THE CLASS OF RANDOM GRAPHS ARISING FROM EXCHANGEABLE RANDOM MEASURES

VICTOR VEITCH AND DANIEL M. ROY

Abstract. We introduce a class of random graphs that we argue meets many of the desiderata one would demand of a model to serve as the foundation for a statistical analysis of real-world networks. The class of random graphs is defined by a probabilistic symmetry: invariance of the distribution of each graph to an arbitrary relabelings of its vertices. In particular, following Caron and Fox, we interpret a symmetric simple point process on $\mathbb{R}_+^2$ as the edge set of a random graph, and formalize the probabilistic symmetry as joint exchangeability of the point process. We give a representation theorem for the class of random graphs satisfying this symmetry via a straightforward specialization of Kallenberg's representation theorem for jointly exchangeable random measures on $\mathbb{R}_+^2$. The distribution of every such random graph is characterized by three (potentially random) components: a nonnegative real $I \in \mathbb{R}_+$, an integrable function $S : \mathbb{R}_+ \to \mathbb{R}_+$, and a symmetric measurable function $W : \mathbb{R}_+^2 \to [0, 1]$ that satisfies several weak integrability conditions. We call the triple (I, S, W) a graphex, in analogy to graphons, which characterize the (dense) exchangeable graphs on $\mathbb{N}$. Indeed, the model we introduce here contains the exchangeable graphs as a special case, as well as the "sparse exchangeable" model of Caron and Fox. We study the structure of these random graphs, and show that they can give rise to interesting structure, including sparse graph sequences. We give explicit equations for expectations of certain graph statistics, as well as the limiting degree distribution. We also show that certain families of graphexes give rise to random graphs that, asymptotically, contain an arbitrarily large fraction of the vertices in a single connected component.

Contents

1. Introduction

Random graph models are a key tool for understanding the structure of real-world networks, especially through data. In particular, a random graph model can serve as the foundation for a statistical analysis: observed link structure is modeled as a realization from the random graph model, whose parameters are in some unknown

configuration. The goal is to then infer the configuration of the parameters, and in doing so, understand properties of the network that gave rise to the observed link structure.

The quality of the inferences we can make depends in part on the fidelity of the model, but building realistic models of networks is challenging: the models must be simple enough to be tractable, yet flexible enough to accurately represent a wide range of phenomena. In the setting of densely connected networks, the well-known exchangeable graph model provides a tractable yet general framework. However, the vast majority of real-world networks are sparsely connected—two nodes chosen at random are very unlikely to be directly connected by a link. Accordingly, for some configuration of their parameters, realistic random graph models for networks must be sparse, exhibiting only a vanishing fraction of all possible edges as they become large. At the same time, the link structure of real-world networks is rich: e.g., in social networks, phenomena such as homophily (informally, friends of friends are more likely to be friends), "small-world" connectivity (two randomly chosen individuals are likely to be connected by a short path of friendship), and power law degree distributions (the number of friends an individual may have varies across many orders of magnitude) are common [New09; Dur06]. It is a remarkable gap in modern statistical practice that there is no general framework for the statistical analysis of real-world networks.

There is no shortage of proposals for random graph models of real-world networks; however, these models tend to be ad hoc, exhibiting certain properties of real-world networks by design, but behaving pathologically in other aspects. It is difficult to assess the statistical applicability of such models.

One approach to identifying large but tractable families of random graphs is to consider the family of all random graphs satisfying a small number of natural assumptions. In this paper, we define a class of random graph models in terms of a single invariance principle: that the distribution of a graph should be invariant to an arbitrary relabeling of its vertices. From this assumption, we derive and study a general class of random graphs suitable for modeling network structures. We show that these random graphs admit a simple, tractable specification and give rise to complex structures of the kinds observed in real world networks. Moreover, our derivation is closely analogous to an approach that has been used to define broadly useful statistical models in other settings. For instance, the classical i.i.d. setting and the graphon setting for densely connected networks are both derived from analogous invariance assumptions [OR15]. Indeed, we show that the exchangeable graph models are a special case of the models we derive here. These observations suggest that the models we identify in this paper may be broadly useful for the statistical analysis of real-world networks.

To explain our approach we begin by reviewing a closely related approach used to define models for the statistical analysis of densely connected networks. In this setting, networks are modeled as random graphs represented by their adjacency matrices; an observed $n \times n$ adjacency matrix is modeled as the leading size-n principal submatrix of some infinite array of random variables. The infinite structure automatically provides consistent models for datasets of different size. The foundational structural assumption by which the dense graph framework is defined is a probabilistic symmetry: *joint exchangeability of the infinite array.* This is the requirement that the distribution of the infinite array is invariant under joint

permutations of the indices of the array; intuitively, this means that the labeling of the vertices of a graph does not carry information about its structure.

The statistical framework can be derived using the Aldous–Hoover representation theorem for jointly exchangeable arrays. Specialized to the case of infinite adjacency matrices, this theorem asserts that the adjacency matrix of a random graph on $\mathbb{N}$ is jointly exchangeable iff its distribution can be written as a mixture over a certain privileged family of distributions (namely, the ergodic measures). Each member of this family is specified in terms of a symmetric, measurable function $W : [0,1]^2 \to [0,1]$, now known as a *graphon*. It follows that the space of probability distributions on $n \times n$ observations of a densely connected networks can be parameterized by the space of graphons. A particular consequence of the theorem is that the expected number of links among every n individuals is $\binom{n}{2}\|W\|_1$; i.e., the graph is either empty or dense. As stated plainly in [OR15], these models are thus misspecified as statistical models for real-world networks.

The derivation of the dense graph framework is a particular instance of a general recipe for constructing statistical models: a probabilistic symmetry is assumed on some infinite random structure and an associated representation theorem characterizes the ergodic measures, forming the foundation of a framework for statistical analysis. The first main contribution of the present paper is the analogous representation theorem for the sparse (and dense) graph setting, which we arrive at by a straightforward adaptation of a result of Kallenberg [Kal90; Kal05]. Our inspiration comes from recent paper of Caron and Fox [CF14] that exploits a connection between random measures and random graphs to exhibit a class of sparse random graphs. In their paper, they observe that their random graphs satisfy a natural analogue of joint exchangeability when considered as a point process and make use of an associated representation theorem to study the model. The present paper reverses this chain of reasoning, beginning with the symmetry on point processes and elucidating the full family of random graphs that arise from the associated representation theorem. In the graph context, joint exchangeability of point processes retains the interpretation that the labels of vertices carry no information about the structure of the graph.

Following Caron and Fox, we represent random graphs as an infinite simple point processes on $\mathbb{R}_+^2$ with finite random graphs given by truncating the support of the point process to a finite set (see Fig. 3). The representation theorem associated to joint exchangeability of point processes is known by the work of Kallenberg [Kal90; Kal05]. We arrive at our representation theorem by a straightforward translation of this result into the random graph setting. The random graphs picked out by our representation theorem have three possible components: isolated edges, infinite stars, and a final piece that provides the interesting graph structure. The basic object for the distributions of these random graphs is a triple (I, S, W) where $I \in \mathbb{R}_+$, $S : \mathbb{R}_+ \to \mathbb{R}_+$ is integrable, and $W : \mathbb{R}_+^2 \to [0,1]$ is a symmetric measurable function satisfying certain weak integrability conditions. (See Theorem 4.9; W integrable is sufficient but not necessary.) We call the triple a *graphex*. In this paper we focus on random graphs without isolated edges or infinite stars, and so we take $I = S = 0$; when there is no risk of confusion, we will use the term *graphex* to refer to the function W alone with the understanding that the triple is then of the form $(0, 0, W)$. The distribution of every such random graph, which we call a Kallenberg exchangeable graph, is characterized by some (possibly random) graphex.

Graphexes are the analogues of graphons and the space of distributions on (sparse) graphs can be parameterized by the space of graphexes.

It remains to explain the construction of the random graph associated with a graphex. Let $\boldsymbol{\theta} = \mathbb{R}_+$ be the space of labels of the graph, $\boldsymbol{\vartheta} = \mathbb{R}_+$ be the space of latent parameters, and Π be a unit rate Poisson process on $\boldsymbol{\theta} \times \boldsymbol{\vartheta}$. Intuitively, the random graph is given by independently randomly including each pair of points in Π as an edge of graph with a probability determined by the graphex W. A point of the Poisson process is included as a vertex of the graph if and only if it participates in at least one edge. The construction of the random graph is explained in Fig. 1. Formally, treating the collection of edges $\{(\theta_i, \theta_j)\}$ as the basic random object of interest the generative model given W and Π is:

$$(\theta_i, \theta_j) \mid W, \Pi \overset{ind}{\sim} \text{Bernoulli}(W(\vartheta_i, \vartheta_j)). \tag{1.1}$$

Finite size graphs are given by restricting to only edges (θ_i, θ_j) such that $\theta_i, \theta_j < \nu$ and including vertices only if they participate in at least one such edge. These distributions are consistent for datasets of different sizes and admit sparse graphs, allowing for the realistic modeling of physical networks. Moreover, in a sense we make precise in Section 3.1, the exchangeable graphs derived from the Aldous–Hoover theory are contained as a subfamily of the Kallenberg exchangeable graphs, and correspond those graphs generated by graphexes of the form $(0, 0, W)$ where W is compactly supported, and therefore equal to the dilation of some graphon. Thus the KEG framework is a generalization of the exchangeable graph framework to the sparse graph regime.

Let G_ν be the random graph given by truncating the label space $\boldsymbol{\theta}$ to $[0, \nu]$ (see Fig. 1); we call the random graph model $(G_\nu)_{\nu \in \mathbb{R}_+}$ the *Kallenberg exchangeable graph* (KEG) associated with W. The bulk of the present paper is devoted to deriving properties of these graphs in terms of the graphex W. For simplicity of presentation we ignore self edges here, giving full statements in the body of the paper. Let $\mu_W(x) = \int_{\mathbb{R}_+} W(x, y)\mathrm{d}x$.

(1) Given a point (θ, ϑ) in the latent Poisson process, the degree of the vertex labeled θ is Poisson distributed with mean $\nu\mu_W(\vartheta)$.

(2) The expected number of edges $e_\nu = |e(G_\nu)|$ is

$$\mathbb{E}[e_\nu] = \frac{1}{2}\nu^2 \iint_{\mathbb{R}_+^2} W(x, y)\mathrm{d}x\mathrm{d}y. \tag{1.2}$$

(3) The expected number of vertices $v_\nu = |v(G_\nu)|$ is

$$\mathbb{E}[v_\nu] = \nu \int_{\mathbb{R}_+} (1 - e^{-\nu\mu_W(x)})\mathrm{d}x. \tag{1.3}$$

(4) Subject to some technical constraints, the scaling limit of the asymptotic degree distribution has an explicit expression in terms of W. Let k_ν be some non-decreasing function of ν and let D_ν be the degree of a randomly selected vertex of G_ν, then

$$P(D_\nu \geq k_\nu \mid G_\nu) \xrightarrow{p} \lim_{\nu \to \infty} \frac{\sum_{k=k_\nu}^{\infty} \frac{\nu^k}{k!} \int \mu_W(x)^k e^{-\nu\mu_W(x)}\mathrm{d}x}{\int_{\mathbb{R}_+} (1 - e^{-\nu\mu_W(x)})\mathrm{d}x}. \tag{1.4}$$

This result establishes that the random graph construction in this paper can give rise to sparse graphs.

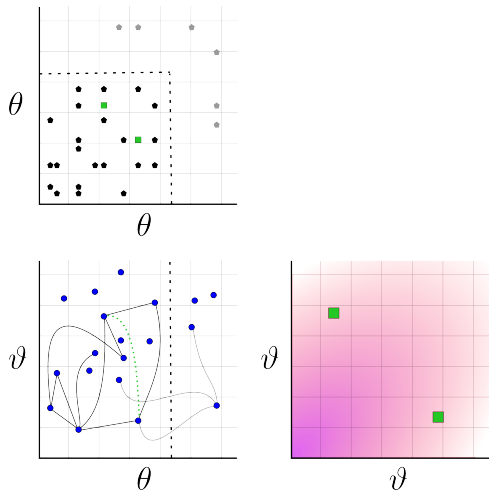

FIGURE 1. (Kallenberg exchangeable graph) Random graphs arising from exchangeable random measures are characterized by three (potentially random) components: a non-negative real $I \in \mathbb{R}_+$, an integrable function $S : \mathbb{R}_+ \to \mathbb{R}_+$, and a symmetric measurable function $W : \mathbb{R}_+^2 \to [0,1]$ satisfying some weak integrability conditions. We call the triple (I, S, W) a graphex. The most interesting structure arises from W. A particular W is illustrated by the magenta heatmap (lower right). Given W, an infinite random graph with a vertex set in $\boldsymbol{\theta}$ is generated in this model according to:

1. Sample a (latent) unit rate Poisson process Π on $\boldsymbol{\theta} \times \boldsymbol{\vartheta}$.
2. For each pair of points $(\theta_i, \vartheta_i), (\theta_j, \vartheta_j) \in \Pi$ include edge (θ_i, θ_j) with probability $W(\vartheta_i, \vartheta_j)$.
3. Include θ_i as a vertex whenever θ_i participates in at least one edge.

Finite subgraphs are given by restricting the space $\boldsymbol{\theta}$ to be less than some finite value. The lower left panel of the figure shows a realization of a latent Poisson process with a realization of the edge structure superimposed. A finite subgraph (black edges) is given by taking only points with $\theta < 4.2$. The edge $(3.2, 2.1)$ (green, dotted squares) is included with probability $W(1.1, 4.7) = W(4.7, 1.1)$; this is shown in the middle panel. Edges that include a point of Π with $\theta > 4.2$ (grey, transparent) are not included in the subgraph. Vertices, such as 2.7, that participate only in edges with a terminus that has $\theta > 4.2$ are not included in the subgraph. The upper left panel shows the pictured graph as a realization of a random measure on $\boldsymbol{\theta} \times \boldsymbol{\theta}$ space.

(5) Certain choices of W admit highly connected graphs. Suppose $W(x,y) = f(x)f(y)$, let $C_1(G_\nu)$ be the largest connected component of G_ν, and let $\epsilon > 0$, then

$$\lim_{\nu\to\infty} P(|C_1(G_\nu)| > (1-\epsilon)|v\,(G_\nu)|) = 1. \tag{1.5}$$

This means that the sparse structure can arise in an interesting way: it is not simply a consequence of having a collection of disjoint dense graphs.

We begin by giving background on random graph modeling and the use of probabilistic symmetry in Section 2. In Section 3, we give a number of illustrative examples of Kallenberg exchangeable graphs to make the construction concrete. In Section 4, we establish the representation theorem and give a formal characterization of the models we derive. In Section 5, we derive the first moments of several graph statistics of G_ν using point process techniques, allowing self edges. An expression for asymptotic degree distribution of these graphs in terms of the graphex is derived in Section 6. Finally, in Section 7, we study the structure of the Kallenberg exchangeable graphs generated by graphexes of the form $W(x,y) = f(x)f(y)1[x \neq y]$ with the goal of establishing the asymptotic connectivity structure. Several other interesting features of these random graphs are uncovered in the course of establishing this result. In particular, we show that degree power law distributions and "small-world" phenomena arise naturally in this framework.

2. Background

In order to relate the Kallenberg exchangeable graph model to a diverse range of existing random graph models, it will be useful to have a general definition for the term 'random graph model'. In this paper, a random graph model is an indexed family of graph-valued random variables $G_{s,\phi}$, where s specifies the "size" of the graph and takes values in a totally ordered set S, and where $\phi \in \Phi$ determines some distributional properties (and so could play the role of a parameter in a statistical model). We will write $\mu_{s,\phi}$ for the distribution of $G_{s,\phi}$.[1] Our definition is deliberately vague about the meaning of 'graph-valued' as different models will naturally be described in terms of different concrete spaces.

For example, the well-known Erdős–Rényi–Gilbert model is the family of simple random graphs $G_{n,p}$ on $n \in \mathbb{N}$ vertices, where each edge appears independently with probability $p \in [0,1]$. Concretely, we can think of $G_{n,p}$ as a random $n \times n$ adjacency matrix, or equivalently, as a symmetric $n \times n$ array of 0/1-valued (i.e., binary) random variables whose diagonal is zero. In a statistical setting, we might model the network of friendships among n individuals as a realization of $G_{n,p}$ for some unknown p. In this case, the goal of statistical analysis would be to make inferences about the parameter p given some particular observed dataset in the form of an adjacency matrix.

The Erdős–Rényi–Gilbert model can be seen as special case of the more general random graph model that arises from the graphon theory or from the Aldous–Hoover representation theorem. In this case, the size again determines the number of vertices, but the parameter is a graphon, i.e., a symmetric, measurable function $W : [0,1]^2 \to [0,1]$. (The Erdős–Rényi–Gilbert model corresponds with constant graphons $W(x,y) = p$ for some $p \in [0,1]$.) This class of random graphs are known

[1] In a statistical setting, the family of distributions $\mu_{s,\phi}$ would be the natural structure to call a model. Here we adopt the language of graph theorists.

as the *exchangeable graphs*, although we will sometimes refer to them as the (dense) exchangeable graphs to distinguish them from the Kallenberg exchangeable graphs.

In the exchangeable graph model, the size parameter is the number of vertices. This is the typical approach to indexing random graph models. In contrast, the size parameter of a Kallenberg exchangeable graph model is a non-negative real ν that is proportional to the square root of the *expected* number of *edges*.

2.1. **Desiderata for random graph models.** For the purpose of modeling real-world networks, one of the key properties of a random graph model is the relationship between the number of edges and vertices. Consider a random graph model $G_{s,\phi}$, fix a parameter ϕ, and let $s_n \uparrow \infty$ be some diverging sequence of sizes. For a graph G, let $|e(G)|$ and $|v(G)|$ denote the number of edges and vertices, respectively. To avoid pathologies, we will assume that $|v(G)| \to \infty$ as $n \to \infty$. Then the sequence $(G_{s_n,\phi})$ is *sparse* or *not dense* if, with probability one,

$$\frac{\sqrt{|e(G)|}}{|v(G)|} \to 0 \qquad \text{as } n \to \infty. \tag{2.1}$$

This condition states that, asymptotically, graphs with v vertices have $o(v^2)$ edges. More generally, it is interesting to identify whether there is a (potentially random) exponent k such that, asymptotically, there are $\Theta(v^k)$ edges.

For statistical applications, it is desirable to impose a desideratum in addition to sparsity. The prototypical statistical network analysis has the following structure: an observed network g_s is modeled as a realization of a random graph $G_{s,\phi}$ for some size s and for some unknown parameter ϕ; the goal is to infer the parameter ϕ. In some random graph models, the sequence $G_{s_1,\phi}, G_{s_2,\phi}, \ldots$ of graphs is a model of the dynamics by which a network grows and evolves. In the statistical problems motivating this paper, however, the size parameter s is akin to sample size in the sense that collecting more data corresponds to choosing larger values of s. It is therefore natural to demand that the distributions associated with different sizes are "consistent" with one another in the sense that moving from $G_{s,\phi}$ to $G_{t,\phi}$, for $t > s$, can be understood as collecting additional data.

One way to formalize this notion of consistency is to demand that the distributions of the random graphs $G_{s,\phi}$ be *projective*. Projectivity is defined in terms of a projective system, i.e., a family of measurable maps $(f_{s,t};\, s \le t \in S)$ where $f_{s,t}$ maps graphs of size t to graphs of size $s \le t$, $f_{s,s}$ is the identity, and $f_{r,t} = f_{r,s} \circ f_{s,t}$ for all $r \le s \le t$. A random graph model is *projective* if, for some projective system $(f_{s,t};\, s \le t \in S)$, it holds that $G_{s,\phi} \stackrel{d}{=} f_{s,t}(G_{t,\phi})$ for every $s < t \in S$ and parameter ϕ.

Intuitively, this is simply the requirement that a data set of size t can be understood as a data set of size $s < t$ augmented with some additional observations. Indeed, if a random graph model $(G_{s,\phi})$ is projective with respect to a projective system $(f_{s,t};\, s \le t \in S)$, then it is possible to construct the random variables $G_{s,\phi}$ in such a way that the identity $G_{s,\phi} = f_{s,t}(G_{t,\phi})$ holds almost surely, and not only in distribution. In view of this, the connection with the idea of s as sample size is clear. The graphs $G_{s_j,\phi}$ for an increasing sequence $s_1, s_2, \ldots$ of sizes are nested.

Both the (dense) exchangeable graph model and the Kallenberg exchangeable graph model are projective. (See Figs. 1 and 2 for illustrations). The (dens) exchangeable graph model is projective with respect to the maps $f_{m,n}$ that take an $n \times n$ adjacency matrix to its principal leading $m \times m$ submatrix. In other words,

dropping the last $n-m$ rows and columns from $G_{n,W}$ produces an array with the same distribution as $G_{m,W}$. The Kallenberg exchangeable graph model is projective with respect to the maps $f_{s,t}$ that take a measure on $[0,t]^2$ to its restriction on $[0,s]^2$. In other words, $G_{s,W} \stackrel{d}{=} G_{t,W}(\cdot \cap [0,s]^2)$ for all $s,t \in \mathbb{R}_+$.

The projectivity of the KEG model sets it apart from random graph models that achieve sparsity by percolating dense random graph models such as the exchangeable graph model, i.e., a sparse graph model is produced by randomly deleting each edge in a dense graph model independently with a probability that grows with the number of vertices. Examples of such models abound [BJR07; BR07; BCCZ14a; BCCZ14b], and in some cases consistent estimators have been developed [WO13; BCCG15; BCS15]. Each of these random graph models is parametrized by a size n that determines the number of vertices, and, for every size n, these random graph models are also jointly exchangeable. It then follows from the Aldous–Hoover and graphon theory, as well as the fact that they are not dense, that these random graph models are not projective.

While dropping projectivity allowed for sparse random graph models, the lack of projectivity complicates the statistical applicability of these models. At the very least, the interpretation of the aforementioned consistency results is not straightforward. Indeed, these models are usually understood to generate the size n graphs independently of each other. Even an adaptation of these models designed to impose some consistency between datasets of different size seems inappropriate for modeling data observation as, for instance, every time a new vertex is observed some fraction of the edges already in the graph will be randomly deleted.

2.2. **Models from symmetries.** Up until this point, we have focused on very general desiderata for random graph models. Merely requiring sparsity and projectivity, however, does not alone lead to a tractable class of models. Indeed, without any restrictions on the model, data will convey no information as to the process that gave rise to it. To enable statistical inference, it is necessary to make some structural assumptions on the parametrization of the random graph model. At the same time, we want a flexible model to serve as the foundation of a broadly applicable framework for the statistical analysis of network data, and so we want to impose as few assumptions as possible.

A general approach towards identifying large tractable families of distributions is to consider the class of all distributions satisfying a particular invariance. The structure of such invariant classes can be understood in general terms using very general results on ergodic decompositions, or, in some cases, via explicit characterizations given by so-called representation theorems. Both (dense) exchangeable graphs and KEGs are examples of such families, but to clarify the idea of defining a class of models by an invariance principle, we will review a fundamental class of examples: the exchangeable sequences. (The following development owes much to [OR15], where the reader can find more details.)

