they share an edge with probability  $1 - (1 - \pi)^m$ .

One way of thinking of this model is as a superposition of locally-defined Erdős-Renyi  $G(n, p)$  graphs. A  $G(n, p)$  graph builds a graph with  $n$  vertices, by including each of the  $\binom{n}{2}$  potential edges with probability  $p$ . In the partially observed clique model, each row of the SB-IBP selects a clique  $\mathcal{C}_n$  of vertices, and builds a subgraph on those vertices according to a  $G(|\mathcal{C}_n|, \pi_n)$  model. Within the clique, we have a dense subgraph almost surely.

## 3 GRAPH PROPERTIES

In this section, we explore the statistical properties of graphs constructed using an SB-IBP distribution over cliques. In particular, we show that we obtain sparse graphs that exhibit densely connected subgraphs, with a non-vanishing local clustering coefficient. We focus on the fully observed setting described in Section 2.1, noting that the partially observed extension described in Section 2.2 will inherit similar properties.

### 3.1 SPARSITY

Let  $|E|$  be the number of edges in a graph, and  $|V|$  be the number of vertices. We can define the density of a fixed graph as the number of edges over the number of possible edges, i.e.  $D = 2|E|/|V|(|V| - 1)$ . We can extend this concept to random graphs, by looking at how the number of edges varies with the number of vertices as we change parameters of the model. We say a random graph is dense if  $|E|$  grows quadratically with  $|V|$ , and sparse if it grows sub-quadratically. To explore the sparsity of the proposed random graph, we will look at how  $|E|$  and  $|V|$  vary with the number of cliques. In our construction, the number of cliques is a Poisson( $\tau$ ) random variable, and so the expected number of cliques can be manipulated by altering the rate  $\tau$  of that Poisson distribution.

**Proposition 1.** *As the number  $N$  of cliques grows to infinity, the number of vertices  $|V|$  grows as  $N^\sigma$ .*

*Proof.* This is a direct consequence of the power-law behavior of the SB-IBP: the number of vertices in a graph with  $N$  generating cliques correspond to the number of non-zero columns in a sample from the SB-IBP with  $N$  rows. Broderick et al. [2012] and Heaukulani and Roy [2018] show that the expected number of non-zero columns grows as  $\frac{\alpha}{\sigma} \frac{\Gamma(c+1)}{\Gamma(c+\sigma)} N^\sigma$ .  $\square$

**Proposition 2.** *As the number  $N$  of cliques grows to infinity, the number of edges  $|E|$  grows as  $O\left(\min(N^{\frac{1+\sigma}{2}}, N^{\frac{3\sigma}{2}})\right)$ . Combined with the result in Proposition 1, this means that the number of edges grows with the number of vertices as*

$O\left(\min(|V|^{\frac{1+\sigma}{2\sigma}}, |V|^{3/2})\right)$ , implying that the graph is sparse.

*Proof.* Conditioned on the probabilities  $\mu_i, \mu_j$  assigned by the stable beta process to vertices  $i$  and  $j$ , the probability of an edge between  $i$  and  $j$  given  $N$  cliques is  $1 - (1 - \mu_i\mu_j)^N$ . By linearity of expectation, the total expected number of edges, given  $\mu$ , is therefore  $\sum_{i>j} (1 - (1 - \mu_i\mu_j)^N)$ . Therefore, we can obtain the expected total number of edges in the graph as

$$\mathbb{E}[|E|] = \frac{1}{2} \int \int (1 - (1 - wv)^N) \nu(dw)\nu(dv) \quad (3)$$

This equation also describes the number of edges in an edge-exchangeable graph based on the stable beta process, proposed by Cai et al. [2016]. Although the two models are different—our model is not edge-exchangeable, and the Cai et al. model is not locally dense—the expected number of edges coincide due to linearity of expectation. Cai et al. show, via a Poissonization approach, that the integral in Equation 3 grows as  $O\left(\min(N^{\frac{1+\sigma}{2}}, N^{\frac{3\sigma}{2}})\right)$ .<sup>1</sup>  $\square$

**Corollary 1.** *The number of edges grows with the number of vertices as  $O\left(\min(|V|^{\frac{1+\sigma}{2\sigma}}, |V|^{3/2})\right)$ , implying that the graph is sparse for all values of  $\sigma$ .*

*Proof.* This follows directly from Propositions 1 and 2, and is also noted by Cai et al. [2016].  $\square$

As we approach the limit where  $\sigma = 1$ , we obtain the trivially sparse graph where each clique contains a  $\text{Poisson}(\alpha)$  number of vertices, and there are no edges connecting cliques. In the supplement, we show simulations that validate these results empirically.

