(SPARSE) EXCHANGEABLE RANDOM GRAPHS

VICTOR VEITCH

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy Graduate Department of Statistical Sciences University of Toronto

ABSTRACT

The common statistical models for network valued datasets (implicitly) assume vertex exchangeability, a multi-dimensional analogue to de Finetti style exchangeability that provides a natural formalization of the requirement that the labels of a random graph carry no information about its structure. This seemingly innocuous assumption implies that the corresponding networks are densely connected: a property that rarely holds for real-world networks of practical interest. This thesis develops the statistical theory for network valued data; in particular, generalizing the dense exchangeable models to the sparse graph regime. The key ingredient is a novel notion of exchangeability for graphs based on a point process representation of networks. The key main results are: The construction of a new modeling framework through a de Finetti style representation theorem. The characterization of several important sampling distribution properties of the resulting random graphs. The introduction of a generic non-parametric estimator (analogous to the empirical measure) for these modes. And, finally, the recasting of the core exchangeability concepts in the language of sampling, which leads to a clear articulation of the sampling design naturally associated with the models, as well as a unification with the theory of limits of large (sparse) graphs.

ATTRIBUTION

The content of this thesis covers several papers developed in collaboration with other researchers, and the results herein were possible only through the effort, insight, and occasional genius of everyone involved.

Chapters 2 and 3 cover work with my advisor Daniel M. Roy on the mathematical foundations of statistical inference for network data. Chapter 2 corresponds to [VR15], posted to the arXiv in 2015, and Chapter 3 corresponds to [VR16], posted to the arXiv in 2016.

Chapter 4 covers work I did with Christian Borgs, Jennifer Chayes, and Henry Cohn at Microsoft Research New England during an internship. At the time of writing, this material is not yet otherwise publicly available. However, I anticipate this material will appear shortly in substantially the same form that it is presented in this thesis.

As far as I can see,
all a priori statements in physics have their origin in symmetry.

— H. Weyl

ACKNOWLEDGEMENTS

My time in Toronto has been both immensely enjoyable and productive. I have benefited enormously from the both keen insights and companionship of my colleagues and fellow students—these include Wei Deng, Bill Huang, Alex Shestopaloff, and many others.

I am grateful to my advisor Daniel Roy, who has gone above and beyond on every occasion. His mentorship and support have been enormously valuable, as has his sustained enthusiasm for really deeply understanding the problems at hand. The work we have done together has turned me into a much stronger researcher. It's been a real joy to be his first student.

I'd also like to thank my friends Benson and Isabel for many fond memories, quite a bit of advice on graph theory, and an awful lot of board games.

Finally, I'd like to thank my family and my partner for their unwavering love and support.

CONTENTS

1 INTRODUCTION

The topic of this thesis is the mathematical foundations for statistical modeling of relational or network data. In this setting, the dataset g is modeled as a graph, idealized as a sample drawn from some probability distribution over graphs, $P(g;\theta)$, whose parameters are in some unknown configuration. The goal is to infer the configuration of the parameters on the basis of the observation, and thus understand the properties of the observed network. In this thesis, we substantially develop the theory for the statistical analysis of network data. Along the way, we uncover some interesting insights about the foundations of statistical inference generally; particularly, on the role of probabilistic symmetries, sampling, empirical measures, and the interplay between these subjects.

The first main challenge is to identify families of probability distributions over graphs that are suitable for modeling network data. A satisfactory family of distributions should, minimally, meet the following basic desiderata:

1. graphs drawn from the distributions should be able to capture the rich structure of real-world networks of interest

2. it should be possible to recover the parameters of the observed network on the basis of the observed data

Remarkably, there is no general approach to the statistical modeling of network data that satisfies these requirements. More precisely, there is already well developed and powerful theory—the *graphon* framework—for the analysis of densely connected networks (that is, networks that contain a constant fraction of all possible edges as the size becomes large), but there is no comparable theory for sparse (that is, not dense) networks. The vast majority of real-world networks are sparsely connected, so the absence of a satisfactory modeling framework is a considerable omission. This thesis addresses the gap, generalizing the graphon framework to the sparse graph setting.

The most popular approaches to statistical network modeling, e.g. [NS01; HRH02; ABFX08; MJG09; LOGR12], fall under the remit of the graphon framework for dense random graphs. Graphon models are the class of generative models for random graphs where each vertex i is assigned some independent latent features ϑ_i, and, conditional on these latent features, each pair of vertices i, j is connected by an edge independently with probability $W(\vartheta_i, \vartheta_j)$ determined by the latent features of i and j; the function W is known as a graphon. This is a

very natural class of models, and models of this type, such as stochastic block models and latent feature models, have a long history in the literature. However, as already alluded to, graphon models give rise to random graphs that are almost surely *dense* (or trivially empty if the graphon is $W = 0$ almost everywhere). Thus, as stated plainly in [OR15], graphon models are misspecified as statistical models for most real-world data.

A key to arriving at the generalized models of this thesis is a second perspective on graphon models rooted in probabilistic symmetry. In this view, we identify a projective family $(G_n)_{n\in\mathbb{N}}$ of random graphs with the upper left $n \times n$ submatrices of an infinite random adjacency matrix A, and then define the class of models to be those such that the distribution of A is invariant under joint permutations of its rows and columns. This *exchangeability* of A is a natural formalization of the requirement that the labels of the vertices of a random graph should be uninformative about the structure of the graph. The fact that the graphon models are the models defined by exchangeability of the infinite adjacency matrix is, essentially, the content of the celebrated Aldous–Hoover theorem [Ald81; Hoo79].

This 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. We follow this broad strategy.

Our inspiration comes from recent paper of Caron and Fox [CF14] that exploits a bijection between random graphs and point processes on $\mathbb{R}_+^2$ to exhibit a class of sparse random graphs. In their paper, they observe that their random graphs satisfy a natural notion of exchangeability exchangeability when considered as a point process, and make use of an associated representation theorem to study the model. We reverse 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.

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. 4). Accordingly, each edge is identified with a label $(\theta_i, \theta_j) \in \mathbb{R}_+^2$, and each vertex with a label $\theta_i \in \mathbb{R}_+$. The representation theorem associated to exchangeable 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 distribution of every random graph satisfying the symmetry is characterized by a triple $\mathcal{W} = (I, S, W)$ where $I \in \mathbb{R}_+$, $S : \mathbb{R}_+ \to \mathbb{R}_+$ is a measurable function such that $\min(S, 1)$ is integrable, and $W : \mathbb{R}_+^2 \to [0, 1]$ is a symmetric measurable function

satisfying certain weak integrability conditions. (See Theorem 2.4.9; W integrable is sufficient but not necessary.) The associated random graphs have three possible components: isolated edges, governed by I, infinite stars, governed by S, and a final piece that provides the interesting graph structure, governed by $\mathcal{W}$.

We give the probability distribution $P(\Gamma; \mathcal{W})$ over (infinite) random graphs Γ associated with each graphex in terms of the associated generative model. See Fig. 1. Let $\theta = \mathbb{R}_+$ be the space of labels of the graph, $\vartheta = \mathbb{R}_+$ be the space of latent parameters, and $\Pi = \{\theta_i, \vartheta_i\}_{i\in\mathbb{N}}$ be a unit rate Poisson process on $\theta \times \vartheta$. For all $i < j$, conditioned on the points of Π, the random graph includes the edge $\theta_i \leftrightarrow \theta_j$ with probability $W(\vartheta_i, \vartheta_j)$, independently of all other edges. This is the interesting component of the graph structure. The infinite stars component is given by, for each point (θ_i, ϑ_i), including edges $\{(\theta_i, \sigma_{ij})\}$, where $\{\sigma_{ij}\}_{j\in\mathbb{N}}$ is distributed as a rate $S(\vartheta_i)$ Poisson process independent of everything else. The isolated edges component is given by including edges $\{(\rho_k, \rho'_k)\} \subset \theta^2$ where $\{(\rho_k, \rho'_k)\}$ are distributed as a rate I Poisson process independent of everything else. To pass from an infinite random graph Γ with vertices labeled in $\mathbb{R}_+$ to a projective sequence $(\Gamma_s)_{s\in\mathbb{R}_+}$ we define Γ_s to be the (labeled) graph given by restricting to vertices with labels $\theta_i < s$, and removing any vertices that are isolated in the induced subgraph. We refer to Γ as a Kallenberg exchangeable graph, overloading this notation to also refer to the family $(\Gamma_s)_{s\in\mathbb{R}_+}$

The contribution of this thesis may now be stated plainly: we introduce and develop the graphex framework for the statistical modeling of network and relational data.

The main content of the thesis is organized into three chapters on different aspects of the theoretical development. The three chapters correspond to the papers [VR15], [VR16] and [BCCV17] respectively. In each case, the content of the chapters very closely follows the associated papers—the exceptions are minimal edits to ensure consistent notation and ease the flow of reading (by, e.g., not recapitulating the representation theorem three times).

The Class of Random Graphs Arising from Exchangeable Random Measures

In Chapter 2, we use the representation theorem to derive the graphex framework and we argue that the models of the framework are good candidates for the statistical modeling of real-world data. We first explain in detail the core idea of defining model classes by probabilistic symmetries, in particular, giving due attention to the key role of projectivity and the distinction between finite and infinite exchangeability. We then apply Kallenberg's representation theorem to derive the representation theorem for sparse exchangeable graphs, thereby defining the graphex framework. It remains to show that the mod-

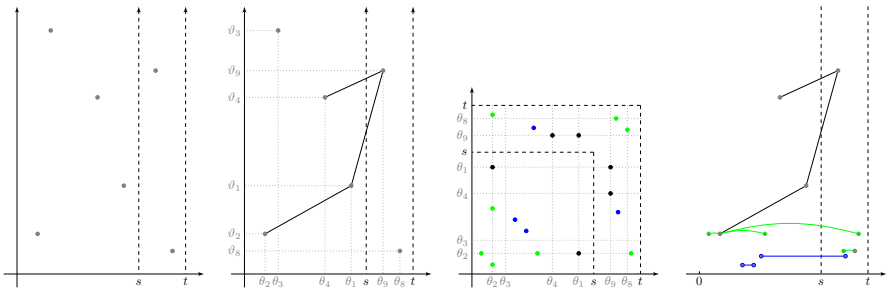

Figure 1: Generative process of a Kallenberg exchangeable graph generated by a graphex $\mathcal{W} = (I, S, W)$ defined on $\mathbb{R}_+$ with Lebesgue measure, observed at sizes s and t. First panel: a (necessarily truncated) realization of the latent Poisson process Π_t on $[0, t] \times \mathbb{R}_+$. A countably infinite number of points lie above the six points visualized. Second panel: Edges due to the graphon component W are sampled by connecting each distinct pair of points $(\theta_i, \vartheta_i), (\theta_j, \vartheta_j) \in \Pi_t$ independently with probability $W(\vartheta_i, \vartheta_j)$. Integrability conditions on W imply that only a finite number of edges will appear, despite there being an infinite number of points in Π_t. Assume the three edges are the only ones. Third panel: The edge set Γ_t represented as an adjacency measure on $[0, t]^2$. The edges in the graphon component appears as (symmetric pairs of) black dots; the edges corresponding to the star component S appear in green; the isolated edges (from the I component) appear in blue. At size s, only the edges in $[0, s]^2$ (inner dashed black line) appear in the graph. The edges $\{\theta_j, \sigma_{jk}\}$ of the star (S) component of the process (green) centered at θ_j are realizations of a rate-$S(\vartheta_j)$ Poisson process $\{\sigma_{jk}\}$ along the line through θ_j (show as green dots along grey dotted lines). Hence, at size t, each point θ_i is the center of $\mathrm{Poi}(t\, S(\vartheta_i))$ star process rays. The edges $\{\rho_i, \rho'_j\}$ generated by the isolated edge (I) component of the process (blue) are a realization of a rate-I Poisson process on the upper (or lower) triangle of $[0, t]^2$, reflected. At size t, there are $\mathrm{Poi}(t^2\, I)$ isolated edges due to this part of the graphex. The final panel shows the graphs corresponding to the sampled adjacency measure at sizes s and t.

els of graphex framework can capture the complex structure of real-world networks. Our first contribution to that end is to show that the dense graphon models are (essentially) contained as a special case of the graphex models. Next, we derive some sampling distribution properties of $(\Gamma_s)_{s\in\mathbb{R}_+}$ in terms of the generating graphex. This allows us to show that graphex models can indeed give rise to sparse graphs and, moreover, that they encompass much of the rich structure required for high fidelity modeling; e.g. they naturally encompass models with extreme degree heterogeneity ("power law" degree distributions), and very short typical path lengths even in the sparse regime ("small world" behaviour).

Sampling and Estimation for (Sparse) Exchangeable Graphs

In Chapter 3, we identify the sampling scheme that is naturally associated with the models of the framework, and we introduce a general consistent estimator for the graphex. The sampling scheme is a modification of independent vertex sampling that throws away vertices that are isolated in the sampled subgraph. The estimator is the analogue of the empirical measure of classical i.i.d. statistics. The estimator is the dilation of the empirical graphon estimator, which is already known to be a consistent estimator in the dense graphon setting.

The results of this chapter are important for a number of reasons: Identifying the sampling scheme associated to the graphex models helps clarify the meaning and applicability of the models. Finding the analogue of the empirical measure in this setting opens the door to the development of non-parametric techniques for graphs. Less obviously, the machinery developed in order to formalize the estimation problem is critical for any further statistical theory; for example, pinning down the right notions of consistent estimation and identifiability in this setting is a subtle problem that is now resolved.

It also worth emphasizing that sampling and estimation and extremely tightly connected. One of the key ideas in the development of our estimator is that a sample from a very large random graph Γ_s generated by $\mathcal{W}$ looks approximately like a (smaller) random graph drawn from $\mathcal{W}$ directly, even conditional on Γ_s. This is the network data analogue of the familiar idea that one can approximate samples from probability measure simply by subsampling a very large dataset. The validity of this approach, in both the sequence and network settings, relies critically on the exchangeability assumption—this is one way in which the tight connection between sampling and exchangeability manifests. Making this precise is one of the major contributions of this chapter.

We also note that Svante Janson has written a followup paper [Jan17] to [VR16] (the original material of this chapter), substantially improving the details of the proofs (and, consequently, some of the presen-

tations), as well as somewhat tightening the results. In particular, the consistent estimation result of Chapter 3 works for observations taken at some countable sequence of sizes $s_1, s_2, \ldots$, [Jan17] improves the result to hold for observations for continuum s. Additionally, the consistent estimation result here only holds in probability when the true underlying graphex is non-integrable; Janson improves this to almost sure convergence in all cases.

Sampling Perspectives on (Sparse) Exchangeable Graphs

Chapter 4 addresses two closely related questions: When is it appropriate to model a network dataset using the graphex framework? And, what is the relationship in this setting between statistical modeling and limits of large sparse graph sequences?

The context for the later question is that, in parallel to [VR15], a group of graph theorists independently developed the graphex framework in the context of graph limit theories [BCCH16]. In the graph limit setting, the central problem is: what is a sensible notion of convergence for a sequence $G_1, G_2, \ldots$ of (non-random) graphs of growing size, and what is the natural limit object? In the dense graph setting, many superficially distinct notions of convergence turn out to coincide, with graphons as the natural limit object. See [Lov13] for a textbook level review. Indeed, the dense graphon models (and the term graphon) were developed by graph theorists independently of the statistical work on exchangeable random graph models. The two perspectives were unified in [DJ08; Aus08]; part of the contribution of Chapter 4 is to repeat this unification in the sparse graph setting.

Very informally, we define a new notion of graph limit—*sampling convergence*—by saying that a graph sequence $G_1, G_2, \ldots$ converges if randomly sampled subgraphs converge in distribution. More precisely, for each G_k we sample H_k^s by including each vertex of G_k independently with probability $\frac{s}{\sqrt{2e(G_k)}}$, and returning the induced subgraph with the isolated vertices removed. The reason for picking each vertex with a probability proportional to $1/\sqrt{2e(G_k)}$ is that this keeps the expected number of edges of H_k^s constant as k grows. We say the sequence $G_1, G_2, \ldots$ is sampling convergent if H_k^s converges in distribution for all $s \in \mathbb{R}_+$. The main result is that whenever a sequence converges in this sense, the (weak) limit of $\Pr(H_k^s \in \cdot)$ is equal to the distribution of a size-s sample from some integrable graphex $\mathcal{W}$. Thus, integrable graphexes are the natural limit objects of sampling convergence. This gives another close connection between sampling and exchangeability (in fact, our proof of this result makes explicit use of Kallenberg's representation theorem). The connection to [BCCH16] is given by our demonstration that convergence in the sense of [BCCH16] implies sampling convergence.

We further show that (at least up to some natural equivalencies) every integrable graphex $\mathcal{W}$ arises as the limit of some sampling convergent graph sequence. This later observation has the nice consequence that we can use the limit theory to define the graphex models without any explicit appeal to the rather strange representation of graphs as point processes; it thus sidesteps questions about the interpretation of this representation and the associated notion of exchangeability. Instead, we can use the limit theory to formalize the idea that a graphex distribution is appropriate for modelling a network that can be thought of as generated by independently choosing nodes from a much larger network, and then throwing away isolated nodes. This idea was also at least implicit in the construction of the estimator of Chapter 3.

This last observation begs the question of whether we could have defined the graphex models in terms of the sampling scheme in the first place, without any appeal to either exchangeability or graph limits. Indeed, this is the case. In Chapter 3 we show that graphex models have the property that if a graph $H_{s,p}$ is generated by first generating a size-s graph Γ_s according to $\mathcal{W}$ and then independently including each vertex of Γ_s with probability p, and finally throwing away the isolated vertices in the sample, then (ignoring labels) $H_{s,p}$ is distributed as a size ps graph generated from $\mathcal{W}$. In Chapter 4 we show that this is a defining property of the graphex models. In particular, this sampling invariance is exactly equivalent to the exchangeability we use to define the graphex models.

2

THE CLASS OF RANDOM GRAPHS ARISING FROM EXCHANGEABLE RANDOM MEASURES

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].

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. It is a remarkable gap in modern statistical practice that there is no general framework for the statistical analysis of real-world networks.

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 chapter, 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 models admit a simple, tractable specification and give rise to complex structures of the kinds observed in real-world networks.

Models from invariance principles

The idea of that invariance principles yield models is an old one. For instance, the classical i.i.d. model and the graphon model for densely connected networks are both derived from 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 chapter may be broadly useful for the statistical analysis of real-world networks.

To explain our approach, we begin by reviewing the closely related approach for modeling 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 symmetric array of binary random variables. The infinite structure automatically provides consistent models for datasets of different size, and captures the statistical assumptions that, while we may be in possession of size-n sample of data, a larger size sample *could have*, in principle, been collected, but the size-n subsample would have been unaffected. The fundamental 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 speaking, exchangeability encodes the assumption that the labeling of the vertices of a graph do not carry information about its structure.

The dense graph 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.

Sparse graph models from exchangeability

This subsection repeats the construction of the graphex models described in the introduction of the thesis, and may be safely skipped by a reader who feels comfortable with that construction.

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 chapter 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 chapter 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. 4). 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 a measurable function such that $\min(S, 1)$ is integrable, and $W : \mathbb{R}_+^2 \to [0, 1]$ is a symmetric measurable function satisfying certain weak integrability conditions. (See Theorem 2.4.9; W integrable is sufficient but not necessary.) We call the triple a *graphex*. In this chapter we focus on random graphs without isolated edges or infinite stars, and so we take $I = S = 0$; We will overload the

term *graphon* to refer to W, using *Kallenberg graphon* when we need to emphasize that we are working with the new, generalized notion of graphon.

The distribution of every such random graph, which we call a Kallenberg exchangeable graph, is characterized by some (possibly random) graphex. Graphexes are the sparse graph analogues of graphons in the dense graph setting, 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. For simplicity, we restrict to the $(0, 0, W)$ case here. Let $\theta = \mathbb{R}_+$ be the space of labels of the graph, $\vartheta = \mathbb{R}_+$ be the space of latent parameters, and Π be a unit rate Poisson process on $\theta \times \vartheta$ with points $\{\theta_i, \vartheta_i\}_{i\in\mathbb{N}}$. For all $i < j$, conditioned on the points of Π, the random graph includes the edge $\theta_i \leftrightarrow \theta_j$ with probability $W(\vartheta_i, \vartheta_j)$, independently of all other edges. The vertices of the random graph are then taken to be the subset of those points $\{\theta_i\}$ that participate in at least one edge. The construction of the random graph is depicted in Fig. 2. One can then construct a nested sequence of (a.s. finite) random subgraphs: the random subgraph of size $s \geqslant 0$ is the graph containing those edges (θ_i, θ_j) such that $\theta_i, \theta_j < s$, with a vertex included if and only if it appears in at least one such edge. By construction, the distributions of these random graphs are consistent with one another by marginalization. We show later that one can also obtain sparse graph sequences from this model, allowing us to build more realistic models physical networks. Moreover, in a sense we make precise in Section 2.3.1, the exchangeable graphs derived from the Aldous–Hoover theory are contained as a subfamily of the Kallenberg exchangeable graphs, and correspond to 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.