Consider the classical setting of statistical inference: a sequence of real-valued measurements $x_1, \ldots, x_n$ are made of a system in some unknown configuration, and this sequence is modeled as a realization from some unknown distribution $\mu_n \in \mathcal{M}_1(\mathbb{R}^n)$. If, in principle, we could have made any number of measurements, then there exists a sequence of distributions $\mu_1, \mu_2, \ldots$ that are projective with respect to the maps $f_{m,n}$ that take length-n sequences to their length-m prefixes. It follows from general results in probability theory that there exists an infinite sequence

$X_1, X_2, \ldots$ of random variables such that μ_n is the distribution of $(X_1, \ldots, X_n)$. Therefore, we are modeling observed length-n sequences $(x_1, \ldots, x_n)$ as realizations of prefixes $(X_1, \ldots, X_n)$ of the infinite random sequence $(X_1, X_2, \ldots)$. Let μ be the unknown distribution of the infinite sequence.

Without making any further assumptions, it would seem that μ is an unknown element of the space $\mathcal{M}_1(\mathbb{R}^\infty)$ of all distributions on infinite sequences of real numbers. However, a finite prefix of a realization drawn from an arbitrary element $\mu \in \mathcal{M}_1(\mathbb{R}^\infty)$ does not convey any information about the generating process μ. However, if we assume that the infinite sequence of random variables $X_1, X_2, \ldots$ is *exchangeable*, i.e.,

$$(X_1, \ldots, X_n) \stackrel{d}{=} (X_{\sigma(1)}, \ldots, X_{\sigma(n)}) \tag{2.2}$$

for every $n \in \mathbb{N}$ and every permutation σ of $[n] = \{1, \ldots, n\}$, then, by de Finetti's representation theorem [Fin30; Fin37; HS55], the random variables $X_1, X_2, \ldots$ are conditionally i.i.d., i.e., there exists a probability measure $\mathcal{P}$ on the space $\mathcal{M}_1(\mathbb{R})$ of probability measures on $\mathbb{R}$ such that

$$M \;\sim\; \mathcal{P} \tag{2.3}$$

$$X_1, X_2, \ldots \mid M \;\stackrel{iid}{\sim}\; M. \tag{2.4}$$

We can express the distribution μ in terms of $\mathcal{P}$: For a distribution m on $\mathbb{R}$, let m^∞ be the distribution of an infinite i.i.d.-m sequence. Then

$$\mu(B) = \int_{\mathcal{M}_1(\mathbb{R})} m^\infty(B)\, \mathcal{P}(\mathrm{d}m), \qquad \text{for measurable } B \subseteq \mathbb{R}^\infty. \tag{2.5}$$

The distribution μ is uniquely determined by $\mathcal{P}$, and vice versa. From Eq. (2.5), we can see that the space of distributions of exchangeable sequences is a convex set. It is known that every such distribution can be written as a unique mixture of the infinite product measures of the form m^∞, which are the extreme points. These extreme points are precisely the *ergodic measures*.

The statistical utility of exchangeability is obvious: it follows from the disintegration theorem [Kal01, Thm. 4.4] and the law of large numbers that

$$M(A) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} 1(X_j \in A) \text{ a.s.} \tag{2.6}$$

On the other hand, even an infinite realization $(x_1, x_2, \ldots)$ gives no information about $\mathcal{P}$. For this reason, in a statistical setting, in addition to assuming that ν is an element in the space of distributions of exchangeable sequences, we assume that ν is ergodic, i.e., ν is an unknown element in the space of distributions of i.i.d. sequences. Since every ν has the form m^∞ for some probability measure m on $\mathbb{R}$, it follows that the natural parameter space is the space $\mathcal{M}_1(\mathbb{R})$, and our model is $\mu_{n,\phi} = \phi^n$.

The statistical utility of exchangeability is not merely a matter of theoretical convenience; the vast majority of statistical practice falls under the remit of this framework. Inference of the kind taught in introductory statistics courses is recovered by restricting $\mathcal{P}$ to have support only on families of models with finite dimensional parameterizations, e.g., the normal distributions. The case where $\mathcal{P}$ has support on distributions without finite dimensional parameterizations are so called non-parametric models, of which there are many practical examples.

It is worth emphasizing that although de Finetti's representation theorem is often characterized as a justification for the use of independence in Bayesian modeling, for our purposes the deeper point is that assuming a probabilistic symmetry characterizes the primitive of random sequence models (M, a probability distribution on $\mathbb{R}$) and gives a simple generative recipe for the data in terms of this primitive. It is this later perspective that is paralleled in the derivation of the KEG model.

2.3. **Models for graphs from symmetries.** We have seen how the assumption that an idealized infinite sequence of observations is exchangeable leads to a considerable simplification of the space of distributions under consideration. Moreover, it is clear that finite samples can be used to make inferences about the generating process. We now turn to related results for networks. In particular, we derive the traditional exchangeable graph model from exchangeability and then connect it to the Kallenberg exchangeable graph model.

Consider a partial observation of a network: an array of measurements $x_{i,j}$, for $1 \leq i, j \leq n$, are made between n entities numbered from 1 to n. We write $x_{i,j} = 1$ if a link exists between i and j, and write $x_{i,j} = 0$ otherwise. We will assume the relationship is symmetric, i.e., $x_{i,j} = x_{j,i}$ and that no entity links to itself, i.e., $x_{i,i} = 0$. In other words, our data is a simple graph over n vertices, and we can model it as a realization from some distribution $\mu_n \in \mathcal{M}_1(\{0,1\}^{n\times n})$ concentrating on symmetric arrays with zeros along the diagonal. If, in principle, we could have collected data on any number of entities, then there exists a sequence of distributions $\mu_1, \mu_2, \ldots$ that are projective with respect to the maps $f_{m,n}$ that take $n \times n$ arrays to their leading $m \times m$ subarrays. Again, from general results in probability theory, there exists an infinite array of random variables $X_{i,j}$, for $i, j \in \mathbb{N}$, such that μ_n is the distribution of $(X_{i,j};\, i, j \leq n)$. Therefore, we model observed $n \times n$ adjacency matrices $(x_1, \ldots, x_n)$ as realizations of prefixes $(X_{i,j};\, i, j \leq n)$ of the infinite adjacency matrix $(X_{i,j};\, i, j \in \mathbb{N})$. Let μ be the distribution of the infinite array matrix.

Let us now consider probabilistic symmetries on this infinite idealized network observation. The class of exchangeable sequences has a literal—if naïve—counterpart in the graph setting: the class of edge-exchangeable graphs. The assumption that the edges are exchangeable is the assumption that

$$(X_{i,j};\, i, j \leq n) \stackrel{d}{=} (X_{\sigma(i,j)};\, i, j \leq n), \tag{2.7}$$

for every $n \in \mathbb{N}$ and every permutation σ of $[n] \times [n]$ that is symmetric, i.e., $\sigma(i, j) = (i', j')$ if and only if $\sigma(j, i) = (j', i')$. This assumption is too severe, however, because it is simply exchangeability of a sequence in disguise.

To see this, let $\mathbb{N}_2$ be the set of pairs $(i, j) \in \mathbb{N}^2$ such that $i < j$ let $\iota : \mathbb{N} \to \mathbb{N}_2$ be an arbitrary bijection, and define $Y_n = X_{\iota(n)}$. Then Eq. (2.7) implies that the sequence of random variables $Y_1, Y_2, \ldots$ are exchangeable and so they are conditionally i.i.d. But then the edges $X_{\iota(n)}$, for $n \in \mathbb{N}$, are also conditionally i.i.d. Therefore, there exists a random variable p in $[0, 1]$ such that, conditioned on p, the edges $X_{i,j}$ are i.i.d. and each edge appears with probability p. This is none other than the Erdős–Rényi–Gilbert model with a random edge probability. The class of ergodic measures in this case is precisely the Erdős–Rényi–Gilbert model.

The natural analogue of exchangeability in the graph setting is to assume that the labels of the vertices are exchangeable. Informally, this is the assumption that the vertex labels carry no information. Given that we are representing an observed

adjacency matrix as a prefix of an idealized infinite symmetric binary array, vertex-exchangeability is formalized as the requirement that distribution of the array is invariant under simultaneous permutation of its rows and columns. More carefully, an array of random variables $X_{i,j}$ is *jointly exchangeable* when

$$(X_{i,j};\, i,j \leq n) \stackrel{d}{=} (X_{\sigma(i),\sigma(j)};\, i,j \leq n) \tag{2.8}$$

for every $n \in \mathbb{N}$ and every permutation σ of $[n]$. A characterization of infinite jointly exchangeable adjacency matrices can be easily derived from the Aldous–Hover representation theorem for general jointly exchangeable arrays [Ald81; Hoo79]. In particular, every ergodic measures is characterized by a symmetric measurable function $W : [0,1]^2 \to [0,1]$, whose diagonal is zero. This same object was later rediscovered independently by graph theorists as the limit object in a theory of limits of dense graphs [LS06; LS07; Lov13]. In this context it was named a graphon, which is the nomenclature we use here. The relationship between the graphon as the defining object for distributions of jointly exchangeable arrays and as the limit object of dense graph theory is explained by [DJ08]. More concretely, the generative model for vertex-exchangeable graphs is (see Fig. 2)

$$W \;\sim\; \mu \tag{2.9}$$

$$\{U_i\} \;\stackrel{iid}{\sim}\; \mathrm{Uni}[0,1] \tag{2.10}$$

$$(X_{ij}) \mid W, U_i, U_j \;\stackrel{ind}{\sim}\; \mathrm{Bernoulli}(W(U_i,U_j)), \tag{2.11}$$

where μ is a measure on the space of symmetric functions from the unit square to the unit interval with zero diagonal. The fact that projective and jointly exchangeable adjacency matrices cannot be sparse is a simple consequence of this generative model and the law of large numbers. In particular, any nondiagonal entry is one with probability $\|W\|_1$. This framework is the exchangeable graph model, whose nomenclature is now self explanatory. Comparing the generative model for the exchangeable graph model with the KEG generative model (see Fig. 1) makes it clear that the distinction that allows for more general graphs in the KEG setting is that the latent variables associated with each vertex are not independent, and the sizes of the graphs are random.

It is possible to construct a sparse and projective random graph model if we drop the requirement that the arrays of each size $n \in \mathbb{N}$ be exchangeable. For example, the preferential attachment model of [BA99] can be understood in these terms, although historically it was developed independently of these concerns for the special purpose of giving a mechanism of graph *growth* that leads to power law behavior in the degree distribution. Ad hoc models of this kind tend to fail to capture certain key elements of real-world network structure. For instance, as shown by [BBCS14], the limiting local structure of preferential attachment graphs is a tree, and so these networks would be pathological models of social networks, which exhibit homophily.

2.4. **Random graphs as random measures.** The key ingredient for generalizing the exchangeable graph model is a correspondence between random graphs and symmetric simple point processes due to Caron and Fox [CF14] (see Fig. 3). Again, restricting ourselves to simple graphs for simplicity of presentation, the edge set of a random graph is a random finite or countable collection of tuples $(x,y) \in \mathbb{R}^2_+$, and the vertex set is the set of those real numbers x such that x participates in at least

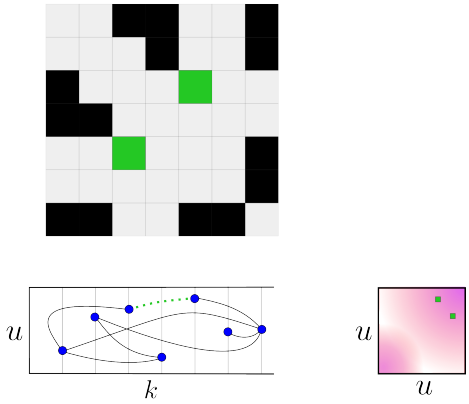

FIGURE 2. Graphon random graph model. In the jointly exchangeable array setting a random graph model is characterized by a (potentially random) symmetric measurable function $W : [0,1]^2 \to [0,1]$ called a graphon. An example graphon is depicted as a magenta heatmap (lower right). Conditional on W, a random graph of size n is generated by independently assigning to each vertex $k \in \{1,\dots,n\}$ a latent random variable $U_k \sim \mathrm{Uni}(0,1)$ (values along vertical axis) and including each edge (k,l) independently with probability $W(U_k, U_l)$. For example, edge $(3,5)$ (green, dotted) is present with probability $W(0.72, 0.9)$; the green boxes in the right square represent the values of (u_3, u_5) and (u_5, u_3). The upper left panel shows the graph realization as an adjacency matrix.

one edge. Concretely, the random graph is represented by a simple point process G on $\mathbb{R}^2_+$ containing a point (x,y) iff there is an edge (x,y) in the random graph.

It will be mathematically convenient to represent simple point processes by simple random measures, i.e., purely atomic random measures whose atoms all have mass one. In this case, each atom in the simple random measure represents a point of the point process. Having made this choice, the idealized infinite observation in this setting is the infinite point process G, and finite observations are the restrictions $G_t = G(\cdot \cap [0,t]^2)$, for $t \in \mathbb{R}_+$, of the infinite point process G to the bounded square subsets $[0,t]^2 \subset \mathbb{R}^2_+$ containing the origin. The distribution of these restrictions of G are automatically projective with respect to the maps $f_{s,t}$ that takes a measure on $[0,t]^2$ to its restriction on $[0,s]^2$. In contrast to the exchangeable graph model, the KEG model has a continuously indexed size parameter and the number of vertices in each finite restriction G_t is itself a random quantity.

It is important to note that the graph corresponding to the restriction G_s to $[0,s]^2$ has as its vertex set only those vertices $x \in [0,s]$ that appear in some edge (x,y) where $y \in [0,s]$. In particular, there will, in general, be vertices in $[0,s]$ that appear for the first time in a restriction $[0,t]$, for $t > s$. This is an essential property of this representation, and is the way that the seeming equivalence between exchangeability and density can be relaxed. The point labeled 2.7 in Fig. 1 provides a concrete example of this phenomena.

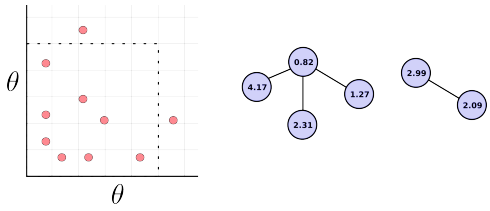

FIGURE 3. Random graphs as point processes. Random point processes on $\mathbb{R}_+^2$ correspond to infinite random graphs, with finite subgraphs given by restricting the point process to a finite square. Points of the process correspond to graph edges and the vertex structure is deduced from the edge structure. Pictured is a realization of a point process and the realization of the random graph that corresponds to truncating at $\theta = 5$.

As observed by Caron and Fox, when random graphs are represented as point processes, vertex-exchangeability corresponds to joint exchangeability for *random measures*. Formally, a random measure ξ on $\mathbb{R}_+^2$ is *jointly exchangeable* when

$$\xi \stackrel{d}{=} \xi \circ (f \otimes f)^{-1} \tag{2.12}$$

for every measure preserving transformation $f : \mathbb{R}_+ \to \mathbb{R}_+$, where $\otimes$ is the tensor product. This probabilistic symmetry was introduced by Aldous, who also conjectured a concrete representation theorem [Ald85, Conj. 15.15], later established rigorously by Kallenberg [Kal90; Kal05]. We will refer to the representation theorem as the Kallenberg representation theorem.

We now describe the Kallenberg exchangeable graph model plainly: It is the random graph model that arises from the symmetry of joint exchangeability of symmetric simple point processes on $\mathbb{R}_+^2$, when these structures are interpreted as the edge sets of random graphs. We give a representation theorem for these structures via a straightforward application of Kallenberg's representation theorem in the specific context of symmetric simple point processes on $\mathbb{R}_+^2$. From this result, we see that every ergodic measure is determined by a triple (I, S, W), which we call a graphex. From a statistical standpoint, the graphexes are the natural parameters, and every random graph is seen to arise via the corresponding generative process (Fig. 1). The KEG model is projective, exchangeable, and admits sparse graphs, thereby providing a statistical framework for network analysis that avoids some of the pitfalls of other random graph models. Both the traditional exchangeable graph model and the Caron–Fox model are special cases, and so the KEG model can be seen as a generalization and unification of these models.

3. EXAMPLES

The aim of this section is to work through the details of several informative examples to build intuition for the structure of the Kallenberg exchangeable graph models we consider here. We focus on those graphexes where $I = S = 0$, and so we will refer to W as the graphex without any risk of confusion. We are particularly interested in the sparsity of these graph models. Theorem 5.3 establishes that (ignoring self edges) for all random graphs G_ν generated by graphex W it holds that $\mathbb{E}[e_\nu] = \frac{1}{2}\nu^2\|W\|_1$; i.e., the expected number of edges scales as ν^2 in all cases.

Intuitively then we expect the sparsity of a random graph model to be determined by $\mathbb{E}[v_\nu] = \nu \int_{\mathbb{R}_+} 1 - e^{-\nu \mu_W(x)} \mathrm{d}x$ (from Theorem 5.4, ignoring self edges). This suggests that the slower $\mu_W(x) = \int_{\mathbb{R}_+} W(x,y)\mathrm{d}y$ decays the sparser the graph will be, an intuition that is borne out by the examples of this section.

3.1. **Graphon models.** The above argument suggests that the most densest graphs will correspond to those W that are compactly supported. Let $\widetilde{W} : [0,1]^2 \to [0,1]$ be a graphon and consider the graphex given by the dilation

$$W(x,y) = \begin{cases} \widetilde{W}(x/c, y/c) & x \leq c, y \leq c \\ 0 & \text{otherwise.} \end{cases} \tag{3.1}$$

In this case, points $(\theta, \vartheta) \in \Pi$ of the latent Poisson process will fail to connect to an edge if $\vartheta > c$, and so such points they never participate in the graph and can be discarded. This means that for finite size graph G_ν given by restricting $\theta \leq \nu$ the relevant underlying process is the unit rate Poisson process on $[0,\nu] \times [0,c]$. The generative model for the graph can be expressed as:

$$N_\nu \sim \mathrm{Poi}(c\,\nu) \tag{3.2}$$

$$\{\theta_i\} \mid N_\nu \overset{iid}{\sim} \mathrm{Uni}[0,\nu] \tag{3.3}$$

$$\{\vartheta_i\} \mid N_\nu \overset{iid}{\sim} \mathrm{Uni}[0,1] \tag{3.4}$$

$$(\theta_i, \theta_j) \mid \widetilde{W}, \vartheta_i, \vartheta_j \overset{ind}{\sim} \mathrm{Bernoulli}(\widetilde{W}(\vartheta_i, \vartheta_j)). \tag{3.5}$$

A little thought shows that this is just a trivial modification of the graphon model. Instead of indexing the family of graphs by the number of vertices ($\mathbb{N}$) we now index them by the continuous parameter ν and have $\mathrm{Poi}(c\,\nu)$ candidate vertices at each stage. The vertices now have i.i.d. uniform labels instead of the integer labels of the traditional graphon model and vertices are only included if they connect to at least one edge. The critical components of the graphon model structure are unchanged: the primitive is still the graphon $\widetilde{W} : [0,1]^2 \to [0,1]$, the conditional independence of the edges is the same, the latent variables are independent, and these graphs are necessarily asymptotically dense (or empty). This is the sense in which the graphon model is a special case of the graphex model derived in this paper.

In fact, these are the only dense KEGs arising from (integrable) graphexes: Theorem 5.6 shows that G is dense iff the generating (integrable) graphex has compact support.

3.2. **Slow Decay.** We next consider a graphex with tails that go to 0 slowly:

$$W(x,y) = \begin{cases} 0 & x = y, \\ (x+1)^{-2}(y+1)^{-2} & \text{otherwise,} \end{cases} \tag{3.6}$$

where the condition $W(x,x) = 0\ \forall x \in \mathbb{R}_+$ forbids self edges. In this case $\mu_W(x) = \frac{1}{3}(x+1)^{-2}$ and by Theorem 5.4

$$\mathbb{E}[v_\nu] = \nu(\sqrt{\pi}\sqrt{\nu/3}\mathrm{erf}(\sqrt{\nu/3}) + e^{-\nu/3} - 1) \tag{3.7}$$

$$\sim \sqrt{\frac{\pi}{3}}\nu^{3/2},\ \nu \to \infty. \tag{3.8}$$

By Theorem 5.5 the number of vertices with degree k has expectation:

$$\mathbb{E}[N_{\nu,k}] = \frac{\nu^{k+1}}{k!}(\frac{1}{3})^k \int_1^\infty x^{-2k} e^{-\frac{1}{3}\nu x^{-2}} \mathrm{d}x \tag{3.9}$$

$$= \frac{\nu^{k+1}}{k!}(\frac{1}{3})^k \int_0^1 x^{2(k-1)} e^{-\frac{1}{3}\nu x^2} \mathrm{d}x \tag{3.10}$$

$$= \frac{\Gamma(-\frac{1}{2}+k) - \Gamma(-\frac{1}{2}+k, \frac{\nu}{3})}{2\sqrt{3}k!}\nu^{3/2} \tag{3.11}$$

$$\sim \frac{\Gamma(-\frac{1}{2}+k)}{2\sqrt{3}k!}\nu^{3/2},\ \nu \to \infty. \tag{3.12}$$

By Theorem 6.1 it follows that the degree D_ν of a uniformly selected vertex of G_ν satisfies

$$\mathrm{P}(D_\nu = k \mid G_\nu) \xrightarrow{p} \frac{\Gamma(-\frac{1}{2}+k)}{2\sqrt{\pi}k!},\ \nu \to \infty, \tag{3.13}$$

so in particular a randomly selected vertex of G_ν will have finite degree even in the infinite graph limit. For large k

$$\frac{\Gamma(-\frac{1}{2}+k)}{2\sqrt{\pi}k!} \sim k^{-\frac{3}{2}},\ k \to \infty, \tag{3.14}$$

so this is an example of a random graph model with power-law degree distribution. Note that, in the limit, while the degree of a randomly chosen vertex is finite almost surely, it is infinite in expectation.

3.3. **Fast Decay.** Next we consider a graphex with quickly decaying tails. Let

$$W(x,y) = \begin{cases} 0 & x = y \\ e^{-x}e^{-y} & \text{otherwise.} \end{cases} \tag{3.15}$$

Then $\mu(x) = e^{-x}$ and so by Theorem 5.4

$$\mathbb{E}[v_\nu] = \nu \int_{\mathbb{R}_+} 1 - e^{-\nu e^{-x}} \mathrm{d}x \tag{3.16}$$

$$= \nu \int_0^1 \frac{1}{x}(1 - e^{-\nu x})\mathrm{d}x \tag{3.17}$$

$$= \nu(\gamma + \Gamma(0,\nu) + \log(\nu)) \tag{3.18}$$

$$\sim \nu \log \nu,\ \nu \to \infty. \tag{3.19}$$

As expected, the rapidly decaying graphex gives rise to a graph that is much more dense than one from the slowly decaying graphex.

By Theorem 5.5 the number of vertices with degree k has expectation:

$$\mathbb{E}[N_{\nu,k}] = \frac{\nu^{k+1}}{k!} \int_0^\infty e^{-kx} e^{-\nu e^{-x}} \mathrm{d}x \tag{3.20}$$

$$= \frac{\nu}{k!}(\Gamma(k) - \Gamma(k,\nu)) \tag{3.21}$$

$$\sim \frac{\nu}{k},\ \nu \to \infty. \tag{3.22}$$

so that for fixed k only a vanishing fraction of the vertices will have degree k as $\nu \to \infty$. More precisely, since $\sum_{k=1}^{\nu^\beta} \frac{\nu}{k} \sim \beta\nu \log \nu,\ \nu \to \infty$ we have by Theorem 6.1 that for $0 < \beta < 1$

$$P(D_\nu \leq \nu^\beta) \xrightarrow{p} \beta,\ \nu \to \infty \tag{3.23}$$

where D_ν is a random vertex of G_ν.