### 3.2 LOCAL DENSITY AND CLUSTERING COEFFICIENTS

By construction, each observed vertex appears in at least one of the  $N$  generating cliques in  $Z$ , with expected clique size  $\alpha$ . As the number of generating cliques a vertex appears in grows, so does the expected size of the largest generating cliques (since  $\mathbb{E}[\max(X_1, X_2)] \geq \max(\mathbb{E}[X_1], \mathbb{E}[X_2])$ ). The expected size of the maximal clique for the vertex (i.e. the largest clique containing that vertex) will be larger again, since three or more overlapping cliques can introduce triangles that do not appear in any of the individual cliques. This ensures that, even though the overall graph is sparse, there will be locally

<sup>1</sup>For results for alternative choices of completely random measure, see also Cai et al. [2016].

dense subgraphs, with expected size lower bounded by  $\alpha$ . We show plots demonstrating this empirically in the supplement.

The local clustering coefficient [Watts and Strogatz, 1998] is a popular tool for measuring the local density of a graph. The local clustering coefficient of vertex  $i$  is given by  $C_i = \lambda_i/k_i(k_i - 1)$ , where  $\lambda_i$  is the number of triangles containing vertex  $i$ , and  $k_i$  is the number of neighbors of vertex  $i$ . In other words,  $C_i$  is a local measure of the ratio of the number of triangles to the number of potential triangles, given a vertex’s neighborhood.

The local clustering coefficient is a measure of the transitivity of the graph: to what extent the presence of the edges  $(i, j)$  and  $(j, k)$  imply the presence of an edge  $(i, k)$ . Intuitively, adding edges in cliques, rather than individually, increases transitivity: knowing that the graph contains edges  $(i, j)$  and  $(j, k)$  increases the likelihood that there is at least one clique containing all three vertices, implying the presence of the third edge.

**Proposition 3.** *The average local clustering coefficient remains non-zero as the number of cliques grows to infinity, provided  $\sigma > 0$ .*

*Proof.* Let  $j$  be the number of cliques to which a given vertex belongs. Vertices belonging to a single clique have a local clustering coefficient of 1. As  $j$  increases, the expected local clustering coefficient will decrease, and will be minimized when it is the only common member of those  $j$  cliques. In this extreme setting, the clustering coefficient will vary as  $j^{-1}$  (since the vertex’s expected neighborhood size grows as  $j\alpha$ , and the expected number of triangles as  $j\alpha^2$ ). For large  $j$  and  $N$ , the distribution over the number of cliques,  $j$ , that a vertex belongs to can be approximated by a  $\text{Zeta}(1+\sigma)$  distribution (Equation 2), and the expectation of  $j^{-1}$  under this distribution is  $\zeta(2+\sigma)/\zeta(1+\sigma) > 0$  for  $\sigma \in (0, 1)$ . This implies that the average clustering coefficient will not tend to zero as the graph grows.  $\square$

### 3.3 DEGREE DISTRIBUTION

The expected size of each generating clique is  $\alpha$ , and the number of cliques an vertex belongs to follows a power law distribution (Equation 2). This leads to a power law distribution over the degree, with the expected degree decreasing with  $\sigma$  and increasing with  $\alpha$ .

The left hand column of Figure 1 shows the degree distribution (in the form of a boxplot) for different values of  $\sigma$  and  $\alpha$ , and different numbers of cliques  $N$ . We see increasingly heavy tails as  $\sigma$  increases.

Since the number of vertices also increases with  $\sigma$ , the maximum degree is higher for larger  $\sigma$ . The right hand

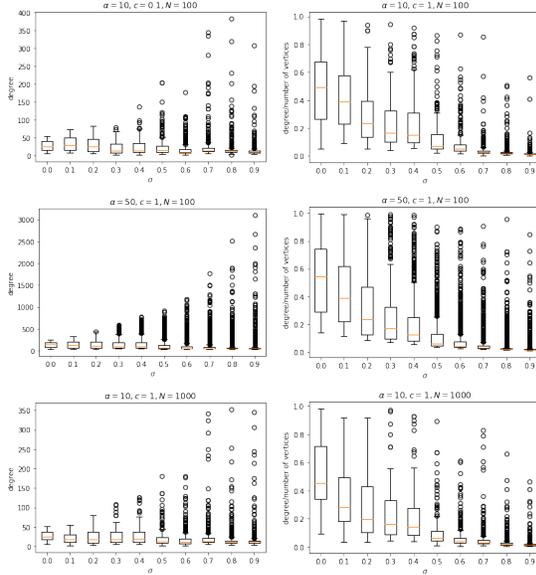


Figure 1: Vertex degrees for different values of  $\sigma$ ,  $\alpha$  and  $N$ , for  $c = 1$ . Left: Raw degrees; Right: degree/number of vertices.

column of Figure 1 shows the degree divided by the total number of vertices, making it easier to compare different values of  $\sigma$ . We see that the average proportion decreases and the distribution grows heavier tailed as  $\sigma$  increases, for all values of  $\alpha$  and  $N$ .

### 3.4 DENSITY OF THE INTERSECTION GRAPH OF GENERATING CLIQUES

The clique graph of a graph is the intersection graph of its maximal cliques [Hamelink, 1968]. Since our construction does not explicitly generate maximal cliques, we consider the generating-clique graph—i.e. the intersection graph of the generating cliques specified in  $Z$ . Perhaps surprisingly given the sparsity of the overall graph, this intersection graph is dense.