Sampling distribution properties

Let Γ_s be the random graph given by truncating the label space θ to $[0, s]$ (see Fig. 2); we call the random graph model $(G_s)_{s\in\mathbb{R}_+}$ the *Kallenberg exchangeable graph* (KEG) associated with W. The bulk of the present chapter is devoted to deriving properties of these graphs in terms of the Kallenberg graphon W. For simplicity of presentation we ignore self edges here, giving full statements in the body of the chapter. Let $\mu_W(x) = \int_{\mathbb{R}_+} W(x, y)\, dy$.

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

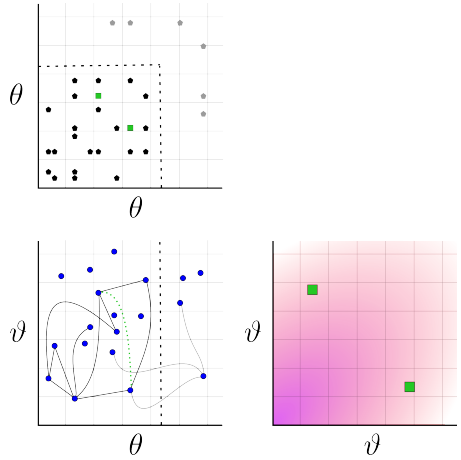

Figure 2: (Kallenberg exchangeable graph) Random graphs arising from exchangeable random measures are characterized by three (potentially random) components: a non-negative real $I \in \mathbb{R}_+$, a measurable function $S : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\min(S, 1)$ is integrable, and a 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 θ is generated in this model according to:

1. Sample a (latent) unit rate Poisson process Π on $\theta \times \vartheta$.
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 θ 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 $\theta \times \theta$ space.

2. The expected number of edges $e_s = |e(\Gamma_s)|$ is

$$\mathbb{E}[e_s] = \frac{1}{2}s^2 \iint_{\mathbb{R}_+^2} W(x,y)\,dx\,dy. \tag{2.1.1}$$

3. The expected number of vertices $v_s = |v(\Gamma_s)|$ is

$$\mathbb{E}[v_s] = s\int_{\mathbb{R}_+} (1 - e^{-s\mu_W(x)})\,dx. \tag{2.1.2}$$

4. Subject to some technical conditions, the scaling limit of the asymptotic degree distribution has an explicit expression in terms of W. Let k_s be some non-decreasing function of s and let D_s be the degree of a randomly selected vertex of Γ_s. Then

$$P(D_s \geqslant k_s \mid \Gamma_s) \xrightarrow{p} \lim_{s\to\infty} \frac{\sum_{k=k_s}^{\infty} \frac{s^k}{k!} \int \mu_W(x)^k e^{-s\mu_W(x)}\,dx}{\int_{\mathbb{R}_+} (1 - e^{-s\mu_W(x)})\,dx}. \tag{2.1.3}$$

   We use this result to show that our random graph construction can give rise to sparse graphs.

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

$$\lim_{s\to\infty} P(|C_1(\Gamma_s)| > (1-\epsilon)\,|v(\Gamma_s)|) = 1. \tag{2.1.4}$$

   Therefore, sparse structure can arise in an interesting way, and not simply as a consequence of there being a collection of disjoint dense graphs (although this is possible to model as well).

Overview

We begin by giving background on random graph models and the role of probabilistic symmetry in Section 2.2. In Section 2.3, we give a number of illustrative examples of Kallenberg exchangeable graphs. In Section 2.4, we establish the representation theorem and give a formal characterization of the models we derive. In Section 2.5 we give several theorems characterizing the properties of the generative models: we derive the first moments of several graph statistics of Γ_s using point process techniques, allowing self edges; we derive an expression for asymptotic degree distribution of these graphs in terms of W; and, we study the structure of the Kallenberg exchangeable graphs generated by graphexes of the form $(I, S, W) = (0, 0, f(x)f(y)\mathbb{1}[x \neq y])$ 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.

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 chapter, 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 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 s that is proportional to the square root of the *expected* number of *edges*.

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 $G_{s,\phi}$,

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.

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,

$$\frac{\sqrt{|e(G)|}}{|v(G)|} \to 0 \qquad \text{as } n \to \infty. \tag{2.2.1}$$

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 chapter, 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 $(f_{s,t};\, s \leqslant t \in S)$ where $f_{s,t}$ maps graphs of size t to graphs of size $s \leqslant t$, $f_{s,s}$ is the identity, and $f_{r,t} = f_{r,s} \circ f_{s,t}$ for all $r \leqslant s \leqslant t$. A random graph model is *projective* if, for some projective system $(f_{s,t};\, s \leqslant t \in S)$, it holds that $G_{s,\phi} \overset{d}{=} f_{s,t}(G_{t,\phi})$ 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 $(f_{s,t};\, s \leqslant t \in S)$, 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. 2 and 3 for illustrations). The (dense) 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, $\Gamma_{s,W} \overset{d}{=} \Gamma_{t,W}(\cdot \cap [0,s]^2)$ for all $s, t \in \mathbb{R}_+$.

The projectivity of the KEG model sets it apart from random graph models that achieve sparsity by 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 an integer 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 allows 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.

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 and standard references.)

Consider the classical setting of statistical inference: a sequence of real-valued measurements $x_1, \dots, x_n$ 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 $\mu_1, \mu_2, \dots$ 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 $X_1, X_2, \dots$ of random variables such that μ_n is the distribution of $(X_1, \dots, X_n)$. Therefore, we are modeling observed length-n sequences $(x_1, \dots, x_n)$ as realizations of prefixes $(X_1, \dots, X_n)$ of the infinite random sequence $(X_1, X_2, \dots)$. 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 $X_1, X_2, \dots$ is *exchangeable*, i.e.,

$$(X_1, \dots, X_n) \overset{d}{=} (X_{\sigma(1)}, \dots, X_{\sigma(n)}) \tag{2.2.2}$$

for every $n \in \mathbb{N}$ and every permutation σ of $[n] = \{1, \dots, n\}$, then, by de Finetti's representation theorem [Fin30; Fin37; HS55], the random variables $X_1, X_2, \dots$ 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

$$M \sim \mathcal{P} \tag{2.2.3}$$

$$X_1, X_2, \dots \mid M \overset{iid}{\sim} M. \tag{2.2.4}$$

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

$$\mu(B) = \int_{\mathcal{M}_1(\mathbb{R})} m^\infty(B)\, \mathcal{P}(\mathrm{d}m), \qquad \text{for measurable } B \subseteq \mathbb{R}^\infty. \tag{2.2.5}$$

The distribution μ is uniquely determined by $\mathcal{P}$, and vice versa. From Eq. (2.2.5), 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

$$M(A) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} 1(X_j \in A) \text{ a.s.} \tag{2.2.6}$$

On the other hand, even an infinite realization $(x_1, x_2, \dots)$ gives no information about $\mathcal{P}$. For this reason, in a statistical setting, in addition to assuming that s is an element in the space of distributions of exchangeable sequences, we assume that s is ergodic, i.e., s is an unknown element in the space of distributions of i.i.d. sequences. Since every s 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 attention to a subset $T \subseteq \mathcal{M}_1(\mathbb{R})$ corresponding to a finite dimensional family of distributions, e.g., the normal distributions. Nonparametric models correspond to restricting to families T that do not admit a finite dimensional parameterization.

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 determines a natural parameter (in the case of real-valued exchangeable sequences, a probability measure on $\mathbb{R}$) and gives a generative recipe for the data in terms of this primitive (i.i.d. sampling in this example). It is this later perspective that is paralleled in the derivation of the KEG model.

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 $1 \leqslant i, j \leqslant n$, 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 $(X_{i,j};\, i, j \leqslant n)$. Therefore, we model observed $n \times n$ adjacency matrices $(x_1, \ldots, x_n)$ as realizations of prefixes $(X_{i,j};\, i, j \leqslant n)$ 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

$$(X_{i,j};\, i, j \leqslant n) \stackrel{d}{=} (X_{\sigma(i,j)};\, i, j \leqslant n), \tag{2.2.7}$$

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. (2.2.7) 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

$$(X_{i,j};\, i, j \leqslant n) \stackrel{d}{=} (X_{\sigma(i),\sigma(j)};\, i, j \leqslant n) \tag{2.2.8}$$

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–Hoover 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. 3)

$$W \sim \mu \tag{2.2.9}$$

$$\{U_i\} \overset{iid}{\sim} \text{Uni}[0,1] \tag{2.2.10}$$

$$(X_{ij}) \mid W, U_i, U_j \overset{ind}{\sim} \text{Bernoulli}(W(U_i, U_j)), \tag{2.2.11}$$

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. 2) 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.

Random graphs as random measures

The key ingredient for generalizing the exchangeable graph model is a correspondence between random graphs and symmetric simple

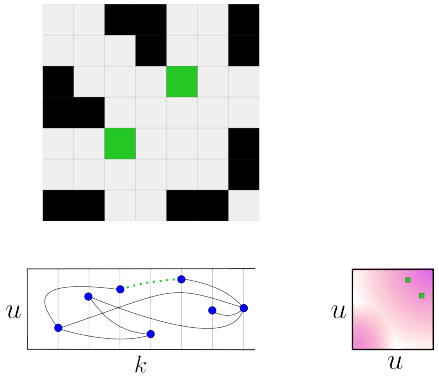

Figure 3: Dense exchangeable random graph model. In the jointly exchangeable array setting a random graph model is characterized by a (potentially random) symmetric measurable function $W : [0,1]^2 \to [0,1]$ called a graphon. An example graphon is depicted as a magenta heatmap (lower right). Conditional on W, a random graph of size n is generated by independently assigning to each vertex $k \in \{1,\ldots,n\}$ 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.

point processes due to Caron and Fox [CF14] (see Fig. 4). 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 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 $\Gamma_t = \Gamma(\cdot \cap [0,t]^2)$, for $t \in \mathbb{R}_+$, of the infinite point process Γ 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 Γ_t is itself a random quantity.

It is important to note that the graph corresponding to the restriction Γ_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,

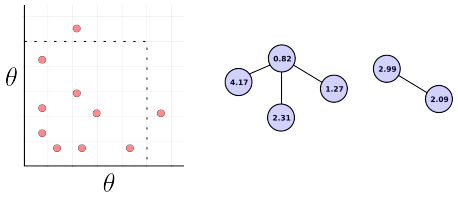

Figure 4: Random graphs as point processes. Random point processes on $\mathbb{R}_+^2$ correspond to infinite random graphs, with finite subgraphs given by restricting the point process to a finite square. Points of the process correspond to graph edges and the vertex structure is deduced from the edge structure. Pictured is a realization of a point process and the realization of the random graph that corresponds to truncating at $\theta = 5$.

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. 2 provides a concrete example of this phenomena.

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

$$\xi \stackrel{d}{=} \xi \circ (f \otimes f)^{-1} \tag{2.2.12}$$

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. 2). 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.

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 $I = S = 0$. We are particularly interested in the sparsity of these graph models. Theorem 2.5.1 establishes that (ignoring self edges) for all random graphs Γ_s generated by graphex $(0, 0, W)$ it holds that $\mathbb{E}[e_s] = \frac{1}{2}s^2\|W\|_1$; i.e., the expected number of edges scales as s^2 in all cases. Intuitively then we expect the sparsity of a random graph model to be determined by $\mathbb{E}[\nu_s] = s\int_{\mathbb{R}_+} 1 - e^{-s\mu_W(x)}\,dx$ (from Theorem 2.5.2, ignoring self edges). This suggests that the slower $\mu_W(x) = \int_{\mathbb{R}_+} W(x, y)\,dy$ decays the sparser the graph will be, an intuition that is borne out by the examples of this section.

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 Kallenberg graphon given by the dilation

$$W(x, y) = \begin{cases} \widetilde{W}(x/c, y/c) & x \leqslant c, y \leqslant c \\ 0 & \text{otherwise.} \end{cases} \tag{2.3.1}$$

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 never participate in the graph and can be discarded. This means that for finite size graph Γ_s given by restricting $\theta \leqslant s$ the relevant underlying process is the unit rate Poisson process on $[0, s] \times [0, c]$. The generative model for the graph can be expressed as:

$$N_s \sim \text{Poi}(c\,s) \tag{2.3.2}$$

$$\{\theta_i\} \mid N_s \overset{iid}{\sim} \text{Uni}[0, s] \tag{2.3.3}$$

$$\{\vartheta_i\} \mid N_s \overset{iid}{\sim} \text{Uni}[0, 1] \tag{2.3.4}$$

$$(\theta_i, \theta_j) \mid \widetilde{W}, \vartheta_i, \vartheta_j \overset{ind}{\sim} \text{Bernoulli}(\widetilde{W}(\vartheta_i, \vartheta_j)). \tag{2.3.5}$$

A little thought shows that this is just a trivial modification of the (dense) graphon model. Instead of indexing the family of graphs by the number of vertices ($\mathbb{N}$) we now index them by the continuous parameter s and have $\text{Poi}(c\,s)$ 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 chapter.

In fact, these are the only dense KEGs arising from (integrable) graphexes: Theorem 2.5.3 shows that G generated by $(0, 0, W)$ is dense iff the generating (integrable) W has compact support.

Slow Decay

We next consider a Kallenberg graphon with tails that go to 0 slowly:

$$W(x,y) = \begin{cases} 0 & x = y, \\ (x+1)^{-2}(y+1)^{-2} & \text{otherwise}, \end{cases} \tag{2.3.6}$$

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 2.5.2

$$\mathbb{E}[\nu_s] = s(\sqrt{\pi}\sqrt{s/3}\operatorname{erf}(\sqrt{s/3}) + e^{-s/3} - 1) \tag{2.3.7}$$

$$\sim \sqrt{\frac{\pi}{3}} s^{3/2},\ s \to \infty. \tag{2.3.8}$$

By Theorem 2.A.3 the number of vertices with degree k has expectation:

$$\mathbb{E}[N_{s,k}] = \frac{s^{k+1}}{k!}(\frac{1}{3})^k \int_1^\infty x^{-2k} e^{-\frac{1}{3}sx^{-2}}\,dx \tag{2.3.9}$$

$$= \frac{s^{k+1}}{k!}(\frac{1}{3})^k \int_0^1 x^{2(k-1)} e^{-\frac{1}{3}sx^2}\,dx \tag{2.3.10}$$

$$= \frac{\Gamma(-\frac{1}{2}+k) - \Gamma(-\frac{1}{2}+k, \frac{s}{3})}{2\sqrt{3}k!} s^{3/2} \tag{2.3.11}$$

$$\sim \frac{\Gamma(-\frac{1}{2}+k)}{2\sqrt{3}k!} s^{3/2},\ s \to \infty. \tag{2.3.12}$$

(Note that Γ here is denoting the incomplete gamma function, and not a Kallenberg Exchangeable Graph.) By Theorem 2.5.4 it follows that the degree D_s of a uniformly selected vertex of Γ_s satisfies

$$\Pr(D_s = k \mid \Gamma_s) \xrightarrow{p} \frac{\Gamma(-\frac{1}{2}+k)}{2\sqrt{\pi}k!},\ s \to \infty, \tag{2.3.13}$$

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

$$\frac{\Gamma(-\frac{1}{2}+k)}{2\sqrt{\pi}k!} \sim k^{-\frac{3}{2}},\ k \to \infty, \tag{2.3.14}$$

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.

Fast Decay

Next we consider a Kallenberg graphon with quickly decaying tails. Let

$$W(x,y) = \begin{cases} 0 & x = y \\ e^{-x}e^{-y} & \text{otherwise.} \end{cases} \tag{2.3.15}$$

Then $\mu(x) = e^{-x}$ and so by Theorem 2.5.2

$$\begin{aligned} \mathbb{E}[\nu_s] &= s\int_{\mathbb{R}_+} 1 - e^{-se^{-x}}\,dx && (2.3.16)\\ &= s\int_0^1 \frac{1}{x}(1-e^{-sx})\,dx && (2.3.17)\\ &= s(\gamma + \Gamma(0,s) + \log(s)) && (2.3.18)\\ &\sim s\log s,\ s\to\infty. && (2.3.19) \end{aligned}$$

As expected, the rapidly decaying Kallenberg grpahon gives rise to a graph that is much more dense than one generated by the slowly decaying Kallenberg graphon.

By Theorem 2.A.3 the number of vertices with degree k has expectation:

$$\begin{aligned} \mathbb{E}[N_{s,k}] &= \frac{s^{k+1}}{k!}\int_0^\infty e^{-kx}e^{-se^{-x}}\,dx && (2.3.20)\\ &= \frac{s}{k!}(\Gamma(k) - \Gamma(k,s)) && (2.3.21)\\ &\sim \frac{s}{k},\ s\to\infty. && (2.3.22) \end{aligned}$$

so that for fixed k only a vanishing fraction of the vertices will have degree k as $s\to\infty$. More precisely, since $\sum_{k=1}^{s^\beta}\frac{s}{k} \sim \beta s\log s,\ s\to\infty$ we have by Theorem 2.5.4 that for $0<\beta<1$

$$P(D_s \leqslant s^\beta) \xrightarrow{p} \beta,\ s\to\infty \tag{2.3.23}$$

where D_s is a random vertex of Γ_s.

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 $\theta \times \theta$ measure they had studied had been a product of completely random measures, that model would have corresponded to a Kallenberg graphon of the form $W(x, y) = f(x)f(y)$. Instead, they actually consider a measure on $\theta \times \theta$ 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 Kallenberg graphon

$$W(x, y) = \begin{cases} 1 - \exp(-g(x)g(y)) & x = y \\ 1 - \exp(-2g(x)g(y)) & x \neq y \end{cases} \tag{2.3.24}$$

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.

REPRESENTATION THEOREM

We now turn to giving formal statements of our construction and proving the representation theorem at the heart of the thesis. 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 2.4.1. An *adjacency measure* is a locally finite symmetric simple measure on $\mathbb{R}_+^2$. The *s-truncation* of an adjacency measure ξ is the adjacency measure $\xi(\cdot \cap [0, s]^2)$ obtained by restricting ξ to $[0, s]^2$.

Definition 2.4.2. 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 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. 4 for an illustration.)