3.4. **Caron and Fox.** As already alluded to, the family of random graph models considered by Caron and Fox in [CF14] is a special case of the one considered here. Indeed, in their paper they prove their model satisfies joint exchangeability when considered as a random measure and use Kallenberg's representation theorem to derive some model properties. Nevertheless, the connection is opaque because their model is constructed from products of completely random measures and they cast their model in terms of Lévy process intensities. If the $\boldsymbol{\theta} \times \boldsymbol{\theta}$ measure they had studied had been a product of completely random measures, that model would have corresponded to a graphex of the form $W(x, y) = f(x)f(y)$. Instead, they actually consider a measure on $\boldsymbol{\theta} \times \boldsymbol{\theta}$ given by using the product of completely random measures as a base measure for a Cox process. This gives rise to a directed multigraph which is then transformed into a simple graph by including edge $\{\theta_i, \theta_j\}$ if and only if there is at least one directed edge between θ_i and θ_j. A little algebra shows this model corresponds to the graphex

$$W(x, y) = \begin{cases} 1 - \exp(-g(x)g(y)) & x = y \\ 1 - \exp(-2g(x)g(y)) & x \neq y \end{cases} \tag{3.24}$$

where $g(x) : \mathbb{R}_+ \to \mathbb{R}_+$. Caron and Fox derive this expression in their paper, and give g in terms of the intensity of the defining Lévy process.

4. Representation Theorem for Random Graphs represented by Exchangeable Symmetric Simple Point Processes

We now turn to giving formal statements of our construction and proving the representation theorem at the heart of the paper. In fact, this mostly amounts to translating Kallenberg's representation theorem for jointly exchangeable random measures on $\mathbb{R}_+$ to the random graph setting.

The central objects of study here are undirected, unweighted graphs whose vertices are labeled with values in $\mathbb{R}_+$. For a graph G, we will write $v(G)$ and $e(G)$ to denote the set of vertices and edges, respectively. We begin by formalizing the idea of a graph represented by a measure.

Definition 4.1. An *adjacency measure* is a locally finite symmetric simple measure on $\mathbb{R}_+^2$. The *ν-truncation* of an adjacency measure ξ is the adjacency measure $\xi(\cdot \cap [0, \nu]^2)$ obtained by restricting ξ to $[0, \nu]^2$.

Definition 4.2. Let G be a simple graph, possibly with loops, whose edge set $e(G)$ is a locally finite subset of $\mathbb{R}_+^2$. Then the *adjacency measure of G* is the adjacency measure $\sum_{(x,y)\in e(G)} \delta_{(x,y)}$.

Note that the adjacency measures of a graphs G and G' coincide if and only if their edge sets do. In particular, vertices that do not participate in an edge are "forgotten". We will be interested in the smallest graph corresponding to an

adjacency measure ξ, which is necessarily the graph with the same edge set and no isolated vertices. (See Fig. 3 for an illustration.)

Definition 4.3. Let $\xi = \sum_{i<\kappa} \delta_{e_i}$ be an adjacency measure, where $\kappa \in \mathbb{Z}_+ \cup \{\infty\}$ and $e_1, e_2, \dots$ is a sequence of distinct elements of $\mathbb{R}^2_+$. Then the *simple graph associated with* ξ is the graph G whose edge set is $\{e_i : i < \kappa\}$ and whose vertex set is $\{x : \exists i < \kappa \, \exists y \in \mathbb{R}_+ \, e_i = (x, y)\}$.

Remark 4.4. This correspondence extends to directed weighted graphs in an obvious way by dropping the requirement that the adjacency measure be symmetric and allowing the adjacency measure to assign a mass other than one to each of its atoms; i.e., a directed weighted adjacency measure is a locally finite purely atomic measure, and so would have the form $\xi = \sum_{ij} \omega_{ij} \delta_{(\theta_i, \theta_j)}$. ◁

A random adjacency measure is an (a.s. locally finite) symmetric simple point process. We will represent random graphs by their random adjacency measures, noting that only nonisolated vertices are captured by this representation.

Informally, we are interested in those simple random graphs embedded in $\mathbb{R}_+$ whose distributions are invariant to every relabeling of the vertices of the random graph. We can formalize this notion of invariance in terms of a symmetry of the corresponding adjacency measure. We begin with a definition of exchangeability for random measures due to Aldous:

Definition 4.5. A random measure ξ on $\mathbb{R}^2_+$ is said to be *jointly exchangeable* if, for every measure preserving transformation f on $\mathbb{R}_+$, we have

$$\xi \circ (f \otimes f)^{-1} \stackrel{d}{=} \xi. \tag{4.1}$$

The following result, due to Kallenberg, characterizes the space of exchangeable measures on $\mathbb{R}^2_+$ as well as its extreme points: Let Λ denote Lebesgue measure on $\mathbb{R}_+$ and let Λ_D denote Lebesgue measure on the diagonal of $\mathbb{R}^2_+$.

Theorem 4.6 (Kallenberg [Kal05; Kal90])**.** *A random measure ξ on $\mathbb{R}^2_+$ is jointly exchangeable iff almost surely*

$$\xi = \sum_{i,j} f(\alpha, \vartheta_i, \vartheta_j, \zeta_{\{i,j\}}) \delta_{\theta_i, \theta_j} \tag{4.2}$$

$$+ \sum_{j,k} (g(\alpha, \vartheta_j, \chi_{jk}) \delta_{\theta_j, \sigma_{jk}} + g'(\alpha, \vartheta_j, \chi_{jk}) \delta_{\sigma_{jk}, \theta_j}) \tag{4.3}$$

$$+ \sum_{k} (l(\alpha, \eta_k) \delta_{\rho_k, \rho'_k} + l'(\alpha, \eta_k) \delta_{\rho'_k, \rho_k}) \tag{4.4}$$

$$+ \sum_{j} (h(\alpha, \vartheta_j)(\delta_{\theta_j} \otimes \Lambda) + h'(\alpha, \vartheta_j)(\Lambda \otimes \delta_{\theta_j})) + \beta \Lambda_D + \gamma \Lambda^2, \tag{4.5}$$

for some measurable function $f \geq 0$ on $\mathbb{R}^4_+$, $g, g' \geq 0$ on $\mathbb{R}^3_+$ and $h, h', l, l' \geq 0$ on $\mathbb{R}^2_+$, some collection of independent uniformly distributed random variables $(\zeta_{\{i,j\}})$ on $[0,1]$, some independent unit rate Poisson processes $\{(\theta_j, \vartheta_j)\}$ and $\{(\sigma_{ij}, \chi_{ij})\}_j$, for $i \in \mathbb{N}$, on $\mathbb{R}^2_+$ and $\{(\rho_j, \rho'_j, \eta_j)\}$ on $\mathbb{R}^3_+$, and some independent set of random variables $\alpha, \beta, \gamma \geq 0$. The latter can be chosen to be non-random iff ξ is extreme.

The task is to translate this into a statement about random graphs, or more specifically, their adjacency measures. Because adjacency measures are purely atomic, all terms with a Lebesgue component (Eq. (4.5)) must have measure zero. The

remaining purely atomic terms underlying a jointly exchangeable random measure have the following interpretation for adjacency measures:

(1) $\sum_{i,j} f(\alpha, \vartheta_i, \vartheta_j, \zeta_{\{i,j\}})\delta_{\theta_i,\theta_j}$: this term contributes most of the interesting structure for the random graph models. The random measure ξ will be symmetric and simple if and only if f is a.e. $\{0,1\}$-valued and symmetric in its second and third arguments, for a.e. fixed first and fourth argument. (It is clear that this can easily be strengthened to hold everywhere.) This leads to the correspondence illustrated in Fig. 1. (General f could be used to model directed, weighted graphs in an obvious way.) The tuples (θ_i, θ_j) are possible edges of the graph and the points θ_i are candidate vertices.

(2) $\sum_{j,k}(g(\alpha, \vartheta_j, \chi_{jk})\delta_{\theta_j,\sigma_{jk}} + g'(\alpha, \vartheta_j, \chi_{jk})\delta_{\sigma_{jk},\theta_j})$: this term contributes stars. To see this, note that each candidate vertex θ_j has an associated Poisson process $\{\sigma_{jk}\}$. The points are a.s. distinct: i.e., $\{\theta_l\} \cap \{\sigma_{jk}\} = \emptyset$ and $\{\sigma_{jk}\} \cap \{\sigma_{lk}\}$ for $j \neq l$ with probability one. This means the candidate vertices $\{\sigma_{jk}\}$ will only ever participate in edges with θ_j, hence the star structure. The random measure ξ will be a.s. symmetric and simple iff $g = g'$ and g is $\{0,1\}$-valued.

(3) $\sum_k (l(\alpha, \eta_k)\delta_{\rho_k,\rho'_k} + l'(\alpha, \eta_k)\delta_{\rho'_k,\rho_k})$: this term contributes isolated edges. To see this, note that, with probability one, $\{\rho_k\} \cap \{\rho'_k\} = \emptyset$ and these candidate vertices do not coincide with any other candidate vertices (e.g., $\{\rho_k\} \cap \{\theta_l\} = \emptyset$). This means that if (ρ_i, ρ_j) is an edge of the graph then with probability 1 (ρ_i, x) will not be an edge for any $x \in \mathbb{R}_+$. Again, the random measure ξ will be a.s. symmetric and simple iff $l = l'$ and l is $\{0,1\}$-valued.

The following theorem characterizes the space of exchangeable adjacency measures as well as its extreme points:

Theorem 4.7 (Random graph representation)**.** *Let ξ be a random adjacency measure. Then ξ is jointly exchangeable iff almost surely*

$$\xi = \sum_{i,j} 1[W(\alpha, \vartheta_i, \vartheta_j) \leq \zeta_{\{i,j\}}]\delta_{\theta_i,\theta_j} \tag{4.6}$$

$$+ \sum_{j,k} 1[\chi_{jk} \leq S(\alpha, \vartheta_j)](\delta_{\theta_j,\sigma_{jk}} + \delta_{\sigma_{jk},\theta_j}) \tag{4.7}$$

$$+ \sum_k 1[\eta_k \leq I(\alpha)](\delta_{\rho_k,\rho'_k} + \delta_{\rho'_k,\rho_k}), \tag{4.8}$$

for some measurable function $S : \mathbb{R}_+^2 \to \mathbb{R}_+$, $I : \mathbb{R}_+ \to \mathbb{R}_+$, $W : \mathbb{R}_+^3 \to [0,1]$, where $W(a, \cdot, \cdot)$ is symmetric for every $a \in \mathbb{R}_+$; some collection of independent uniformly distributed random variables $(\zeta_{\{i,j\}})$ in $[0,1]$; some independent unit rate Poisson processes $\{(\theta_j, \vartheta_j)\}$ and $\{(\sigma_{ij}, \chi_{ij})\}_j$, for $i \in \mathbb{N}$, on $\mathbb{R}_+^2$ and $\{(\rho_j, \rho'_j, \eta_j)\}$ on $\mathbb{R}_+^3$; and an independent random variable $\alpha \geq 0$. The latter can be chosen to be non-random iff ξ is extreme.

The second term of this measure corresponds to stars centered at the points $\{\theta_j\}$ and the third term corresponds to isolated edges that do not connect to the rest of the graph.

Proof. Most of this result is immediate from the text preceding the theorem. One direction of the correspondence is immediate: the random measure ξ is obviously jointly exchangeable.

In the other direction, let f, α, $\{\theta_i, \vartheta_i\}$, and $\{\zeta_{\{i,j\}}\}$ be as in Theorem 4.6, and let

$$\xi_{\{i,j\}} := f(\alpha, \vartheta_i, \vartheta_j, \zeta_{\{i,j\}}), \tag{4.9}$$

which is well-defined because f is symmetric in its second and third arguments. Define $W : \mathbb{R}_+^3 \to \mathbb{R}_+$ by

$$W(a,t,t') = \Lambda\{z \in [0,1] \,:\, f(a,t,t',z) = 1\} = \Lambda f(a,t,t',\cdot\,), \tag{4.10}$$

and write W_a for $W(a,\cdot,\cdot)$. Note that W_a is symmetric. Let $\mathcal{F} := \sigma(\alpha, \{(\vartheta_i, \theta_i)\}_{i\in\mathbb{N}})$. Then the random variables $\xi_{\{i,j\}}$, for $\{i,j\} \in \tilde{\mathbb{N}}_2$, are independent given $\mathcal{F}$ and satisfy

$$\mathbb{E}[\xi_{\{i,j\}}|\mathcal{F}] \overset{\text{a.s.}}{=} W_\alpha(\vartheta_i, \vartheta_j). \tag{4.11}$$

Let $\{\zeta'_{\{i,j\}}\}$ be an i.i.d. uniform array on $\tilde{\mathbb{N}}_2$, independent from $\mathcal{F}$, and define, for $\{i,j\} \in \tilde{\mathbb{N}}_2$,

$$\xi'_{\{i,j\}} = 1(W_\alpha(\vartheta_i, \vartheta_j) \le \zeta'_{\{i,j\}}). \tag{4.12}$$

Then it is clear that

$$(\alpha, ((\theta_i,\vartheta_i)_{i\in\mathbb{N}}), (\xi'_{\{i,j\}})_{\{i,j\}\in\tilde{\mathbb{N}}_2}) \overset{d}{=} (\alpha, ((\theta_i,\vartheta_i)_{i\in\mathbb{N}}), (\xi_{\{i,j\}})_{\{i,j\}\in\tilde{\mathbb{N}}_2}) \tag{4.13}$$

and so, by a transfer argument Kallenberg [Kal01, Cor 6.11], there exists an i.i.d. uniform array $\{\zeta''_{\{i,j\}}\}$ on $\tilde{\mathbb{N}}_2$ independent also from $\mathcal{F}$ such that

$$\xi_{\{i,j\}} \overset{\text{a.s.}}{=} 1(W_\alpha(\vartheta_i, \vartheta_j) \le \zeta''_{\{i,j\}}). \tag{4.14}$$

Similarly, letting g and l be as in Theorem 4.6, define

$$S(a,t) := \Lambda\{z \in \mathbb{R}_+ \,:\, g(a,t,z) = 1\} = \Lambda g(a,t,\cdot\,) \tag{4.15}$$

and

$$I(a) := \Lambda\{z \in \mathbb{R}_+ \,:\, l(a,z) = 1\} = \Lambda l(a,\cdot\,). \tag{4.16}$$

A similar argument to above can be used to show that the terms involving S and I agree with their counterparts in Theorem 4.6. □

From the representation theorem, we learn that the extreme members, from which all other can be recovered as mixtures, are naturally defined in terms of a triple (I, S, W), where $I \in \mathbb{R}_+$ and $S : \mathbb{R}_+ \to \mathbb{R}_+$ and $W : \mathbb{R}_+^2 \to \mathbb{R}_+$ are measurable, and W is symmetric.

In general, an exchangeable simple point process ξ of the form above may not be finite when restricted to a finite region $[0,t]^2$. We want finite restrictions of the adjacency measure to correspond to finite size observations, and so we must isolate conditions on the triple (I, S, W) so that the random measure is a.s. finite on bounded sets. The following result, due to Kallenberg, gives necessary and sufficient conditions for a jointly exchangeable measure to be a.s. locally finite.

Theorem 4.8 (local summability [Kal05, Prop. 9.25])**.** *Let ξ be as in Theorem 4.6, write $\hat{f} = f \wedge 1$, and let*

$$f_1 = \Lambda^2_{23}\hat{f}, \qquad f_2 = \Lambda^2_{13}\hat{f}, \qquad g_1 = \Lambda_2\hat{g}, \tag{4.17}$$

where Λ^2_{23} denotes two-dimensional Lebesgue measure in the second and third coordinates, and similarly for Λ^2_{13} and Λ_2. For fixed α, the random measure ξ is a.s. locally finite iff these five conditions are fulfilled:

(i) $\Lambda(\hat{l} + \hat{h} + \hat{h}') < \infty$,
(ii) $\Lambda(\hat{g}_1 + \hat{g}'_1) < \infty$,
(iii) $\Lambda\{f_i = \infty\} = 0$ *and* $\Lambda\{f_i > 1\} < \infty$ *for* $i = 1, 2$,
(iv) $\Lambda^2[\hat{f}; f_1 \vee f_2 \leq 1] < \infty$,
(v) $\Lambda\hat{l}' + \Lambda_D\Lambda\hat{f} < \infty$.

(Note that we have corrected a typo in part (iv), where the integral was taking w.r.t. Λ not Λ^2.) The consequences for adjacency measures is as follows:

Theorem 4.9 (locally finite graphex)**.** *Let ξ be as in Theorem 4.7 for fixed α, and drop the first coordinate from the definitions of I, S, and W. Let $\mu_W(t) = \Lambda W(t, \cdot) = \int_{\mathbb{R}_+} W(t, t')\,\mathrm{d}t'$. The random measure ξ is a.s. locally finite iff these four conditions are fulfilled:*

(i) $I < \infty$,
(ii) $\Lambda S = \int_{\mathbb{R}_+} S(t)\,\mathrm{d}t < \infty$,
(iii) $\Lambda\{\mu_W = \infty\} = 0$ *and* $\Lambda\{\mu_W > 1\} < \infty$,
(iv) $\Lambda^2[W; \mu_W \vee \mu_W \leq 1] = \int_{\mathbb{R}^2_+} W(x, y)\,1[\mu_W(x) \leq 1]\,1[\mu_W(y) \leq 1]\mathrm{d}x\mathrm{d}y < \infty$,
(v) $\int_{\mathbb{R}_+} W(x, x)\,\mathrm{d}x < \infty$.

In particular, ξ is a.s. locally finite if S and W are integrable and $I < \infty$.

Remark 4.10. An example showing that there are nonintegrable W admitting a.s. locally finite exchangeable adjacency measures is the function $W(x, y) = 1[xy \leq 1]$. Its marginal is $\mu_W(x) = \frac{1}{x}$, which obviously satisfies (iii). Moreover, $W = 0$ a.e. on the set $\{(x, y) : \mu_W(x) \wedge \mu_W(y) \leq 1\} = \{(x, y) : x, y \geq 1\}$, satisfying (iv). ◁

These conditions leads us to the following definition:

Definition 4.11. A *graphex* is a triple (I, S, W), where $I \geq 0$ is a non-negative real, $S : \mathbb{R}_+ \to \mathbb{R}_+$ is integrable, and $W : \mathbb{R}^2_+ \to [0, 1]$ is symmetric, and satisfies parts (iii)–(v) of Theorem 4.9.

In situations where there is no risk of confusion, we will abuse nomenclature and use the term graphex to refer to the W component alone, with the understanding that the corresponding triple is $(0, 0, W)$.

The name graphex is chosen in analogy to graphon, the limit object in the dense graph setting, and graphing, the limit objects in the bounded degree graph setting [Lov13].

The marginal μ_W of the graphex component W arises in the characterization of a.s. finite undirected graph point processes. This function will turn out to be an important quantity in a number of different contexts.

Definition 4.12. The *graphex marginal* is $\mu_W(x) = \int_{\mathbb{R}_+} W(x, y)\mathrm{d}y$.

Theorem 4.7 gives us a precise picture of the structure of random graphs corresponding to jointly exchangeable simple point processes: First, the potential vertices are the points of a collection of Poisson processes. For the graph component corresponding to W, there is a Poisson process on $\boldsymbol{\theta} \times \boldsymbol{\vartheta} = \mathbb{R}_+^2$, and each pair of vertices $(\theta_i, \vartheta_i), (\theta_j, \vartheta_j)$ of the process are connected independently with probability $W(\vartheta_i, \vartheta_j)$. For each vertex (θ_i, ϑ_i) in this component, there is a corresponding Poisson process on $\mathbb{R}_+$ with rate $S(\vartheta_i)$. Every point of this Poisson process connects to the vertex (θ_i, ϑ_i) and no other point. Finally, a Poisson process on $\mathbb{R}_+^2$ with rate I produces pairs $(x, y) \in \mathbb{R}_+^2$ of vertices that are connected to each other but no other vertices.

We now define the class of Kallenberg exchangeable graphs:

Definition 4.13. A *Kallenberg exchangeable graph (KEG) associated with graphex* (I, S, W) is the random graph G associated with an exchangeable adjacency measure ξ of the form given in Eq. (4.2). The Kallenberg exchangeable graph model is the family of ν-truncations $G_\nu = \xi(\cdot \cap [0, \nu]^2)$, for $\nu \in \mathbb{R}_+$. When the graphex is clear from context, we will simply refer to G as the Kallenberg exchangeable graph.

The first term of Eq. (4.2) gives essentially all of the interesting graph structure, and so for the rest of the paper, we will restrict attention to models that take $S = I = 0$. Before doing so, we note that the natural analogue of Erdős–Rényi–Gilbert graphs in the KEG model corresponds to graphs for which $I \geq 0$, $S = 0$, and W is constant on a set of the form $[0, c]^2$ and 0 otherwise. In this case, if W is not identically zero, then later results will imply that the truncated graph sequence is dense.

Consider now the structure arising from W alone. Because $I = S = 0$, we will refer to W as the graphex without any risk of confusion. Let Π be a unit rate Poisson process on $\boldsymbol{\theta} \times \boldsymbol{\vartheta}$ as in Theorem 4.7. A Kallenberg exchangeable graph G associated with W has vertex set

$$v(G) = \{\theta_i \mid (\theta_i, \vartheta_i) \in \Pi \wedge \exists \theta_j \in \Pi : W(\vartheta_i, \vartheta_j) > \zeta_{\{i,j\}}\} \tag{4.18}$$

and edge set

$$e(G) = \{\{\theta_i, \theta_j\} \mid (\theta_i, \vartheta_i), (\theta_j, \vartheta_j) \in \Pi \wedge W(\vartheta_i, \vartheta_j) > \zeta_{\{i,j\}}\}. \tag{4.19}$$

Remark 4.14. A graphex with $W(\vartheta, \vartheta) = 0$ for all $\vartheta \in \mathbb{R}_+$ generates a KEG with no self edges. ◁

Remark 4.15. Notice that if G is a KEG associated to W and G_ν is G restricted to $[0, \nu]$ then G_ν is *not* the same as the induced subgraph of G given by restricting to vertices of G with labels $\leq \nu$. The reason for this is that the induced subgraph includes an (infinite) collection of vertices that do not connect to any edges. However, it is true that $G_\nu \uparrow G$ in the sense that $v(G_\nu) \uparrow v(G)$ and $e(G_\nu) \uparrow e(G)$ as $\nu \uparrow \infty$. ◁

Remark 4.16. The model can be extended to weighted graphs by replacing the indicator term $1[W(\alpha, \vartheta_i, \vartheta_j) \leq \zeta_{\{i,j\}}]$ by a general random variable parameterized by $W(\alpha, \vartheta_i, \vartheta_j)$. The model can be extended to directed graphs by mimicking the 4-graphon approach used by [CAF15] to extend the exchangeable graph model to directed graphs. ◁

Definition 4.17. We will often refer to Π as the *latent Poisson process*. For a point of the latent Poisson process $(\theta_i, \vartheta_i) \in \Pi$ the *label* of the point is θ_i and the *latent value* is ϑ_i.