**Proposition 4.** *The intersection graph of the maximal cliques is dense, and the expected number of edges connecting two generating cliques is  $\alpha \frac{1-\sigma}{1+c}$ .*

*Proof.* The expected number of vertices by which two cliques overlap is given by  $\sum_{i=1}^{\infty} \mu_i^2$ . Campbell’s theorem tells us that  $E[\sum_{x \in \Pi} f(x)] = \int_S f(x) \nu(dx)$ , where  $\Pi$  is a Poisson process on  $S$  with rate measure  $\nu(dx)$ —in our case, the Lévy measure  $\nu(d\mu)$  of the stable beta process (Equation 1)—and  $f$  is a measurable function. Therefore the expected overlap between two cliques is

$$\int_0^1 \nu(d\mu) = \alpha \frac{1-\sigma}{1+c}.$$

This indicates that two cliques overlap with positive probability, meaning that the resulting intersection graph is dense and that there is a path between any two vertices with positive probability, even if they do not belong to the same clique.  $\square$

## 4 RELATED WORK

The idea of a random edge clique cover was explored by Barber [2008], under the name Clique Matrices. As in this paper, a binary matrix  $Z$  is used to represent cliques within the graph; however the number of vertices is fixed and the entries are i.i.d. Bernoulli random variables. Such a graph is a special case of the model proposed here, but it lacks the sparsity, power-law degree, and unbounded number of vertices that are discussed in Section 3. Barber [2008] also proposed a noisily-observed Clique Matrix model, where two vertices are connected with probability  $\sigma(Z^T Z)$  (where  $\sigma(\cdot)$  represents the sigmoid function). Unlike the noisy-OR formulation for partially observed cliques proposed in Section 2.2, this model precludes sparse graphs, since it allows edges between vertices with no cliques in common.

Kronecker graphs [Leskovec et al., 2010] are a family of random graph models that construct adjacency matrices by starting with a low-dimensional initiator matrix, and iteratively applying the Kronecker product. The resulting deterministic (given the initiator matrix) binary graph can be used to generate a stochastic Kronecker graph, by randomly including edges from the binary graph. Kronecker graphs exhibit a number of realistic graph properties. In particular, their degree distributions follow a multinomial distribution and can exhibit power law behavior depending on the initiator matrix, and for certain parametrizations they yield sparse graphs. However, they do not generally yield high clustering coefficients [Pinar et al., 2012, Durak et al., 2012], as we see in our experimental analysis (Section 5.1). Further, even in the stochastic setting, Kronecker graphs make very strong self-similarity assumptions about the graph, which may be overly restrictive.

Neural networks have been used to learn generative models for sampling from random graphs [You et al., 2018, Simonovsky and Komodakis, 2018, Li et al., 2018]. Typically, these highly parametrized models are learned using multiple graphs. The sparsity properties of these graphs have not been explored in the literature.

Preferential attachment models [Barabási and Albert, 1999, Albert and Barabási, 2002] are another family of random graphs yield graphs with power-law degree distribution, where graphs are constructed by adding edges according to a rich-get-richer rule. However, in stan-

standard preferential attachment models, the average clustering coefficient tends to zero as the graph grows [Albert and Barabási, 2002, Bollobás and Riordan, 2003]. This can be avoided if we have spatial or temporal information about our graph, by incorporating spatial information into the attachment rule [Flaxman et al., 2007, Cooper et al., 2014, Jacob and Mörters, 2015].

Caron and Fox [2017] introduce a family of random graphs or multigraphs parametrized by a generalized gamma process on the space  $\mathcal{V}$  of potential vertices, and create either a multigraph by sampling edges as  $\text{Poisson}(\pi_i \pi_j)$ , or a graph by sampling edges as Bernoulli  $(1 - \exp\{-\pi_i \pi_j\})$  (i.e. truncating the multigraph). This can be generalized further by specifying the model in terms of a function  $W$  applied to a Poisson process on the space  $\mathcal{V} \times \mathbb{R}_+$  [Veitch and Roy, 2015, Borgs et al., 2018]. Under certain choices of  $W$  (including those corresponding to the original Caron and Fox [2017] paper), the resulting graphs are sparse and exhibit power law degree distribution. This framework has been extended to incorporate clustering structure [Herlau et al., 2016, Todeschini et al., 2016].

The edge-exchangeable graph framework [Cai et al., 2016, Crane and Dempsey, 2018] constructs multigraphs that are based on a latent atomic measure  $\mathcal{M} = \sum_i \pi_i \delta_{\theta_i}$ , where edges can be thought of arriving sequentially according to a distribution parametrized by  $\mathcal{M} \times \mathcal{M}$ . Under certain conditions on  $\mathcal{M}$ , the number of vertices grows sub-quadratically with the number of edges as the sequence progresses, and the graph has power law degree distribution. In the specific case where  $\mathcal{M}$  is a normalized generalized gamma process, the edge-exchangeable graph is equivalent to the Caron & Fox multigraph when conditioned on the number of edges. Structure can be incorporated using a mixture of edge-exchangeable graphs [Williamson, 2016].