Definition 2.4.3. Let $\xi = \sum_{i<\kappa} \delta_{e_i}$ be an adjacency measure, where $\kappa \in \mathbb{Z}_+ \cup \{\infty\}$ and $e_1, e_2, \ldots$ 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 2.4.4. 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 2.4.5. 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

$$\xi \circ (f \otimes f)^{-1} \stackrel{d}{=} \xi. \tag{2.4.1}$$

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

$$\begin{aligned}
\xi = & \sum_{i,j} f(\alpha, \vartheta_i, \vartheta_j, \zeta_{\{i,j\}}) \delta_{\theta_i, \theta_j} && (2.4.2)\\
& + \sum_{j,k} (g(\alpha, \vartheta_j, \chi_{jk}) \delta_{\theta_j, \sigma_{jk}} + g'(\alpha, \vartheta_j, \chi_{jk}) \delta_{\sigma_{jk}, \theta_j}) && (2.4.3)\\
& + \sum_{k} (l(\alpha, \eta_k) \delta_{\rho_k, \rho'_k} + l'(\alpha, \eta_k) \delta_{\rho'_k, \rho_k}) && (2.4.4)\\
& + \sum_{j} (h(\alpha, \vartheta_j)(\delta_{\theta_j} \otimes \Lambda) + h'(\alpha, \vartheta_j)(\Lambda \otimes \delta_{\theta_j})) + \beta \Lambda_D + \gamma \Lambda^2, && (2.4.5)
\end{aligned}$$

for some measurable function $f \geqslant 0$ *on* $\mathbb{R}_+^4$, $g, g' \geqslant 0$ *on* $\mathbb{R}_+^3$ *and* $h, h', l, l' \geqslant 0$ *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* $\alpha, \beta, \gamma \geqslant 0$. *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. (2.4.5)) 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. 2. (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 2.4.7 (Random graph representation). *Let ξ be a random adjacency measure. Then ξ is jointly exchangeable iff almost surely*

$$\xi = \sum_{i,j} 1[\zeta_{\{i,j\}} \leqslant W(\alpha, \vartheta_i, \vartheta_j)]\delta_{\theta_i,\theta_j} \tag{2.4.6}$$

$$+ \sum_{j,k} 1[\chi_{jk} \leqslant S(\alpha, \vartheta_j)](\delta_{\theta_j,\sigma_{jk}} + \delta_{\sigma_{jk},\theta_j}) \tag{2.4.7}$$

$$+ \sum_{k} 1[\eta_k \leqslant I(\alpha)](\delta_{\rho_k,\rho'_k} + \delta_{\rho'_k,\rho_k}), \tag{2.4.8}$$

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 $\alpha \geqslant 0$. 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 2.4.6, and let

$$\xi_{\{i,j\}} := f(\alpha, \vartheta_i, \vartheta_j, \zeta_{\{i,j\}}), \tag{2.4.9}$$

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

$$W(a,t,t') = \Lambda\{z \in [0,1] : f(a,t,t',z) = 1\} = \Lambda f(a,t,t',\cdot), \tag{2.4.10}$$

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

$$\mathbb{E}[\xi_{\{i,j\}}|\mathcal{F}] \overset{a.s.}{=} W_\alpha(\vartheta_i,\vartheta_j). \tag{2.4.11}$$

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

$$\xi'_{\{i,j\}} = 1(W_\alpha(\vartheta_i,\vartheta_j) \leqslant \zeta'_{\{i,j\}}). \tag{2.4.12}$$

Then it is clear that

$$(\alpha, ((\theta_i,\vartheta_i)_{i\in\mathbb{N}}), (\xi'_{\{i,j\}})_{\{i,j\}\in\tilde{\mathbb{N}}_2}) \overset{d}{=} (\alpha, ((\theta_i,\vartheta_i)_{i\in\mathbb{N}}), (\xi_{\{i,j\}})_{\{i,j\}\in\tilde{\mathbb{N}}_2}) \tag{2.4.13}$$

and so, by a transfer argument Kallenberg [Kalo1, 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

$$\xi_{\{i,j\}} \overset{\text{a.s.}}{=} 1(W_\alpha(\vartheta_i, \vartheta_j) \leqslant \zeta''_{\{i,j\}}). \tag{2.4.14}$$

Similarly, letting g and l be as in Theorem 2.4.6, define

$$S(a,t) := \Lambda\{z \in \mathbb{R}_+ : g(a,t,z) = 1\} = \Lambda g(a,t,\cdot) \tag{2.4.15}$$

and

$$I(a) := \Lambda\{z \in \mathbb{R}_+ : l(a,z) = 1\} = \Lambda l(a,\cdot). \tag{2.4.16}$$

A similar argument to above can be used to show that the terms involving S and I agree with their counterparts in Theorem 2.4.6. □

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 2.4.8 (local summability [Kalo5, Prop. 9.25]). *Let ξ be as in Theorem 2.4.6, write $\hat{f} = f \wedge 1$, and let*

$$f_1 = \Lambda_{23}^2 \hat{f}, \qquad f_2 = \Lambda_{13}^2 \hat{f}, \qquad g_1 = \Lambda_2 \hat{g}, \tag{2.4.17}$$

where Λ_{23}^2 denotes two-dimensional Lebesgue measure in the second and third coordinates, and similarly for Λ_{13}^2 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) $\Lambda^2[\hat{f}; f_1 \vee f_2 \leqslant 1] < \infty$,

(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 2.4.9 (locally finite graphex). *Let ξ, be as in Theorem 2.4.7 for fixed α, and drop the first coordinate from the definitions of I, S, and W. Let $\mu_W(t) = \Lambda W(t, \cdot) = \int_{\mathbb{R}_+} W(t, t')\, dt'$. The random measure ξ, is a.s. locally finite iff these four conditions are fulfilled:*

(i) $I < \infty$,

(ii) $\Lambda \min(S, 1) = \int_{\mathbb{R}_+} \min(S(t), 1)\, dt < \infty$,

(iii) $\Lambda\{\mu_W = \infty\} = 0$ *and* $\Lambda\{\mu_W > 1\} < \infty$,

(iv) $\Lambda^2[W; \mu_W \vee \mu_W \leqslant 1] = \int_{\mathbb{R}_+^2} W(x, y)\, 1[\mu_W(x) \leqslant 1]\, 1[\mu_W(y) \leqslant 1]\, dx\, dy < \infty$,

(v) $\int_{\mathbb{R}_+} W(x, x)\, dx < \infty$.

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

Remark 2.4.10. An example showing that there are nonintegrable W admitting a.s. locally finite exchangeable adjacency measures is the function $W(x, y) = 1[xy \leqslant 1]$. Its marginal is $\mu_W(x) = \frac{1}{x}$, which obviously satisfies (iii). Moreover, $W = 0$ a.e. on the set $\{(x, y) : \mu_W(x) \wedge \mu_W(y) \leqslant 1\} = \{(x, y) : x, y \geqslant 1\}$, satisfying (iv).

These conditions leads us to the following definition:

Definition 2.4.11. A *graphex* is a triple (I, S, W), where $I \geqslant 0$ is a non-negative real, $S : \mathbb{R}_+ \to \mathbb{R}_+$ is a measurable function such that $\min(S, 1)$ is integrable, and $W : \mathbb{R}_+^2 \to [0, 1]$, the *Kallenberg graphon*, is symmetric, and satisfies parts (iii)–(v) of Theorem 2.4.9.

In situations where there is no risk of confusion, we will abuse nomenclature and use the term graphon to refer to the W component.

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 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 2.4.12. The *Kallenberg graphon marginal* is $\mu_W(x) = \int_{\mathbb{R}_+} W(x, y)\, dy$.

Theorem 2.4.7 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 $\theta \times \vartheta = \mathbb{R}_+^2$, 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 2.4.13. 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. (2.4.6). The Kallenberg exchangeable graph model is the family of s-truncations $\Gamma_s = \xi(\cdot \cap [0, s]^2)$, for $s \in \mathbb{R}_+$. When the graphex is clear from context, we will simply refer to G as the Kallenberg exchangeable graph.

The first term of Eq. (2.4.6) gives essentially all of the interesting graph structure, and so for the rest of the chapter, 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 $I \geqslant 0$, $S = 0$, and W is constant on a set of the form $[0, c]^2$ and o 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. Let Π be a unit rate Poisson process on $\theta \times \vartheta$ as in Theorem 2.4.7. A Kallenberg exchangeable graph G associated with W has vertex set

$$v(G) = \{\theta_i : (\theta_i, \vartheta_i) \in \Pi \wedge \exists \theta_j \in \Pi : W(\vartheta_i, \vartheta_j) > \zeta_{\{i,j\}}\} \quad (2.4.18)$$

and edge set

$$e(G) = \{\{\theta_i, \theta_j\} \mid (\theta_i, \vartheta_i), (\theta_j, \vartheta_j) \in \Pi \wedge W(\vartheta_i, \vartheta_j) > \zeta_{\{i,j\}}\}. \quad (2.4.19)$$

Remark 2.4.14. A Kallenberg graphon with $W(\vartheta, \vartheta) = 0$ for all $\vartheta \in \mathbb{R}_+$ generates a KEG with no self edges.

Remark 2.4.15. Notice that if G is a KEG associated to W and Γ_s is G restricted to $[0, s]$ then Γ_s is *not* the same as the induced subgraph of G given by restricting to vertices of G with labels $\leqslant s$. 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 $\Gamma_s \uparrow G$ in the sense that $v(\Gamma_s) \uparrow v(G)$ and $e(\Gamma_s) \uparrow e(G)$ as $s \uparrow \infty$.

Remark 2.4.16. The model can be extended to weighted graphs by replacing the indicator term $\mathbb{1}[W(\alpha, \vartheta_i, \vartheta_j) \leqslant \zeta_{\{i,j\}}]$ 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 2.4.17. 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 2.4.18. Point processes are central to our construction. For a point process $\mathbf{P}$ we will often refer to points $p_i \in \mathbf{P}$ where the index i is given by some unspecified measurable function of $\mathbf{P}$. For example, if $\mathbf{P}$ 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 $\mathbf{P}$). 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.

KEY PROPERTIES

It remains to show that the models derived in the previous section are sufficiently powerful to model the structure of real-world networks. To that end, we derive a number of the generative model properties of the Kallenberg exchangeable graphs.

Expected Number of Edges and Vertices

We begin by deriving the expected number of edges and vertices for Γ_s generated by $(0, 0, W)$ for integrable W. In our statements, we implicitly assume W is non-random; in the case of random W the results can be understood as conditional statements. Proofs are deferred to Section 2.A.

Theorem 2.5.1. *The expected number of edges* $e_s = |e(\Gamma_s)|$ *is*

$$\mathbb{E}[e_s] = \frac{1}{2}s^2 \iint_{\mathbb{R}_+^2} W(x, y)\, dx\, dy + s \int_{\mathbb{R}_+} W(x, x)\, dx. \tag{2.5.1}$$

Theorem 2.5.2. *The expected number of vertices* $v_s = |v(\Gamma_s)|$ *is*

$$\mathbb{E}[v_s] = s \int_{\mathbb{R}_+} (1 - e^{-s\mu_W(x)})\, dx + s \int_{\mathbb{R}_+} e^{-s\mu_W(x)} W(x, x)\, dx. \tag{2.5.2}$$

Intuitively speaking, the expected number of edges grows as s^2 for any choice of integrable W, but the expected number of vertices grows at a rate determined by μ_W, with fatter tails corresponding to a faster rate of growth. It's easy to see that the expected number of vertices is $\Theta(s)$ only if W is compactly supported, which suggests that a Kallenberg exchangeable graph can be dense only if it is generated by compactly supported W. Indeed this is so, as captured by the following theorem:

Theorem 2.5.3. *Let $(\Gamma_s)_{s\in\mathbb{R}_+}$ be Kallenberg exchangeable graph with graphex $(0,0,W)$. If W is compactly supported, then $(\Gamma_s)_{s\in\mathbb{R}_+}$ is dense with probability 1. Conversely, if W is integrable and not compactly supported, then $(\Gamma_s)_{s\in\mathbb{R}_+}$ is sparse with probability 1.*

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 $I = S = 0$. To formalize the notion of limiting degree distribution, let Γ_s be a Kallenberg exchangeable graph on $[0,s)$ with graphex $(0,0,W)$, and let D_s be the degree of a vertex chosen uniformly at random from $v(\Gamma_s)$. The central object of study is then the random distribution function $k \mapsto P(D_s \leqslant k \mid \Gamma_s)$ and its scaling limit. The main theorem about the limiting degree distribution is:

Theorem 2.5.4. *Let $(0,0,W)$ be a graphex with W integrable such that*

1. *There exist some constants $C, T > 0$ such that for all λ and $\omega > T$ it holds that $\int W(\lambda,x)W(\omega,x)\,dx \leqslant C\mu_W(\lambda)\mu_W(\omega)$.*
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 $\frac{\mu_W(x)}{\mu_W'(x)}\frac{1}{x} \geqslant -1$.*

Let $k_s = o(s)$. Then,

$$\Pr(D_s > k_s \mid \Gamma_s) \xrightarrow{p} \lim_{s\to\infty} \frac{\sum_{n=k_{s+1}}^{\infty} \int \frac{1}{n!} e^{-s\mu_W(x)} (s\mu_W(x))^n \,dx}{\int 1 - e^{-s\mu_W(x)}\,dx}. \tag{2.5.3}$$

The proof is deferred to Section 2.B, which also contains a detailed explanation of the motivation and need for the various technical assumptions.

In the case $\mu_W(\lambda) = (1+\lambda)^{-2}$ the right hand side of this expression is in $(0,1)$ for $k_s = k$ for any choice of k. That is, even in the infinite graph limit a constant fraction of the vertices will have degree $\leqslant k$ 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, $\Pr(D_s > k \mid \Gamma_s) \xrightarrow{p} 1$. However, we saw that $\Pr(D_s > s^\beta \mid \Gamma_s) \to 1-\beta$ for $\beta \in (0,1)$; i.e., taking $k_s = s^\beta$ 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 W is.

Connectivity for Product Form 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. To resolve this concern, we show that, at least for some families of graphexes, almost all of the vertices belong to a single component asymptotically:

Theorem 2.5.5. *Let* G *be a KEG generated by graphex* $(0, 0, W)$ *with* $W = f(x)f(y)1[x \neq y]$*, let* $C_1(\Gamma_s)$ *be the largest connected component of* Γ_s*, and let* $\epsilon > 0$*, then*

$$\lim_{s\to\infty} \Pr(|C_1(\Gamma_s)| > (1-\epsilon)\,|\nu(\Gamma_s)|) = 1. \tag{2.5.4}$$

The proof is deferred to Section 2.C.

The proof reveals some interesting further structure of KEGs generated by Kallenberg graphons of the product form $W = f(x)f(y)1[x \neq y]$. 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. Additionally, it is 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. Our proof Theorem 2.5.5 establishes that KEGs generated by graphexes with product form have both of these properties.

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 $[0, s]$, in terms of the generating graphex. We focus on those graphexes where $I = 0$ and $S = 0$. 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 Γ_s by projecting the latent Poisson process Π_s along its second coordinate on to a random point set in $\vartheta \simeq \mathbb{R}_+$ as $\Pi_s^P = \{\vartheta_i \mid (\theta_i, \vartheta_i) \in \Pi_s\}$, which is then a rate s 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:

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

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,

$$D(\vartheta, \Pi_s^P, (\zeta_{\{i,j\}})) \tag{2.A.2}$$

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

For any $\lambda \in \mathbb{R}_+$ the probability that $\lambda \in \Pi_s^P$ is 0 and so $D(\lambda, \Pi_s^P, \zeta_{\{i,j\}})$ 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 o 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 $D(\lambda, \Pi_s^P \cup \{\lambda\}, \zeta_{\{i,j\}})$. Recalling the Kallenberg graphon marginal $\mu_W(x) = \int_{\mathbb{R}_+} W(x,y)\,dy$:

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

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

$$D(\lambda, \Pi_s^P \cup \{\lambda\}, \zeta_{\{i,j\}}) = \sum_{p \in \Pi_s^P} 1[W(\lambda,p) \geqslant \zeta_{i(\lambda)i(p)}] + 2 \cdot 1[W(\lambda,\lambda) \geqslant \zeta_{i(\lambda)i(\lambda)}]. \tag{2.A.3}$$

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

$$D_{\text{self}} = 2 \cdot 1[W(\lambda,\lambda) \geqslant \zeta_{i(\lambda)i(\lambda)}] \tag{2.A.4}$$

and

$$D_{ext} = \sum_{p \in \Pi_s^P} 1[W(\lambda, p) \geqslant \zeta_{i(\lambda)i(p)}] \tag{2.A.5}$$

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

We have that

$$\int_{\mathbb{R}_+}\int_{[0,1]} 1\left[u \leqslant W(\lambda, y)\right] s\, du\, dy = s\int_{\mathbb{R}_+} W(\lambda, y)\, dy < \infty \text{ a.s.}, \tag{2.A.6}$$

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

$$\mathbb{E}[\exp(itD_{ext})] = \mathbb{E}[\exp(it\sum_{p \in \Pi_s^P} 1\left[\zeta_{i(\lambda)i(p)} \leqslant W(\lambda, p)\right])] \tag{2.A.7}$$

$$= \exp\{\int_{\mathbb{R}_+}\int_{[0,1]} (1 - e^{it1[u \leqslant W(\lambda,y)]} s\, du\, dy)\} \tag{2.A.8}$$

$$= \exp\{s\sum_{n=1}^{\infty} \frac{(it)^n}{n!}\int_{\mathbb{R}_+}\int_{[0,1]} 1\left[u \leqslant W(\lambda, y)\right] du\, dy\} \tag{2.A.9}$$

$$= \exp\{s\mu_W(\lambda)(e^{it} - 1)\}. \tag{2.A.10}$$

Hence, D_{ext} is a $\text{Poi}(s\mu_W(\lambda))$ 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

$$e_s \overset{d}{=} \frac{1}{2}\sum_{\vartheta \in \Pi_s^P} D(\vartheta, \Pi_s^P, (\zeta_{\{i,j\}})), \tag{2.A.11}$$

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 2.4.7.

Standard point process formulas deal with computing expressions of the form

$$\mathbb{E}[\sum_{\lambda \in \xi} h(\lambda, \xi)] \tag{2.A.12}$$

where ξ 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 2.A.2 (Extended Slivnyak–Mecke)**.** *Let Φ be a rate s 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*

$$\mathbb{E}[\sum_{p\in\Phi} f(p,\Phi,U)] = s\int_{\mathbb{R}_+} \mathbb{E}[f(x,\Phi\cup\{x\},U)]\,dx. \tag{2.A.13}$$

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

$$\mathbb{E}[\sum_{p\in\Phi} f(p,\Phi,U)] = \int_0^1 \mathbb{E}[\sum_{p\in\Phi} f(p,\Phi,u)]\,du. \tag{2.A.14}$$

By the usual Palm calculus, the inner expectation satisfies

$$\mathbb{E}[\sum_{p\in\Phi} f(p,\Phi,u)] = \int_{\mathbb{R}_+}\int_{\mathbb{M}} f(x,\varphi,u)P_x(\,d\varphi)s\,dx, \tag{2.A.15}$$

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:

$$\int_{\mathbb{M}} f(x,\varphi,u)P_x(\,d\varphi) = \int_{\mathbb{M}} f(x,\varphi\cup\{x\},u)P(\,d\varphi). \tag{2.A.16}$$

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

The main results now follow easily:

Theorem 2.5.1. *The expected number of edges $e_s = |e(\Gamma_s)|$ is*

$$\mathbb{E}[e_s] = \frac{1}{2}s^2\iint_{\mathbb{R}_+^2} W(x,y)\,dx\,dy + s\int_{\mathbb{R}_+} W(x,x)\,dx. \tag{2.5.1}$$

Proof. By Lemmas 2.A.1 and 2.A.2,

$$\mathbb{E}[e_s] = \frac{1}{2}\mathbb{E}[\sum_{\vartheta\in\Pi_s^P} D(\vartheta,\Pi_s^P,(\zeta_{\{i,j\}}))] \tag{2.A.17}$$

$$= \frac{1}{2}s\int_{\mathbb{R}_+} \mathbb{E}[D(x,\Pi_s^P\cup\{x\},(\zeta_{\{i,j\}}))]\,dx \tag{2.A.18}$$

$$= \frac{1}{2}s\int_{\mathbb{R}_+} s\mu_W(x) + 2W(x,x)\,dx \tag{2.A.19}$$

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

Theorem 2.5.2. *The expected number of vertices $v_s = |v(\Gamma_s)|$ is*

$$\mathbb{E}[v_s] = s\int_{\mathbb{R}_+}(1-e^{-s\mu_W(x)})\,dx + s\int_{\mathbb{R}_+} e^{-s\mu_W(x)}W(x,x)\,dx. \tag{2.5.2}$$

Proof. By Lemmas 2.A.1 and 2.A.2,

$$\mathbb{E}[\nu_s] = \mathbb{E}[\sum_{\vartheta \in \Pi_s^P} 1\left[D(\vartheta, \Pi_s^P, (\zeta_{\{i,j\}})) \geqslant 1\right]] \tag{2.A.20}$$

$$= s\int_{\mathbb{R}_+} \Pr(D(x, \Pi_s^P \cup \{x\}, (\zeta_{\{i,j\}})) \geqslant 1)\, dx \tag{2.A.21}$$

$$= s\int_{\mathbb{R}_+} 1 - \Pr(D_{ext} = 0)\Pr(D_{self} = 0)\, dx \tag{2.A.22}$$

$$= s\int_{\mathbb{R}_+} 1 - e^{-s\mu_W(x)}(1 - W(x,x))\, dx, \tag{2.A.23}$$

where D_{ext} and D_{self} are defined as in Lemma 2.A.1. Splitting up the integral is justified since $1-\exp(-s\mu_W(x)) \geqslant 0$ and $\exp(-s\mu_W(x))W(x,x) \geqslant 0$ 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 2.B.