We close this section with a word of warning about point process notation:

Remark 4.18. Point processes are central to our construction. For a point process $\boldsymbol{P}$ we will often refer to points $p_i \in \boldsymbol{P}$ where the index i is given by some unspecified measurable function of $\boldsymbol{P}$. For example, if $\boldsymbol{P}$ is a Poisson process then the points could be indexed by the ordering of their Euclidean distances to the origin. This is convenient for writing summations across the point process and for unambiguously associating dimensions when the points are multidimensional (e.g., $p_i = (a_i, b_i)$ then we understand a_i and b_i are part of the same tuple in $\boldsymbol{P}$). However, there is a small subtlety here: any choice of indexing function will be informative about the value of the point of the process. For example, if the points of a Poisson process are indexed by their distance to the origin then the value of the index is informative about the value of the point. As a result, some care must be taken when making statements of (conditional) independence. ◁

5. Expected Number of Edges and Vertices

In this section we derive the expected values of the number of vertices and edges of Kallenberg exchangeable graphs restricted to $[0, \nu]$, in terms of their underlying graphex. We focus on those graphexes where $I = 0$ and $S = 0$ so we refer to W as the graphex without any risk of confusion. Throughout this section we implicitly assume W is non-random; in the case of random W the results can be understood as conditional statements.

The intuition for the main proof idea is to find the distribution of the degree of a single point in the latent Poisson process, write the statistics of interest as sums of functions of the degrees of the points and appeal to the linearity of expectation to evaluate these expressions. For example, the number of edges in a graph is the sum of the degrees of all of the vertices divided by 2. This perspective allows the use of powerful techniques for computing expectations of sums over point processes.

Because the θ labels of the graph carry no information it is easiest to treat G_ν by projecting the latent Poisson process Π_ν along its second coordinate on to a random point set in $\boldsymbol{\vartheta} \simeq \mathbb{R}_+$ as $\Pi_\nu^P = \{\vartheta_i \mid (\theta_i, \vartheta_i) \in \Pi_\nu\}$, which is then a rate ν Poisson process. For φ a locally finite, simple sequence and $\{z_{\{i,j\}}\}$ a sequence of values in $[0, 1]$ such that $z_{ij} = z_{ji}$, then for $x \in \varphi$ define the degree function:

$$D(x, \varphi, \{z_{ij}\}) = \sum_{p \in \varphi \setminus \{x\}} 1[W(x, p) \geq z_{i(x)i(p)}] + 2 \cdot 1[W(x, x) \geq z_{i(x)i(x)}] \tag{5.1}$$

where $_{i(x)=i(x,\varphi)}$ gives the index of the point $x \in \varphi$ with respect to the natural ordering on $\mathbb{R}_+$. Intuitively speaking, for a symmetric array $\zeta_{\{i,j\}}$ of uniform $[0, 1]$ random variables,

$$D(\vartheta, \Pi_\nu^P, (\zeta_{\{i,j\}})) \tag{5.2}$$

is the degree of a point $(\theta, \vartheta) \in \Pi_\nu$ under a KEG process, conditional on $(\theta, \vartheta) \in \Pi_\nu$.

For any $\lambda \in \mathbb{R}_+$ the probability that $\lambda \in \Pi_\nu^P$ is 0 and so $D(\lambda, \Pi_\nu^P, \zeta_{\{i,j\}})$ is ill defined. We wish to derive the distribution of the degree of a point λ under the promise that it's in the point process. Because this is a measure 0 event the conditioning is in general somewhat tricky. The idea is formalized by Palm theory, which for a measure P on point sequences defines a Palm measure P_λ that behaves as the required conditional distribution; see [CSKM13] for an accessible introduction. The Slivnyak–Mecke theorem asserts that a Poisson process Π with

a promise $\lambda \in \Pi$ (in the Palm sense) is equal in distribution to $\Pi \cup \{\lambda\}$, so the correct object to work with is $D(\lambda, \Pi_\nu^P \cup \{\lambda\}, \zeta_{\{i,j\}})$. Recalling the graphex marginal $\mu_W(x) = \int_{\mathbb{R}_+} W(x,y)\mathrm{d}y$:

Lemma 5.1. *Let $x \in \mathbb{R}_+$. Then $D(\lambda, \Pi_\nu^P \cup \{\lambda\}, (\zeta_{\{i,j\}})) \stackrel{d}{=} D_{ext} + D_{self}$ where $D_{ext} \sim \mathrm{Poi}(\nu\mu_W(\lambda))$ and $\frac{1}{2}D_{self} \sim \mathrm{Bernoulli}(W(\lambda,\lambda))$ independently.*

Proof. With probability 1, $\lambda \notin \Pi_\nu^P$ so

$$D(\lambda, \Pi_\nu^P \cup \{\lambda\}, \zeta_{\{i,j\}}) = \sum_{p \in \Pi_\nu^P} 1[W(\lambda,p) \geq \zeta_{i(\lambda)i(p)}] + 2 \cdot 1[W(\lambda,\lambda) \geq \zeta_{i(\lambda)i(\lambda)}]. \tag{5.3}$$

Since $\zeta_{i(\lambda)i(\lambda)} \sim U[0,1]$ independent of everything else letting

$$D_{\mathrm{self}} = 2 \cdot 1[W(\lambda,\lambda) \geq \zeta_{i(\lambda)i(\lambda)}] \tag{5.4}$$

and

$$D_{\mathrm{ext}} = \sum_{p \in \Pi_\nu^P} 1[W(\lambda,p) \geq \zeta_{i(\lambda)i(p)}] \tag{5.5}$$

establishes the independence of the two terms and that $\frac{1}{2}D_{\mathrm{self}} \sim \mathrm{Bernoulli}(W(\lambda,\lambda))$.

We have that

$$\int_{\mathbb{R}_+}\int_{[0,1]} 1\left[u \leq W(\lambda,y)\right] \nu \mathrm{d}y\mathrm{d}u = \nu \int_{\mathbb{R}_+} W(\lambda,y)\mathrm{d}y < \infty \text{ a.s.}, \tag{5.6}$$

where the a.s. finiteness is one of the defining conditions of the graphex W. It then follows by a version of Campbell's theorem [Kin93, §5.3], the characteristic function of D_{ext} is

$$\begin{aligned}
\mathbb{E}[\exp(itD_{\mathrm{ext}})] &= \mathbb{E}[\exp(it \sum_{p \in \Pi_\nu^P} 1\left[\zeta_{i(\lambda)i(p)} \leq W(\lambda,p)\right])] && (5.7)\\
&= \exp\{\int_{\mathbb{R}_+}\int_{[0,1]} (1 - e^{it1[u \leq W(\lambda,y)]}\nu \mathrm{d}u\mathrm{d}y)\} && (5.8)\\
&= \exp\{\nu \sum_{n=1}^{\infty} \frac{(it)^n}{n!} \int_{\mathbb{R}_+}\int_{[0,1]} 1\left[u \leq W(\lambda,y)\right] \mathrm{d}u\mathrm{d}y\} && (5.9)\\
&= \exp\{\nu\mu_W(\lambda)(e^{it} - 1)\}. && (5.10)
\end{aligned}$$

Hence, D_{ext} is a $\mathrm{Poi}(\nu\mu_W(\lambda))$ distributed random variable, completing the proof. □

We would now like to access the first moments of various graph quantities by writing them as sums of (functions of) the degree and exploiting the linearity of expectation to circumvent dependencies. For example, the total number of edges of the graph is

$$e_\nu \stackrel{d}{=} \frac{1}{2} \sum_{\vartheta \in \Pi_\nu^P} D(\vartheta, \Pi_\nu^P, (\zeta_{\{i,j\}})), \tag{5.11}$$

where the equality is in distribution (as opposed to almost sure) because the indexing $_{i(x)}$ of the latent Poisson process used by the degree function is not the same as the indexing used in Theorem 4.7.

Standard point process formulas deal with computing expressions of the form

$$\mathbb{E}[\sum_{\lambda\in\Gamma} h(\lambda,\Gamma)] \tag{5.12}$$

where Γ is a simple point process. Sums across the degrees of points of the process do not immediately have this form because the degree depends on the i.i.d. uniform array $(\zeta_{\{i,j\}})$, so we will need a slight extension. Let $\mathbb{M}$ denote the family of all sets of points φ in $\mathbb{R}_+$ that are both locally finite and simple, then:

Lemma 5.2 (Extended Slivnyak–Mecke)**.** *Let Φ be a rate ν Poisson process on $\mathbb{R}_+$, U an independent uniform random variable, and $f:\mathbb{R}_+\times\mathbb{M}\times[0,1]\to\mathbb{R}_+$ a measurable non-negative function. Then*

$$\mathbb{E}[\sum_{p\in\Phi} f(p,\Phi,U)] = \nu\int_{\mathbb{R}_+}\mathbb{E}[f(x,\Phi\cup\{x\},U)]\mathrm{d}x. \tag{5.13}$$

Proof. By the independence of U and Φ, the non-negativity of f, and Tonelli's theorem, we have

$$\mathbb{E}[\sum_{p\in\Phi} f(p,\Phi,U)] = \int_0^1\mathbb{E}[\sum_{p\in\Phi} f(p,\Phi,u)]\mathrm{d}u. \tag{5.14}$$

By the usual Palm calculus, the inner expectation satisfies

$$\mathbb{E}[\sum_{p\in\Phi} f(p,\Phi,u)] = \int_{\mathbb{R}_+}\int_{\mathbb{M}} f(x,\varphi,u)P_x(\mathrm{d}\varphi)\nu\mathrm{d}x, \tag{5.15}$$

where P_x is the local Palm distribution of a unit rate Poisson process. Letting P be the distribution of a unit rate Poisson process, the Slivnyak–Mecke theorem gives:

$$\int_{\mathbb{M}} f(x,\varphi,u)P_x(\mathrm{d}\varphi) = \int_{\mathbb{M}} f(x,\varphi\cup\{x\},u)P(\mathrm{d}\varphi). \tag{5.16}$$

The result then follows by a second application of Tonelli's theorem to change the order of integration. □

The main results of this section now follow easily:

Theorem 5.3. *The expected number of edges $e_\nu = |e(G_\nu)|$ is*

$$\mathbb{E}[e_\nu] = \frac{1}{2}\nu^2\iint_{\mathbb{R}_+^2} W(x,y)\mathrm{d}x\mathrm{d}y + \nu\int_{\mathbb{R}_+} W(x,x)\mathrm{d}x. \tag{5.17}$$

Proof. By Lemmas 5.1 and 5.2,

$$\mathbb{E}[e_\nu] = \frac{1}{2}\mathbb{E}[\sum_{\vartheta\in\Pi_\nu^P} D(\vartheta,\Pi_\nu^P,(\zeta_{\{i,j\}}))] \tag{5.18}$$

$$= \frac{1}{2}\nu\int_{\mathbb{R}_+}\mathbb{E}[D(x,\Pi_\nu^P\cup\{x\},(\zeta_{\{i,j\}}))]\mathrm{d}x \tag{5.19}$$

$$= \frac{1}{2}\nu\int_{\mathbb{R}_+}\nu\mu_W(x) + 2W(x,x)\mathrm{d}x \tag{5.20}$$

By assumption, $\|\mu_W\|_1 = \|W\|_1 < \infty$ and $\int_{\mathbb{R}_+} W(\lambda,\lambda)\mathrm{d}\lambda < \infty$, and so $\mathbb{E}[e_\nu] < \infty$ and the result follows by the linearity of integration. □

Theorem 5.4. *The expected number of visible vertices $v_\nu = |v(G_\nu)|$ is*

$$\mathbb{E}[v_\nu] = \nu \int_{\mathbb{R}_+} (1 - e^{-\nu\mu_W(x)})\mathrm{d}x + \nu \int_{\mathbb{R}_+} e^{-\nu\mu_W(x)} W(x,x)\mathrm{d}x. \tag{5.21}$$

Proof. By Lemmas 5.1 and 5.2,

$$\mathbb{E}[v_\nu] = \mathbb{E}[\sum_{\vartheta \in \Pi_\nu^P} 1\left[D(\vartheta, \Pi_\nu^P, (\zeta_{\{i,j\}})) \geq 1\right]] \tag{5.22}$$

$$= \nu \int_{\mathbb{R}_+} \mathrm{P}(D(x, \Pi_\nu^P \cup \{x\}, (\zeta_{\{i,j\}})) \geq 1)\mathrm{d}x \tag{5.23}$$

$$= \nu \int_{\mathbb{R}_+} 1 - \mathrm{P}(D_{\text{ext}} = 0)\mathrm{P}(D_{\text{self}} = 0)\mathrm{d}x \tag{5.24}$$

$$= \nu \int_{\mathbb{R}_+} 1 - e^{-\nu\mu_W(x)}(1 - W(x,x))\mathrm{d}x, \tag{5.25}$$

where D_{ext} and D_{self} are defined as in Lemma 5.1. Splitting up the integral is justified since $1 - \exp(-\nu\mu_W(x)) \geq 0$ and $\exp(-\nu\mu_W(x))W(x,x) \geq 0$ for all x. □

A nearly identical argument can be used to find the expected number of vertices of a specified degree. This result is interesting in its own right and is used as a lemma in Section 6.

Theorem 5.5. *The expected number of vertices of degree k in G_ν, $N_{\nu,k}$, is*

$$\begin{aligned}\mathbb{E}[N_{\nu,k}] = \nu^{k+1} \int_{\mathbb{R}_+} \Bigg[&\frac{\mu_W(x)^k}{k!} e^{-\nu\mu_W(x)} \\ &+ \frac{1}{\nu^2} \frac{\mu_W(x)^{k-2}}{(k-2)!} e^{-\nu\mu_W(x)} (1 - \frac{(\nu\mu_W(x))^2}{k(k-1)}) W(x,x) \Bigg] \mathrm{d}x\end{aligned} \tag{5.26}$$

Proof. The result follows from essentially the same argument as the previous two theorems and some straightforward algebraic manipulations. □

Notice that in the limit as $\nu \to \infty$ the contribution of self edges ($W(\lambda,\lambda) \neq 0$) is negligible in the sense that terms due to the edges between distinct vertices dominate asymptotically for Theorems 5.3 to 5.5.

We end this section by applying our results on the expected number of vertices and edges to show that a KEG is dense iff the generating graphex is compactly supported.

Theorem 5.6. *Let G be Kallenberg exchangeable graph with graphex $(0, 0, W)$. If W is compactly supported, then G is dense with probability 1. Conversely, if W is integrable and not compactly supported, then G is sparse with probability 1.*

Proof. We have already shown in Section 3.1 that if W is compactly supported then the corresponding KEG is dense (or empty) with probability 1 because these models correspond exactly to graphon models.

Conversely, suppose that the KEG G generated by W is dense with positive probability. This means that there are constants $c, p > 0$ such that

$$\liminf_{\nu\to\infty} \mathrm{P}(e_\nu > cv_\nu^2) > p, \tag{5.27}$$

where $e_\nu = e(G_\nu)$ and $v_\nu = v(G_\nu)$. With

$$\mathbb{E}[e_\nu] \geq \mathrm{P}(e_\nu > cv_\nu^2)\mathbb{E}[cv_\nu^2] \tag{5.28}$$

and Jensen's inequality, this implies $\mathbb{E}[e_\nu] = \Omega(\mathbb{E}[v_\nu]^2)$.

Now, by Theorem 5.4,

$$\mathbb{E}[v_\nu] = \nu \int_{\mathbb{R}_+} 1 - e^{-\nu\mu_W(x)}\mathrm{d}x + \nu \int_{\mathbb{R}_+} e^{-\nu\mu_W(x)} W(x,x)\mathrm{d}x, \tag{5.29}$$

and monotone convergence shows $\int_{\mathbb{R}_+} 1 - e^{-\nu\mu_W(x)}\mathrm{d}x \uparrow \infty$ iff μ_W is not compactly supported. Thus for G dense with positive probability and W not compactly supported it holds that

$$\mathbb{E}[e_\nu] = \omega(\nu^2). \tag{5.30}$$

However, by Theorem 5.3, $\mathbb{E}[e_\nu] = \Theta(\nu^2)$. This contradiction completes the proof. □

6. Degree Distribution in the Asymptotic Limit

One of the major advantage of KEGs over previous exchangeable graph models is that they allow for sparse graphs of the kind typically seen in application; in particular this means the KEG models should allow for a variety of degree (scaling) behaviours. Caron and Fox [CF14] characterized the degree distribution in the large graph limit for the particular case of directed graphs based on generalized gamma processes. We now describe the limiting degree distribution of Kallenberg exchangeable graphs. We focus on those graphexes where $I = S = 0$ so we refer to W as the graphex without any risk of confusion. To formalize the notion of limiting degree distribution, let G_ν be a Kallenberg exchangeable graph on $[0, \nu)$ with graphex W, and let D_ν be the degree of a vertex chosen uniformly at random from $v(G_\nu)$. The central object of study is then the random distribution function $k \mapsto P(D_\nu \leq k \mid G_\nu)$ and its scaling limit. The primary aim of this section is to prove the following theorem:

Theorem 6.1. *Let W be an integrable graphex such that*

1. *There exist some constants $C, T > 0$ such that for all λ and $\omega > T$ it holds that $\int W(\lambda, x)W(\omega, x)\mathrm{d}x \leq C\mu_W(\lambda)\mu_W(\omega)$.*
2. *μ_W is monotonically decreasing.*
3. *μ_W is differentiable.*
4. *There is some $\chi > 0$ such that for all $x > \chi$ holds that $\frac{\mu_W(x)}{\mu_W'(x)}\frac{1}{x} \geq -1$.*

Let $k_\nu = o(\nu)$. Then,

$$P(D_\nu > k_\nu \mid G_\nu) \xrightarrow{p} \lim_{\nu\to\infty} \frac{\sum_{n=k_{\nu+1}}^{\infty} \int \frac{1}{n!} e^{-\nu\mu_W(x)} (\nu\mu_W(x))^n \mathrm{d}x}{\int 1 - e^{-\nu\mu_W(x)}\mathrm{d}x}. \tag{6.1}$$

In the case $\mu_W(\lambda) = (1+\lambda)^{-2}$ the right hand side of this expression is in $(0,1)$ for $k_\nu = k$ for any choice of k. That is, even in the infinite graph limit a constant fraction of the vertices will have degree $\leq k$ for a fixed integer k. By contrast, for $\mu_W(\lambda) = e^{-\lambda}$ the degree of a randomly chosen vertex goes to ∞ so, for fixed k, $\mathrm{P}(D_\nu > k \mid G_\nu) \xrightarrow{p} 1$. However, we saw that $\mathrm{P}(D_\nu > \nu^\beta \mid G_\nu) \to 1 - \beta$ for $\beta \in (0,1)$; i.e., taking $k_\nu = \nu^\beta$ results in a non-trivial limit on the right hand side. That is, this theorem can be understood intuitively as characterizing the rate of growth of the degree of a typical vertex. This scaling limit affords a precise notion of "how dense" the graph associated to a particular graphex is.

Let $n^{(\nu)}_{>l}$ denote the number of vertices of G_ν with degree greater than l. It is immediate that

$$\mathrm{P}(D_\nu \geq k_\nu \mid G_\nu) = \frac{n^{(\nu)}_{>k_\nu}}{n^{(\nu)}_{>0}}, \tag{6.2}$$

i.e., the probability of choosing a vertex of degree greater than k_ν is the proportion of such vertices among all vertices. Notice that, even for fixed l, the random variable $n^{(\nu)}_{>l}$ grows with ν. Further notice that like D_ν the random variable $n^{(\nu)}_{>k_\nu}/n^{(\nu)}_{>0}$ is ill defined for the event $n^{(\nu)}_{>0} = 0$; however this is a measure 0 event in the limit $\nu \to \infty$. The content of Theorem 6.1 can be understood as saying that the limit of the ratio $\frac{n^{(\nu)}_{>l}}{n^{(\nu)}_{>0}}$ is the limit of the ratio of the expectations,

$$\frac{n^{(\nu)}_{>l}}{n^{(\nu)}_{>0}} \xrightarrow{p} \lim_{\nu\to\infty} \frac{\mathbb{E}[n^{(\nu)}_{>l}]}{\mathbb{E}[n^{(\nu)}_{>0}]},\ \nu \to \infty. \tag{6.3}$$

Reasoning about the degree of a randomly selected vertex is substantially simplified by selecting only from those with label $\theta \in [0,1]$ and ignoring the contribution of edges (θ_i, θ_j) with $\theta_i, \theta_j \leq 1$. The reason for this is that it allows us to eliminate one form of dependence between the degrees of distinct points; namely the dependence arising from the requirement that each terminus attached to a vertex has a matching terminus attached to some other vertex in the set. Intuitively, studying this simplification is valid because the θ labels of the points of the latent Poisson process are independent of their degrees and as the graph becomes large only a negligible number of edges have both termini with labels $\theta \leq 1$. Let $N^{(\nu)}_{>l}$ be the number of vertices of G_ν with label $\theta_i < 1$ and greater than l neighbours $\{\theta_j\}$ where $\theta_j > 1$. The following lemma establishes the claimed equivalence:

Lemma 6.2. *The limiting distribution of $n^{(\nu)}_{>l}/n^{(\nu)}_{>0}$ is the same as the limiting distribution of the ratio that considers only vertices with label $\theta_i \leq 1$ and counts only edges (θ_i, θ_j) with $\theta_j > 1$,*

$$\lim_{\nu\to\infty} \frac{n^{(\nu)}_{>l}}{n^{(\nu)}_{>0}} \overset{d}{=} \lim_{\nu\to\infty} \frac{N^{(\nu)}_{>l}}{N^{(\nu)}_{>0}}. \tag{6.4}$$

Proof. The validity of this equality is a consequence of the following three observations:

(1) $\lim_{\nu\to\infty} \mathrm{P}(N^{(\nu)}_{>0} = 0) = 0$ so $\lim_{\nu\to\infty} \frac{N^{(\nu)}_{>l}}{N^{(\nu)}_{>0}}$ is well defined.

(2) The θ label of a point of the latent Poisson process is independent of its degree. Let $\tilde{D}_\nu$ be the degree of a vertex chosen uniformly at random from those members of $v(G_\nu)$ with label $\theta < 1$ and let $\tilde{N}^{(\nu)}_{>l}$ be the number of such vertices with degree greater than l. Because the degree of a point $(\theta_i, \vartheta_i) \in \Pi$ is independent of the value of θ_i it holds that, conditional on $\{\tilde{N}^{(\nu)}_{>0} > 0\}$,

$$\mathrm{P}(D_\nu > l \mid G_\nu) \overset{d}{=} \mathrm{P}(\tilde{D}_\nu > l \mid G_\nu). \tag{6.5}$$

This immediately implies

$$\frac{n^{(\nu)}_{>l}}{n^{(\nu)}_{>0}} \overset{d}{=} \frac{\tilde{N}^{(\nu)}_{>l}}{\tilde{N}^{(\nu)}_{>0}}. \tag{6.6}$$

(3) The number of edges (θ_i, θ_j) with $\theta_i, \theta_j \leq 1$ is almost surely finite and $N^{(\nu)}_{>0} \uparrow \infty$ almost surely, so the probability of randomly choosing a vertex that participates in at least one of the neglected edges goes to 0 as $\nu \to \infty$, thus

$$\lim_{\nu\to\infty} \mathrm{P}(\tilde{D}_\nu > l \mid G_\nu) \overset{\text{a.s.}}{=} \lim_{\nu\to\infty} \frac{N^{(\nu)}_{>l}}{N^{(\nu)}_{>0}}. \tag{6.7}$$

□

To treat the limiting distribution of this ratio we introduce

$$\Pi_0 = \{\vartheta \mid (\theta, \vartheta) \in \Pi_{\nu+1}, \theta \leq 1\} \tag{6.8}$$

$$\Pi_{(1,\nu+1]} = \{(\theta, \vartheta) \mid (\theta, \vartheta) \in \Pi_{\nu+1}, \theta > 1\}, \tag{6.9}$$

i.e., we break the latent Poisson process into the component with $\theta \leq 1$ and the component with $\theta > 1$ and then project out the θ value of Π_0 since it contains no useful information. Notice that Π_0 and $\Pi_{(1,\nu+1]}$ are independent Poisson processes.