Due to their exchangeable construction, neither the sparse exchangeable graphs nor the edge-exchangeable graphs are able to obtain graphs with the sort of clique structure shown here. This is because the exchangeable constructions do not allow transitivity: conditioned on the latent representations, the presence of edges  $(i, j)$  and  $(j, k)$  does not increase the probability of the edge  $(i, k)$  also being present. In sparse exchangeable setting, while the resulting graphs will contain dense subgraphs, most of the vertices will not belong to such a subgraph Borgs et al. [2018], Veitch and Roy [2015].

One variant of the edge-exchangeable graphs that is particularly relevant to our work is the multiple edges per step edge-exchangeable graph proposed by Cai et al. [2016]. Here, at each step a random number of edges are added, with edge  $(i, j)$  selected with probability  $\mu_i \mu_j$  in-

dependently of all other edges, leading to multiple edges per step. If—as is given as an example in Cai et al. [2016]—the  $\mu_i$  are the atom sizes of a stable beta process, then the expected total number of edges and vertices, and the expected number of edges between two vertices conditioned on their associated  $\mu_i$ , coincide with the model described in this paper, due to linearity of expectation. The difference arises because the multiple edge-exchangeable model generates edges *independently* at each step. In our model, within each clique, the edges are dependent – giving rise to dense subgraphs. Despite having similar expected sparsity, the multi-step edge-exchangeable model does not have the same local density properties as our model. Further, by representing edge probabilities using a product of beta stable processes, this edge-exchangeable model lacks the conjugacy that is present when we represent clique inclusion probabilities using a single stable beta process. As a result, posterior inference in the multiple-edge stable beta model poses a significant challenge, and has not yet been addressed in the literature.

## 5 EXPERIMENTAL EVALUATION

We explore properties of our distribution over graphs in both the fully observed and the partially observed setting. We infer the clique matrix underlying a random graph using a reversible jump MCMC algorithm that proposes either splitting or merging cliques. In the partially observed setting, we augment this split/merge procedure with Gibbs steps that help learn the fine-grained structure of the clique cover; we found that adding Gibbs steps had little benefit in the fully observed setting. Full inference details are given in the supplement.

In our experiments, we optimize the hyperparameters using gradient descent; an alternative approach would be to place a prior on them and infer their posterior distribution using Metropolis-Hastings proposals.

### 5.1 FULLY OBSERVED GRAPH SETTING

We begin by ascertaining how well the fully observed model can capture statistical properties of real-world graphs. We consider the following graph statistics: ratio of triangles to vertices; density  $D = 2|E|/|V|(|V| - 1)$ ; average degree; average maximal clique per vertex; average local clustering coefficient. We also look at the degree distribution and the distribution over the maximal clique per vertex. We use the `NetworkX` python package [Hagberg et al., 2008] to calculate clustering coefficients and maximal cliques.

To determine how well-suited our model is for modeling real graphs, we consider four real-world graphs,

	Truth	RCC	BNP	Kron
triang./vertex	1.34	1.99(.13)	0.01(.01)	1.21(.06)
density ( $\times 1k$ )	1.29	1.74(.01)	1.28(.01)	5.03(.11)
av. degree	4.16	4.40(.14)	3.47(.18)	7.70(.09)
max. clique	3.50	4.20(.09)	2.03(.01)	2.49(.03)
cluster. coeff.	0.59	0.69(.02)	0.00(.00)	0.06(.01)

(a) NeurIPS collaborations.

	Truth	RCC	BNP	Kron
triang./vertex	1.91	4.14(.27)	0.01(.01)	0.01(.01)
density ( $\times 1k$ )	2.00	2.70(.01)	2.00(.00)	0.75(.08)
av. degree	4.57	6.05(.17)	4.58(.12)	2.60(.02)
max. clique	4.28	5.52(.14)	2.03(.01)	2.03(.01)
cluster. coeff.	0.81	0.78(.02)	0.00(.00)	0.00(0.01)

(b) IMDB comedies.

	Truth	RCC	BNP	Kron
triang./vertex	9.21	3.25(.40)	0.07(.01)	0.01(.01)
density ( $\times 1k$ )	1.06	1.29(.01)	1.07(.00)	0.47(.01)
av. degree	5.53	6.53(.47)	5.61(.16)	3.25(.01)
max. clique	5.13	4.64(.15)	2.11(.01)	2.03(.01)
cluster. coeff.	0.56	0.64(.02)	0.00(.01)	0.00(.01)

(c) ArXiv GR-QC.

	Truth	RCC	BNP	Kron
triang./vertex	2.76	2.26(.53)	2.27(.80)	2.71(.11)
density ( $\times 1k$ )	6.30	7.58(.53)	6.45(.44)	6.05(.12)
av. degree	7.33	7.58(.64)	7.62(.68)	9.67(.09)
max. clique	3.17	3.21(.10)	2.58(.01)	2.74(.04)
cluster. coeff.	0.22	0.25(.02)	0.05(.01)	0.09(.01)

(d) ENRON emails.