Theorem 2.A.3. *The expected number of vertices of degree k in Γ_s, $N_{s,k}$, is*

$$\mathbb{E}[N_{s,k}] = s^{k+1}\int_{\mathbb{R}_+}\left[\frac{\mu_W(x)^k}{k!}e^{-s\mu_W(x)} + \frac{1}{s^2}\frac{\mu_W(x)^{k-2}}{(k-2)!}e^{-s\mu_W(x)}(1 - \frac{(s\mu_W(x))^2}{k(k-1)})W(x,x)\right] dx \tag{2.A.24}$$

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 $s \to \infty$ 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 2.A.3, 2.5.1 and 2.5.2.

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 graphon is compactly supported.

Theorem 2.5.3. *Let $(\Gamma_s)_{s\in\mathbb{R}_+}$ be Kallenberg exchangeable graph with graphex $(0,0,W)$. If W is compactly supported, then $(\Gamma_s)_{s\in\mathbb{R}_+}$ is dense with probability 1. Conversely, if W is integrable and not compactly supported, then $(\Gamma_s)_{s\in\mathbb{R}_+}$ is sparse with probability 1.*

Proof. We have already shown in Section 2.3.1 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

$$\liminf_{s\to\infty} \Pr(e_s > cv_s^2) > p, \tag{2.A.25}$$

where $e_s = e(\Gamma_s)$ and $v_s = v(\Gamma_s)$. With

$$\mathbb{E}[e_s] \geqslant \Pr(e_s > cv_s^2)\mathbb{E}[cv_s^2] \tag{2.A.26}$$

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

Now, by Theorem 2.5.2,

$$\mathbb{E}[v_s] = s\int_{\mathbb{R}_+} 1 - e^{-s\mu_W(x)}\,dx + s\int_{\mathbb{R}_+} e^{-s\mu_W(x)}W(x,x)\,dx, \tag{2.A.27}$$

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

$$\mathbb{E}[e_s] = \omega(s^2). \tag{2.A.28}$$

However, by Theorem 2.5.1, $\mathbb{E}[e_s] = \Theta(s^2)$. This contradiction completes the proof.

□

DEGREE DISTRIBUTION IN THE ASYMPTOTIC LIMIT

The primary aim of this section is to prove the following theorem:

Theorem 2.5.4. *Let $(0, 0, W)$ be a graphex with W integrable such that*

1. *There exist some constants $C, T > 0$ such that for all λ and $\omega > T$ it holds that $\int W(\lambda, x)W(\omega, x)\,dx \leqslant C\mu_W(\lambda)\mu_W(\omega)$.*
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 $\frac{\mu_W(x)}{\mu_W'(x)}\frac{1}{x} \geqslant -1$.*

Let $k_s = o(s)$. Then,

$$\Pr(D_s > k_s \mid \Gamma_s) \xrightarrow{p} \lim_{s\to\infty} \frac{\sum_{n=k_{s+1}}^{\infty} \int \frac{1}{n!} e^{-s\mu_W(x)}(s\mu_W(x))^n\,dx}{\int 1 - e^{-s\mu_W(x)}\,dx}. \tag{2.5.3}$$

Let $n_{>l}^{(s)}$ denote the number of vertices of Γ_s with degree greater than l. It is immediate that

$$\Pr(D_s \geqslant k_s \mid \Gamma_s) = \frac{n_{>k_s}^{(s)}}{n_{>0}^{(s)}}, \tag{2.B.1}$$

i.e., the probability of choosing a vertex of degree greater than k_s is the proportion of such vertices among all vertices. Notice that, even for fixed l, the random variable $n_{>l}^{(s)}$ grows with s. Further notice that like D_s the random variable $n_{>k_s}^{(s)}/n_{>0}^{(s)}$ is ill defined for the event $n_{>0}^{(s)} = 0$; however this is a measure 0 event in the limit $s \to \infty$. The content of Theorem 2.5.4 can be understood as saying that the limit of the ratio $\frac{n_{>l}^{(s)}}{n_{>0}^{(s)}}$ is the limit of the ratio of the expectations,

$$\frac{n_{>l}^{(s)}}{n_{>0}^{(s)}} \xrightarrow{p} \lim_{s\to\infty} \frac{\mathbb{E}[n_{>l}^{(s)}]}{\mathbb{E}[n_{>0}^{(s)}]}, \; s \to \infty. \tag{2.B.2}$$

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 $\theta_i, \theta_j \leqslant 1$. 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 $\theta \leqslant 1$. Let $N_{>l}^{(s)}$ be the number of vertices of Γ_s with label $\theta_i < 1$ and greater than l neighbours $\{\theta_j\}$ where $\theta_j > 1$. The following lemma establishes the claimed equivalence:

Lemma 2.B.1. *The limiting distribution of $n_{>l}^{(s)}/n_{>0}^{(s)}$ is the same as the limiting distribution of the ratio that considers only vertices with label $\theta_i \leqslant 1$ and counts only edges (θ_i, θ_j) with $\theta_j > 1$,*

$$\lim_{s\to\infty} \frac{n_{>l}^{(s)}}{n_{>0}^{(s)}} \overset{d}{=} \lim_{s\to\infty} \frac{N_{>l}^{(s)}}{N_{>0}^{(s)}}. \tag{2.B.3}$$

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

1. $\lim_{s\to\infty} \Pr(N_{>0}^{(s)} = 0) = 0$ so $\lim_{s\to\infty} \frac{N_{>l}^{(s)}}{N_{>0}^{(s)}}$ is well defined.

2. The θ label of a point of the latent Poisson process is independent of its degree. Let $\tilde{D}_s$ be the degree of a vertex chosen uniformly at random from those members of $\nu(\Gamma_s)$ with label $\theta < 1$ and let $\tilde{N}_{>l}^{(s)}$ 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 $\{\tilde{N}^{(s)}_{>0} > 0\}$,

$$\Pr(D_s > l \mid \Gamma_s) \stackrel{d}{=} \Pr(\tilde{D}_s > l \mid \Gamma_s). \tag{2.B.4}$$

This immediately implies

$$\frac{n^{(s)}_{>l}}{n^{(s)}_{>0}} \stackrel{d}{=} \frac{\tilde{N}^{(s)}_{>l}}{\tilde{N}^{(s)}_{>0}}. \tag{2.B.5}$$

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

$$\lim_{s\to\infty} \Pr(\tilde{D}_s > l \mid \Gamma_s) \stackrel{a.s.}{=} \lim_{s\to\infty} \frac{N^{(s)}_{>l}}{N^{(s)}_{>0}}. \tag{2.B.6}$$

□

To treat the limiting distribution of this ratio we introduce

$$\Pi_0 = \{\vartheta \mid (\theta, \vartheta) \in \Pi_{s+1}, \theta \leqslant 1\} \tag{2.B.7}$$

$$\Pi_{(1,s+1]} = \{(\theta, \vartheta) \mid (\theta, \vartheta) \in \Pi_{s+1}, \theta > 1\}, \tag{2.B.8}$$

i.e., we break the latent Poisson process into the component with $\theta \leqslant 1$ 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 $\Pi_{(1,s+1]}$ 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

$$D_s(x, \bar{u}, \{(\phi_i, \varphi_i)\}) = \sum_i 1[W(x, \varphi_i) > u_i] 1[\phi_i \leqslant s + 1]. \tag{2.B.9}$$

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

$$D_s(\lambda, \bar{\zeta}_i, \Pi_{(1,\infty)}) \tag{2.B.10}$$

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

$$D_{j,s}(x) = D_s(x, \bar{U}_j, \Pi_{(1,\infty)}). \tag{2.B.11}$$

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

It follows by mimicking the proof of Lemma 2.A.1 that

$$D_{j,s}(x) \sim \mathrm{Poi}(s\mu_W(x)) \tag{2.B.12}$$

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

$$N^{(s)}_{>l} = \sum_i 1[D_s(\lambda_i, \bar{\zeta}_i, \Pi_{(1,\infty)}) > l]. \tag{2.B.13}$$

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

$$\mathbb{E}[N^{(s)}_{>l}] = \int \Pr(D_{1,s}(x) > l)\, dx. \tag{2.B.14}$$

The idea of the proof of Theorem 2.5.4 is to show that

$$N^{(s)}_{>k_s}/\mathbb{E}[N^{(s)}_{>0}] \xrightarrow{p} \lim_{s\to\infty} \mathbb{E}[N^{(s)}_{>k_s}]/\mathbb{E}[N^{(s)}_{>0}],\ s \to \infty. \tag{2.B.15}$$

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

$$\frac{N^{(s)}_{>k_s}}{N^{(s)}_{>0}} \xrightarrow{p} \lim_{s\to\infty} \mathbb{E}[N^{(s)}_{>k_s}]/\mathbb{E}[N^{(s)}_{>0}],\ s \to \infty. \tag{2.B.16}$$

Using Chebyshev's inequality, a sufficient condition for Eq. (2.B.15) to hold is

$$\operatorname{var}\left[N^{(s)}_{>k_s}\right] = o(\mathbb{E}[N^{(s)}_{>0}]^2). \tag{2.B.17}$$

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

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

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

We do not know of any examples of an integrable graphon W 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 $(l, \lambda), (w, \omega) \in \Pi_s$ under Γ_s and observe that for a graphon W that is 0 on the diagonal (i.e., forbidding self-edges),

$$L(\lambda, \omega) \sim \operatorname{Poi}(s \int W(\lambda, x)W(\omega, x)\, dx), \tag{2.B.18}$$

with respect to the Palm measure $P_{\lambda,\omega}$[2]. This can be shown by an argument very similar to Lemma 2.A.1. Thus the assumption can be

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

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 $\frac{\mu_W(x)}{\mu'_W(x)}\frac{1}{x} \geqslant -1$. 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 graphon W if there is some other graphon 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 2.5.3). A particular consequence of this last assumption is that for any function $l(s) \to 0$ as $s \to \infty$ it holds that $\mu_W^{-1}(l(s)) \to \infty$, 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 $\left[N^{(s)}_{>k_s}\right]$.

Lemma 2.B.2. *Let* $k_s = o(s)$, *then*

$$\begin{aligned} \mathit{var}\left[N^{(s)}_{>k_s}\right] &= \mathbb{E}[N^{(s)}_{>k_s}] && \text{(2.B.19)}\\ &\quad + \iint \Pr(D_{1,s}(x) > k_s, D_{2,s}(y) > k_s) && \text{(2.B.20)}\\ &\quad - \Pr(D_{1,s}(x) > k_s)\Pr(D_{2,s}(y) > k_s)\,dx\,dy && \text{(2.B.21)} \end{aligned}$$

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

$$D_s(\lambda_i, \bar{\zeta}_i, \Pi_{(1,\infty)}) \tag{2.B.22}$$

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

$$N^{(s)}_{>k_s} \mid \Pi_{(1,\infty)} \;\sim\; \mathrm{Poi}(\mathbb{E}[N^{(s)}_{>k_s} \mid \Pi_{(1,\infty)}]). \tag{2.B.23}$$

Using this, the formula for conditional variance is

$$\begin{aligned} \operatorname{var}\left[N^{(s)}_{>k_s}\right] &= \mathbb{E}[\operatorname{var}\left[N^{(s)}_{>k_s} \mid \Pi_{(1,\infty)}\right]] + \operatorname{var}\left[\mathbb{E}[N^{(s)}_{>k_s} \mid \Pi_{(1,\infty)}]\right] && \text{(2.B.24)}\\ &= \mathbb{E}[N^{(s)}_{>k_s}] + \operatorname{var}\left[\mathbb{E}[N^{(s)}_{>k_s} \mid \Pi_{(1,\infty)}]\right]. && \text{(2.B.25)} \end{aligned}$$

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

$$\mathbb{E}[N^{(s)}_{>k_s} \mid \Pi_{(1,\infty)}] = \int_{\mathbb{R}_+} \mathbb{E}[1[D_s(x, \bar{U}, \Pi_{(1,\infty)}) > k_s] \mid \Pi_{(1,\infty)}]\, dx \tag{2.B.26}$$

$$= \int_{\mathbb{R}_+} \Pr(D_{1,s}(x) > k_s \mid \Pi_{(1,\infty)})\, dx, \tag{2.B.27}$$

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

$$\iint_{\mathbb{R}^2_+} \Pr(D_{1,s}(x) > k_s \wedge D_{2,s}(y) > k_s \mid \Pi_{(1,\infty)})\, dx\, dy. \tag{2.B.28}$$

By Tonelli's theorem,

$$\mathbb{E}[\mathbb{E}[N^{(s)}_{>k_s} \mid \Pi_{(1,\infty)}]^2] = \iint_{\mathbb{R}^2_+} \Pr(D_{1,s}(x) > k_s \wedge D_{2,s}(y) > k_s)\, dx\, dy \tag{2.B.29}$$

whence

$$\operatorname{var}\left[\mathbb{E}[N^{(s)}_{>k_s} \mid \Pi_{(1,\infty)}]\right] = \iint_{\mathbb{R}^2_+} \Pr(D_{1,s}(x) > k_s \wedge D_{2,s}(y) > k_s)\, dx\, dy \tag{2.B.30}$$

$$- \iint_{\mathbb{R}^2_+} \Pr(D_{1,s}(x) > k_s) \Pr(D_{2,s}(y) > k_s)\, dx\, dy \tag{2.B.31}$$

and the claimed result follows. □

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

$$\Pr(D_{1,s}(x) > k_s, D_{2,s}(y) > k_s) - \Pr(D_{1,s}(x) > k_s) \Pr(D_{2,s}(y) > k_s) \tag{2.B.32}$$

$$\leqslant \Pr(D_{1,s}(x) > k_s) g(y) \tag{2.B.33}$$

so that

$$\operatorname{var}\left[N^{(s)}_{>k_s}\right] \leqslant \mathbb{E}[N^{(s)}_{>k_s}] + \iint \Pr(D_{1,s}(x) > k_s) g(y)\, dx\, dy \tag{2.B.34}$$

$$= \mathbb{E}[N^{(s)}_{>k_s}](1 + \int g(y)\, dy). \tag{2.B.35}$$

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

Lemma 2.B.3. *Let* T *be a value such that for* $y > T$ *it holds that*

$$\int W(x,z)W(y,z)\,dz \leqslant C\mu_W(x)\mu_W(y) \tag{2.B.36}$$

and

$$2C\mu(y) \leqslant 1 - \log 2. \tag{2.B.37}$$

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

$$g(y) = \begin{cases} \Pr(D_{2,s}(y) \leqslant k_s) & y \leqslant T \\ \Pr(D_{2,s}(y) + B(y) > k_s \wedge D_{2,s}(y) \leqslant k_s) & y > T. \end{cases} \tag{2.B.38}$$

Then,

$$\Pr(D_{1,s}(x) > k_s, D_{2,s}(y) > k_s) - \Pr(D_{1,s}(x) > k_s)\Pr(D_{2,s}(y) > k_s) \tag{2.B.39}$$

$$\leqslant \Pr(D_{1,s}(x) > k_s)g(y) \tag{2.B.40}$$

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

$$D_a = D_{1,s}(x) \tag{2.B.41}$$

$$D_b = D_{2,s}(y). \tag{2.B.42}$$

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

$$\Pr(D_a > k_s, D_b > k_s) = \Pr(D_a > k_s)\Pr(D_b > k_s | D_a > k_s), \tag{2.B.43}$$

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

$$M_i = 1[W(x, \vartheta_i) > U_i^1] \tag{2.B.44}$$

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

$$N_a = \{\vartheta_i \mid (\theta_i, \vartheta_i) \in \Pi_{(1,s+1]},\ M_i = 1\}, \tag{2.B.45}$$

the neighbours of a, and

$$\bar{N}_a = \{\vartheta_i \mid (\theta_i, \vartheta_i) \in \Pi_{(1,s+1]},\ M_i = 0\}, \tag{2.B.46}$$

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

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

$$\vartheta_i \overset{iid}{\sim} \frac{W(x, \vartheta_i)}{\mu_W(x)}. \tag{2.B.47}$$

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

$$D_b = D_b^{(N_a)} + D_b^{(\bar{N}_a)}, \tag{2.B.48}$$

where, by an application of Campbell's theorem,

$$D_b^{(\bar{N}_a)} \sim \text{Poi}(s(\mu_W(y) - \int W(x, z)W(y, z)\, dz)) \tag{2.B.49}$$

and

$$D_b^{(N_a)} \mid D_a \sim \text{Bin}(D_a, p_{x,y}) \tag{2.B.50}$$

independently, with

$$p_{x,y} = \frac{1}{\mu_W(x)} \int W(x, z)W(y, z)\, dz. \tag{2.B.51}$$

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 $D_a > k_s$ we introduce random variables L_1, L_2 such that on the event $\{D_a > k_s\}$

$$L_1 + L_2 = D_a \tag{2.B.52}$$

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

$$L_1 \sim \text{Poi}(s\mu_W(x)) \tag{2.B.53}$$

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

$$D_b^{(N_a)} = B_1 + B_2, \tag{2.B.54}$$

and

$$B_1 \mid L_1 \sim \text{Bin}(L_1, p_{x,y}) \tag{2.B.55}$$

$$B_2 \mid L_2 \sim \text{Bin}(L_2, p_{x,y}) \tag{2.B.56}$$

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

$$B_1 \sim \mathrm{Poi}(s \int W(x,z)W(y,z)\,dz) \tag{2.B.57}$$

and so

$$(D_b^{(\tilde{N}_a)} + B_1) \mid \{D_a > k_s\} \sim \mathrm{Poi}(s\mu_W(y)). \tag{2.B.58}$$

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

$$H = D_b^{(\tilde{N}_a)} + B_1 \tag{2.B.59}$$

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 $D_b = H + B_2$,

$$\Pr(D_b > k_s | D_a > k_s) = \mathbb{E}[\Pr(H + B_2 > k_s \mid L_1, L_2) \mid D_a > k_s]. \tag{2.B.60}$$

Then,

$$\Pr(H + B_2 > k_s \mid L_1, L_2) = \tag{2.B.61}$$
$$\Pr(H > k_s \mid L_1) + \Pr(H + B_2 > k_s \wedge H \leqslant k_s \mid L_1, L_2), \tag{2.B.62}$$

and L_1 has been defined so that

$$\mathbb{E}[\Pr(H > k_s \mid L_1) \mid D_a > k_s] = \Pr(D_b > k_s). \tag{2.B.63}$$

We have now arrived at

$$\Pr(D_a > k_s, D_b > k_s) = \Pr(D_a > k_s)[\Pr(D_b > k_s) + R], \tag{2.B.64}$$

where the remainder term is

$$R = \mathbb{E}[\Pr(H + B_2 > k_s \wedge H \leqslant k_s \mid L_1, L_2) \mid D_a > k_s] \tag{2.B.65}$$
$$= \Pr(H + B_2 > k_s \wedge H \leqslant k_s \mid D_a > k_s). \tag{2.B.66}$$

Note that

$$\Pr(D_a > k_s, D_b > k_s) - \Pr(D_a > k_s)\Pr(D_b > k_s) = \Pr(D_a > k_s)R \tag{2.B.67}$$

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

$$R \leqslant \Pr(H \leqslant k_s \mid D_a > k_s) \tag{2.B.68}$$
$$= \Pr(D_{2,s}(y) \leqslant k_s) \tag{2.B.69}$$

suffices. This establishes the claim for $y \leqslant T$ in the lemma statement. The remaining task is to find a good bound in the regime of y where $s\mu_W(y)$ 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

$$\Pr(B_2 > b \mid H) = \mathbb{E}[\Pr(B_2 > b \mid L_2) \mid H)]. \tag{2.B.70}$$

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

$$\Pr(B_2 > b \mid L_2) = 1 - \sum_{n=0}^{b} \binom{L_2}{n} {p_{x,y}}^n (1 - p_{x,y})^{L_2 - n}. \tag{2.B.71}$$

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

$$\Pr(B_2 > b \mid H) \leqslant \Pr(\tilde{B} > b \mid H), \tag{2.B.72}$$

where $\tilde{B} \mid H \sim \text{Bin}(\mathbb{E}[L_2 \mid H], p_{x,y})$. 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 $\mathbb{E}[L_2 \mid H] \leqslant 5k_s$ 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

$$Q \stackrel{d}{=} D_a \mid \{D_a > k_s\} \tag{2.B.73}$$

and Q independent of L_1 such that

$$\begin{aligned} L_2 &= \mathbb{1}[L_1 \leqslant k_s](Q - L_1) \\ \implies \mathbb{E}[L_2 \mid H] &\leqslant \Pr(L_1 \leqslant k_s \mid H)\mathbb{E}[Q]. \end{aligned} \tag{2.B.74}$$

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

1. Draw l_1 from the Poisson distribution. If $l_1 > k_s$ 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:

$$L_1 \mid H \sim \text{Bin}(H, \frac{\int W(x,z)W(y,z)\,dx}{\mu_W(y)}) + Z \tag{2.B.75}$$

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

$$\Pr(L_1 \leqslant k_s \mid H)\mathbb{E}[Q] \leqslant \Pr(Z \leqslant k_s)\mathbb{E}[Q]. \tag{2.B.76}$$

Further,

$$\mathbb{E}[Q] < k_s + s\mu_W(x), \tag{2.B.77}$$

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

$$G \sim \text{Gamma}(k_s, 1) \mid G < s\mu_W(x) \tag{2.B.78}$$
$$Q = k_s + \text{Poi}(s\mu_W(x) - G). \tag{2.B.79}$$

For $s\mu_W(x) \leqslant 2k_s$, it immediately follows that

$$\Pr(Z \leqslant k_s)\mathbb{E}[Q] \leqslant 5k_s \tag{2.B.80}$$

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

$$\Pr(Z \leqslant k_s)\mathbb{E}[Q] \leqslant k_s + 2\Pr(Z = k_s)s\mu_W(x) \tag{2.B.81}$$
$$= k_s + 2\frac{1}{k_s!}e^{-s\mu_W(x)(1-p_{x,y})}(s\mu_W(x)(1-p_{x,y}))^{k_s}(s\mu_W(x)) \tag{2.B.82}$$
$$\leqslant k_s + 2\frac{1}{k_s!}e^{-s\mu_W(x)(1-C\mu_W(y))}(s\mu_W(x)(1-C\mu_W(y)))^{k_s}(s\mu_W(x)) \tag{2.B.83}$$

The second term satisfies

$$\frac{2}{k_s!}e^{-s\mu_W(x)(1-p_{x,y})}(s\mu_W(x)(1-p_{x,y}))^{k_s}(s\mu_W(x)) \tag{2.B.84}$$
$$= 2\frac{k_s+1}{1-C\mu_W(y)}\Pr(\tilde{Z} = k_s + 1), \tag{2.B.85}$$

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

$$2\frac{k_s+1}{1-C\mu_W(y)}\Pr(\tilde{Z} = k_s+1) \leqslant 2(1-C\mu_W(y))^{k_s}\frac{1}{k_s!}e^{-2k_s(1-C\mu_W(y))}(2k_s)^{k_s+1} \tag{2.B.86}$$
$$\leqslant 4k_s 2^{k_s}e^{-k_s(1-2C\mu_W(y))}\left(\frac{1}{k_s!}{k_s}^{k_s}e^{-k_s}\right) \tag{2.B.87}$$
$$\leqslant 4k_s, \tag{2.B.88}$$

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

$$\Pr(Z \leqslant k_s)\mathbb{E}[Q] \leqslant 5k_s \tag{2.B.89}$$

for all values of x.