For $x \in \mathbb{R}_+$, $\bar{u} = (u_i)$ a sequence of values in $[0, 1]$ and $\{(\phi_i, \varphi_i)\}$ a locally finite, simple sequence with elements in $(1, \infty) \times \mathbb{R}_+$ we define

$$D_\nu(x, \bar{u}, \{(\phi_i, \varphi_i)\}) = \sum_i 1[W(x, \varphi_i) > u_i] 1[\phi_i \leq \nu + 1]. \tag{6.10}$$

There exists a marking $(\lambda_i, \bar{\zeta}_i)$ of Π_0 where each $\bar{\zeta}_i = (\zeta^i_j)$ is a sequence of independent $U[0, 1]$ random variables such that

$$D_\nu(\lambda, \bar{\zeta}_i, \Pi_{(1,\infty)}) \tag{6.11}$$

is the degree of the point $\lambda \in \Pi_0$. Let $\bar{U}_j = (U^j_i)$ be independent sequences of independent $U[0, 1]$ random variables and define

$$D_{j,\nu}(x) = D_\nu(x, \bar{U}_j, \Pi_{(1,\infty)}). \tag{6.12}$$

These random variables will arise naturally in the course of the proof.

It follows by mimicking the proof of Lemma 5.1 that

$$D_{j,\nu}(x) \;\sim\; \mathrm{Poi}(\nu \mu_W(x)) \tag{6.13}$$

marginally. The importance of $D_\nu(\lambda, \bar{\zeta}_i, \Pi_{(1,\infty)})$ in the context of the present section comes from the relation

$$N^{(\nu)}_{>l} = \sum_i 1[D_\nu(\lambda_i, \bar{\zeta}_i, \Pi_{(1,\infty)}) > l]. \tag{6.14}$$

where $(U_i)^\lambda$ is a marking of Π_0. We will make heavy use of the observation that, by Campbell's formula,

$$\mathbb{E}[N^{(\nu)}_{>l}] = \int \mathrm{P}(D_{1,\nu}(x) > l) \mathrm{d}x. \tag{6.15}$$

The idea of the proof of Theorem 6.1 is to show that

$$N^{(\nu)}_{>k_\nu} / \mathbb{E}[N^{(\nu)}_{>0}] \xrightarrow{p} \lim_{\nu\to\infty} \mathbb{E}[N^{(\nu)}_{>k_\nu}] / \mathbb{E}[N^{(\nu)}_{>0}], \; \nu \to \infty. \tag{6.16}$$

The special case $k_\nu = 0$ gives $N^{(\nu)}_{>0} / \mathbb{E}[N^{(\nu)}_{>0}] \xrightarrow{p} 1$ and an application Slutsky's theorem then establishes

$$\frac{N^{(\nu)}_{>k_\nu}}{N^{(\nu)}_{>0}} \xrightarrow{p} \lim_{\nu\to\infty} \mathbb{E}[N^{(\nu)}_{>k_\nu}] / \mathbb{E}[N^{(\nu)}_{>0}], \; \nu \to \infty. \tag{6.17}$$

Using Chebyshev's inequality, a sufficient condition for Eq. (6.16) to hold is

$$\text{var}\left[N^{(\nu)}_{>k_\nu}\right] = o(\mathbb{E}[N^{(\nu)}_{>0}]^2). \tag{6.18}$$

The majority of the proof is aimed at characterizing the growth rate of var $\left[N^{(\nu)}_{>k_\nu}\right]$.

In order to do this, we will need to make an assumption about the graphex W that controls the average dependence between the degrees of different vertices of G_ν:

Assumption 1. There exist some constants $C, T > 0$ such that for all λ and $\omega > T$ it holds that $\int W(\lambda, x)W(\omega, x)\mathrm{d}x \leq C\mu_W(\lambda)\mu_W(\omega)$.

We do not know of any examples of an integrable graphex that violates this assumption, although $W(x, y) = 1[xy < 1]$ does. To understand what the assumption means, let $L(\lambda, \omega)$ be the number of common neighbours of points $(l, \lambda), (w, \omega) \in \Pi_\nu$ under G_ν and observe that for a graphex W that is 0 on the diagonal (i.e., forbidding self-edges),

$$L(\lambda, \omega) \;\sim\; \text{Poi}(\nu \int W(\lambda, x)W(\omega, x)\mathrm{d}x), \tag{6.19}$$

with respect to the Palm measure $P_{\lambda,\omega}$[2]. This can be shown by an argument very similar to Lemma 5.1. Thus the assumption can be understood as requiring that the average number of common neighbours between a pair of vertices is at most a constant factor larger than it would be in the case $W(x, y) = \mu_W(x)\mu_W(y)$.

We further assume for simplicity that $\mu_W(x)$ is strictly monotonically decreasing, differentiable and that there is some $\chi > 0$ such that for all $x > \chi$ holds that $\frac{\mu_W(x)}{\mu'_W(x)}\frac{1}{x} \geq -1$. It is not clear which, if any, of these assumptions are necessary for the result to hold. The last condition in particular may already be implied by the other assumptions. Moreover, the result will hold automatically for a graphex W if there is some other graphex W' such that W' satisfies the conditions of the theorem and the KEGs corresponding to W and W' are equal in distribution.

Invertibility implies that W does not have compact support; i.e., the graph is sparse (Theorem 5.6). A particular consequence of this last assumption is that for any function $l(\nu) \to 0$ as $\nu \to \infty$ it holds that $\mu_W^{-1}(l(\nu)) \to \infty$, a fact that will be used heavily in this section and the next.

Subject to these assumptions we may now begin the argument to bound var $\left[N^{(\nu)}_{>k_\nu}\right]$.

Lemma 6.3. *Let $k_\nu = o(\nu)$, then*

$$\textit{var}\left[N^{(\nu)}_{>k_\nu}\right] = \mathbb{E}[N^{(\nu)}_{>k_\nu}] \tag{6.20}$$

$$+ \iint P(D_{1,\nu}(x) > k_\nu, D_{2,\nu}(y) > k_\nu) - P(D_{1,\nu}(x) > k_\nu)P(D_{2,\nu}(y) > k_\nu)\mathrm{d}x\mathrm{d}y \tag{6.21}$$

Proof. Let $\{(\lambda_i, \bar{\zeta}_i)\}$ be a marking of Π_0 such that each $\bar{\zeta}_i = (\zeta^i_j)$ is a sequence of independent identically distributed $U[0, 1]$ random variables and

$$D_\nu(\lambda_i, \bar{\zeta}_i, \Pi_{(1,\infty)}) \tag{6.22}$$

[2]Recall this is just the measure that guarantees that λ, ω are elements of the point process.

is the degree of point λ. Conditional on $\Pi_{(1,\infty)}$ the degrees $D_\nu(\lambda, \bar{\zeta}_i, \Pi_{(1,\infty)})$ of each point $\lambda \in \Pi_0$ are a marking of Π_0 so

$$N^{(\nu)}_{>k_\nu} \mid \Pi_{(1,\infty)} \;\sim\; \mathrm{Poi}(\mathbb{E}[N^{(\nu)}_{>k_\nu} \mid \Pi_{(1,\infty)}]). \tag{6.23}$$

Using this, the formula for conditional variance is

$$\begin{aligned}\mathrm{var}\left[N^{(\nu)}_{>k_\nu}\right] &= \mathbb{E}[\mathrm{var}\left[N^{(\nu)}_{>k_\nu} \mid \Pi_{(1,\infty)}\right]] + \mathrm{var}\left[\mathbb{E}[N^{(\nu)}_{>k_\nu} \mid \Pi_{(1,\infty)}]\right] &(6.24)\\ &= \mathbb{E}[N^{(\nu)}_{>k_\nu}] + \mathrm{var}\left[\mathbb{E}[N^{(\nu)}_{>k_\nu} \mid \Pi_{(1,\infty)}]\right]. &(6.25)\end{aligned}$$

An application of Campbell's formula to the second term gives:

$$\begin{aligned}\mathbb{E}[N^{(\nu)}_{>k_\nu} \mid \Pi_{(1,\infty)}] &= \int_{\mathbb{R}_+} \mathbb{E}[1[D_\nu(x, \bar{U}, \Pi_{(1,\infty)}) > k_\nu] \mid \Pi_{(1,\infty)}]\mathrm{d}x &(6.26)\\ &= \int_{\mathbb{R}_+} \mathrm{P}(D_{1,\nu}(x) > k_\nu \mid \Pi_{(1,\infty)})\mathrm{d}x, &(6.27)\end{aligned}$$

where $\bar{U}$ is a sequence of $U[0,1]$ random variables independent of $\Pi_{(1,\infty)}$. Then $\mathbb{E}[N^{(\nu)}_{>k_\nu} \mid \Pi_{(1,\infty)}]^2$ is

$$\iint_{\mathbb{R}^2_+} \mathrm{P}(D_{1,\nu}(x) > k_\nu \wedge D_{2,\nu}(y) > k_\nu \mid \Pi_{(1,\infty)})\mathrm{d}x\mathrm{d}y. \tag{6.28}$$

By Tonelli's theorem,

$$\mathbb{E}[\mathbb{E}[N^{(\nu)}_{>k_\nu} \mid \Pi_{(1,\infty)}]^2] = \iint_{\mathbb{R}^2_+} \mathrm{P}(D_{1,\nu}(x) > k_\nu \wedge D_{2,\nu}(y) > k_\nu)\mathrm{d}x\mathrm{d}y \tag{6.29}$$

whence

$$\begin{aligned}\mathrm{var}\left[\mathbb{E}[N^{(\nu)}_{>k_\nu} \mid \Pi_{(1,\infty)}]\right] &= \iint_{\mathbb{R}^2_+} \mathrm{P}(D_{1,\nu}(x) > k_\nu \wedge D_{2,\nu}(y) > k_\nu)\mathrm{d}x\mathrm{d}y &(6.30)\\ &\quad - \iint_{\mathbb{R}^2_+} \mathrm{P}(D_{1,\nu}(x) > k_\nu)\mathrm{P}(D_{2,\nu}(y) > k_\nu)\mathrm{d}x\mathrm{d}y &(6.31)\end{aligned}$$

and the claimed result follows. □

Bounding the variance requires controlling the average dependence between $D_{1,\nu}(x)$ and $D_{2,\nu}(y)$, as captured by the second term in the lemma above. The degree of a point λ gives information about the degree of a point ω only through $\Pi_{(1,\nu+1]}$. Intuitively, as $\nu \to \infty$, the degree of λ gives very little information about $\Pi_{(1,\nu+1]}$ so the pairwise dependence between degrees is weak and the variance of $N^{(\nu)}_{>l}$ is small. Formalizing this intuition proves to be somewhat tricky. Essentially, the strategy is to find a bound of the form

$$\begin{aligned}&\mathrm{P}(D_{1,\nu}(x) > k_\nu, D_{2,\nu}(y) > k_\nu) - \mathrm{P}(D_{1,\nu}(x) > k_\nu)\mathrm{P}(D_{2,\nu}(y) > k_\nu) &(6.32)\\ &\qquad\qquad \leq \mathrm{P}(D_{1,\nu}(x) > k_\nu)g(y) &(6.33)\end{aligned}$$

so that

$$\begin{aligned}\mathrm{var}\left[N^{(\nu)}_{>k_\nu}\right] &\leq \mathbb{E}[N^{(\nu)}_{>k_\nu}] + \iint \mathrm{P}(D_{1,\nu}(x) > k_\nu)g(y)\mathrm{d}x\mathrm{d}y &(6.34)\\ &= \mathbb{E}[N^{(\nu)}_{>k_\nu}](1 + \int g(y)\mathrm{d}y). &(6.35)\end{aligned}$$

The goal is then to find a bounding function $g(y)$ such that $\int g(y)\mathrm{d}y$ is small. The next lemma provides such an expression.

Lemma 6.4. *Let T be a value such that for $y > T$ it holds that*

$$\int W(x,z)W(y,z)\mathrm{d}z \leq C\mu_W(x)\mu_W(y) \tag{6.36}$$

and

$$2C\mu(y) \leq 1 - \log 2. \tag{6.37}$$

Further, let $B(y) \sim \mathrm{Bin}(5k_\nu, C\mu_W(y))$ independently of $D_{2,\nu}(y)$ and define

$$g(y) = \begin{cases} P(D_{2,\nu}(y) \leq k_\nu) & y \leq T \\ P(D_{2,\nu}(y) + B(y) > k_\nu \wedge D_{2,\nu}(y) \leq k_\nu) & y > T. \end{cases} \tag{6.38}$$

Then,

$$P(D_{1,\nu}(x) > k_\nu, D_{2,\nu}(y) > k_\nu) - P(D_{1,\nu}(x) > k_\nu)P(D_{2,\nu}(y) > k_\nu) \tag{6.39}$$

$$\leq P(D_{1,\nu}(x) > k_\nu)g(y) \tag{6.40}$$

Proof. Let $x, y \in \mathbb{R}_+$ and define

$$D_a = D_{1,\nu}(x) \tag{6.41}$$

$$D_b = D_{2,\nu}(y). \tag{6.42}$$

It is conceptually helpful to think of a, b as points of the latent Poisson process with ϑ values x, y respectively, but the proof does not make formal use of this. The expression

$$\mathrm{P}(D_a > k_\nu, D_b > k_\nu) = \mathrm{P}(D_a > k_\nu)\mathrm{P}(D_b > k_\nu | D_a > k_\nu), \tag{6.43}$$

makes it clear that $g(y)$ is a bound on $\mathrm{P}(D_b > k_\nu | D_a > k_\nu) - \mathrm{P}(D_b > k_\nu)$. The focus will be on bounding $\mathrm{P}(D_b > k_\nu | D_a > k_\nu)$. To do this, introduce a marking $\{((\theta_i, \vartheta_i), M_i)\}$ of $\Pi_{(1,\infty)}$ where

$$M_i = 1[W(x, \vartheta_i) > U_i^1] \tag{6.44}$$

indicates whether each point connects to a. This induces the obvious marking[3] on $\Pi_{(1,\nu+1]}$ that breaks $\Pi_{(1,\nu+1]}$ into two independent sets:

$$N_a = \{\vartheta_i \mid (\theta_i, \vartheta_i) \in \Pi_{(1,\nu+1]},\ M_i = 1\}, \tag{6.45}$$

the neighbours of a, and

$$\bar{N}_a = \{\vartheta_i \mid (\theta_i, \vartheta_i) \in \Pi_{(1,\nu+1]},\ M_i = 0\}, \tag{6.46}$$

the non-neighbours of a. By construction $|N_a| = D_a$ and the neighbours $N_a = \{\vartheta_i\}_{i=1}^{D_a}$ are, conditional on D_a, independently and identically distributed with probability density

$$\vartheta_i \overset{iid}{\sim} \frac{W(x, \vartheta_i)}{\mu_W(x)}. \tag{6.47}$$

The non-neighbours $\bar{N}_a$ are a Poisson process on $\mathbb{R}_+$ with intensity $\nu(1 - W(x, \vartheta))$. The degree of the point b may be written as the sum of its connections to the neighbours and non-neighbours of a,

$$D_b = D_b^{(N_a)} + D_b^{(\bar{N}_a)}, \tag{6.48}$$

[3] the full marking is defined on $\Pi_{(1,\infty)}$ for consistency of the indices of the points (θ_i, ϑ_i).

where, by an application of Campbell's theorem,

$$D_b^{(\bar{N}_a)} \sim \text{Poi}(\nu(\mu_W(y) - \int W(x,z)W(y,z)\mathrm{d}z)) \tag{6.49}$$

and

$$D_b^{(N_a)} \mid D_a \sim \text{Bin}(D_a, p_{x,y}) \tag{6.50}$$

independently, with

$$p_{x,y} = \frac{1}{\mu_W(x)} \int W(x,z)W(y,z)\mathrm{d}z. \tag{6.51}$$

It is now clear that the dependence of D_b on D_a comes in only through the number of trials of $D_b^{(N_a)} \mid D_a$.

To treat $D_b^{(N_a)}$ conditional on the event $D_a > k_\nu$ we introduce random variables L_1, L_2 such that on the event $\{D_a > k_\nu\}$

$$L_1 + L_2 = D_a \tag{6.52}$$

and implicitly specify the joint distribution of L_1, L_2 by requiring L_1 to have marginal distribution

$$L_1 \sim \text{Poi}(\nu\mu_W(x)) \tag{6.53}$$

conditional on $\{D_a > k_\nu\}$. Intuitively, L_1 is the number of neighbours of a that would exist without conditioning on $D_a > k_\nu$ and L_2 is the number of additional neighbours that are present as a result of the conditioning. Therefore on the event $\{D_a > k_\nu\}$ there are random variables B_1, B_2 such that:

$$D_b^{(N_a)} = B_1 + B_2, \tag{6.54}$$

and

$$B_1 \mid L_1 \sim \text{Bin}(L_1, p_{x,y}) \tag{6.55}$$

$$B_2 \mid L_2 \sim \text{Bin}(L_2, p_{x,y}) \tag{6.56}$$

independently conditional on L_1, L_2. The point of introducing these auxiliary random now becomes clear as:

$$B_1 \sim \text{Poi}(\nu \int W(x,z)W(y,z)\mathrm{d}z) \tag{6.57}$$

and so

$$(D_b^{(\bar{N}_a)} + B_1) \mid \{D_a > k_\nu\} \sim \text{Poi}(\nu\mu_W(y)). \tag{6.58}$$

Intuitively, conditional on $\{D_a > k_\nu\}$, D_b splits into a term

$$H = D_b^{(\bar{N}_a)} + B_1 \tag{6.59}$$

with the unconditional distribution of D_b plus a term B_2 that accounts for the 'extra' neighbours of b that one expects to see as a result of learning that the degree of a is large.

As $D_b = H + B_2$,

$$\mathrm{P}(D_b > k_\nu | D_a > k_\nu) = \mathbb{E}[\mathrm{P}(H + B_2 > k_\nu \mid L_1, L_2) \mid D_a > k_\nu]. \tag{6.60}$$

Then,

$$\mathrm{P}(H + B_2 > k_\nu \mid L_1, L_2) = \tag{6.61}$$

$$\mathrm{P}(H > k_\nu \mid L_1) + \mathrm{P}(H + B_2 > k_\nu \wedge H \leq k_\nu \mid L_1, L_2), \tag{6.62}$$

and L_1 has been defined so that

$$\mathbb{E}[\mathrm{P}(H > k_\nu \mid L_1) \mid D_a > k_\nu] = \mathrm{P}(D_b > k_\nu). \tag{6.63}$$

We have now arrived at

$$\mathrm{P}(D_a > k_\nu, D_b > k_\nu) = \mathrm{P}(D_a > k_\nu)[\mathrm{P}(D_b > k_\nu) + R], \tag{6.64}$$

where the remainder term is

$$R = \mathbb{E}[\mathrm{P}(H + B_2 > k_\nu \wedge H \leq k_\nu \mid L_1, L_2) \mid D_a > k_\nu] \tag{6.65}$$

$$= \mathrm{P}(H + B_2 > k_\nu \wedge H \leq k_\nu \mid D_a > k_\nu). \tag{6.66}$$

Note that

$$\mathrm{P}(D_a > k_\nu, D_b > k_\nu) - \mathrm{P}(D_a > k_\nu)\mathrm{P}(D_b > k_\nu) = \mathrm{P}(D_a > k_\nu)R \tag{6.67}$$

so that to complete the proof it remains to show that $R \leq g(y)$. For $\nu\mu_W(y)$ large the crude bound

$$R \leq \mathrm{P}(H \leq k_\nu \mid D_a > k_\nu) \tag{6.68}$$

$$= \mathrm{P}(D_{2,\nu}(y) \leq k_\nu) \tag{6.69}$$

suffices. This establishes the claim for $y \leq T$ in the lemma statement. The remaining task is to find a good bound in the regime of y where $\nu\mu_W(y)$ is not large. In particular, it suffices to find a bound for B_2 independent of H with a distribution that does not depend on x. To that end, let $b > 0$ and write

$$\mathrm{P}(B_2 > b \mid H) = \mathbb{E}[\mathrm{P}(B_2 > b \mid L_2) \mid H)]. \tag{6.70}$$

As $B_2 \mid L_2 \sim \mathrm{Bin}(L_2, p_{x,y})$,

$$\mathrm{P}(B_2 > b \mid L_2) = 1 - \sum_{n=0}^{b} \binom{L_2}{n} {p_{x,y}}^n (1 - p_{x,y})^{L_2 - n}. \tag{6.71}$$

The salient fact here is that $\varphi(l) = \sum_{n=0}^{b} \binom{l}{n} p(x,y)^n (1 - p(x,y)^{l-n}$ is a convex function in l and so by a conditional Jensen's inequality

$$\mathrm{P}(B_2 > b \mid H) \leq \mathrm{P}(\tilde{B} > b \mid H), \tag{6.72}$$

where $\tilde{B} \mid H \sim \mathrm{Bin}(\mathbb{E}[L_2 \mid H], p_{x,y})$. The task is then to find a bound for the conditional expectation that is independent of H, which we accomplish by demonstrating a constant bound $\mathbb{E}[L_2 \mid H] \leq 5k_\nu$ for y sufficiently large. L_2 is independent of H conditional on L_1 so bounding the conditional expectation can be accomplished by understanding the distribution of $L_2 \mid L_1$ and $L_1 \mid H$. There exists Q with

$$Q \stackrel{d}{=} D_a \mid \{D_a > k_\nu\} \tag{6.73}$$

and Q independent of L_1 such that

$$L_2 = 1[L_1 \leq k_\nu](Q - L_1)$$

$$\implies \mathbb{E}[L_2 \mid H] \leq \mathrm{P}(L_1 \leq k_\nu \mid H)\mathbb{E}[Q]. \tag{6.74}$$

This can be understood as the following sampling scheme for a truncated Poisson distribution:

(1) Draw l_1 from the Poisson distribution. If $l_1 > k_\nu$ stop.
(2) Otherwise sample y from the truncated distribution, so that $l_1 + (y - l_1)$ is a trivially a correct sample.