	Truth	RCC	BNP	Kron
triang./vertex	356	211(71)	35.1(7.0)	4.77(.50)
density ( $\times 1k$ )	10.8	12.0(.58)	11.1(.41)	8.7(.04)
av. degree	43.7	48.0(2.6)	44.9(2.3)	35.5(.01)
max. clique	15.9	7.27(.50)	3.63(.09)	3.11(.03)
cluster. coeff.	0.60	0.42(.04)	0.05(.01)	0.02(.01)

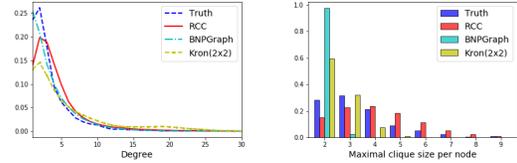
(e) Facebook circles.

Table 1: Empirical evaluation of statistical properties of random graphs generated from the parameters learned from real-world graphs (standard error in parentheses). The parameters are taken from posterior MCMC draws.

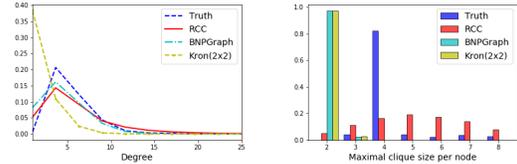
- The co-authorship graph for NeurIPS publications from 1987-2003 (2715 vertices, 4733 edges).<sup>2</sup>
- The largest connected component of the co-appearance graph generated from IMDB cast lists of comedies from 2000-2002 (2288 vertices, 5232 vertices).<sup>3</sup>
- The co-authorship graph for arXiv publications in the category GR-QC (General Relativity and Quantum Cosmology) from 2003 (5241 vertices, 14484

<sup>2</sup><https://www.kaggle.com/benhamner/nips-papers>

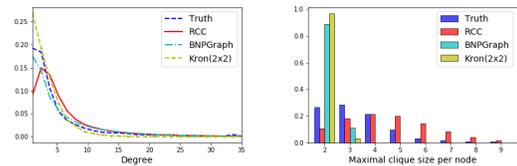
<sup>3</sup>[www.imdb.com](http://www.imdb.com)



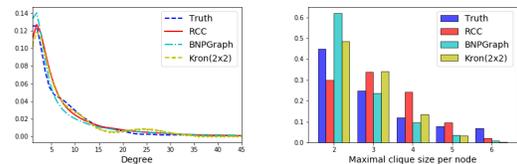
(a) NeurIPS collaborations (1987-2003).



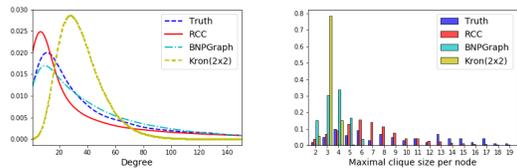
(b) IMDB comedies cast (2000-2002).



(c) ArXiv GR-QC (up to 2003).



(d) ENRON (1 month of data)



(e) Facebook circles [McAuley and Leskovec, 2012]

Figure 2: Empirical degree distributions and maximal clique distributions.

edges) [Leskovec et al., 2007].

- The full interaction graph from a single month (January 2000) of the ENRON dataset (1172 vertices, 4293 edges) [Klimt and Yang, 2004].
- Anonymous data collected from Facebook social circles,<sup>4</sup> comprising 10 interconnected friend lists of people obtained using an external app (4039 vertices, 88234 edges) [McAuley and Leskovec, 2012].

We model each graph using our proposed (fully ob-

<sup>4</sup>The computation of the maximal clique per vertex for the Facebook social circles was omitted due to the size of this graph.

Clique	Mean dist.	Authors
1	1.61	Weare(2), Platt(2), Burges(2), Liblit(1), Aiken(1), Zheng(1), Swenson(2), Crisp(3)
2	1.86	Sundararajan(3), Ghaoui(1), Bhattacharyya(1), Lanckriet(1), Keerthi(2), Nilim(2)
3	1.91	Kim(1), Sastry(1), Xing(1), Weiss(1), Yedidia(2), Ng(1), Russell(1), Freeman(2), Karp(1), Yanover(2)
4	2.16	Fukumizu(1), Amari(2), Bartlett(1), Wu(3), McAuliffe(1), Nakahara(3), Murata(3), Akaho(2), Bach(1)
5	1.82	Blei(1), Stromsten(2), DeSilva(2), Kemp(2), Ng(1), Steyvers(2), Griffiths(1), Danks(2), Tenenbaum(1), Sanjana(2)
6	2.36	Chichilnisky(3), Nguyen(1), Simoncelli(2), Pillow(3), Schwartz(3), Wang(3), Wainwright(1), Todorov(1), Paninski(3)

Table 2: Meta-analysis of the six cliques of Michael Jordan. Shortest-path distance to Michael Jordan is shown in parentheses; connections not present in the original graph are shown in red. Mean dist. is the mean path length (along the original graph) between clique members.

served) random clique cover model—labeled RCC in tables and figures—and learn the hyperparameters using maximum likelihood. We then sample 25 graphs using these hyperparameters, and compare the average statistics of the sampled graph with those of the original graph in Table 1 and Figure 2. We consider two comparison models:

- **BNPgraph**: The sparse exchangeable model of Caron and Fox [2017] (implemented using code published by the authors), which can capture graph sparsity, power-law degree distribution, and an unbounded number of vertices.
- **Kron**: A stochastic Kronecker graph, with  $2 \times 2$  initiator matrix inferred and sampled using the `kronfit` and `krongen` software packages introduced by Leskovec et al. [2010].  $2 \times 2$  initiator matrices have been shown to be a good fit for real-world graph models [Leskovec et al., 2010].