Putting together Eqs. (2.B.72), (2.B.74), (2.B.80) and (2.B.89):

$$\Pr(H + B_2 > k_s \wedge H \leqslant k_s \mid D_a > k_s) \tag{2.B.90}$$
$$\leqslant \Pr(H + B(y) > k_s \wedge H \leqslant k_s \mid D_a > k_s) \tag{2.B.91}$$

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

$$H \mid \{D_a > k_s\} \stackrel{d}{=} D_{1,s}(y) \tag{2.B.92}$$
$$B(y) \sim \text{Bin}(5k_s, C\mu_W(x)). \tag{2.B.93}$$

This completes the proof of the lemma. □

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

$$\text{var}\left[N^{(s)}_{>k_s}\right] \leqslant \mathbb{E}[N^{(s)}_{>k_s}](1 + \int g(y)\,dy). \tag{2.B.94}$$

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

Lemma 2.B.4. *Let $g(y)$ be as in Lemma 2.B.3 and suppose W is integrable. If the sequence k_s is bounded then*

$$\int g(y)\,dy = o(\mathbb{E}[N^{(s)}_{>k_s}]). \tag{2.B.95}$$

Proof. Let $T_s = \sqrt{\mathbb{E}[N^{(s)}_{>k_s}]}$ so that by Lemma 2.B.3 for s large enough

$$\int_{\mathbb{R}_+} g(y)\,dy \leqslant T_s + \int_{T_s}^{\infty} \Pr(D_{2,s}(y) + B(y) > k_s \wedge D_{2,s}(y) \leqslant k_s)\,dy. \tag{2.B.96}$$

Moreover

$$\Pr(D_{2,s}(y) + B(y) > k_s \wedge D_{2,s}(y) \leqslant k_s) \leqslant \Pr(\tilde{B}(y) > 1), \tag{2.B.97}$$

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

$$\Pr(\tilde{B}(y) > 1) \leqslant 5kC\mu_W(y) \tag{2.B.98}$$

so that

$$\int_{T_s}^{\infty} \Pr(\tilde{B}(y) > 1)\,dy \leqslant 5kC \int_{T_s}^{\infty} \mu_W(y)\,dy \tag{2.B.99}$$
$$= o(1), \tag{2.B.100}$$

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

Region of $\mathbb{R}_+$	Upper bound for $g(y)$
$[0, \mu^{-1}((1+\epsilon)\frac{k_s}{s})]$	$\Pr(D_{2,s}(y) \leqslant k_s)$
$(\mu^{-1}((1+\epsilon)\frac{k_s}{s}), \mu^{-1}((1-\epsilon)\frac{k}{s}))$	1
$(\mu^{-1}((1-\epsilon)\frac{k_s}{s}), \infty)$	$\Pr(B(y) > \frac{\epsilon}{2}k_s) + \Pr(D_{2,s}(y) > (1-\frac{\epsilon}{2})k_s)$

Table 1: Upper bounds on $g(y)$

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

$$\mathbb{E}[N_{>0}^{(s)}] = \Omega(\mu_W^{-1}(\frac{1}{s})), \tag{2.B.101}$$

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

Lemma 2.B.5. *For $0 < \epsilon < 1$,*

$$\int_0^{\mu_W^{-1}((1+\epsilon)\frac{k_s}{s})} \Pr(D_{2,s}(y) \leqslant k_s)\,dy \leqslant \frac{1+\epsilon}{\epsilon}(\frac{1+\epsilon}{e^\epsilon})^{k_s}\mu_W^{-1}(\frac{k_s}{s}). \tag{2.B.102}$$

Proof. Because $\Pr(D_{2,s}(y) \leqslant k_s)$ is monotonically increasing in y over the domain of integration, the integral is bounded by

$$\mu_W^{-1}((1+\epsilon)\frac{k_s}{s})\Pr(D_{2,s}(\mu_W^{-1}((1+\epsilon)\frac{k_s}{s})) \leqslant k_s). \tag{2.B.103}$$

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

$$\Pr(D_{2,s}(\mu_W^{-1}((1+\epsilon)\frac{k_s}{s})) \leqslant k_s) \leqslant (1+\frac{1}{\epsilon})\Pr(D_{2,s}(y) = k_s) \tag{2.B.104}$$

$$= (1+\frac{1}{\epsilon})\frac{1}{k_s!}((1+\epsilon)k_s)^{k_s}e^{-(1+\epsilon)k_s} \tag{2.B.105}$$

$$\leqslant \frac{1}{e}\frac{1+\epsilon}{\epsilon}(\frac{1+\epsilon}{e^\epsilon})^{k_s}. \tag{2.B.106}$$

□

For $y > \mu_W^{-1}((1-\epsilon)\frac{k_s}{s})$ we can bound $g(y)$ (and thus $\int_{\mu_W^{-1}((1-\epsilon)\frac{k_s}{s})}^{\infty} g(y)\,dy$) by

$$\Pr(D_{2,s}(y) + B(y) > k_s \wedge D_{2,s}(y) \leqslant k_s) \tag{2.B.107}$$
$$\leqslant \Pr(D_{2,s}(y) + B(y) > k_s \wedge D_{2,s}(y) \leqslant (1-\frac{\epsilon}{2})k_s) \tag{2.B.108}$$
$$+ \Pr(D_{2,s}(y) > (1-\frac{\epsilon}{2})k_s) \tag{2.B.109}$$
$$\leqslant \Pr(B(y) > \frac{\epsilon}{2}k_s) + \Pr(D_{2,s}(y) > (1-\frac{\epsilon}{2})k_s). \tag{2.B.110}$$

The next lemma controls the second term in this bound.

Lemma 2.B.6. *Suppose there is some $\chi > 0$ such that for all $x > \chi$ it holds that*

$$\frac{\mu_W(x)}{x\mu_W'(x)} \geqslant -1, \tag{2.B.111}$$

then, for s sufficiently large such that $\frac{k_s}{s} \leqslant \mu_W(\chi)$ and ϵ such that $0 < \epsilon < 1$,

$$\int_{\mu_W^{-1}((1-\epsilon)\frac{k_s}{s})}^{\infty} \Pr(D_{2,s}(y) > (1-\frac{\epsilon}{2})k_s)\,dy \leqslant \frac{2}{k_s\epsilon + 2}\mu_W^{-1}(\frac{k_s}{s}) \tag{2.B.112}$$

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

$$\Pr(D_{2,s}(y) > (1-\frac{\epsilon}{2})k_s) \tag{2.B.113}$$
$$\leqslant (\frac{1-\epsilon/2+1/k_s}{\epsilon/2+1/k_s})\frac{1}{\lfloor(1-\frac{\epsilon}{2})k_s\rfloor!}e^{-s\mu_W(y)}(s\mu_W(y))^{(1-\frac{\epsilon}{2})k_s}. \tag{2.B.114}$$

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

$$\int_{\mu_W^{-1}((1-\epsilon)\frac{k_s}{s})}^{\infty} e^{-s\mu_W(y)}(s\mu_W(y))^{(1-\frac{\epsilon}{2})k_s}\,dy = -\int_0^{(1-\epsilon)k_s} e^{-x}x^{(1-\frac{\epsilon}{2})k_s}\frac{1}{s}f'(\frac{x}{s})\,dx. \tag{2.B.115}$$

Let $z = f(x)$ and write

$$\frac{\mu_W(z)}{z\mu_W'(z)} = \frac{f'(x)x}{f(x)} \tag{2.B.116}$$

so by assumption for $x \leqslant \mu_W(\chi)$ holds that $x\frac{f'(x)}{f(x)} \geqslant -1$. Thus for s sufficiently large that $\frac{k_s}{s} \leqslant \mu_W(\chi)$ it holds that

$$-\int_0^{(1-\epsilon)k_s} e^{-x}x^{(1-\frac{\epsilon}{2})k_s}\frac{1}{s}f'(\frac{x}{s})\,dx \leqslant \int_0^{(1-\epsilon)k_s} e^{-x}x^{(1-\frac{\epsilon}{2})k_s-1}f(\frac{x}{s})\,dx. \tag{2.B.117}$$

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

$$(xf(x))' = f(x) + xf'(x) \tag{2.B.118}$$
$$= f(x)(1 + x\frac{f'(x)}{f(x)}) \tag{2.B.119}$$
$$\geqslant 0. \tag{2.B.120}$$

This implies

$$\int_0^{(1-\epsilon)k_s} e^{-x} x^{(1-\frac{\epsilon}{2})k_s - 1} f(\frac{x}{s})\, dx \leqslant (1-\epsilon)k_s f(\frac{k_s}{s}) \int_0^{(1-\epsilon)k_s} e^{-x} x^{(1-\frac{\epsilon}{2})k_s - 2}\, dx \tag{2.B.121}$$
$$\leqslant (1-\epsilon)k_s f(\frac{k_s}{s}) \Gamma((1-\frac{\epsilon}{2})k_s - 1) \tag{2.B.122}$$
$$= f(\frac{k_s}{s}) \Gamma((1-\frac{\epsilon}{2})k_s). \tag{2.B.123}$$

This establishes

$$\int_{\mu_W^{-1}((1-\epsilon)\frac{k_s}{s})}^{\infty} \Pr(D_{2,s}(y) > (1-\frac{\epsilon}{2})k_s)\, dy \leqslant (\frac{1-\epsilon/2+1/k_s}{\epsilon/2+1/k_s}) f(\frac{k_s}{s}) \frac{\Gamma((1-\frac{\epsilon}{2})k_s)}{\Gamma((1-\frac{\epsilon}{2})k_s+1)} \tag{2.B.124}$$
$$= \frac{1}{\epsilon/2+1/k_s} \frac{1}{k_s} f(\frac{k_s}{s}) \tag{2.B.125}$$

as claimed. □

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

Lemma 2.B.7. *Suppose there is some $\chi > 0$ such that, for all $x > \chi$,*

$$\frac{\mu_W(x)}{x\mu_W'(x)} \geqslant -1, \tag{2.B.126}$$

and let B and C be as in Lemma 2.B.3. For s sufficiently large such that $\frac{k_s}{s} \leqslant \mu_W(\chi)$ and ϵ such that $10C\frac{k_s}{s} \leqslant \epsilon < 1$,

$$\int_{\mu_W^{-1}((1-\epsilon)\frac{k_s}{s})}^{\infty} \Pr(B > \frac{\epsilon}{2}k_s)\, dy \leqslant (\frac{C}{10}\frac{\epsilon}{1-\epsilon}\frac{k_s}{s})^{\epsilon k_s/2} \frac{1}{\epsilon k_s/2 - 1} \mu_W^{-1}(\frac{(1-\epsilon)k_s}{s}). \tag{2.B.127}$$

Proof. The condition $10C\frac{k_s}{s} \leqslant \epsilon$ ensures that

$$C\mu_W(y) < \frac{\epsilon/2k_s}{5k_s}, \tag{2.B.128}$$

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

$$\Pr(B > \frac{\epsilon}{2}k_s) \leqslant \exp(-5k_s S(\frac{\epsilon/2k_s}{5k_s} \| C\mu_W(y))), \tag{2.B.129}$$

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).

$$S(\frac{\epsilon}{10}\|C\mu_W(y)) \geqslant \frac{\epsilon}{10}\log\frac{10}{C\epsilon}\frac{1}{\mu_W(y)}, \tag{2.B.130}$$

whence

$$\Pr(B > \frac{\epsilon}{2}k_s) \leqslant (\frac{C}{10}\epsilon)^{\epsilon k/2}\mu_W(y)^{\frac{k}{2}}. \tag{2.B.131}$$

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

$$\int_{\mu_W^{-1}((1-\epsilon)\frac{k_s}{s})}^{\infty} \mu_W(y)^{\epsilon k_s/2}\,dy = s^{-\epsilon k/2}\int_0^{(1-\epsilon)k_s} x^{\epsilon k_s/2}\frac{1}{s}f'(\frac{x}{s})\,dx. \tag{2.B.132}$$

Following the same reasoning as in the proof of Lemma 2.B.6,

$$x^2\frac{1}{s}f'(\frac{x}{s}) \leqslant (1-\epsilon)k_s f(\frac{(1-\epsilon)k_s}{s}) \tag{2.B.133}$$

on the domain of integration so,

$$s^{-\epsilon k/2}\int_0^{(1-\epsilon)k_s} x^{\epsilon k/2}\frac{1}{s}f'(\frac{x}{s})\,dx \tag{2.B.134}$$

$$\leqslant s^{-\epsilon k/2}(1-\epsilon)k_s f(\frac{(1-\epsilon)k_s}{s})[\frac{1}{\epsilon k_s/2-1}((1-\epsilon)k_s)^{\epsilon k_s/2-1}] \tag{2.B.135}$$

$$= (\frac{k_s}{s})^{\epsilon k_s/2}(1-\epsilon)^{\epsilon k_s/2}\frac{1}{\epsilon k_s/2-1}f(\frac{(1-\epsilon)k_s}{s}). \tag{2.B.136}$$

□

In particular, the last several lemmas combine to show that for $\epsilon_s \leqslant 1$ such that $\epsilon_s = \omega(\frac{1}{k_s})$ and $\epsilon_s = \omega(\frac{k_s}{s})$ it holds that

$$\int_0^{\mu_W^{-1}((1+\epsilon)\frac{k_s}{s})} g(y)\,dy + \int_{\mu_W^{-1}((1-\epsilon)\frac{k_s}{s})}^{\infty} g(y)\,dy = o(\mu_W^{-1}(\frac{1}{s})). \tag{2.B.137}$$

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

$$(\mu_W^{-1}(1+\epsilon)\frac{k_s}{s}, \mu_W^{-1}(1-\epsilon)\frac{k_s}{s}) \tag{2.B.138}$$

as a possible foil to $\int g(y)\,dy = o(\mathbb{E}[N^{(s)}_{>0}])$. In this regime we expect

$$g(y) = \Pr(D_{2,s}(y) + B(y) > k_s \wedge D_{2,s}(y) \leqslant k_s) \tag{2.B.139}$$

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

Lemma 2.B.8. *Suppose that μ_W is differentiable and that there is some $\chi > 0$ such that for all $x > \chi$ it holds that*

$$\frac{\mu_W(x)}{x\mu'_W(x)} \geqslant -1. \tag{2.B.140}$$

Then for $\epsilon > 0$ and s sufficiently large such that $(1+\epsilon)\frac{k_s}{s} \leqslant \mu_W(\chi)$, it holds that

$$\mu_W^{-1}((1-\epsilon)\frac{k_s}{s}) - \mu_W^{-1}((1+\epsilon)\frac{k_s}{s}) \leqslant 2\frac{\epsilon}{1-\epsilon}\mu_W^{-1}((1-\epsilon)\frac{k_s}{s}) \tag{2.B.141}$$

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 $(1-\epsilon)\frac{k_s}{s} \leqslant x^* \leqslant (1+\epsilon)\frac{k_s}{s}$ such that

$$\begin{aligned} f((1-\epsilon)\frac{k_s}{s}) - f((1+\epsilon)\frac{k_s}{s}) &= -2\epsilon\frac{k_s}{s}f'(x^*) && \text{(2.B.142)}\\ &= -2\epsilon\frac{k_s}{s}\frac{1}{x^*}x^*f'(x^*) && \text{(2.B.143)}\\ &\leqslant 2\frac{\epsilon}{1-\epsilon}f((1-\epsilon)\frac{k_s}{s}), && \text{(2.B.144)} \end{aligned}$$

where the final line follows as in Lemma 2.B.6. □

We can now complete our intermediate goal:

Lemma 2.B.9. *Let $g(y)$, T and C be as in Lemma 2.B.3. Suppose $k_s \uparrow \infty$ and $k_s = o(s)$. Suppose that μ_W is differentiable and that there is some $\chi > 0$ such that for all $x > \chi$ it holds that*

$$\frac{\mu_W(x)}{x\mu'_W(x)} \geqslant -1. \tag{2.B.145}$$

Then

$$\int g(y)\,dy = o(\mathbb{E}[N^{(s)}_{>0}]) \tag{2.B.146}$$

Proof. Let $\epsilon_s \downarrow 0$ such that $\epsilon_s = \omega(\sqrt{\frac{1}{k_s}})$ and $\epsilon_s = \omega(\sqrt{\frac{k_s}{s}})$. Let

$$h(y) = \begin{cases} \Pr(D_{2,s}(y) \leqslant k_s) & y \leqslant \mu_W^{-1}((1+\epsilon_s)\frac{k_s}{s}) \\ 1 & y \in (\mu_W^{-1}((1+\epsilon_s)\frac{k_s}{s}), \mu_W^{-1}((1-\epsilon_s)\frac{k_s}{s})) \\ \Pr(B(y) > \frac{\epsilon}{2}k_s) + \Pr(D_{2,s}(y) > (1-\frac{\epsilon}{2})k_s) & y \geqslant \mu_W^{-1}((1-\epsilon_s)\frac{k_s}{s}). \end{cases} \tag{2.B.147}$$

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

$$g(y) \leqslant h(y). \tag{2.B.148}$$

Moreover, it is straightforward to verify that the conditions on ϵ_s with Lemmas 2.B.5 to 2.B.8 imply

$$\int h(y)\,dy = o(\mu_W^{-1}((1-\epsilon_s)\frac{k_s}{s})). \tag{2.B.149}$$

(For Lemma 2.B.5 it suffices to consider the worst case $\epsilon_s = \sqrt{\frac{1}{k_s}}$.)