The definitions above can be used to derive:

$$L_1 \mid H \;\sim\; \mathrm{Bin}(H, \frac{\int W(x,z)W(y,z)\mathrm{d}x}{\mu_W(y)}) + Z \tag{6.75}$$

where $Z \;\sim\; \mathrm{Poi}(\nu\mu_W(x)(1-p_{x,y}))$ is independent of the first term. Thus,

$$\mathrm{P}(L_1 \leq k_\nu \mid H)\mathbb{E}[Q] \leq \mathrm{P}(Z \leq k_\nu)\mathbb{E}[Q]. \tag{6.76}$$

Further,

$$\mathbb{E}[Q] < k_\nu + \nu\mu_W(x), \tag{6.77}$$

which can be seen by noting that there is some random variable G such that

$$G \;\sim\; \mathrm{Gamma}(k_\nu, 1) \mid G < \nu\mu_W(x) \tag{6.78}$$

$$Q = k_\nu + \mathrm{Poi}(\nu\mu_W(x) - G). \tag{6.79}$$

For $\nu\mu_W(x) \leq 2k_\nu$, it immediately follows that

$$\mathrm{P}(Z \leq k_\nu)\mathbb{E}[Q] \leq 5k_\nu \tag{6.80}$$

For $\nu\mu_W(x) > 2k_\nu$ the assumption $2C\mu(y) \leq 1 - \log 2$ for large enough y implies $\mathbb{E}[Z] \geq k_\nu$ so a Poisson tail bound [Gly87] may be applied to Z to find

$$\mathrm{P}(Z \leq k_\nu)\mathbb{E}[Q] \leq k_\nu + 2\mathrm{P}(Z = k_\nu)\nu\mu_W(x) \tag{6.81}$$

$$= k_\nu + 2\frac{1}{k_\nu!}e^{-\nu\mu_W(x)(1-p_{x,y})}(\nu\mu_W(x)(1-p_{x,y}))^{k_\nu}(\nu\mu_W(x)) \tag{6.82}$$

$$\leq k_\nu + 2\frac{1}{k_\nu!}e^{-\nu\mu_W(x)(1-C\mu_W(y))}(\nu\mu_W(x)(1-C\mu_W(y)))^{k_\nu}(\nu\mu_W(x)) \tag{6.83}$$

The second term satisfies

$$\frac{2}{k_\nu!}e^{-\nu\mu_W(x)(1-p_{x,y})}(\nu\mu_W(x)(1-p_{x,y}))^{k_\nu}(\nu\mu_W(x)) \tag{6.84}$$

$$= 2\frac{k_\nu+1}{1-C\mu_W(y)}\mathrm{P}(\tilde{Z} = k_\nu + 1), \tag{6.85}$$

where $\tilde{Z} \;\sim\; \mathrm{Poi}(\nu\mu_W(x)(1-C\mu_W(y)))$. This term is maximized over $\nu\mu_W(x) \geq 2k_\nu$ when $\mathbb{E}[\tilde{Z}]$ is minimal, i.e., when $\nu\mu_W(x) = 2k_\nu$. Subbing in,

$$2\frac{k_\nu+1}{1-C\mu_W(y)}\mathrm{P}(\tilde{Z} = k_\nu + 1) \leq 2(1-C\mu_W(y))^{k_\nu}\frac{1}{k_\nu!}e^{-2k_\nu(1-C\mu_W(y))}(2k_\nu)^{k_\nu+1} \tag{6.86}$$

$$\leq 4k_\nu 2^{k_\nu}e^{-k_\nu(1-2C\mu_W(y))}(\frac{1}{k_\nu!}{k_\nu}^{k_\nu}e^{-k_\nu}) \tag{6.87}$$

$$\leq 4k_\nu, \tag{6.88}$$

where the final line uses $2C\mu_W(y) \leq 1 - \log 2$. It then follows that

$$\mathrm{P}(Z \leq k_\nu)\mathbb{E}[Q] \leq 5k_\nu \tag{6.89}$$

for all values of x.

Putting together Eqs. (6.72), (6.74), (6.80) and (6.89):

$$\mathrm{P}(H + B_2 > k_\nu \wedge H \leq k_\nu \mid D_a > k_\nu) \leq \mathrm{P}(H + B(y) > k_\nu \wedge H \leq k_\nu \mid D_a > k_\nu) \tag{6.90}$$

where, conditional on $D_a > k_\nu$, H and $B(y)$ are independent with

$$H \mid \{D_a > k_\nu\} \stackrel{d}{=} D_{1,\nu}(y) \tag{6.91}$$

$$B(y) \sim \mathrm{Bin}(5k_\nu, C\mu_W(x)). \tag{6.92}$$

This completes the proof of the lemma. □

Roughly speaking, the content of the previous two lemmas amounts to

$$\mathrm{var}\left[N^{(\nu)}_{>k_\nu}\right] \leq \mathbb{E}[N^{(\nu)}_{>k_\nu}](1 + \int g(y)\mathrm{d}y). \tag{6.93}$$

That is, the growth of the variance with ν is controlled by $\int g(y)\mathrm{d}y$. Recalling that our aim is to show $\mathrm{var}\left[N^{(\nu)}_{>k_\nu}\right] = o(\mathbb{E}[N^{(\nu)}_{>k_\nu}]^2)$ we must establish that $\int g(y)\mathrm{d}y = o(\mathbb{E}[N^{(\nu)}_{>0}])$. The remainder of the proof is devoted to showing this. It turns out that the appropriate way to do this depends on whether k_ν goes to infinity.

Lemma 6.5. *Let $g(y)$ be as in Lemma 6.4 and suppose W is integrable. If the sequence k_ν is bounded then*

$$\int g(y)\mathrm{d}y = o(\mathbb{E}[N^{(\nu)}_{>k_\nu}]). \tag{6.94}$$

Proof. Let $T_\nu = \sqrt{\mathbb{E}[N^{(\nu)}_{>k_\nu}]}$ so that by Lemma 6.4 for ν large enough

$$\int_{\mathbb{R}_+} g(y)\mathrm{d}y \leq T_\nu + \int_{T_\nu}^{\infty} \mathrm{P}(D_{2,\nu}(y) + B(y) > k_\nu \wedge D_{2,\nu}(y) \leq k_\nu)\mathrm{d}y. \tag{6.95}$$

Moreover

$$\mathrm{P}(D_{2,\nu}(y) + B(y) > k_\nu \wedge D_{2,\nu}(y) \leq k_\nu) \leq \mathrm{P}(\tilde{B}(y) > 1), \tag{6.96}$$

where, letting $k = \lim_{\nu\to\infty} k_\nu$, $\tilde{B}(y) \sim \mathrm{Bin}(5k, C\mu_W(y))$. By Markov's inequality

$$\mathrm{P}(\tilde{B}(y) > 1) \leq 5kC\mu_W(y) \tag{6.97}$$

so that

$$\int_{T_\nu}^{\infty} \mathrm{P}(\tilde{B}(y) > 1)\mathrm{d}y \leq 5kC \int_{T_\nu}^{\infty} \mu_W(y)\mathrm{d}y \tag{6.98}$$

$$= o(1), \tag{6.99}$$

where the final line follows by the integrability of μ_W. Thus $\int_{\mathbb{R}_+} g(y)\mathrm{d}y = O(\sqrt{\mathbb{E}[N^{(\nu)}_{>k_\nu}]})$. □

The case $k\uparrow\infty$ is substantially trickier. Essentially the strategy here is to break up to domain of y into three components and use a different tractable and reasonably tight bound on $g(y)$ in each region, see Table 1. An important intermediate step is the observation

$$\mathbb{E}[N^{(\nu)}_{>0}] = \Omega(\mu_W^{-1}(\frac{1}{\nu})), \tag{6.100}$$

Region of $\mathbb{R}_+$	Upper bound for $g(y)$
$[0, \mu^{-1}((1+\epsilon)\frac{k_\nu}{\nu})]$	$\mathrm{P}(D_{2,\nu}(y) \leq k_\nu)$
$(\mu^{-1}((1+\epsilon)\frac{k_\nu}{\nu}), \mu^{-1}((1-\epsilon)\frac{k}{\nu}))$	1
$(\mu^{-1}((1-\epsilon)\frac{k_\nu}{\nu}), \infty)$	$\mathrm{P}(B(y) > \frac{\epsilon}{2}k_\nu) + \mathrm{P}(D_{2,\nu}(y) > (1-\frac{\epsilon}{2})k_\nu)$

TABLE 1. Upper bounds on $g(y)$

which will eventually allow us to show $\int g(y)\mathrm{d}y = o(\mathbb{E}[N_{>0}^{(\nu)}])$ by establishing bounds on the integral in terms of $\mu_W^{-1}(\frac{1}{\nu})$. For instance, the next lemma can be understood as establishing that $\int_0^{\mu_W^{-1}((1+\epsilon)\frac{k_\nu}{\nu})} \mathrm{P}(D_{2,\nu}(y) \leq k_\nu)\mathrm{d}y$ is at most an exponentially vanishing (in k_ν) fraction of $\mathbb{E}[N_{>0}^{(\nu)}]$.

Lemma 6.6. *For $0 < \epsilon < 1$,*

$$\int_0^{\mu_W^{-1}((1+\epsilon)\frac{k_\nu}{\nu})} P(D_{2,\nu}(y) \leq k_\nu)\mathrm{d}y \leq \frac{1+\epsilon}{\epsilon}(\frac{1+\epsilon}{e^\epsilon})^{k_\nu}\mu_W^{-1}(\frac{k_\nu}{\nu}). \tag{6.101}$$

Proof. Because $\mathrm{P}(D_{2,\nu}(y) \leq k_\nu)$ is monotonically increasing in y over the domain of integration, the integral is bounded by

$$\mu_W^{-1}((1+\epsilon)\frac{k_\nu}{\nu})\mathrm{P}(D_{2,\nu}(\mu_W^{-1}((1+\epsilon)\frac{k_\nu}{\nu})) \leq k_\nu). \tag{6.102}$$

As $\mathbb{E}[D_{2,\nu}(\mu_W^{-1}((1+\epsilon)\frac{k_\nu}{\nu}))] = (1+\epsilon)k_\nu > k_\nu$ a tail bound [Gly87] applies:

$$\begin{aligned}
\mathrm{P}(D_{2,\nu}(\mu_W^{-1}((1+\epsilon)\frac{k_\nu}{\nu})) \leq k_\nu) &\leq (1+\frac{1}{\epsilon})\mathrm{P}(D_{2,\nu}(y) = k_\nu) && (6.103)\\
&= (1+\frac{1}{\epsilon})\frac{1}{k_\nu!}((1+\epsilon)k_\nu)^{k_\nu}e^{-(1+\epsilon)k_\nu} && (6.104)\\
&\leq \frac{1}{e}\frac{1+\epsilon}{\epsilon}(\frac{1+\epsilon}{e^\epsilon})^{k_\nu}. && (6.105)
\end{aligned}$$

□

For $y > \mu_W^{-1}((1-\epsilon)\frac{k_\nu}{\nu})$ we can bound $g(y)$ (and thus $\int_{\mu_W^{-1}((1-\epsilon)\frac{k_\nu}{\nu})}^\infty g(y)\mathrm{d}y$) by

$$\begin{aligned}
&\mathrm{P}(D_{2,\nu}(y) + B(y) > k_\nu \wedge D_{2,\nu}(y) \leq k_\nu) && (6.106)\\
&\leq \mathrm{P}(D_{2,\nu}(y) + B(y) > k_\nu \wedge D_{2,\nu}(y) \leq (1-\frac{\epsilon}{2})k_\nu) && (6.107)\\
&+ \mathrm{P}(D_{2,\nu}(y) > (1-\frac{\epsilon}{2})k_\nu) && (6.108)\\
&\leq \mathrm{P}(B(y) > \frac{\epsilon}{2}k_\nu) + \mathrm{P}(D_{2,\nu}(y) > (1-\frac{\epsilon}{2})k_\nu). && (6.109)
\end{aligned}$$

The next lemma controls the second term in this bound.

Lemma 6.7. *Suppose there is some $\chi > 0$ such that for all $x > \chi$ it holds that*

$$\frac{\mu_W(x)}{x\mu_W'(x)} \geq -1, \tag{6.110}$$

then, for ν sufficiently large such that $\frac{k_\nu}{\nu} \leq \mu_W(\chi)$ and ϵ such that $0 < \epsilon < 1$,

$$\int_{\mu_W^{-1}((1-\epsilon)\frac{k_\nu}{\nu})}^\infty P(D_{2,\nu}(y) > (1-\frac{\epsilon}{2})k_\nu)\mathrm{d}y \leq \frac{2}{k_\nu\epsilon + 2}\mu_W^{-1}(\frac{k_\nu}{\nu}) \tag{6.111}$$

Proof. For $y \in [\mu_W^{-1}((1-\epsilon)\frac{k_\nu}{\nu}), \infty)$ it holds that $\mathbb{E}[D_{2,\nu}(y)] < (1-\epsilon/2)k_\nu$ so a tail bound [Gly87] applies:

$$\mathrm{P}(D_{2,\nu}(y) > (1-\frac{\epsilon}{2})k_\nu) \leq (\frac{1-\epsilon/2+1/k_\nu}{\epsilon/2+1/k_\nu})\frac{1}{\lfloor(1-\frac{\epsilon}{2})k_\nu\rfloor!}e^{-\nu\mu_W(y)}(\nu\mu_W(y))^{(1-\frac{\epsilon}{2})k_\nu}. \tag{6.112}$$

Because $\mu_W(y)$ is strictly monotonic the component of the bound that depends on y may be integrated by substitution. For notational simplicity, let $f(x) = \mu_W^{-1}(y)$, then

$$\int_{\mu_W^{-1}((1-\epsilon)\frac{k_\nu}{\nu})}^{\infty} e^{-\nu\mu_W(y)}(\nu\mu_W(y))^{(1-\frac{\epsilon}{2})k_\nu}\mathrm{d}y = -\int_0^{(1-\epsilon)k_\nu} e^{-x}x^{(1-\frac{\epsilon}{2})k_\nu}\frac{1}{\nu}f'(\frac{x}{\nu})\mathrm{d}x. \tag{6.113}$$

Let $z = f(x)$ and write

$$\frac{\mu_W(z)}{z\mu_W'(z)} = \frac{f'(x)x}{f(x)} \tag{6.114}$$

so by assumption for $x \leq \mu_W(\chi)$ holds that $x\frac{f'(x)}{f(x)} \geq -1$. Thus for ν sufficiently large that $\frac{k_\nu}{\nu} \leq \mu_W(\chi)$ it holds that

$$-\int_0^{(1-\epsilon)k_\nu} e^{-x}x^{(1-\frac{\epsilon}{2})k_\nu}\frac{1}{\nu}f'(\frac{x}{\nu})\mathrm{d}x \leq \int_0^{(1-\epsilon)k_\nu} e^{-x}x^{(1-\frac{\epsilon}{2})k_\nu-1}f(\frac{x}{\nu})\mathrm{d}x. \tag{6.115}$$

Moreover, $xf(x)$ is a monotonically non-decreasing function on $x \leq \mu_W(\chi)$, which may be established by:

$$(xf(x))' = f(x) + xf'(x) \tag{6.116}$$

$$= f(x)(1 + x\frac{f'(x)}{f(x)}) \tag{6.117}$$

$$\geq 0. \tag{6.118}$$

This implies

$$\int_0^{(1-\epsilon)k_\nu} e^{-x}x^{(1-\frac{\epsilon}{2})k_\nu-1}f(\frac{x}{\nu})\mathrm{d}x \leq (1-\epsilon)k_\nu f(\frac{k_\nu}{\nu})\int_0^{(1-\epsilon)k_\nu} e^{-x}x^{(1-\frac{\epsilon}{2})k_\nu-2}\mathrm{d}x \tag{6.119}$$

$$\leq (1-\epsilon)k_\nu f(\frac{k_\nu}{\nu})\Gamma((1-\frac{\epsilon}{2})k_\nu - 1) \tag{6.120}$$

$$= f(\frac{k_\nu}{\nu})\Gamma((1-\frac{\epsilon}{2})k_\nu). \tag{6.121}$$

This establishes

$$\int_{\mu_W^{-1}((1-\epsilon)\frac{k_\nu}{\nu})}^{\infty} \mathrm{P}(D_{2,\nu}(y) > (1-\frac{\epsilon}{2})k_\nu)\mathrm{d}y \leq (\frac{1-\epsilon/2+1/k_\nu}{\epsilon/2+1/k_\nu})f(\frac{k_\nu}{\nu})\frac{\Gamma((1-\frac{\epsilon}{2})k_\nu)}{\Gamma((1-\frac{\epsilon}{2})k_\nu+1)} \tag{6.122}$$

$$= \frac{1}{\epsilon/2+1/k_\nu}\frac{1}{k_\nu}f(\frac{k_\nu}{\nu}) \tag{6.123}$$

as claimed. □

The next lemma establishes the other half of the tail bound for $g(y)$:

Lemma 6.8. *Suppose there is some $\chi > 0$ such that, for all $x > \chi$,*

$$\frac{\mu_W(x)}{x\mu_W'(x)} \geq -1, \tag{6.124}$$

and let B and C be as in Lemma 6.4. For ν sufficiently large such that $\frac{k_\nu}{\nu} \leq \mu_W(\chi)$ and ϵ such that $10C\frac{k_\nu}{\nu} \leq \epsilon < 1$,

$$\int_{\mu_W^{-1}((1-\epsilon)\frac{k_\nu}{\nu})}^{\infty} P(B > \frac{\epsilon}{2}k_\nu)\mathrm{d}y \leq (\frac{C}{10}\frac{\epsilon}{1-\epsilon}\frac{k_\nu}{\nu})^{\epsilon k_\nu/2}\frac{1}{\epsilon k_\nu/2 - 1}\mu_W^{-1}(\frac{(1-\epsilon)k_\nu}{\nu}). \tag{6.125}$$

Proof. The condition $10C\frac{k_\nu}{\nu} \leq \epsilon$ ensures that

$$C\mu_W(y) < \frac{\epsilon/2k_\nu}{5k_\nu}, \tag{6.126}$$

for $y > \mu_W^{-1}((1-\epsilon)\frac{k_\nu}{\nu})$. Recalling $B \sim \mathrm{Bin}(5k_\nu, C\mu_W(y))$, this allows a large deviation bound [AG89] to be applied:

$$\mathrm{P}(B > \frac{\epsilon}{2}k_\nu) \leq \exp(-5k_\nu S(\frac{\epsilon/2k_\nu}{5k_\nu}\|C\mu_W(y))), \tag{6.127}$$

where $S(q\|p) = q\log\frac{q}{p} + (1-q)\log\frac{1-q}{1-p}$ is the relative entropy between Bernoulli(q) and Bernoulli(p).

$$S(\frac{\epsilon}{10}\|C\mu_W(y)) \geq \frac{\epsilon}{10}\log\frac{10}{C\epsilon}\frac{1}{\mu_W(y)}, \tag{6.128}$$

whence

$$\mathrm{P}(B > \frac{\epsilon}{2}k_\nu) \leq (\frac{C}{10}\epsilon)^{\epsilon k/2}\mu_W(y)^{\frac{\epsilon\mathfrak{k}}{2}}. \tag{6.129}$$

It remains to integrate this bound. Let $f(x) = \mu_W^{-1}(x)$ then

$$\int_{\mu_W^{-1}((1-\epsilon)\frac{k_\nu}{\nu})}^{\infty} \mu_W(y)^{\epsilon k_\nu/2}\mathrm{d}y = \nu^{-\epsilon k/2}\int_0^{(1-\epsilon)k_\nu} x^{\epsilon k_\nu/2}\frac{1}{\nu}f'(\frac{x}{\nu})\mathrm{d}x. \tag{6.130}$$

Following the same reasoning as in the proof of Lemma 6.7,

$$x^2\frac{1}{\nu}f'(\frac{x}{\nu}) \leq (1-\epsilon)k_\nu f(\frac{(1-\epsilon)k_\nu}{\nu}) \tag{6.131}$$

on the domain of integration so,

$$\nu^{-\epsilon k/2}\int_0^{(1-\epsilon)k_\nu} x^{\epsilon k/2}\frac{1}{\nu}f'(\frac{x}{\nu})\mathrm{d}x \leq \nu^{-\epsilon k/2}(1-\epsilon)k_\nu f(\frac{(1-\epsilon)k_\nu}{\nu})[\frac{1}{\epsilon k_\nu/2-1}((1-\epsilon)k_\nu)^{\epsilon k_\nu/2-1}] \tag{6.132}$$

$$= (\frac{k_\nu}{\nu})^{\epsilon k_\nu/2}(1-\epsilon)^{\epsilon k_\nu/2}\frac{1}{\epsilon k_\nu/2-1}f(\frac{(1-\epsilon)k_\nu}{\nu}). \tag{6.133}$$

□

In particular, the last several lemmas combine to show that for $\epsilon_\nu \leq 1$ such that $\epsilon_\nu = \omega(\frac{1}{k_\nu})$ and $\epsilon_\nu = \omega(\frac{k_\nu}{\nu})$ it holds that

$$\int_0^{\mu_W^{-1}((1+\epsilon)\frac{k_\nu}{\nu})} g(y)\mathrm{d}y + \int_{\mu_W^{-1}((1-\epsilon)\frac{k_\nu}{\nu})}^{\infty} g(y)\mathrm{d}y = o(\mu_W^{-1}(\frac{1}{\nu})). \tag{6.134}$$

With the observation that $\mathbb{E}[N_{>0}^{(\nu)}] = \Omega(\mu_W^{-1}(\frac{1}{\nu}))$ this leaves only the region

$$(\mu_W^{-1}(1+\epsilon)\frac{k_\nu}{\nu}, \mu_W^{-1}(1-\epsilon)\frac{k_\nu}{\nu}) \tag{6.135}$$

as a possible foil to $\int g(y)\mathrm{d}y = o(\mathbb{E}[N_{>0}^{(\nu)}])$. In this regime we expect

$$g(y) = \mathrm{P}(D_{2,\nu}(y) + B(y) > k_\nu \wedge D_{2,\nu}(y) \leq k_\nu) \tag{6.136}$$

to be approximately constant because $\mathbb{E}[D_{2,\nu}(y)] \approx k_\nu$ so we make due with the bound $g(y) \leq 1$.

Lemma 6.9. *Suppose that μ_W is differentiable and that there is some $\chi > 0$ such that for all $x > \chi$ it holds that*

$$\frac{\mu_W(x)}{x\mu_W'(x)} \geq -1. \tag{6.137}$$

Then for $\epsilon > 0$ and ν sufficiently large such that $(1+\epsilon)\frac{k_\nu}{\nu} \leq \mu_W(\chi)$, it holds that

$$\mu_W^{-1}((1-\epsilon)\frac{k_\nu}{\nu}) - \mu_W^{-1}((1+\epsilon)\frac{k_\nu}{\nu}) \leq 2\frac{\epsilon}{1-\epsilon}\mu_W^{-1}((1-\epsilon)\frac{k_\nu}{\nu}) \tag{6.138}$$

Proof. Let $f(x) = \mu_W^{-1}(x)$. Since μ_W is differentiable so is f. By the mean value theorem there is some point $(1-\epsilon)\frac{k_\nu}{\nu} \leq x^* \leq (1+\epsilon)\frac{k_\nu}{\nu}$ such that

$$f((1-\epsilon)\frac{k_\nu}{\nu}) - f((1+\epsilon)\frac{k_\nu}{\nu}) = -2\epsilon\frac{k_\nu}{\nu}f'(x^*) \tag{6.139}$$

$$= -2\epsilon\frac{k_\nu}{\nu}\frac{1}{x^*}x^*f'(x^*) \tag{6.140}$$

$$\leq 2\frac{\epsilon}{1-\epsilon}f((1-\epsilon)\frac{k_\nu}{\nu}), \tag{6.141}$$

where the final line follows as in Lemma 6.7. □

We can now complete our intermediate goal:

Lemma 6.10. *Let $g(y)$, T and C be as in Lemma 6.4. Suppose $k_\nu \uparrow \infty$ and $k_\nu = o(\nu)$. Suppose that μ_W is differentiable and that there is some $\chi > 0$ such that for all $x > \chi$ it holds that*

$$\frac{\mu_W(x)}{x\mu_W'(x)} \geq -1. \tag{6.142}$$

Then

$$\int g(y)\mathrm{d}y = o(\mathbb{E}[N_{>0}^{(\nu)}]) \tag{6.143}$$

Proof. Let $\epsilon_\nu \downarrow 0$ such that $\epsilon_\nu = \omega(\sqrt{\frac{1}{k_\nu}})$ and $\epsilon_\nu = \omega(\sqrt{\frac{k_\nu}{\nu}})$. Let

$$h(y) = \begin{cases} \mathrm{P}(D_{2,\nu}(y) \leq k_\nu) & y \leq \mu_W^{-1}((1+\epsilon_\nu)\frac{k_\nu}{\nu}) \\ 1 & y \in (\mu_W^{-1}((1+\epsilon_\nu)\frac{k_\nu}{\nu}), \mu_W^{-1}((1-\epsilon_\nu)\frac{k_\nu}{\nu})) \\ \mathrm{P}(B(y) > \frac{\epsilon}{2}k_\nu) + \mathrm{P}(D_{2,\nu}(y) > (1-\frac{\epsilon}{2})k_\nu) & y \geq \mu_W^{-1}((1-\epsilon_\nu)\frac{k_\nu}{\nu}). \end{cases} \tag{6.144}$$

Because μ_W is not compactly supported, for ν sufficiently large $\mu_W^{-1}((1+\epsilon_\nu)\frac{k_\nu}{\nu}) > T$ and in this regime it is immediate that

$$g(y) \leq h(y). \tag{6.145}$$

Moreover, it is straightforward to verify that the conditions on ϵ_ν with Lemmas 6.6 to 6.9 imply

$$\int h(y)\mathrm{d}y = o(\mu_W^{-1}((1-\epsilon_\nu)\frac{k_\nu}{\nu})). \tag{6.146}$$

(For Lemma 6.6 it suffices to consider the worst case $\epsilon_\nu = \sqrt{\frac{1}{k_\nu}}$.)