We see that both the Caron & Fox model and our model models do a good job at capturing sparsity and degree distribution. However, the Caron & Fox model does not capture the locally dense structure. While simulations from our model have similar clustering coefficients to the real graphs, the Caron & Fox model has near-zero clustering coefficients. Our simulations also exhibit larger maximal cliques. While the maximal clique distribution for the IMDB dataset seems to overestimate the number of cliques, we believe this is an artifact of the data. The dataset only includes the four highest billed actors for each movie, artificially deflating clique sizes. The Kronecker graph generally does a worse job at capturing the graph sparsity and degree, which we hypothesize is due to the restrictive self-similarity of the underlying graph.

We further inspected the differences between models by

looking at the adjacency matrices and visual representations of the random graphs generated from the best-fitted parameters. Figure 3 shows the results. For ease of visualization, we only include the top-left  $100 \times 100$  block of the adjacency matrices; the full matrices are included in the supplement. The sparse exchangeable graph is sparse everywhere with no community structure, whereas the Kronecker graph shows a structure unrelated to the original graph. In contrast, our model produces a graph that looks more similar to the original graph, and with a similar community structure shown by the adjacency matrix. We used the Python package `NetworkX` [Hagberg et al., 2008] to draw the graphs with a spring layout; we also used its implementation of the Girvan-Newman algorithm [Girvan and Newman, 2002] to arrange the adjacency matrices by communities.

## 5.2 PARTIALLY OBSERVED GRAPH SETTING

The partially observed setting offers the ability to infer a latent graph—and corresponding set of latent cliques—underlying an observed graph. Intuitively, this latent graph is likely to capture latent community-type structure. To evaluate this empirically, we modeled the NeurIPS co-authorship dataset described above using a partially observed model. We consider publications within the period 1999-2003, assume a shared clique parameter  $\pi$ , and infer hyperparameters using Metropolis-Hastings. Figure 4 shows the latent structure found (based on a single sample from the posterior).

Posterior draws for  $\pi$  concentrated around 0.4, robustly with respect to hyperparameter specification. Thus, roughly 60% of the edges in the inferred graph are not in the original graph. As expected given the short period of analysis, 491 out of the 553 authors in consideration belonged to only one clique, and only eleven of

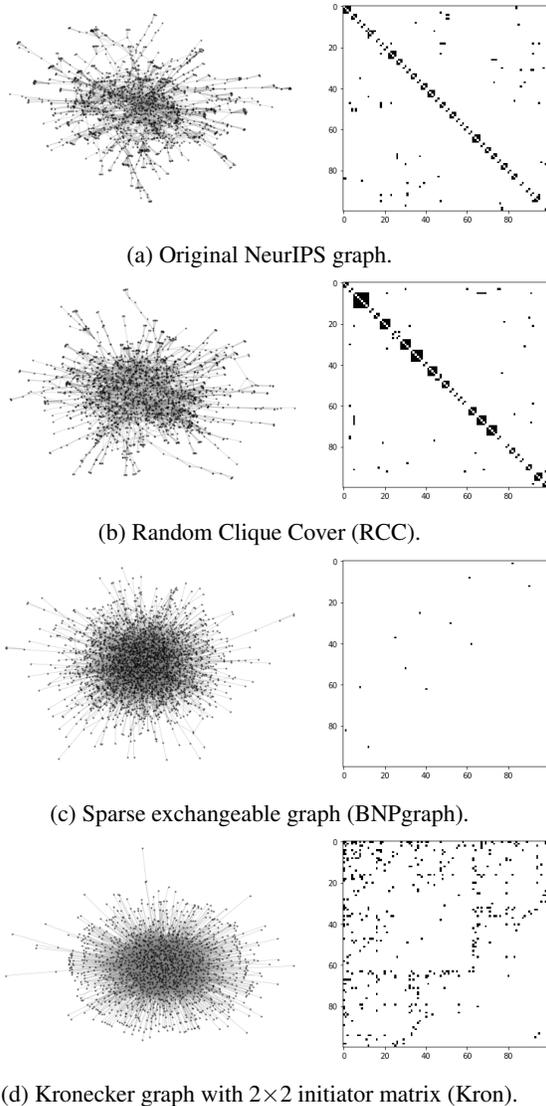


Figure 3: Co-authorship graph for NeurIPS publications (1987-2003) and graphs sampled using (b) our model, and (c, d) comparison methods. Left: graph visualization; right: top-left  $100 \times 100$  block of the adjacency matrix arranged by communities.

them to more than three. The average clique size was 8.93. The author belonging to the most cliques was (perhaps unsurprisingly) Michael Jordan. The members of the six cliques he belonged to are shown in Table 2. It is worth mentioning that case-by-case inspection show that in some cases, inferred edges were in fact collaborations outside the period of observation.