Next,

$$\begin{aligned}
\mathbb{E}[N^{(s)}_{>0}] &= \int_{\mathbb{R}_+} 1 - e^{-s\mu_W(y)}\,dy && (2.B.150)\\
&\geqslant \int_0^{\mu_w^{-1}(\frac{1}{s})} 1 - e^{-1}\,dy && (2.B.151)\\
&= \Omega(\mu_W^{-1}(\frac{1}{s})). && (2.B.152)
\end{aligned}$$

Thus $\mathbb{E}[N^{(s)}_{>0}] = \Omega(\mu_W^{-1}((1-\epsilon_s)\frac{k_s}{s}))$, completing the proof. □

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

Proof of Theorem 2.5.4. By Lemma 2.B.1 it suffices to show $\operatorname{var}\left[N^{(s)}_{>k_s}\right] = o(\mathbb{E}[N^{(s)}_{>0}]^2)$. By Lemmas 2.B.2 and 2.B.3,

$$\operatorname{var}\left[N^{(s)}_{>k_s}\right] \leqslant \mathbb{E}[N^{(s)}_{>k_s}](1 + \int g(y)\,dy), \tag{2.B.153}$$

where $g(y)$ is as defined in Lemma 2.B.3. Lemma 2.B.4, for bounded k_s, and Lemma 2.B.9, for $k_s \uparrow \infty$, establish

$$\int g(y)\,dy = o(\mathbb{E}[N^{(s)}_{>0}]), \tag{2.B.154}$$

completing the proof. □

CONNECTIVITY FOR PRODUCT FORM KEGS

Definition 2.C.1. We call a KEG *product form* if the associated graphex has $I = S = 0$ and W of the form

$$W(x,y) = \begin{cases} 0 & x = y \\ f(x)f(y) & \text{otherwise.} \end{cases} \tag{2.C.1}$$

The purpose of this section is to prove Theorem 2.5.5: product form KEGs have an arbitrarily large fraction of the vertices contained in a single connected component in the large graph limit.

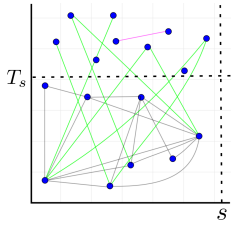

Figure 5: The basic structure of product form KEGs. The induced subgraph below T_s in gray is fully connected. Above T_s the vast majority of the vertices of the graph connect to the below threshold subgraph, in green. This leaves only the very small number of vertices connected only to vertices that lie entirely above T_s, in magenta.

Remark 2.C.2. Product form 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 s.

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. [Bolo1]) 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. 5. The first piece intuition is that for fixed s we can set the threshold T_s such that nearly every point of the latent Poisson process with ϑ below T_s will have an edge connected to it; because of this we can treat the connectivity of the below T_s induced subgraph using the traditional random graph machinery. The connectivity of vertices lying above T_s that participate in at least one edge connecting below T_s then follows straightforwardly. This leaves only the vertices in the induced subgraph above T_s that do not connect to a point below T_s 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 $\Pi_s = \{(\theta_i, \vartheta_i) \in \Pi \mid \theta_i \leqslant s\}$ be the restriction of this process to label-space $\leqslant s$. Let the Poisson process below a cutoff value x in ϑ space be $\Pi_{s,\leqslant x} = \{(\theta_i, \vartheta_i) \in \Pi_s \mid \vartheta_i < x\}$ and let the process above the cutoff be $\Pi_{s,>x} = \{(\theta_i, \vartheta_i) \in \Pi_s \mid \vartheta_i > x\}$.

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

Lemma 2.C.3. *Let $W(x,y) = f(x)f(y)1[x \neq y]$ be a product form (Kallenberg) graphon, then there is some other product form (Kallenberg) graphon $W' = h(x)h(y)1[x \neq y]$ such that h is monotone decreasing and the KEGs generated by 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 $f(x)$ is compactly supported (i.e., the dense graph case) 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 $f(x)$ is not compactly supported. Because f is integrable, 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. Let c_k be the measure of A_k, and put $s_k = \sum_{j=1}^{k} c_j$. By [Ker14, Thm. A.20], for each k such that $c_k > 0$, there is a measure preserving transformation $\tilde{\tau}_k$ such that $\tilde{\tau}_k[0, c_k] = A_k$. When $c_k = 0$, define $\tilde{\tau}_k$ on $\{0\}$ arbitrarily. Let $C_k = s_{k-1} + [0, c_k)$, let $Z = \{x \in \mathbb{R}_+ : f(x) > 0\}$, and define $\tau : \mathbb{R}_+ \to Z$ by $\tau(x) = \tilde{\tau}_k(x - s_{k-1})$ when $x \in C_k$. By construction, τ is measure-preserving. Moreover, τ imposes a pseudo-monotonicity where $f(\tau(x)) < \frac{1}{k}$ and $f(\tau(y)) \geqslant \frac{1}{k}$ implies $x > y$. By [Lov13, Prop. A19], for each k such that $c_k > 0$, there is a measure preserving transformation $\varphi_k : C_k \to C_k$ and a monotonically decreasing function h_k on C_k such that $1_{C_k} f \circ \tau = h_k \circ \varphi_k$. Taking $\varphi = \sum_{k=1}^{\infty} 1_{C_k} \varphi_k$ and $h = \sum_{k=1}^{\infty} 1_{C_k} h_k$, completes the proof. □

We take f to be monotone decreasing for the remainder of the section. Because the result is trivial for f compactly supported (the KEG is dense), we also take f to not have compact support. 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 $l_s \in o(1)$ it holds that $f^{-1}(l_s) \in \omega(1)$. Let G be a Kallenberg Exchangeable Graph associated with W and let Γ_s be the restriction to $[0, s]$.

Definition 2.C.4. Let t_s be a function of s such that $t_s \in o(1)$ and $t_s \in \omega(\frac{1}{s})$ and define the *threshold* $T_s = f^{-1}(\frac{1}{s} + t_s)$.

Remark 2.C.5. This notation for the threshold suppresses the dependence on t_s, which should be thought of as going to 0 as quickly as possible consistent with $t_s \in \omega(\frac{1}{s})$.

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 2.C.6)

2. We show that nearly every point of $\Pi_{s,\leqslant T_s}$ participates in an edge connecting to the connected core (Lemma 2.C.7)

3. We lower bound the number of points of $\Pi_{s,>T_s}$ that connect to the connected core (Lemma 2.C.8)

4. We consider the induced subgraph of Γ_s given by $\{\theta_i \in \nu(\Gamma_s) \mid \vartheta_i > T_s\}$ 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 2.C.10)

The first step of the proof is to show that there is an induced subgraph P_s 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, $\sum_{p\in P_s} f(p)$, is an arbitrarily large fraction of the total expected mass in the entire graph: $\mathbb{E}[\sum_{\vartheta_i\in\Pi_s} f(\vartheta_i)] = s\|f\|_1$. The critical fact for use in later parts of the argument turns out to be that the mass of the popularity subgraph scales as s.

Lemma 2.C.6. *Suppose f does not have compact support. Let $T_{s,pop} = f^{-1}(\sqrt{\frac{\log s}{s}})$ and let P_s be the induced subgraph of Γ_s given by including only vertices in $\Pi_{s,<T_{s,}pop}$, then:*

1. *Every element of $\Pi_{s,\leqslant T_{s,}pop}$ connects to an edge; $\lim_{s\to\infty} \left|\Pi_{s,\leqslant T_{s,}pop} \setminus \nu(P_s)\right| = 0$ a.s.*

2. *P_s is almost surely connected; let $C(P_s) = 1$ if P_s is connected and 0 otherwise, then $\lim_{s\to\infty} C(P_s) = 1$ a.s.*

3. *P_s is "ultra-popular" almost surely; letting $S_s = \sum_{p\in P_s} f(p)$ we have for $\epsilon > 0$ that $\lim_{s\to\infty} \frac{S_s}{s} \geqslant (1-\epsilon)\|f\|_1$ a.s.*

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

$$\lim_{s\to\infty} \frac{p_s}{\log N_s/N_s} > 1+\delta \text{ a.s.} \tag{2.C.2}$$

For arbitrary $\gamma > 0$, it holds that $\lim_{s\to\infty} N_s/sT_{s,}pop \geqslant (1-\gamma)$ a.s. and so we have that:

$$\lim_{s\to\infty} \frac{p_s}{\log N_s/N_s} \geqslant \lim_{s\to\infty} \frac{\log s/s}{\log(1-\gamma)sT_{s,}pop/(1-\gamma)sT_{s,}pop} \text{ a.s.} \tag{2.C.3}$$

$$= \infty. \tag{2.C.4}$$

Thus in the limit as $s \to \infty$, the random graph with vertices $\Pi_{s,\leqslant T_{s,}\mathrm{pop}}$ 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_s grows as claimed. For $\gamma > 0$, by Hoeffding's inequality we have:

$$\Pr(S_s < (1-\gamma)\mathbb{E}[S_s \mid N_s] \mid N_s) \leqslant \Pr(|S_s - \mathbb{E}[S_s \mid N_s]| < \gamma\mathbb{E}[S_s \mid N_s] \mid N_s) \tag{2.C.5}$$

$$\leqslant 2\exp(-2\gamma^2 \frac{\mathbb{E}[S_s \mid N_s]^2}{N_s}) \tag{2.C.6}$$

$$= 2\exp(-2\gamma^2 \frac{N_s}{T_{s,}^2\mathrm{pop}}(\int_0^{T_{s,}\mathrm{pop}} f(x)\,dx)^2) \tag{2.C.7}$$

$$\leqslant 2\exp(-2\gamma^2 \frac{N_s}{T_{s,}^2\mathrm{pop}}(1-\gamma)^2\|f\|_1^2), \tag{2.C.8}$$

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

$$\Pr(\frac{S_s}{s(1-\gamma)^2\|f\|_1} < 1-\gamma \mid N_s \geqslant (1-\gamma)sT_{s,}\mathrm{pop}) \tag{2.C.9}$$

$$\leqslant \Pr(\frac{S_s}{\mathbb{E}[S_s \mid N_s]} < 1-\gamma \mid N_s \geqslant (1-\gamma)sT_{s,}\mathrm{pop}) \tag{2.C.10}$$

$$\leqslant 2\exp(-2\gamma^2 \frac{s}{T_{s,}\mathrm{pop}}(1-\gamma)^3\|f\|_1^2). \tag{2.C.11}$$

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

$$\Pr(\frac{S_s}{s(1-\gamma)^2\|f\|_1} < 1-\gamma \mid N_s \geqslant (1-\gamma)sT_{s,}\mathrm{pop}) \tag{2.C.12}$$

$$\leqslant 2\exp(-2\gamma^2(s\log s)^{1/2}(1-\gamma)^3\|f\|_1^2). \tag{2.C.13}$$

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

$$\lim_{s\to\infty} \frac{S_{\lfloor s \rfloor}}{\lfloor s \rfloor + 1} \geqslant (1-\gamma)^3\|f\|_1 \text{ a.s.} \tag{2.C.14}$$

$$\implies \lim_{s\to\infty} \frac{S_s}{s} \geqslant (1-\gamma)^3\|f\|_1 \text{ a.s.} \tag{2.C.15}$$

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

We now have a promise that every point of the latent Poisson process $\Pi_{s,\leqslant T_{s,}\mathrm{pop}}$ participates in the graph. We now establish that, with

high probability, as $s \to \infty$ an arbitrarily large fraction of the points in $\Pi_{s,\leqslant T_s}$ connect to the popular connected core P_s. In particular, this means an arbitrarily large fraction of the points of $\Pi_{s,\leqslant T_s}$ participate in a single connected component of Γ_s.

Lemma 2.C.7. *Suppose f does not have compact support. Let a point $(\theta_i, \vartheta_i) \in \Pi_{s,\leqslant T_s}$ be visible if $\theta_i \in \nu(\Gamma_s)$ and it participates in an edge connecting to P_s, and call a point invisible otherwise. Let $N_{\text{invis},\leqslant T_s}$ be the number of points in $\Pi_{s,\leqslant T_s}$ that are invisible and let $N_{vis,\leqslant T_s}$ be the number of points in $\Pi_{s,\leqslant T_s}$ that are visible, then for $\epsilon > 0$*

$$\lim_{s\to\infty} P(N_{\text{invis},<T_s} > \epsilon N_{vis,<T_s}) = 0. \tag{2.C.16}$$

Proof. By Lemma 2.C.6 it follows that as $s \to \infty$ there are no invisible vertices below $T_{s,\text{pop}} = f^{-1}(\sqrt{\frac{\log s}{s}})$ so it suffices to bound the number of invisible vertices between $T_{s,\text{pop}}$ and T_s. Conditional on P_s, each point $(\theta_i, \vartheta_i) \in \Pi_{s,>T_{s,\text{pop}}}$ connects to P_s independently with probability $1 - \prod_{p\in P_s}(1 - f(\vartheta_i)f(p)) \geqslant 1 - e^{-f(\vartheta_i)S_s}$ where $S_s = \sum_{p\in P_s} f(p)$. Since labeling each point of the Poisson process $\Pi_{s,>T_{s,\text{pop}}}$ by whether or not it connects to $T_{s,\text{pop}}$ is, conditional on P_s, a marking of the Poisson process, we immediately have that the number of visible and invisible points in $\{(\theta_i, \vartheta_i) \in \Pi_s \mid T_{s,\text{pop}} < T_s\}$ are independent random variables and that there exists random variables $N_{s,\text{ub}}$ and $N_{s,\text{vis}}$ such that,

$$N_{s,\text{ub}} \sim \text{Poi}(s \int_{T_{s,\text{pop}}}^{T_s} e^{-f(x)S_s}\, dx) \tag{2.C.17}$$

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

$$N_{s,\text{vis}} \sim \text{Poi}(s \int_{T_{s,\text{pop}}}^{T_s} 1 - e^{-f(x)S_s}\, dx) \tag{2.C.18}$$

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

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

$$\frac{s\int_{T_{s,\text{pop}}}^{T_s} e^{-f(x)S_s}\, dx}{s\int_{T_{s,\text{pop}}}^{T_s} 1 - e^{-f(x)S_s}\, dx} \leqslant \frac{(T_s - T_{s,\text{pop}})e^{-f(T_s)S_s}}{(T_s - T_{s,\text{pop}})} \tag{2.C.19}$$

$$\leqslant e^{-(\frac{1}{s}+t_s)S_s}. \tag{2.C.20}$$

Invoking $\lim_{s\to\infty} S_s/s \geqslant \frac{1}{2}\|f\|_1 = 1$ a.s. from Lemma 2.C.6 completes the result since this means $\lim_{s\to\infty} t_s S_s = \infty$ a.s. □

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

Lemma 2.C.8. *Suppose f does not have compact support. Let*

$$N_{s,>T_s} = |\{(\theta_i, \vartheta_i) \in \Pi_{s,>T_s} \mid \exists p \in \nu(P_s) \text{ such that } \{\theta_i, p\} \in e(\Gamma_s)\}|, \tag{2.C.21}$$

be the number of points above T_s that connect to P_s. Then there exists a random variable $N_{s,+}$ such that $N_{s,+} \leqslant N_{s,>T_s}$ and

$$N_{s,+} \mid S_s \sim \text{Poi}(s \int_{T_s}^{\infty} 1 - e^{-f(x)S_s}\, dx) \tag{2.C.22}$$

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

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

Lemma 2.C.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 $f^{-1}(\frac{1}{t}) \geqslant ct$ 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 $\left|B_{t_i} \cap B_{t_{i+1}}\right| \leqslant \epsilon$, 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_s as a cutoff below which every candidate vertex participates in the graph, the requirement $T_s = o(s)$ is obvious. Suppose otherwise, then there would be $\Omega(s^2)$ visible vertices in the graph and $\Theta(s^2)$ expected edges, pushing the graph into the ultra-sparse regime where $|e(\Gamma_s)| = O(|\nu(\Gamma_s)|)$. The above lemma shows that $T_s = o(s)$ does indeed hold, since $T_s = o(f^{-1}(1/s))$ and $f^{-1}(1/s) = o(s)$. With this result in hand,

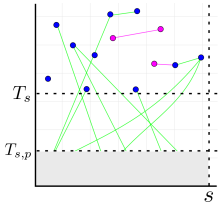

Figure 6: The structure of negligible vertices above T_s. Vertices with distance > 2 to P_s (below $T_{s,p}$) are ignored, these are marked in magenta.

Lemma 2.C.10. *Suppose f does not have compact support. Call a vertex $\theta_i \in \nu(\Gamma_s)$ ignored if $(\theta_i, \vartheta_i) \in \Pi_{s,>T_s}$ and its distance to P_s is greater than 2. Let N_{ignore} be the number of ignored vertices; then fixing $\epsilon > 0$,*

$$\lim_{s\to\infty} \Pr\left(\frac{N_{ignore}}{|\nu(\Gamma_s)|} > \epsilon\right) \to 0. \tag{2.C.23}$$

Proof. We mark each point in the Poisson process $\Pi_{s,>T_s}$ above T_s by whether it participates in an edge with a terminus in P_s. As in Lemma 2.C.8, this forms a marking of the Poisson process conditional on P_s so that the random subset of $\Pi_{s,>T_s}$ that is at distance one (**c**lose) to P_s,

$$C_s = \{(\theta_i, \vartheta_i) \in \Pi_{s,>T_s} \mid \exists p \in \nu(P_s) \text{ such that } \{\theta_i, p\} \in e(\Gamma_s)\}, \tag{2.C.24}$$

and the remaining subset $\Pi_{s,>T_s} \setminus C_s$ are independent Poisson processes conditional on P_s.

Let $e_{s,\text{ignore}}$ be the the number of edges in the induced subgraph of Γ_s given by restricting the vertex set to $\Pi_{s,>T_s} \setminus C_s$. It is immediate that $N_{\text{ignore}} \leqslant 2e_{s,\text{ignore}}$ (see Fig. 6). Obviously $|\nu(\Gamma_s)| > |C_s|$ and by Lemma 2.C.8 $|C_s| > N_{s,+}$ so

$$\Pr\left(\frac{N_{\text{ignore}}}{|\nu(\Gamma_s)|} > \epsilon\right) \leqslant \Pr\left(\frac{2e_{s,\text{ignore}}}{N_{s,+}} > \epsilon\right), \tag{2.C.25}$$

where in particular $e_{s,\text{ignore}}$ and $N_{s,+}$ are independent conditional on $\Pi_{s,\leqslant T_s}$.

We have very little distributional information about $e_{s,\text{ignore}}$ so we use Markov's inequality. Since $\Pi_{s,>T_s} \setminus C_s$ is a Poisson process with rate at most $se^{-f(x)S_s}$ we may repeat the argument of Theorem 2.5.1 to bound $\mathbb{E}[e_{s,\text{ignore}} \mid P_s]$ so that

$$\mathbb{E}\left[\frac{e_{s,\text{ignore}}}{N_{s,+}} \mid \Pi_{s,\leqslant T_s}\right] = \frac{\mathbb{E}[e_{s,\text{ignore}} \mid \Pi_{s,\leqslant T_s}]}{\mathbb{E}[N_{s,+} \mid \Pi_{s,\leqslant T_s}]} \tag{2.C.26}$$

$$\leqslant 2\frac{s^2\left(\int_{T_s}^{\infty} e^{-2S_s f(x)} f(x)\,dx\right)^2}{s\int_{T_s}^{\infty} 1 - e^{-f(x)S_s}\,dx}. \tag{2.C.27}$$

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

$$\mathbb{E}[\frac{e_{s,\text{ignore}}}{N_{s,+}} \mid (1-\gamma) \leqslant \frac{S_s}{s\|f\|_1} \leqslant (1+\gamma)] \leqslant 2s\frac{(\int_{T_s}^{\infty} e^{-2(1+\gamma)\|f\|_1 s f(x)} f(x)\,dx)^2}{\int_{T_s}^{\infty} 1 - e^{-(1-\gamma)\|f\|_1 s f(x)}\,dx}. \tag{2.C.28}$$

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

$$\int_{T_s}^{T_{s,u}} e^{-2(1+\gamma)\|f\|_1 s f(x)} f(x)\,dx + \int_{T_{s,u}}^{\infty} e^{-2(1+\gamma)\|f\|_1 s f(x)} f(x)\,dx \tag{2.C.29}$$

$$\leqslant O(\frac{T_{s,u} - T_s}{s}) + O(\int_{T_{s,u}}^{\infty} f(x)\,dx), \tag{2.C.30}$$

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

$$\int_{T_s}^{T_{s,u}} 1 - e^{-(1-\gamma)\|f\|_1 s f(x)}\,dx + \int_{T_{s,u}}^{\infty} 1 - e^{-(1-\gamma)\|f\|_1 s f(x)}\,dx \tag{2.C.31}$$

$$\geqslant \Omega(T_{s,u} - T_s) + \Omega(s\int_{T_{s,u}}^{\infty} f(x)\,dx), \tag{2.C.32}$$

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 $1 - e^{-cx} \geqslant L_c x$ for $x < 1$. Thus, in particular,

$$\mathbb{E}[\frac{e_{s,\text{ignore}}}{N_{s,+}} \mid (1-\gamma) \leqslant \frac{S_s}{s\|f\|_1} \leqslant (1+\gamma)] \tag{2.C.33}$$

$$= O\left(s(\frac{T_{s,u} - T_s}{s})^2 \frac{1}{T_{s,u} - T_s}, s(\int_{T_{s,u}}^{\infty} f(x)\,dx)^2 \frac{1}{s\int_{T_{s,u}}^{\infty} f(x)\,dx}\right), \tag{2.C.34}$$

$$\tag{2.C.35}$$

and this goes to 0 as $s \to \infty$; the left term because $T_{s,u} = o(s)$ by Lemma 2.C.9 and the right term because f is integrable and $T_{s,u} \to \infty$.