Next,

$$\mathbb{E}[N_{>0}^{(\nu)}] = \int_{\mathbb{R}_+} 1 - e^{-\nu\mu_W(y)}\mathrm{d}y \tag{6.147}$$

$$\geq \int_0^{\mu_w^{-1}(\frac{1}{\nu})} 1 - e^{-1}\mathrm{d}y \tag{6.148}$$

$$= \Omega(\mu_W^{-1}(\frac{1}{\nu})). \tag{6.149}$$

Thus $\mathbb{E}[N_{>0}^{(\nu)}] = \Omega(\mu_W^{-1}((1-\epsilon_\nu)\frac{k_\nu}{\nu}))$, completing the proof. □

We are now equipped to give the proof of the main result:

Proof of Theorem 6.1. By Lemma 6.2 it suffices to show $\mathrm{var}\left[N_{>k_\nu}^{(\nu)}\right] = o(\mathbb{E}[N_{>0}^{(\nu)}]^2)$. By Lemmas 6.3 and 6.4,

$$\mathrm{var}\left[N_{>k_\nu}^{(\nu)}\right] \leq \mathbb{E}[N_{>k_\nu}^{(\nu)}](1 + \int g(y)\mathrm{d}y), \tag{6.150}$$

where $g(y)$ is as defined in Lemma 6.4. Lemma 6.5, for bounded k_ν, and Lemma 6.10, for $k_\nu \uparrow \infty$, establish

$$\int g(y)\mathrm{d}y = o(\mathbb{E}[N_{>0}^{(\nu)}]), \tag{6.151}$$

completing the proof. □

7. Connectivity for Separable KEGs

A serious omission in the results presented thus far is that they give virtually no information about the global structure of the KEGs. In particular, we have as yet made no statements about the connectivity structure of these graphs. The sparse structure that we explore here could, in principle, arise from graphs that consist of large numbers of disconnected dense components. If this were to be the case then these graphs would be uninteresting for physical applications. Our aim in this section is to give a preliminary result showing that this is not the case.

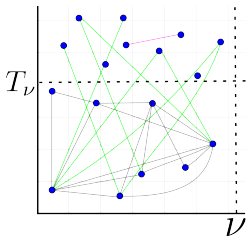

FIGURE 4. The basic structure of separable KEGs. The induced subgraph below T_ν in gray is fully connected. Above T_ν the vast majority of the vertices of the graph connect to the below threshold subgraph, in green. This leaves only the very small number of vertices connected only to vertices that lie entirely above T_ν, in magenta.

Definition 7.1. We call a KEG *separable* if the associated graphex has $I = S = 0$ and W of the form

$$W(x,y) = \begin{cases} 0 & x = y \\ f(x)f(y) & \text{otherwise.} \end{cases} \tag{7.1}$$

We prove that separable KEGs have an arbitrarily large fraction of the vertices contained in a single connected component in the large graph limit. (As usual, because there is no risk of confusion, we will use the term graphex to refer to the function W.)

Remark 7.2. Separability in combination with the graphex integrability conditions immediately implies that f and hence W is integrable and thus that this result only applies for graphs that have a finite expected number of edges when restricted to finite support ν. ◁

The main obstacle to the study of connectivity in the KEG setting is that the graphs are naturally defined in terms of the infinite collection of points in the latent Poisson process with only a finite number of these participating as points in a sampled graph. The difficulty is that traditional tools (e.g. [Bol01]) for studying connectivity begin with a fixed set of vertices of the graph and examine how they become connected as edges are randomly introduced, an approach that is apparently futile in the present setting where we must specify the edge set in order to specify the vertex set. The tactic we use to circumvent this problem hinges on the division of the KEG into three parts based on the latent ϑ values of the vertices: the induced subgraph below some threshold value, the induced subgraph above this threshold and the bi-graph between them; see Fig. 4. The first piece intuition is that for fixed ν we can set the threshold T_ν such that nearly every point of the latent Poisson process with ϑ below T_ν will have an edge connected to it; because of this we can treat the connectivity of the below T_ν induced subgraph using the traditional random graph machinery. The connectivity of vertices lying above T_ν that participate in at least one edge connecting below T_ν then follows straightforwardly. This leaves only the vertices in the induced subgraph above T_ν that do not connect to a point below T_ν and it will turn out that these constitute a negligible fraction of the graph.

We fix some notation that we will need for the rest of this section: Let Π be the unit rate Poisson process on $\mathbb{R}_+^2$ and let $\Pi_\nu = \{(\theta_i, \vartheta_i) \in \Pi \mid \theta_i \leq \nu\}$ be the restriction of this process to label-space $\leq \nu$. Let the Poisson process below a cutoff value x in ϑ space be $\Pi_{\nu,\leq x} = \{(\theta_i, \vartheta_i) \in \Pi_\nu \mid \vartheta_i < x\}$ and let the process above the cutoff be $\Pi_{\nu,>x} = \{(\theta_i, \vartheta_i) \in \Pi_\nu \mid \vartheta_i > x\}$.

We begin by showing we can take $f(x)$ to be monotone decreasing without loss of generality:

Lemma 7.3. *Let $W(x,y) = f(x)f(y)1[x \neq y]$ be a separable graphex, then there is some other separable graphex $W' = h(x)h(y)1[x \neq y]$ such that h is monotone decreasing and the KEGs associated to W and W' are equal in distribution.*

Proof. Because the distribution of a KEG is invariant under measure preserving transformations of the generating graphon, it suffices to show that there are some measure preserving transformations $\tau, \varphi : \mathbb{R}_+ \to \mathbb{R}_+$ and a monotonically decreasing function h such that $f \circ \tau = h \circ \varphi$

If $f(x)$ has bounded domain (i.e., W is a graphon) then the result follows immediately from [Lov13, Prop.A19], which shows that for any bounded f with compact support there is some measure preserving transformation φ on the domain of f and monotone decreasing h such that $f = h \circ \varphi$.

Assume $f(x)$ has unbounded domain. Because f is integrable and measurable the sets $A_k = \{x \mid f(x) \in [\frac{1}{k}, \frac{1}{k+1})\}$ for $k \in \mathbb{N}$ are Borel sets of finite measure. This means in particular ([Ker14, Thm. A.20]) that for A_k with measure c_k there is some measure preserving transformation $\tilde{\tau}$ such that $\tilde{\tau}(A_k) = [0, c_k]$. From this it immediately follows that there exists a measure preserving transformation τ such that $\tau(A_k) = [c_{k-1}, c_k]$ with $c_0 = 0$. That is, τ imposes a pseudo-monotonicity where $f(\tau(x)) < \frac{1}{k}$ and $f(\tau(y)) \geq \frac{1}{k}$ implies $\tau(x) > \tau(y)$. By [Lov13, Prop.A19] there is a measure preserving transformation φ_k and a monotonically decreasing h_k with support $\tau(A_k)$ such that $1_{\tau(A_k)} f \circ \tau = h_k \otimes \varphi_k$. Letting $\varphi = \bigotimes_i \varphi_i$ and $h = \bigotimes_i h_i$ completes the proof. □

We take f to be monotone decreasing for the remainder of the section. Because the result is trivial for f with bounded domain (the KEG is dense) we also take f to have unbounded domain. Denote the left continuous inverse of f by $f^{-1}(t) = \inf\{\lambda : f(\lambda) = t\}$. We will make frequent use of the observation that for $l_\nu \in o(1)$ it holds that $f^{-1}(l_\nu) \in \omega(1)$. Let G be a Kallenberg Exchangeable Graph associated with W and let G_ν be the restriction to $[0, \nu]$.

Definition 7.4. Let t_ν be a function of ν such that $t_\nu \in o(1)$ and $t_\nu \in \omega(\frac{1}{\nu})$ and define the *threshold* $T_\nu = f^{-1}(\frac{1}{\nu} + t_\nu)$.

Remark 7.5. This notation for the threshold suppresses the dependence on t_ν, which should be thought of as going to 0 as quickly as possible consistent with $t_\nu \in \omega(\frac{1}{\nu})$. ◁

The proof now proceeds roughly as follows:

(1) We establish the existence of a connected core that we will show nearly every vertex of the graph connects to (Lemma 7.6)
(2) We show that nearly every point of $\Pi_{\nu,\leq T_\nu}$ participates in an edge connecting to the connected core (Lemma 7.7)
(3) We lower bound the number of points of $\Pi_{\nu,>T_\nu}$ that connect to the connected core (Lemma 7.8)

(4) We consider the induced subgraph of G_ν given by $\{\theta_i \in v(G_\nu) \mid \vartheta_i > T_\nu\}$ and show that the number of points in this subgraph that fail to connect to the connected core is an arbitrarily small fraction of the number of vertices in the graph (Lemma 7.10)

The first step of the proof is to show that there is an induced subgraph P_ν that is both connected and very popular in the sense that every other vertex of the graph will connect to it with high probability. The notion of popularity that we use is the that total mass in the subgraph, $\sum_{p\in P_\nu} f(p)$, is an arbitrarily large fraction of the total expected mass in the entire graph: $\mathbb{E}[\sum_{\vartheta_i\in\Pi_\nu} f(\vartheta_i)] = \nu\|f\|_1$. The critical fact for use in later parts of the argument turns out to be that the mass of the popularity subgraph scales as ν.

Lemma 7.6. *Suppose f does not have compact support. Let $T_{\nu,pop} = f^{-1}(\sqrt{\frac{\log\nu}{\nu}})$ and let P_ν be the induced subgraph of G_ν given by including only vertices in $\Pi_{\nu,<T_\nu,pop}$, then:*

(1) Every element of $\Pi_{\nu,\leq T_\nu,pop}$ connects to an edge; $\lim_{\nu\to\infty}|\Pi_{\nu,\leq T_\nu,pop}\backslash v(P_\nu)| = 0$ a.s.
(2) P_ν is almost surely connected; let $C(P_\nu) = 1$ if P_ν is connected and 0 otherwise, then $\lim_{\nu\to\infty} C(P_\nu) = 1$ a.s.
(3) P_ν is "ultra-popular" almost surely; letting $S_\nu = \sum_{p\in P_\nu} f(p)$ we have for $\epsilon > 0$ that $\lim_{\nu\to\infty} \frac{S_\nu}{\nu} \geq (1-\epsilon)\|f\|_1$ a.s.

Proof. The key insight is that the connection probabilities below $T_{\nu,\mathrm{pop}}$ are lower bounded by $p_\nu = f(T_{\nu,\mathrm{pop}})^2 = \frac{\log\nu}{\nu}$ so that a sufficient condition for claims 1 and 2 is that the Erdős–Rényi–Gilbert random graph $G(N_\nu, p_\nu)$ with $N_\nu \sim \mathrm{Poi}(\nu T_{\nu,\mathrm{pop}})$ is almost surely connected in the limit. A sufficient condition [Bol01] for this is that there exists some $\delta > 0$ such that

$$\lim_{\nu\to\infty} \frac{p_\nu}{\log N_\nu/N_\nu} > 1+\delta \text{ a.s.} \tag{7.2}$$

For arbitrary $\gamma > 0$, it holds that $\lim_{\nu\to\infty} N_\nu/\nu T_{\nu,\mathrm{pop}} \geq (1-\gamma)$ a.s. and so we have that:

$$\lim_{\nu\to\infty} \frac{p_\nu}{\log N_\nu/N_\nu} \geq \lim_{\nu\to\infty} \frac{\log\nu/\nu}{\log(1-\gamma)\nu T_{\nu,\mathrm{pop}}/(1-\gamma)\nu T_{\nu,\mathrm{pop}}} \text{ a.s.} \tag{7.3}$$

$$= \infty. \tag{7.4}$$

Thus in the limit as $\nu\to\infty$, the random graph with vertices $\Pi_{\nu,\leq T_\nu,\mathrm{pop}}$ and independent edge probabilities $f(\vartheta_i)f(\vartheta_j)$ is connected and, in particular, every vertex is contained in an edge, thereby establishing claims 1 and 2.

It remains to show that S_ν grows as claimed. For $\gamma > 0$, by Hoeffding's inequality we have:

$$\mathrm{P}(S_\nu < (1-\gamma)\mathbb{E}[S_\nu \mid N_\nu] \mid N_\nu) \leq \mathrm{P}(|S_\nu - \mathbb{E}[S_\nu \mid N_\nu]| < \gamma \mathbb{E}[S_\nu \mid N_\nu] \mid N_\nu) \tag{7.5}$$

$$\leq 2\exp(-2\gamma^2 \frac{\mathbb{E}[S_\nu \mid N_\nu]^2}{N_\nu}) \tag{7.6}$$

$$= 2\exp(-2\gamma^2 \frac{N_\nu}{T_{\nu,\mathrm{pop}}^2}(\int_0^{T_{\nu,\mathrm{pop}}} f(x)\mathrm{d}x)^2) \tag{7.7}$$

$$\leq 2\exp(-2\gamma^2 \frac{N_\nu}{T_{\nu,\mathrm{pop}}^2}(1-\gamma)^2\|f\|_1^2), \tag{7.8}$$

for ν sufficiently large since $T_{\nu,\mathrm{pop}} \to \infty$ as $\nu \to \infty$. Whence,

$$\mathrm{P}(\frac{S_\nu}{\nu(1-\gamma)^2\|f\|_1} < 1-\gamma \mid N_\nu \geq (1-\gamma)\nu T_{\nu,\mathrm{pop}}) \tag{7.9}$$

$$\leq \mathrm{P}(\frac{S_\nu}{\mathbb{E}[S_\nu \mid N_\nu]} < 1-\gamma \mid N_\nu \geq (1-\gamma)\nu T_{\nu,\mathrm{pop}}) \tag{7.10}$$

$$\leq 2\exp(-2\gamma^2 \frac{\nu}{T_{\nu,\mathrm{pop}}}(1-\gamma)^3\|f\|_1^2). \tag{7.11}$$

Using that $f(x)$ is monotonic and must be integrable we have that $f(x) = o(\frac{1}{x})$ so $\nu/T_{\nu,\mathrm{pop}} \geq (\nu \log \nu)^{1/2}$ and

$$\mathrm{P}(\frac{S_\nu}{\nu(1-\gamma)^2\|f\|_1} < 1-\gamma \mid N_\nu \geq (1-\gamma)\nu T_{\nu,\mathrm{pop}}) \leq 2\exp(-2\gamma^2(\nu\log\nu)^{1/2}(1-\gamma)^3\|f\|_1^2). \tag{7.12}$$

Finally, using $\lim_{\nu\to\infty} \frac{N_\nu}{\nu T_{\nu,\mathrm{pop}}} \geq (1-\gamma)$ a.s. and the Borel–Cantelli lemma establishes

$$\lim_{\nu\to\infty} \frac{S_{\lfloor\nu\rfloor}}{\lfloor\nu\rfloor+1} \geq (1-\gamma)^3\|f\|_1 \text{ a.s.} \tag{7.13}$$

$$\implies \lim_{\nu\to\infty} \frac{S_\nu}{\nu} \geq (1-\gamma)^3\|f\|_1 \text{ a.s.} \tag{7.14}$$

and the result follows since $\gamma > 0$ is arbitrary. □

We now have a promise that every point of the latent Poisson process $\Pi_{\nu,\leq T_{\nu,\mathrm{pop}}}$ participates in the graph. We now establish that, with high probability, as $\nu \to \infty$ an arbitrarily large fraction of the points in $\Pi_{\nu,\leq T_\nu}$ connect to the popular connected core P_ν. In particular, this means an arbitrarily large fraction of the points of $\Pi_{\nu,\leq T_\nu}$ participate in a single connected component of G_ν.

Lemma 7.7. *Suppose f does not have compact support. Let a point $(\theta_i, \vartheta_i) \in \Pi_{\nu,\leq T_\nu}$ be visible if $\theta_i \in v(G_\nu)$ and it participates in an edge connecting to P_ν, and call a point invisible otherwise. Let $N_{invis,\leq T_\nu}$ be the number of points in $\Pi_{\nu,\leq T_\nu}$ that are invisible and let $N_{vis,\leq T_\nu}$ be the number of points in $\Pi_{\nu,\leq T_\nu}$ that are visible, then for $\epsilon > 0$*

$$\lim_{\nu\to\infty} P(N_{invis,<T_\nu} > \epsilon N_{vis,<T_\nu}) = 0. \tag{7.15}$$

Proof. By Lemma 7.6 it follows that as $\nu \to \infty$ there are no invisible vertices below $T_{\nu,\text{pop}} = f^{-1}(\sqrt{\frac{\log \nu}{\nu}})$ so it suffices to bound the number of invisible vertices between $T_{\nu,\text{pop}}$ and T_ν. Conditional on P_ν, each point $(\theta_i, \vartheta_i) \in \Pi_{\nu,>T_{\nu,\text{pop}}}$ connects to P_ν independently with probability $1 - \prod_{p\in P_\nu}(1 - f(\vartheta_i)f(p)) \geq 1 - e^{-f(\vartheta_i)S_\nu}$ where $S_\nu = \sum_{p\in P_\nu} f(p)$. Since labeling each point of the Poisson process $\Pi_{\nu,>T_{\nu,\text{pop}}}$ by whether or not it connects to $T_{\nu,\text{pop}}$ is, conditional on P_ν, a marking of the Poisson process, we immediately have that the number of visible and invisible points in $\{(\theta_i, \vartheta_i) \in \Pi_\nu \mid T_{\nu,\text{pop}} < T_\nu\}$ are independent random variables and that there exists random variables $N_{\nu,\text{ub}}$ and $N_{\nu,\text{vis}}$ such that,

$$N_{\nu,\text{ub}} \sim \text{Poi}(\nu \int_{T_{\nu,\text{pop}}}^{T_\nu} e^{-f(x)S_\nu} \mathrm{d}x) \tag{7.16}$$

is a upper bound for $N_{\text{invis},<T_\nu}$ and

$$N_{\nu,\text{vis}} \sim \text{Poi}(\nu \int_{T_{\nu,\text{pop}}}^{T_\nu} 1 - e^{-f(x)S_\nu} \mathrm{d}x) \tag{7.17}$$

is an independent lower bound for $N_{\text{vis},<T_\nu}$).

Thus a sufficient condition for the claim is $\text{P}(\frac{N_{\nu,\text{ub}}}{N_{\nu,\text{vis}}} > \epsilon) \to 0,\ \nu \to \infty$. Conditional on S_ν, this is a ratio of independent Poisson random variables and this condition will hold if the ratio of their means goes to 0:

$$\frac{\nu \int_{T_{\nu,\text{pop}}}^{T_\nu} e^{-f(x)S_\nu} \mathrm{d}x}{\nu \int_{T_{\nu,\text{pop}}}^{T_\nu} 1 - e^{-f(x)S_\nu} \mathrm{d}x} \leq \frac{(T_\nu - T_{\nu,\text{pop}})e^{-f(T_\nu)S_\nu}}{(T_\nu - T_{\nu,\text{pop}})} \tag{7.18}$$

$$\leq e^{-(\frac{1}{\nu}+t_\nu)S_\nu}. \tag{7.19}$$

Invoking $\lim_{\nu\to\infty} S_\nu/\nu \geq \frac{1}{2}\|f\|_1 = 1$ a.s. from Lemma 7.6 completes the result since this means $\lim_{\nu\to\infty} t_\nu S_\nu = \infty$ a.s. □

The next step is to determine the total number of vertices above T_ν that connect to the popular connected core:

Lemma 7.8. *Suppose f does not have compact support. Let*

$$N_{\nu,>T_\nu} = |\{(\theta_i, \vartheta_i) \in \Pi_{\nu,>T_\nu} \mid \exists p \in v\,(P_\nu) \text{ such that } \{\theta_i, p\} \in e\,(G_\nu)\}|, \tag{7.20}$$

be the number of points above T_ν that connect to P_ν. Then there exists a random variable $N_{\nu,+}$ such that $N_{\nu,+} \leq N_{\nu,>T_\nu}$ and

$$N_{\nu,+} \mid S_\nu \sim \text{Poi}(\nu \int_{T_\nu}^{\infty} 1 - e^{-f(x)S_\nu} \mathrm{d}x) \tag{7.21}$$

Proof. Conditional on P_ν, each point $(\theta_i, \vartheta_i) \in \Pi_{\nu,>T_\nu}$ connects to P_ν independently with probability $1 - \prod_{p\in P_\nu}(1 - f(\vartheta_i)f(p)) \geq 1 - e^{-f(\vartheta_i)S_\nu}$. This is a marking of the Poisson process so the random subset of $\Pi_{\nu,>T_\nu}$ that connects to P_ν is itself a Poisson process with rate $\nu(1 - \prod_{p\in P_\nu}(1 - f(\vartheta_i)f(p)))$. We may then further independently mark the points of this process such that the new random subset will be, conditional on S_ν, a Poisson process with rate $\nu \int_{T_\nu}^{\infty} 1 - e^{-f(x)S_\nu} \mathrm{d}x$. Let the number of points in this process be $N_{\nu,+}$ then it follows immediately that $N_{\nu,+}$ is a lower bound $N_{\nu,>T_\nu}$ and that $N_{\nu,+} \mid S_\nu \sim \text{Poi}(\nu \int_{T_\nu}^{\infty} 1 - e^{-f(x)S_\nu} \mathrm{d}x)$. □

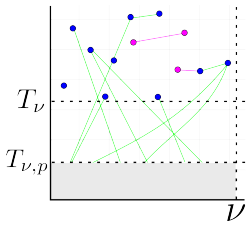

FIGURE 5. The structure of negligible vertices above T_ν. Vertices with distance > 2 to P_ν (below $T_{\nu,p}$) are ignored, these are marked in magenta.