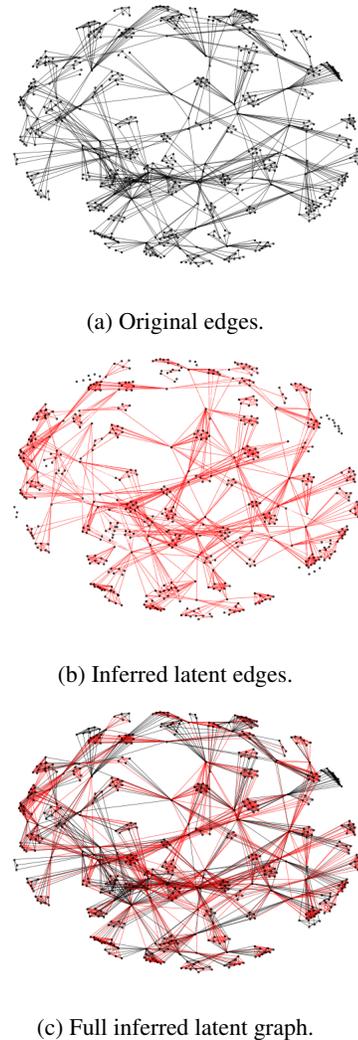


Figure 4: Partially observed graph inference for NeurIPS co-authorships (1999-2003), edges appearing in the latent cliques but not the original graph are shown in red

## 6 DISCUSSION

We have presented a new class of Bayesian nonparametric prior for graphs, based on random clique selection mechanisms, that are appropriate for many real-world graphs. We have presented some preliminary work in a modeling context; we hope to see this form of random graph-based model explored further in future.

While we base our model on the stable beta process, alternative subset selection mechanisms could be proposed. For example, a restricted stable beta IBP [Williamson et al., 2013, Utkovski et al., 2018] could give more explicit control over the number of cliques. We leave exploration of alternative mechanisms as future work.

## References

- Edoardo M Airolidi, David M Blei, Stephen E Fienberg, and Eric P Xing. Mixed membership stochastic block-models. *Journal of Machine Learning Research*, 9: 1981–2014, 2008.
- Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1):47, 2002.
- Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.
- David Barber. Clique matrices for statistical graph decomposition and parameterising restricted positive definite matrices. In *Uncertainty in Artificial Intelligence*, pages 26–33, 2008.
- Benjamin BloemReddy and Peter Orbanz. Randomwalk models of network formation and sequential Monte Carlo methods for graphs. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(5):871–898, 2018.
- Béla Bollobás and Oliver M Riordan. Mathematical results on scale-free random graphs. *Handbook of graphs and networks: from the genome to the internet*, pages 1–34, 2003.
- Christian Borgs, Jennifer T Chayes, Henry Cohn, and Nina Holden. Sparse exchangeable graphs and their limits via graphon processes. *Journal of Machine Learning Research*, 18(210):1–71, 2018.
- Tamara Broderick, Michael I Jordan, and Jim Pitman. Beta processes, stick-breaking and power laws. *Bayesian Analysis*, 7(2):439–476, 2012.
- Diana Cai, Trevor Campbell, and Tamara Broderick. Edge-exchangeable graphs and sparsity. In *Advances in Neural Information Processing Systems*, pages 4249–4257, 2016.
- François Caron and Emily B Fox. Sparse graphs using exchangeable random measures. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(5):1295–1366, 2017.
- Colin Cooper, Alan Frieze, and Paweł Prałat. Some typical properties of the spatial preferred attachment model. *Internet Mathematics*, 10(1-2):116–136, 2014.
- Harry Crane and Walter Dempsey. Edge exchangeable models for interaction networks. *Journal of the American Statistical Association*, 113(523):1311–1326, 2018.
- Chavdar Dangalchev. Generation models for scale-free networks. *Physica A: Statistical Mechanics and its Applications*, 338(3-4):659–671, 2004.
- Nurcan Durak, Ali Pinar, Tamara G Kolda, and C Seshadhri. Degree relations of triangles in real-world networks and graph models. In *International Conference on Information and Knowledge Management*, pages 1712–1716, 2012.
- Paul Erdős and Alfréd Rényi. On the evolution of random graphs. *Publication of the Mathematical Institute of the Hungarian Academy of Sciences*, 5(1):17–60, 1960.
- Abraham D Flaxman, Alan M Frieze, and Juan Vera. A geometric preferential attachment model of networks ii. *Internet Mathematics*, 4(1):87–111, 2007.
- Edgar N Gilbert. Random graphs. *The Annals of Mathematical Statistics*, 30(4):1141–1144, 1959.
- Michelle Girvan and Mark E. J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12): 7821–7826, 2002.
- Aric Hagberg, Pieter Swart, and Daniel S Chult. Exploring network structure, dynamics, and function using NetworkX. Technical report, Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2008.
- Ronald C Hamelink. A partial characterization of clique graphs. *Journal of Combinatorial Theory*, 5(2):192–197, 1968.
- Creighton Heaukulani and Daniel M Roy. Gibbs-type Indian buffet processes. *Bayesian Analysis, Advance publication*, 2018.
- Tue Herlau, Mikkel N Schmidt, and Morten Mørup. Completely random measures for modelling block-structured sparse networks. In *Advances in Neural Information Processing Systems 29*, pages 4260–4268, 2016.
- Nils Lid Hjort. Nonparametric Bayes estimators based on beta processes in models for life history data. *The Annals of Statistics*, 18(3):1259–1294, 1990.
- Paul W Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social Networks*, 5(2):109–137, 1983.
- Emmanuel Jacob and Peter Mörters. Spatial preferential attachment networks: Power laws and clustering coefficients. *The Annals of Applied Probability*, 25(2): 632–662, 2015.
- Brian Karrer and Mark EJ Newman. Stochastic blockmodels and community structure in networks. *Physical Review E*, 83(1):016107, 2011.
- John Kingman. Completely random measures. *Pacific Journal of Mathematics*, 21(1):59–78, 1967.