Putting all of this together and using that $(1-\gamma) \leqslant \lim_{s\to\infty} \frac{S_s}{s\|f\|_1} \leqslant (1+\gamma)$ a.s. by Lemma 2.C.6 we have that:

$$\lim_{s\to\infty} \Pr(\frac{2e_{s,\text{ignore}}}{N_{s,+}} > \epsilon) = \lim_{s\to\infty} \Pr(\frac{2e_{s,\text{ignore}}}{N_{s,+}} > \epsilon \mid (1-\gamma) \leqslant \frac{S_s}{s\|f\|_1} \leqslant (1+\gamma)) \tag{2.C.36}$$

$$\leqslant \lim_{s\to\infty} 2\mathbb{E}[\frac{e_{s,\text{ignore}}}{N_{s,+}} > \epsilon \mid (1-\gamma) \leqslant \frac{S_s}{s\|f\|_1} \leqslant (1+\gamma)] \tag{2.C.37}$$

$$= 0, \tag{2.C.38}$$

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

□

We can now put all of this together:

Theorem 2.5.5. *Let* G *be a KEG generated by graphex* $(0, 0, W)$ *with* $W = f(x)f(y)1[x \neq y]$, *let* $C_1(\Gamma_s)$ *be the largest connected component of* Γ_s, *and let* $\epsilon > 0$, *then*

$$\lim_{s\to\infty} \Pr(|C_1(\Gamma_s)| > (1-\epsilon)\,|v(\Gamma_s)|) = 1. \tag{2.5.4}$$

Proof. For f with compact support this is a trivial consequence of Theorem 2.5.3, 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 product form 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 product form KEGs beyond connectivity, in particular:

1. If two points of a product form 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_s by paths of length at most 2.

2. Although vertices of Γ_s chosen uniformly at random are overwhelmingly likely to follow a degree distribution of the type given in Theorem 2.5.4 there are a vanishingly small fraction of the vertices (those in P_s) with much higher degree.

3

SAMPLING AND ESTIMATION FOR (SPARSE) EXCHANGEABLE GRAPHS

INTRODUCTION

In this chapter, we continue our study of sparse exchangeable graphs, i.e., random graphs whose vertices may be identified with nonnegative reals, $\mathbb{R}_+$, and whose edge sets are then modeled by exchangeable point processes on $\mathbb{R}_+^2$. In a pioneering paper, Caron and Fox [CF14] introduced the notion of sparse exchangeable graphs in the context of nonparametric Bayesian analysis. Building on this work, the general family of all sparse exchangeable graphs was characterized by Veitch and Roy [VR15] and Borgs, Chayes, Cohn, and Holden [BCCH16], and shown to generalize the graphon models for dense graphs to include the sparse graph regime. Sparse exchangeable graphs have a number of desirable properties, including that they define a natural projective family of subgraphs of growing size, which can be used to model the process of observing a larger and larger fraction of a fixed underlying network. This property also provides a firm foundation for the study of various asymptotics, as demonstrated by [VR15; BCCH16] and the results here. In the previous chapter, we characterized the asymptotic degree distribution and connectedness of sparse exchangeable graphs, demonstrating that sparse exchangeable graphs allow for sparsity and admit the rich graph structure (such as small-world connectivity and power law degree distributions) found in large real-world networks. On this basis, we argued that sparse exchangeable graphs can serve as a general statistical model for network data. Despite our understanding of these models, the statistical meaning remains somewhat opaque. Put simply, when would it be natural to use the sparse exchangeable graph model?

The present chapter further develops this framework for statistical network analysis by answering two fundamental questions:

1. What is the notion of sampling naturally associated with this statistical network model? and

2. How can we use an observed dataset to consistently estimate the statistical network model?

The answers to these questions significantly clarify both the meaning of the modeling framework, and its connection to the dense graph framework and the classical i.i.d. sequence framework at the foundation of classical statistics. These questions may be viewed as specific

examples of a general approach to formalizing the problem of statistical analysis on network data being carried out by Orbanz [Orb17].

Although the results of the present chapter hold for general graphexes, for simplicity of exposition, we will temporarily restrict our attention to graphexes of the form $\mathcal{W} = (0, 0, W)$, giving a full treatment in subsequent sections. For a finite labeled graph G, such as each Γ_s, for $s \in \mathbb{R}_+$, we will write $\mathcal{G}(G)$ to denote the *unlabeled*[1] graph corresponding to G.

The first contribution of the present chapter is the identification of a sampling scheme that is naturally associated with the Kallenberg exchangeable graphs:

Definition 3.1.1. A p-*sampling* of an unlabeled graph G is obtained by selecting each vertex of G independently with probability $p \in [0, 1]$, and then returning the edge set of the random vertex-induced subgraph of G.

It is important to note that only the edge set of the vertex-induced subgraph is returned; in other words, vertices that are isolated from the other sampled vertices are thrown away. The key fact about this sampling scheme is that: For $s > 0$ and $r \in [0, s]$, if G_r is an r/s-sampling of $\mathcal{G}(\Gamma_s)$ then $G_r \stackrel{d}{=} \mathcal{G}(\Gamma_r)$. This result justifies the interpretation of the parameter s as a sample size.

In the estimation problem for the Kallenberg exchangeable graph, the observed dataset is a realization of the random sequence of graphs $G_1, G_2, \ldots$ such that $G_k = \mathcal{G}(\Gamma_{s_k})$, and $s_1, s_2, \ldots$ is some sequence of sizes such that $s_k \uparrow \infty$ as $k \to \infty$. The task is to take such an observation and return an estimate for $\mathcal{W}$, where $\mathcal{W}$ is the graphex that generated $(\Gamma_s)_{s \in \mathbb{R}_+}$. Both the formulation and solution of this problem depend on whether the sizes s_k are included as part of the observations.

We first treat the simpler case where the sizes are known. To formalize the estimation problem we must introduce a notion of when one graphex is a good approximation for another. Intuitively, our notion is that, for any fixed s, a size-s random graph generated by an estimator should be close in distribution to a size-s random graph generated by the true graphex. Let $\mathrm{GPD}(\mathcal{W}, s)$ be the distribution of an unlabeled size-s Kallenberg exchangeable graph, i.e., the distribution of $\mathcal{G}(\Gamma_s)$ where Γ is generated by $\mathcal{W}$. Approximation is then formalized by the following notion of convergence:

Definition 3.1.2. Write $\mathcal{W}_k \to_{\mathrm{GP}} \mathcal{W}$ as $k \to \infty$, when $\mathrm{GPD}(\mathcal{W}_k, s) \to \mathrm{GPD}(\mathcal{W}, s)$ weakly as $k \to \infty$, for all $s \in \mathbb{R}_+$.

1 The unlabelled graph corresponding to a labelled graph G is the equivalence class of graphs isomorphic to G. Restricting ourselves to finite unlabelled graphs, we can represent the unlabelled graphs formally in terms of their homomorphism counts, (N_F), where F ranges over the countable set of all finite simple graphs whose vertex set is $[n]$ for some $n \in \mathbb{N}$, and N_F is the number of homomorphisms from F to G.

Our goal in the estimation problem is then to take a sequence of observations and use these to produce a sequence of graphexes $\mathcal{W}_1, \mathcal{W}_2, \ldots$ that are consistent in the sense that $\mathcal{W}_k \to_{GP} \mathcal{W}$ as $k \to \infty$. This is a natural analogue of the definition of consistent estimation used for the convergence of the empirical cumulative distribution function in the i.i.d. sequence setting, and of the definition of consistent estimation used for the convergence of the empirical graphon in the dense graph setting.

Let $v(G)$ denote the number of vertices of graph G. Our estimator is the *dilated empirical graphon*

$$\hat{W}_{(G_k, s_k)} : [0, v(G_k)/s_k)^2 \to \{0, 1\}, \tag{3.1.1}$$

defined by transforming the adjacency matrix of G_k into a step function on $[0, v(G_k)/s_k)^2$ where each pixel has size $1/s_k \times 1/s_k$; see Fig. 7. Intuitively, when the generating graphex is $\mathcal{W} = (0, 0, W)$, we have $s_k \uparrow \infty$ as $k \to \infty$, and the estimator is an increasingly higher and higher resolution pixel picture of the generating graphon. Formally, given a non-empty finite graph G with n vertices labeled $1, \ldots, n$, we define the empirical graphon $\tilde{W}_G : [0, 1]^2 \to \{0, 1\}$ by partitioning $[0, 1]$ into adjacent intervals $I_1, \ldots, I_n$ each of length $1/n$ and taking $\tilde{W}_G = 1$ on $I_i \times I_j$ if i and j are connected in G, and taking $\tilde{W} = 0$ otherwise. The dilated empirical graphon with dilation s is then defined by $\hat{W}_{(G,s)}(x, y) = \tilde{W}_G(x/s, y/s)$. To map an unlabeled graph to a (dilated) empirical graphon we must introduce a labeling of the vertices. Notice that if $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ is a measure-preserving transformation, $\phi \otimes \phi$ is the map $(\phi \otimes \phi)(x, y) = (\phi(x), \phi(y))$, $\mathcal{W} = (I, S, W)$, and $\mathcal{W}' = (I, S \circ \phi, W \circ (\phi \otimes \phi))$, then $\mathrm{GPD}(\mathcal{W}, s) = \mathrm{GPD}(\mathcal{W}', s)$ for all $s \in \mathbb{R}_+$. In particular, the dilated empirical graphon functions corresponding to different labelings of the vertices of G are related by obvious measure-preserving transformations in this way. For the purposes of this chapter, graphexes that give rise to the same distributions over graphs are equivalent. We then define the empirical graphon of an unlabeled graph to be the empirical graphon of that graph with some arbitrary labeling, and we define the dilated empirical graphon similarly. These functions may be thought as arbitrary representatives of the equivalence class on graphons given by equating two graphons whenever they correspond to isomorphic graphs.

The first main estimation result is that $\hat{W}_{(G_k, s_k)} \to_{GP} W$ in probability as $k \to \infty$. That is, for every infinite sequence $N \subseteq \mathbb{N}$ there is a further infinite subsequence $N' \subseteq N$ such that $\hat{W}_{(G_k, s_k)} \to_{GP} W$ almost surely along N'. Subject to an additional technical constraint (implied by integrability of W) the convergence in probability may be replaced by convergence almost surely. Note that consistency holds for observations generated by an arbitrary graphex $\mathcal{W} = (I, S, W)$, not just those with the form $\mathcal{W} = (0, 0, W)$; see Fig. 8.

We now turn to the setting where the observation sizes $s_1, s_2, \ldots$ are not included as part of the observations. In this case, we study two

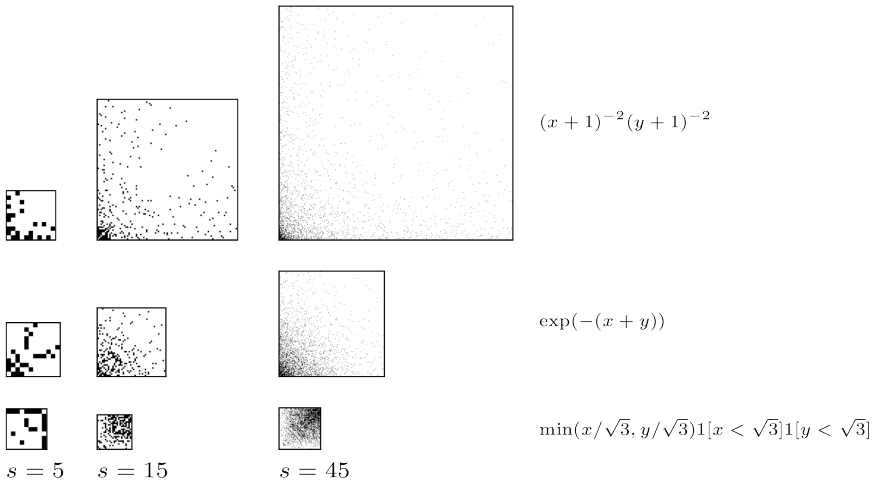

Figure 7: Realizations of dilated empirical graphons of Kallenberg exchangeable graphs generated by $(0, 0, W)$ for W given in the rightmost column, at observation sizes given in the bottom row. Note that the ordering of the vertices used to define the estimator is arbitrary. Here we have suggestively ordered the vertices according to the latent values from the process simulations; with this ordering the dilated empirical graphons are approximate pixel pictures of the generating graphon where the resolution becomes finer as the observation size grows. All three graphons satisfy $\|W\|_1 = 1$, and thus the expected number of edges (black pixels) at each size s is $\frac{1}{2}s^2$ in each column. Note that the rate of dilation is faster for sparser graphs; as established in [VR15], the topmost Kallenberg exchangeable graph used for this example is sparser than the middle Kallenberg exchangeable graph, and the graphon generating the bottom Kallenberg exchangeable graph is compactly supported and thus corresponds to a dense graph.

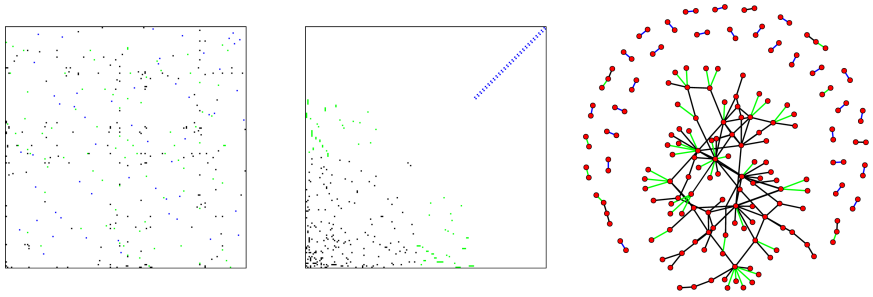

Figure 8: Realization of unlabeled Kallenberg exchangeable graph generated by $\mathcal{W} = (I, S, W)$ at size $s = 15$ (right panel), and associated dilated empirical graphon (left and center panels). The generating graphex is $W = (x+1)^{-2}(y+1)^{-2}$, $S = 1/2\exp(-(x+1))$, and $I = 0.1$. The observation size is $s = 15$. The dilated empirical graphex is pictured as two equivalent representations $\hat{W}_{(G,15)}$ and $\hat{W}'_{(G,15)}$, each with support $[0, 12)^2$ (180 vertices at size 15). Edges from the W component are shown in black, edges from the S component are shown in green, and edges from the I component are shown in blue. Recall that the ordering of the dilated empirical graphon is arbitrary, so the left and center panels depict different representations of the same estimator. The leftmost panel shows the dilated empirical graphon with a random ordering. The middle panel shows the dilated empirical graphon sorted to group the I, S, and W edges, with the W edges sorted as in Fig. 7. The middle panel gives some intuition for why the dilated empirical graphon is able to estimate the entire graphex triple: When a Kallenberg exchangeable graph is generated according to $\hat{W}_{(G,15)}$ with latent Poisson process Π, the disjoint structure of the dilated graphon regions due to the I, S, and W components induces a natural partitioning of Π into independent Poisson processes that reproduce the independence structure used in the full generative model Eq. (2.4.6).

natural models for the dataset. The first is to treat the observed graphs G_k as realizations of $\mathcal{G}(\Gamma_{s_k})$ for some (unknown) sequence $s_k \uparrow \infty$ as $k \to \infty$ that is independent of Γ. Another natural model is to take $G_1, G_2, \dots$ to be the sequence of all distinct graph structures taken on by $(\Gamma_s)_{s\in\mathbb{R}_+}$; in this case, for all k, we take $G_k = \mathcal{G}(\Gamma_{\tau_k})$, where τ_k is the latent size at the kth occasion that the graph structure changes. In this later case, we call $\mathscr{G}(\Gamma) = (\mathcal{G}(\Gamma_{\tau_1}), \mathcal{G}(\Gamma_{\tau_2}), \dots)$ the graph sequence of Γ. (We define the graph sequence formally in Section 4.2.)

Intuitively, $\mathscr{G}(\Gamma)$ is the Kallenberg exchangeable graph Γ with the size information stripped away. In this sense, the graph sequence of Γ is the random object naturally associated to $\mathcal{W}$ when the sizes are unobserved. Thus, in this setting, convergence in distribution of the graph sequences induced by the estimators is a natural notion of consistency.

Definition 3.1.3. Write $\mathcal{W}_k \to_{GS} \mathcal{W}$ as $k \to \infty$ when $\mathscr{G}(\Gamma^k) \xrightarrow{d} \mathscr{G}(\Gamma)$ as $k \to \infty$, for Γ^k generated by $\mathcal{W}_k$ and Γ generated by $\mathcal{W}$.

The notion of consistent estimation corresponding to this convergence is that, for any fixed $\ell \in \mathbb{N}$, the distribution of the length ℓ prefix of the graph sequence generated by the estimator should be close to the distribution of the length ℓ prefix of the graph sequence generated by $\mathcal{W}$. Convergence in distribution of every finite-size prefix is equivalent to convergence in distribution of the entire sequence.

To explain our estimator for this setting, we will need the following concept:

Definition 3.1.4. Let $c \in \mathbb{R}_+$ and let $\mathcal{W} = (I, S, W)$ be a graphex. A *c-dilation of* $\mathcal{W}$ is the graphex $\mathcal{W}^c = (c^2 I, cS(\cdot/c), W(\cdot/c, \cdot/c))$.

The key fact about c-dilations is that $\mathrm{GPD}(\mathcal{W}, s) = \mathrm{GPD}(\mathcal{W}^c, s/c)$ for all $s \in \mathbb{R}_+$, and thus also $\mathscr{G}(\Gamma) \overset{d}{=} \mathscr{G}(\Gamma^c)$ whenever Γ is generated by $\mathcal{W}$ and Γ^c is generated by $\mathcal{W}^c$. That is, the law of the graph sequence is invariant to dilations of the generating graphex. This means, in particular, that the dilation of a graphex is not an identifiable parameter when the observation sizes are not included as part of the observation. The obvious guess for the estimator in this setting is then the estimator for the known-sizes setting with the dilation information stripped away. That is, our estimator is the dilated empirical graphon modulo dilation; i.e., it is simply the empirical graphon $\tilde{W}_{G_k} : [0,1]^2 \to [0,1]$ defined above. In this setting, the empirical graphon is acting as a representative of its equivalence class under the relation that equates graphons that generate graph sequences with the same laws.

The main estimation result is that if either

1. There is some (possibly random) sequence (s_k), independent from Γ, such that $s_k \uparrow \infty$ a.s. and $G_k = \mathcal{G}(\Gamma_{s_k})$ for all $k \in \mathbb{N}$, or
2. $(G_1, G_2, \dots) = \mathscr{G}(\Gamma)$,

then $\tilde{W}_{G_k} \to_{GS} W$ in probability as $k \to \infty$. Subject to an additional technical constraint (implied by integrability), the convergence in probability may be strengthened to convergence almost surely.

Our estimation results are inspired by Kallenberg's development of the theory of estimation for exchangeable arrays [Kal99]. Restricted to the graph setting (that is, 2-dimensional arrays interpreted as adjacency matrices), and translated into modern language, that paper introduced the empirical graphon (although not named as such) and formalized consistency in terms of the weak topology: $W_k \to W$ as $k \to \infty$ when the graphs generated by W_k converge in distribution to the graphs generated by W. The estimation results of the present chapter may be seen as generalizations of [Kal99] to the sparse graph regime.