The final step is to bound the number of vertices above T_ν that will be neglected. These are the vertices that participate in edges lying entirely above T_ν and have a minimum distance greater than 2 to the popular subgraph P_ν. Note that they may be part of the giant component, but their contribution is negligible. We begin with a small technical lemma:

Lemma 7.9. *Let $f : \mathbb{R}_+ \to [0,1]$ be monotonically decreasing and integrable, then $f^{-1}(\frac{1}{t}) = o(t)$.*

Proof. Suppose otherwise so that $\exists c > 0$ such that $f^{-1}(\frac{1}{t}) \geq ct$ infinitely often. Let $\{t_i\}_{i=1}^\infty$ be a strictly increasing sequence of such ts, then for each t_i there exists a box B_{t_i} of area at least c that lies under the graph: namely the box $[0, ct] \times [0, f(ct)]$. For $\epsilon > 0$ we may choose a subsequence $\{\tilde{t}_j\}_{j=1}^\infty \subset \{t_i\}_{i=1}^\infty$ such that $|B_{t_i} \cap B_{t_{i+1}}| \leq \epsilon$, so that the area below f is bounded below by an infinite sum where each term has value at least $c - \epsilon > 0$ thereby arriving at a contradiction. □

Following our interpretation of T_ν as a cutoff below which every candidate vertex participates in the graph, the requirement $T_\nu = o(\nu)$ is obvious. Suppose otherwise, then there would be $\Omega(\nu^2)$ visible vertices in the graph and $\Theta(\nu^2)$ expected edges, pushing the graph into the ultra-sparse regime where $|e(G_\nu)| = O(|v(G_\nu)|)$. The above lemma shows that $T_\nu = o(\nu)$ does indeed hold, since $T_\nu = o(f^{-1}(1/\nu))$ and $f^{-1}(1/\nu) = o(\nu)$. With this result in hand,

Lemma 7.10. *Suppose f does not have compact support. Call a vertex $\theta_i \in v(G_\nu)$ ignored if $(\theta_i, \vartheta_i) \in \Pi_{\nu,>T_\nu}$ and its distance to P_ν is greater than 2. Let N_{ignore} be the number of ignored vertices; then fixing $\epsilon > 0$,*

$$\lim_{\nu\to\infty} P\left(\frac{N_{ignore}}{|v(G_\nu)|} > \epsilon\right) \to 0. \tag{7.22}$$

Proof. We mark each point in the Poisson process $\Pi_{\nu,>T_\nu}$ above T_ν by whether it participates in an edge with a terminus in P_ν. As in Lemma 7.8, this forms a marking of the Poisson process conditional on P_ν so that the random subset of $\Pi_{\nu,>T_\nu}$ that is at distance one (**c**lose) to P_ν,

$$C_\nu = \{(\theta_i, \vartheta_i) \in \Pi_{\nu,>T_\nu} \mid \exists p \in v(P_\nu) \text{ such that } \{\theta_i, p\} \in e(G_\nu)\}, \tag{7.23}$$

and the remaining subset $\Pi_{\nu,>T_\nu} \backslash C_\nu$ are independent Poisson processes conditional on P_ν.

Let $e_{\nu,\text{ignore}}$ be the the number of edges in the induced subgraph of G_ν given by restricting the vertex set to $\Pi_{\nu,>T_\nu} \backslash C_\nu$. It is immediate that $N_{\text{ignore}} \leq 2e_{\nu,\text{ignore}}$

(see Fig. 5). Obviously $|v(G_\nu)| > |C_\nu|$ and by Lemma 7.8 $|C_\nu| > N_{\nu,+}$ so

$$\mathrm{P}(\frac{N_{\text{ignore}}}{|v(G_\nu)|} > \epsilon) \leq \mathrm{P}(\frac{2e_{\nu,\text{ignore}}}{N_{\nu,+}} > \epsilon), \tag{7.24}$$

where in particular $e_{\nu,\text{ignore}}$ and $N_{\nu,+}$ are independent conditional on $\Pi_{\nu,\leq T_\nu}$.

We have very little distributional information about $e_{\nu,\text{ignore}}$ so we use Markov's inequality. Since $\Pi_{\nu,>T_\nu}\backslash C_\nu$ is a Poisson process with rate at most $\nu e^{-f(x)S_\nu}$ we may repeat the argument of Theorem 5.3 to bound $\mathbb{E}[e_{\nu,\text{ignore}} \mid P_\nu]$ so that

$$\mathbb{E}[\frac{e_{\nu,\text{ignore}}}{N_{\nu,+}} \mid \Pi_{\nu,\leq T_\nu}] = \frac{\mathbb{E}[e_{\nu,\text{ignore}} \mid \Pi_{\nu,\leq T_\nu}]}{\mathbb{E}[N_{\nu,+} \mid \Pi_{\nu,\leq T_\nu}]} \tag{7.25}$$

$$\leq 2\frac{\nu^2(\int_{T_\nu}^\infty e^{-2S_\nu f(x)} f(x)\mathrm{d}x)^2}{\nu\int_{T_\nu}^\infty 1 - e^{-f(x)S_\nu}\mathrm{d}x}. \tag{7.26}$$

From this we see that the bound is S_ν measurable. Taking $\gamma > 0$ and working in the regime where $(1-\gamma) \leq \frac{S_\nu}{\nu\|f\|_1} \leq (1+\gamma)$ we have:

$$\mathbb{E}[\frac{e_{\nu,\text{ignore}}}{N_{\nu,+}} \mid (1-\gamma) \leq \frac{S_\nu}{\nu\|f\|_1} \leq (1+\gamma)] \leq 2\nu\frac{(\int_{T_\nu}^\infty e^{-2(1+\gamma)\|f\|_1\nu f(x)} f(x)\mathrm{d}x)^2}{\int_{T_\nu}^\infty 1 - e^{-(1-\gamma)\|f\|_1\nu f(x)}\mathrm{d}x}. \tag{7.27}$$

This can be treated by breaking up the integrals into the contributions above and below and upper threshold $T_{\nu,u} = f^{-1}(\frac{1}{\nu})$. The numerator breaks up as,

$$\int_{T_\nu}^{T_{\nu,u}} e^{-2(1+\gamma)\|f\|_1\nu f(x)} f(x)\mathrm{d}x + \int_{T_{\nu,u}}^\infty e^{-2(1+\gamma)\|f\|_1\nu f(x)} f(x)\mathrm{d}x \tag{7.28}$$

$$\leq O(\frac{T_{\nu,u} - T_\nu}{\nu}) + O(\int_{T_{\nu,u}}^\infty f(x)\mathrm{d}x), \tag{7.29}$$

where we have bounded the left term by the maximum of its integrand. The denominator breaks up as,

$$\int_{T_\nu}^{T_{\nu,u}} 1 - e^{-(1-\gamma)\|f\|_1\nu f(x)}\mathrm{d}x + \int_{T_{\nu,u}}^\infty 1 - e^{-(1-\gamma)\|f\|_1\nu f(x)}\mathrm{d}x \tag{7.30}$$

$$\geq \Omega(T_{\nu,u} - T_\nu) + \Omega(\nu\int_{T_{\nu,u}}^\infty f(x)\mathrm{d}x), \tag{7.31}$$

where the bound on the right term follows from the fact that for constant $c > 0$ there exists L_C depending only on c such that $1 - e^{-cx} \geq L_c x$ for $x < 1$. Thus, in particular,

$$\mathbb{E}[\frac{e_{\nu,\text{ignore}}}{N_{\nu,+}} \mid (1-\gamma) \leq \frac{S_\nu}{\nu\|f\|_1} \leq (1+\gamma)] \tag{7.32}$$

$$= O\left(\nu(\frac{T_{\nu,u} - T_\nu}{\nu})^2\frac{1}{T_{\nu,u} - T_\nu}, \nu(\int_{T_{\nu,u}}^\infty f(x)\mathrm{d}x)^2\frac{1}{\nu\int_{T_{\nu,u}}^\infty f(x)\mathrm{d}x}\right), \tag{7.33}$$

$$\tag{7.34}$$

and this goes to 0 as $\nu \to \infty$; the left term because $T_{\nu,u} = o(\nu)$ by Lemma 7.9 and the right term because f is integrable and $T_{\nu,u} \to \infty$.

Putting all of this together and using that $(1-\gamma) \leq \lim_{\nu\to\infty} \frac{S_\nu}{\nu\|f\|_1} \leq (1+\gamma)$ a.s. by Lemma 7.6 we have that:

$$\lim_{\nu\to\infty} \mathrm{P}(\frac{2e_{\nu,\text{ignore}}}{N_{\nu,+}} > \epsilon) = \lim_{\nu\to\infty} \mathrm{P}(\frac{2e_{\nu,\text{ignore}}}{N_{\nu,+}} > \epsilon \mid (1-\gamma) \leq \frac{S_\nu}{\nu\|f\|_1} \leq (1+\gamma)) \tag{7.35}$$

$$\leq \lim_{\nu\to\infty} 2\mathbb{E}[\frac{e_{\nu,\text{ignore}}}{N_{\nu,+}} > \epsilon \mid (1-\gamma) \leq \frac{S_\nu}{\nu\|f\|_1} \leq (1+\gamma)] \tag{7.36}$$

$$= 0, \tag{7.37}$$

where the second line follows by Markov's inequality. This establishes our claim. □

We can now put all of this together:

Theorem 7.11. *Let G be the KEG generated by $W = f(x)f(y)1[x \neq y]$, let $C_1(G_\nu)$ be the largest connected component of G_ν, and let $\epsilon > 0$, then*

$$\lim_{\nu\to\infty} P(|C_1(G_\nu)| > (1-\epsilon)|v\,(G_\nu)|) = 1. \tag{7.38}$$

Proof. For f with compact support this is a trivial consequence of Theorem 5.6, which shows that the graph is dense. For f without compact support this is an immediate consequence of the lemmas of this section. □

A couple of concluding remarks are in order. Notice that the result extends trivially to allow separable graphs that include self edges because only a vanishing fraction of the vertices have a self edge. The proofs in this section reveal some further interesting structure of separable KEGs beyond connectivity, in particular:

(1) If two points of a separable KEG are chosen at random there will be a very short path between them with high probability, even for very sparse random graphs. This is because both vertices very likely connect to the very dense subgraph P_ν by paths of length at most 2.
(2) Although vertices of G_ν chosen uniformly at random are overwhelmingly likely to follow a degree distribution of the type given in Theorem 6.1 there are a vanishingly small fraction of the vertices (those in P_ν) with much higher degree.

Applied networks folk wisdom [New09; Dur06] holds that real-world graphs often exhibit "small world" behaviour, with very short paths between random vertices even for sparse graphs. Similarly, it's common to observe that real-world graphs tend to follow power law degree distribution except for the highest degree vertices, which have much higher degree than would be expected from such a law. It's interesting that both of these features arise as emergent behaviour of the simple random graph model considered in this section.

8. Discussion

This work was motivated by the need for a statistical framework for the analysis of the sparse graph structure of real-world networks. The Kallenberg random graph model provides such a framework, although the applicability and suitability of this framework—from either empirical or theoretical perspectives—is still to be determined. Our work characterizing the limiting degree distribution and connectivity

establish that these models possess at least some of the properties of real-world networks we might hope to model. The pioneering work of Caron and Fox yields further evidence.

The Kallenberg exchangeable graph model is a natural generalization of the (dense) exchangeable graph model: not only does the defining probabilistic symmetry still retain the interpretation that the vertex labels do not carry any information about the structure of the random graph, but graphons, which parametrize the exchangeable graphs, correspond with compactly-supported graphexes. There are many deep results in the graphon theory for which it is desirable to find sparse graph analogues. Several immediate goals worth pursuing are: identifying the sampling scheme that gives rise to KEGs; finding consistent estimators for a graphex, and identifying their properties; and determining the graph limit theory corresponding to graphexes and its connection with existing graph limit theories for sparse graph sequences. We now discuss these three directions in more detail.

A basic missing piece preventing us from confidently applying KEGs to real-world network data is a characterization of the processes that they model. In particular, consider the problem of studying the properties of a very large graph by sampling a small subgraph according to some random sampling design. Clearly any particular design licenses certain inferences and may even prevent others. In this case the natural question is: what sampling schemes for subgraphs give rise to KEGs? It is well understood that a size-n (dense) exchangeable graph model corresponds to the process of observing the subgraph induced on n vertices sampled uniformly at random from a large (even continuum-sized) graph. One can see this interpretation in the work of Kallenberg [Kal99] and the later independent work within graph theory, beginning with Lovász and Szegedy [LS06]. The generative process for a KEG suggests the following sampling scheme for a finite graph H corresponding to a KEG restricted to $[0, \nu]$:

(1) Sample a Poisson number N of vertices uniformly at random with replacement from H, where the mean of N is $c\,\nu$.
(2) Return the induced edge set, implicitly dropping isolated vertices.

The corresponding graphex is $(0, 0, c \cdot H)$ where $c \cdot H$ denotes the c-dilation of the empirical graphon associated with the finite graph H. (See Section 3.1.) The norm of the dilation is $\|c \cdot H\|_1 = c^2 \|H\|_1$, which we expect to approach zero as the graph H becomes increasingly sparse. This suggests normalizing, by taking the dilation c to be proportional to $\|H\|_1^{-\frac{1}{2}}$. Such a renormalization bears some resemblance to that of the L^p theory discussed below, and is likely to feature in a graph limit theory. This sampling scheme immediately suggests a notion of an empirical graphex, which one would expect to feature prominently in an estimation theory. Identifying other sampling scheme(s) would provide both a sharp understanding of the applicability of our models and substantive guidance on how to subsample large networks.

In the absence of theoretical guidelines to the applicability of the KEG model, a pragmatic approach is to simply fit KEG models to data and assess their appropriateness by empirical evaluations, e.g., of their predictive performance. In practice, this entails identifying classes of KEGs that both admit computationally tractable inference procedures and are flexible enough to capture the structure of real-world networks. The first step in this direction was taken by Caron and Fox [CF14] with Bayesian non-parametric models defined in terms of products of completely random measures. The carefully crafted structure of their model allowed them to develop an

efficient Markov Chain Monte Carlo algorithm to fit their model to sparse graph data comprised of tens of thousands of vertices. More recently, [HSM15] have extended the work of Caron and Fox to obtain an analogue of the well-known stochastic block model. The analogue is easily seen to also be a KEG. Going forward, the close connection between graphexes and graphons suggests that many of the existing models in the (dense) exchangeable graph framework will have natural analogues in KEG framework. This includes many popular models in the literature, e.g., [NS01; HRH02; ABFX08; MGJ09; LOGR12]; see [OR15] for a review.

Finally, it is interesting to consider the connection with graph limit theory. There are at least two distinct contexts in which graphons arise: First, as we have already described in detail, is as the structures characterizing the extreme elements among the exchangeable graphs. Second, is as the limit objects for dense graph sequences [LS06; LS07; Lov13]. The connection between the two perspectives is explained by [DJ08]. The focus of the present paper is the generalization of the first perspective to the sparse regime. Recent work [BCCZ14a; BCCZ14b] has generalized the limit theory to the sparse regime by introducing a new notion of convergence and class of limit objects called L^p graphons, which are symmetric integrable functions $W : [0,1]^2 \to \mathbb{R}_+$. The corresponding W-sparse random graph model is not projective, in contrast to the Kallenberg exchangeable graph model. Understanding the link between the graphex theory and the L^p graphon theory could provide new insights in both graph theory and the statistical analysis of networks.

Acknowledgements

The authors would like to thank Nate Ackerman, Cameron Freer, Benson Joeris, and Peter Orbanz for helpful discussions. The authors would also like to thank Mihai Nica for suggesting the proof of Lemma 7.9. This work was supported by U.S. Air Force Office of Scientific Research grant #FA9550-15-1-0074.

References

[ABFX08] E. M. Airoldi, D. M. Blei, S. E. Fienberg, and E. P. Xing. *Mixed Membership Stochastic Blockmodels.* Journal of machine learning research : JMLR 9 (Sept. 2008), pp. 1981–2014.

[AG89] R. Arratia and L. Gordon. *Tutorial on large deviations for the binomial distribution.* English. Bulletin of Mathematical Biology 51.1 (1989), pp. 125–131.

[Ald81] D. J. Aldous. *Representations for partially exchangeable arrays of random variables.* J. Multivariate Anal. 11.4 (1981), pp. 581–598.

[Ald85] D. J. Aldous. "Exchangeability and Related Topics". In: *École d'Été de Probabilités de Saint-Flour XIII - 1983.* Ed. by P. L. Hennequin. Lecture Notes in Mathematics 1117. Springer, 1985, pp. 1–198.

[BA99] A.-L. Barabási and R. Albert. *Emergence of Scaling in Random Networks.* Science 286.5439 (1999), pp. 509–512. eprint: http://www.sciencemag.org/content/286/5439/509.full.pdf.

[BBCS14] N. Berger, C. Borgs, J. T. Chayes, and A. Saberi. *Asymptotic behavior and distributional limits of preferential attachment graphs.* ArXiv e-prints (Jan. 2014). arXiv: 1401.2792 [math.PR].

[BC09] P. J. Bickel and A. Chen. *A nonparametric view of network models and Newman-Girvan and other modularities*. Proceedings of the National Academy of Sciences 106.50 (2009), pp. 21068–21073. eprint: http://www.pnas.org/content/106/50/21068.full.pdf.

[BCCG15] C. Borgs, J. T. Chayes, H. Cohn, and S. Ganguly. *Consistent nonparametric estimation for heavy-tailed sparse graphs*. ArXiv e-prints (Aug. 2015). arXiv: 1508.06675 [math.ST].

[BCCZ14a] C. Borgs, J. T. Chayes, H. Cohn, and Y. Zhao. *An L^p theory of sparse graph convergence I: limits, sparse random graph models, and power law distributions*. ArXiv e-prints (Jan. 2014). arXiv: 1401.2906 [math.CO].

[BCCZ14b] C. Borgs, J. T. Chayes, H. Cohn, and Y. Zhao. *An L^p theory of sparse graph convergence II: LD convergence, quotients, and right convergence*. ArXiv e-prints (Aug. 2014). arXiv: 1408.0744 [math.CO].

[BCS15] C. Borgs, J. T. Chayes, and A. Smith. *Private Graphon Estimation for Sparse Graphs*. ArXiv e-prints (June 2015). arXiv: 1506.06162 [math.ST].

[BJR07] B. Bollobás, S. Janson, and O. Riordan. *The phase transition in inhomogeneous random graphs*. Random Struct. Alg. 31.1 (2007), pp. 3–122.

[Bol01] B. Bollobás. *Random Graphs*. Cambridge University Press, 2001.

[BR07] B. Bollobás and O. Riordan. *Metrics for sparse graphs* (Aug. 2007). eprint: 0708.1919v3.

[CAF15] D. Cai, N. Ackerman, and C. Freer. *Priors on exchangeable directed graphs*. ArXiv e-prints (Oct. 2015). arXiv: 1510.08440 [math.ST].

[CF14] F. Caron and E. B. Fox. *Sparse graphs using exchangeable random measures*. ArXiv e-prints (Jan. 2014). arXiv: 1401.1137 [stat.ME].

[CSKM13] S. N. Chiu, D. Stoyan, W. S. Kendall, and J. Mecke. *Stochastic Geometry and Its Applications*. 3rd ed. Wiley, 2013.

[DJ08] P. Diaconis and S. Janson. *Graph limits and exchangeable random graphs*. Rendiconti di Matematica, Serie VII 28 (2008), pp. 33–61. eprint: 0712.2749.

[Dur06] R. Durrett. *Random Graph Dynamics*. Cambridge University Press, 2006.

[Fin30] B. de Finetti. *Funzione caratteristica di un fenomeno aleatorio*. Atti Reale Accademia Nazionale dei Lincei VI.4 (1930), pp. 86–133.

[Fin37] B. de Finetti. *La prévision: ses lois logiques, ses sources subjectives*. Ann. Inst. H. Poincaré 7.1 (1937), pp. 1–68.

[Gly87] P. W. Glynn. *Upper bounds on Poisson tail probabilities*. English. Operations Research Letters 6.1 (1987), pp. 9–14.

[Hoo79] D. N. Hoover. *Relations on probability spaces and arrays of random variables*. Tech. rep. Institute of Advanced Study, Princeton, 1979.

[HRH02] P. D. Hoff, A. E. Raftery, and M. S. Handcock. *Latent Space Approaches to Social Network Analysis*. Journal of the American Statistical Association 97.460 (Dec. 2002), pp. 1090–1098.

[HS55] E. Hewitt and L. J. Savage. *Symmetric Measures on Cartesian Products*. Trans. Am. Math. Soc. 80.2 (1955), pp. 470–501.

[HSM15] T. Herlau, M. Schmidt, and M. Mørup. *Completely random measures for modelling block-structured networks*. ArXiv e-prints (July 2015). arXiv: 1507.02925 [stat.ML].

[Kal01] O. Kallenberg. *Foundations of Modern Probability*. 2nd. Springer, 2001.

[Kal05] O. Kallenberg. *Probabilistic Symmetries and Invariance Principles*. Springer, 2005.

[Kal90] O. Kallenberg. *Exchangeable random measures in the plane*. English. Journal of Theoretical Probability 3.1 (1990), pp. 81–136.

[Kal99] O. Kallenberg. *Multivariate sampling and the estimation problem for exchangeable arrays*. J. Theoret. Probab. 12.3 (1999), pp. 859–883.

[Ker14] D. Kerr. *Ergodic Theory: Independence and Dichotomies*. 2014. URL: http://www.math.tamu.edu/~kerr/book/ (visited on 10/28/2015).

[Kin93] J. F. C. Kingman. *Poisson Processes*. Oxford University Press, 1993.

[LOGR12] J. R. Lloyd, P. Orbanz, Z. Ghahramani, and D. M. Roy. "Random function priors for exchangeable arrays". In: *Adv. Neural Inform. Process. Syst. (NIPS) 25*. 2012, pp. 1007–1015.

[Lov13] L. Lovász. *Large Networks and Graph Limits*. American Mathematical Society, 2013.

[LS06] L. Lovász and B. Szegedy. *Limits of dense graph sequences*. J. Combin. Theory Ser. B 96 (2006), pp. 933–957.

[LS07] L. Lovász and B. Szegedy. *Szemerédi's Lemma for the Analyst*. Geom. Func. Anal. 17.1 (2007), pp. 252–270.

[MGJ09] K. T. Miller, T. L. Griffiths, and M. I. Jordan. "Nonparametric latent feature models for link prediction". In: *Adv. Neural Inform. Process. Syst. (NIPS) 20*. 2009, pp. 1276–1284.

[New09] M. Newman. *Networks. An Introduction*. Oxford University Press, 2009.

[NS01] K. Nowicki and T. A. B. Snijders. *Estimation and prediction for stochastic blockstructures*. J. Amer. Statist. Assoc. 96.455 (2001), pp. 1077–1087.

[OR15] P. Orbanz and D. Roy. *Bayesian Models of Graphs, Arrays and Other Exchangeable Random Structures*. Pattern Analysis and Machine Intelligence, IEEE Transactions on 37.2 (2015), pp. 437–461.

[WO13] P. J. Wolfe and S. C. Olhede. *Nonparametric graphon estimation*. ArXiv e-prints (Sept. 2013). arXiv: 1309.5936 [math.ST].

University of Toronto, Department of Statistical Sciences, Sidney Smith Hall, 100 St George Street, Toronto, Ontario, M5S 3G3, Canada