- Bryan Klimt and Yiming Yang. The Enron corpus: A new dataset for email classification research. In *European Conference on Machine Learning*, pages 217–226, 2004.
- Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. Graph evolution: Densification and shrinking diameters. *ACM Transactions on Knowledge Discovery from Data*, 1(1):2, 2007.
- Jure Leskovec, Deepayan Chakrabarti, Jon Kleinberg, Christos Faloutsos, and Zoubin Ghahramani. Kronecker graphs: An approach to modeling networks. *Journal of Machine Learning Research*, 11:985–1042, 2010.
- Yujia Li, Oriol Vinyals, Chris Dyer, Razvan Pascanu, and Peter Battaglia. Learning deep generative models of graphs. *arXiv preprint arXiv:1803.03324*, 2018.
- James Lloyd, Peter Orbanz, Zoubin Ghahramani, and Daniel M Roy. Random function priors for exchangeable arrays with applications to graphs and relational data. In *Advances in Neural Information Processing Systems*, pages 998–1006, 2012.
- Julian McAuley and Jure Leskovec. Learning to discover social circles in ego networks. In *Advances in Neural Information Processing Systems*, pages 539–547, 2012.
- Jaroslav Neetil and Patrice Ossona de Mendez. *Sparsity: Graphs, Structures, and Algorithms*. Springer, 2012.
- James Orlin. Contentment in graph theory: Covering graphs with cliques. *Indagationes Mathematicae (Proceedings)*, 80(5):406–424, 1977.
- Ali Pinar, Comandur Seshadhri, and Tamara G Kolda. The similarity between stochastic Kronecker and Chung-Lu graph models. In *SIAM International Conference on Data Mining*, pages 1071–1082, 2012.
- Martin Simonovsky and Nikos Komodakis. GraphVAE: Towards generation of small graphs using variational autoencoders. In *International Conference on Artificial Neural Networks*, pages 412–422, 2018.
- Yee W Teh and Dilan Görür. Indian buffet processes with power-law behavior. In *Advances in Neural Information Processing Systems*, pages 1838–1846, 2009.
- Romain Thibaux and Michael I Jordan. Hierarchical beta processes and the Indian buffet process. In *Artificial Intelligence and Statistics*, pages 564–571, 2007.
- Adrien Todeschini, Xenia Miscouridou, and François Caron. Exchangeable random measures for sparse and modular graphs with overlapping communities. *arXiv preprint arXiv:1602.02114*, 2016.
- Zoran Utkovski, Melanie F Pradier, Viktor Stojkoski, Fernando Perez-Cruz, and Ljupco Kocarev. Economic complexity unfolded: Interpretable model for the productive structure of economies. *PLoS one*, 13(8): e0200822, 2018.
- Victor Veitch and Daniel M Roy. The class of random graphs arising from exchangeable random measures. *arXiv preprint arXiv:1512.03099*, 2015.
- Duncan J Watts and Steven H Strogatz. Collective dynamics of small-world networks. *Nature*, 393(6684): 440, 1998.
- Sinead A Williamson. Nonparametric network models for link prediction. *The Journal of Machine Learning Research*, 17(1):7102–7121, 2016.
- Sinead A Williamson, Steve N MacEachern, and Eric P Xing. Restricting exchangeable nonparametric distributions. In *Advances in Neural Information Processing Systems*, pages 2598–2606, 2013.
- Jiaxuan You, Rex Ying, Xiang Ren, William Hamilton, and Jure Leskovec. GraphRNN: Generating realistic graphs with deep auto-regressive models. In *International Conference on Machine Learning*, pages 5708–5717, 2018.