The present chapter is also closely related to the recent paper [BCCH16]. Specialized to the case $\vartheta = \mathbb{R}_+$ equipped with Lebesgue measure, that chapter extends the cut distance between compactly supported graphons—a core tool in the limit theory of dense graphs—to arbitrary integrable graphons. Convergence in the cut distance then gives a notion of limit for sequences of graphons. This is extended to a notion of convergence for sequences of (sparse) graphs by saying that a sequence $G_1, G_2, \ldots$ converges in the stretched cut distance sense if and only if $\hat{W}_{(G_1,\sqrt{e(G_1)})}, \hat{W}_{(G_2,\sqrt{e(G_2)})}, \ldots$ converges with respect to the cut distance. That is, each graph G_k is mapped to the empirical graphon dilated by $v(G_k)/\sqrt{e(G_k)}$. The same paper also establishes that $e(G_k)/s_k^2 \to \|W\|_1$ a.s. whenever $(G_k)_{k\in\mathbb{N}}$ are unlabeled size-s KEGs as above. Thus, in the $\|W\|_1 = 1$ case, these dilated empirical graphons, considered as pixel pictures, will look asymptotically identical to the $v(G_k)/s_k$-dilated empirical graphons that we use as estimators in the known sizes case. This suggests a close connection between consistent estimation and convergence in cut distance. Indeed, in the dense graph setting these notions of convergence are known to be equivalent (in the dense setting, the convergence $W_k \to_{GP} W$ as $k \to \infty$ is equivalent to left convergence [DJ08], and left convergence is equivalence to convergence in the cut norm [BCLS+08a]). An analogous result in the sparse graph setting would allow for a very different approach to proving our convergence result in the known size setting, restricted to the special case that the generating graphex is an integrable graphon.

The chapter is organized as follows: In Section 4.2 we give formal definitions for the basic tools of the chapter. The sampling result is derived in Section 3.3. In Section 3.4 we prove the estimation result for the setting where observation sizes are included as part of the observation. We build on this in Section 3.5 to prove the estimation result for the setting where the true underlying observation sizes are not observed.

PRELIMINARIES

We will often have occasion to refer to the unlabeled finite graph associated with a finite adjacency measure.

Definition 3.2.1. Let ξ be a finite adjacency measure. The *unlabelled graph associated with* ξ is $\mathcal{G}(\xi)$.

A particularly important case is the graph associated to the size-s Kallenberg exchangeable graph Γ_s, which is almost surely finite. We will have frequent occasion to refer to the distributions of both the labeled and unlabeled graphs:

Definition 3.2.2. Let $(\Gamma_s)_{s\in\mathbb{R}_+}$ be a Kallenberg exchangeable graph-generated by $\mathcal{W}$. The *finite Kallenberg exchangeable graph distribution* with parameters $\mathcal{W}$ and s is $\mathrm{KEG}(\mathcal{W}, s) = \Pr(\Gamma_s \in \cdot \mid \mathcal{W}, s)$, and $\mathrm{KEG}(\mathcal{W}) = \mathrm{KEG}(\mathcal{W}, \infty)$. The *finite unlabeled Kallenberg exchangeable graph distribution* with parameters $\mathcal{W}$ and s is $\mathrm{GPD}(\mathcal{W}, s) = \Pr(\mathcal{G}(\Gamma_s) \in \cdot \mid \mathcal{W}, s)$.

In order to pass from $\mathcal{G}(\xi)$ back to some adjacency measure ξ' such that $\mathcal{G}(\xi') = \mathcal{G}(\xi)$, we must reintroduce labels. A simple scheme is to produce labels independently and uniformly in some range:

Definition 3.2.3. Let G be an unlabeled graph with edge set E, and let $s > 0$. A *random labeling of* G *into* $[0, s]$, $\mathsf{Lbl}_s(G, \{U_i\})$, is a random adjacency measure $\mathsf{Lbl}_s(G, \{U_i\}) = \sum_{(i,j)\in E} \delta_{(U_i, U_j)}$, where $U_i \overset{iid}{\sim} \mathrm{Uni}[0, s]$, for $i \in \mathbb{N}$. Where there is no risk of confusion, we will write $\mathsf{Lbl}_s(G)$ for $\mathsf{Lbl}_s(G, \{U_i\})$ where $U_i \overset{iid}{\sim} \mathrm{Uni}[0, t]$, for $i \in \mathbb{N}$, independently of everything else.

Because our notion of consistent estimation is a requirement of distributional convergence, the distributions of these random labelings will play a large role. Clearly, the distribution of $\mathsf{Lbl}_s(G)$ is a measurable function of G and s.

Definition 3.2.4. We write $\mathsf{embed}(G, s)(\cdot) = \Pr(\mathsf{Lbl}_s(G) \in \cdot)$ for the distribution of $\mathsf{Lbl}_s(G)$. When G is itself random, a random embedding of G into $[0, s]$ is defined by $\mathsf{embed}(G, s) = \Pr[\mathsf{Lbl}_s(G) | G]$.

We typically think of Kallenberg exchangeable graphs as defining a nested collection of $\mathbb{R}_+$-labeled graph valued random variables $(\Gamma_s)_{s\in\mathbb{R}_+}$. In modeling situations where the labeling is irrelevant, it is natural to instead look at the (countable) collection of all distinct graph structures taken on by $(\Gamma_s)_{s\in\mathbb{R}_+}$; this is the graph sequence associated with Γ. We now turn to formally defining the graph sequence associated with an arbitrary adjacency measure ξ. To that end, define $E : \mathbb{R}_+ \to \mathbb{N}$ by

$$E(s) = \frac{1}{2}\xi[0, s]^2 \qquad \text{for } s \in \mathbb{R}_+. \tag{3.2.1}$$

In the absence of self loops, $E(s)$ is the number of edges present between vertices with labels in $[0, s]$. In general, the jumps of E correspond with the appearance of edges.

Definition 3.2.5. Let ξ be an adjacency measure. The jump times of ξ, written as $\tau(\xi)$, is the sequence $\tau_1, \tau_2, \ldots$ of jumps of E in order of appearance.

Note that the map $\xi \mapsto \tau(\xi)$ is measureable. Intuitively, $\tau_1, \tau_2, \ldots$ are the sample sizes at which edges are added to the unlabeled graph associated with the adjacency measure.

Let χ_s denote the operation of restricting an adjacency measure to those vertices with labels in $[0, s]$, in the sense that $\chi_s\xi(\cdot) = \xi(\cdot \cap [0, s]^2)$. We now formalize the sequence of all distinct unlabeled graphs associated with $(\chi_s\xi)_{s\in\mathbb{R}_+}$:

Definition 3.2.6. The *graph sequence associated with* ξ, written $\mathcal{G}(\xi)$, is the sequence $\mathcal{G}(\chi_{\tau_1}\xi), \mathcal{G}(\chi_{\tau_2}\xi), \ldots$, where $\tau_1, \tau_2, \ldots$ are the jump times of ξ.

SAMPLING

Γ_r, a Kallenberg exchangeable graph of size r, may be generated from Γ_s, a Kallenberg exchangeable graph of size $s > r$, by restricting Γ_s to $[0, r]^2$. In this section we show that this restriction has a natural relation to p-sampling: $\mathcal{G}(\Gamma_r)$ may be generated as an r/s-sampling of $\mathcal{G}(\Gamma_s)$.

The first result we need is that random labelings preserve the law of exchangeable adjacency measures. Intuitively, the labels of the size-s Kallenberg exchangeable graph can be invented by labeling each vertex i.i.d. Uni$[0, s]$.

Lemma 3.3.1. *Let $s > 0$ and let Γ_s be a size-s Kallenberg exchangeable graph generated by $\mathcal{W}$. Then,* $\mathrm{KEG}(\mathcal{W}, s) = \mathbb{E}[\mathsf{embed}(\mathcal{G}(\Gamma_s), s)]$.

Proof. It suffices to show that $\mathsf{Lbl}_s(\mathcal{G}(\Gamma_s)) \stackrel{d}{=} \Gamma_s$.

Suppose Γ_s is generated as in Eq. (2.4.6). For simplicity of exposition, suppose that the generating graphex is $(0, 0, W)$, and the associated latent Poisson process is Π_s. Let $\{\theta'_i\}_{i\in\mathbb{N}} \overset{iid}{\sim} \mathrm{Uni}[0, s]$, and let $\Pi'_s = \{(\theta'_i, \vartheta_i) : (\theta_i, \vartheta_i) \in \Pi_s\}$. By a property of the Poisson process, $\Pi'_s \stackrel{d}{=} \Pi_s$. Let Γ'_s be a size-s Kallenberg exchangeable graph generated using the same latent variables as Γ_s, but with Π'_s replacing Π_s. Then, by construction, $\Gamma'_s \stackrel{d}{=} \mathsf{Lbl}_s(\mathcal{G}(\Gamma_s))$. Moreover, Γ'_s is distributed as a size-s Kallenberg exchangeable graph, so $\Gamma'_s \stackrel{d}{=} \Gamma_s$.

An essentially identical argument proves the result for a Kallenberg exchangeable graph generated by the full graphex. □

The main sampling result is:

Theorem 3.3.2. *Let $\mathcal{W}$ be a graphex, let $s > 0$ and $r \in [0, s]$, let $G_s \sim \mathrm{GPD}(\mathcal{W}, s)$, and let G_r be an r/s-sampling of G_s. Then, $G_r \sim \mathrm{GPD}(\mathcal{W}, s)$.*

Proof. Let $\xi_s = \mathsf{Lbl}_s(G_s)$. It is an obvious consequence of Lemma 3.3.1 that ξ_s is equal in distribution to a size-s Kallenberg exchangeable graph generated by $\mathcal{W}$. Let ξ_r be the restriction of ξ_s to $[0, r]^2$, so $\mathcal{G}(\xi_r) \sim \mathrm{GPD}(\mathcal{W}, r)$. Each vertex of ξ_s has a label in $[0, r]$ independently with probability r/s; thus, $\mathcal{G}(\xi_r) \stackrel{d}{=} G_r$. □

ESTIMATION WITH KNOWN SIZES

This section explains our estimation results for the case where the observations are $(G_1, s_1), (G_2, s_2), \ldots$, where $G_k = \mathcal{G}(\Gamma_{s_k})$ for some Kallenberg exchangeable graph Γ generated according to a graphex $\mathcal{W}$ and some sequence $s_k \uparrow \infty$ in $\mathbb{R}_+$. We consider both the case of an arbitrary non-random divergent sequence and the case where the sizes are taken to be the jumps of the Kallenberg exchangeable graph (that is, the sizes at which new edges enter the graph), in which case we denote the sequence as $\tau_1, \tau_2, \ldots$ As motivated in the introduction, our notion of estimation is formalized as:

Definition 3.4.1. Let $\mathcal{W}_1, \mathcal{W}_2, \ldots$ be a sequence of graphexes. Write $\mathcal{W}_n \to_{\mathrm{GP}} \mathcal{W}$ as $n \to \infty$ when, for all $s \in \mathbb{R}_+$, it holds that $\mathrm{GPD}(\mathcal{W}_n, s) \to \mathrm{GPD}(\mathcal{W}, s)$ weakly as $n \to \infty$.

The goal of estimation is: given a sequence of observations $(G_1, s_1), (G_2, s_2), \ldots$, produce

$$\hat{W}_{(G_k, s_k)} : \mathbb{R}_+^2 \to [0, 1] \tag{3.4.1}$$

such that $\hat{W}_{(G_k, s_k)} \to_{\mathrm{GP}} \mathcal{W}$ as $k \to \infty$, where the convergence may be almost sure or merely in probability.

The main result of this section is that the dilated empirical graphons $\hat{W}_{(G_k, s_k)} \to_{\mathrm{GP}} \mathcal{W}$ for $(G_1, s_1), (G_2, s_2), \ldots$ generated by a graphex $\mathcal{W}$; i.e. the dilated empirical graphon is a consistent estimator for $\mathcal{W}$.

We now turn to an intuitive description of the broad structure of the argument. Conditional on G_k, let $\xi^k = \mathsf{Lbl}_{s_k}(G_k)$ and let $\mathsf{embed}(G_k, s_k)$ be the distribution of ξ^k conditional on G_k. The first convergence result, Theorem 3.4.3, is that, almost surely, the random distributions $\mathsf{embed}(G_k, s_k)$ converge weakly to $\mathcal{L}(\Gamma) = \mathrm{KEG}(\mathcal{W})$. That is, for almost every realization of a Kallenberg exchangeable graph, the point processes defined by randomly labeling the observed finite graphs converge in distribution to the original Kallenberg exchangeable graph. The analogous statement in the i.i.d. sequence setting is that, given some $(X_1, X_2, \ldots)$ where $X_k \overset{\mathrm{iid}}{\sim} P$, and σ_n a random permutation on $[1, \ldots, n]$, the random distributions $\Pr(X_{\sigma_n(1)}, \ldots, X_{\sigma_n(n)} \in \cdot \mid X_1, \ldots, X_n)$ converge weakly almost surely to $\Pr((X_1, X_2, \ldots) \in \cdot)$ as $n \to \infty$.

The convergence in distribution of the point processes on $\mathbb{R}_+^2$ is equivalent to convergence in distribution of the point processes restricted to $[0,r]^2$ for every finite $r \in \mathbb{R}_+$. This perspective lends itself naturally to the interpretation of the limit result as a qualitative approximation theorem: intuitively, $\Pr(\xi^k([0,r]^2 \cap \cdot) \in \cdot \mid G_k)$ approximates $\mathrm{KEG}(\mathcal{W}, r)$, with the approximation becoming exact in the limit $r/s_k \to 0$. This perspective also makes clear the first critical connection between estimation and sampling: conditional on G_k, $\mathcal{G}(\xi^k([0,r]^2 \cap \cdot))$ has the same distribution as an r/s_k-sampling of G_k.

The second key observation is that, conditional on G_k, a sample from $\mathrm{GPD}(\hat{W}_{(G_k,s_k)}, r)$ may be generated by sampling $\mathrm{Poi}(r/s_k \nu(G_k))$ vertices with replacement from G_k and returning the induced edge set. The second step in the proof is to show that this sampling scheme is asymptotically equivalent to r/s_k-sampling in the limit of $s_k \uparrow \infty$; this is the role of Lemmas 3.4.5 and 3.4.7.

Theorem 3.4.8 then puts together these results to conclude that, almost surely, $\mathrm{KEG}(\hat{W}_{(G_k,s_k)}) \to \mathrm{KEG}(\mathcal{W})$ weakly as $k \to \infty$. Some additional technical rigmarole is required to show that this also gives convergence of the (unlabeled) random graphs. This later convergence is the main result of this section, and is established in Theorem 3.4.12.

Convergence in Distribution of Random Embeddings

This subsection uses results from the theory of distributional convergence of point processes to show that, almost surely, $\mathsf{embed}(G_k, s_k) \to \mathrm{GPD}(\mathcal{W}, \infty)$ weakly as $k \to \infty$.

We will need the following definition and technical lemma: A separating class for a locally compact second countable Hausdorff space S is a class $\mathcal{U} \subset S$ such that for any compact open sets with $K \subset G$ there is some $U \in \mathcal{U}$ with $K \subset U \subset G$.

Lemma 3.4.2. *Let $\phi, \phi_1, \phi_2, \ldots$ be simple point processes on a locally compact second countable Hausdorff space S. If*

$$\phi_n(U) \xrightarrow{d} \phi(U), \; n \to \infty \tag{3.4.2}$$

weakly for all U in some separating class for S then

$$\phi_n \xrightarrow{d} \phi, \; n \to \infty \tag{3.4.3}$$

weakly.

Proof. By [Kal01, Thms. 16.28 and 16.29], it suffices to check that $\Pr(\phi_n(U) = 0) \to \Pr(\phi(U) = 0)$ and that $\limsup_n \Pr(\phi_n(U) > 1) \leqslant \Pr(\phi(U) > 1)$. Because $\phi_n(U)$ is a non-negative integer a.s., both conditions are implied by $\phi_n(U) \xrightarrow{d} \phi(U)$. □

Theorem 3.4.3. *Let Γ be a Kallenberg exchangeable graph generated by a non-trivial graphex $\mathcal{W}$, let $s_1, s_2, \ldots$ be some sequence in $\mathbb{R}_+$ such that $s_k \uparrow \infty$ as $k \to \infty$ and let $G_k = \mathcal{G}(\Gamma_{s_k})$ for all k. Then $\mathsf{embed}(G_k, s_k) \to \mathrm{KEG}(\mathcal{W})$ weakly almost surely.*

Proof. For each $k \in \mathbb{N}$, conditional on G_k, let ξ^k be a point process with law $\mathsf{embed}(G_k, s_k)$. Note that $\Gamma \sim \mathrm{KEG}(\mathcal{W})$. Observe that the collection $\mathcal{U}$ of finite unions of rectangles with rational end points is a separating class for $\mathbb{R}_+^2$. Further, ξ^k is simple for all $k \in \mathbb{N}$, as is Γ. Thus by Lemma 3.4.2, to show the claimed result it will suffice to show that, for all $U \in \mathcal{U}$, $\Pr(\xi^k(U) \in \cdot \mid G_k) \to \Pr(\Gamma(U) \in \cdot)$ weakly as $k \to \infty$.

Fix U. To establish this condition we first show that for all bounded continuous functions f, it holds that $\lim_{k\to\infty} \mathbb{E}[f(\xi^k(U)) \mid G_k] = \mathbb{E}[f(\Gamma(U))]$ a.s. Let $\mathcal{F}_{-s}$ be the partially labelled graph derived from Γ by forgetting the labels of all nodes with label $\theta_i < s$. Take $r \in \mathbb{R}_+$ large enough so that $U \subset [0, r]^2$. Then for $s_k > r$,

$$\mathbb{E}[f(\Gamma(U)) \mid \mathcal{F}_{-s_k}] = \mathbb{E}[f(\xi^k(U)) \mid G_k]. \tag{3.4.4}$$

Define $U_t = U + (t, t)$ for $t \in \mathbb{R}_+$ and let

$$X_s^{(r)} = \frac{1}{s-r} \int_0^{s-r} f(\Gamma(U_t))\, dt. \tag{3.4.5}$$

Observe that for s such that $t \leqslant s - r$, the joint exchangeability of Γ implies

$$\mathbb{E}[f(\Gamma(U_t)) \mid \mathcal{F}_{-s}] = \mathbb{E}[f(\Gamma(U)) \mid \mathcal{F}_{-s}]. \tag{3.4.6}$$

Moreover, by the linearity of conditional expectation, for $s > r$, it holds that $\mathbb{E}[X_s^{(r)} \mid \mathcal{F}_{-s}] = \mathbb{E}[f(\Gamma(U)) \mid \mathcal{F}_{-s}]$.

A standard result [Dur10, Ex. 5.6.2] shows that $\lim_{k\to\infty} \mathbb{E}[X_{s_k}^{(r)} \mid \mathcal{F}_{-s_k}] = \mathbb{E}[X_\infty^{(r)} \mid \mathcal{F}_{-\infty}]$ a.s. if $X_{s_k}^{(r)} \to X_\infty^{(r)}$ a.s. and there is some integrable random variable that dominates $X_{s_k}^{(r)}$ for all k; the second condition holds because f is bounded. Notice that $Y_t = f(\Gamma(B_t))$ is a stationary stochastic process. Moreover, it's easy to see from the Kallenberg exchangeable graph construction that Y_t and $Y_{t'}$ are independent whenever $|t - t'| > r$, so (Y_t) is mixing. The ergodic theorem then gives $\lim_{k\to\infty} X_{s_k}^{(r)} = \mathbb{E}[f(\Gamma(U))]$ a.s. This means

$$\lim_{k\to\infty} \mathbb{E}[f(\xi^k(U)) \mid G_k] \to \mathbb{E}[f(\Gamma(U))] \text{ a.s.}, \tag{3.4.7}$$

as promised.

For $l \in \mathbb{Z}_+$, let $f_l(\cdot) = \mathbb{1}[\cdot \leqslant l]$, let $A_l^{(U)}$, for each $U \in \mathcal{U}$, be the set on which

$$\lim_{k\to\infty} \mathbb{E}[f_l(\xi^k(U)) \mid G_k] = \mathbb{E}[f_l(\Gamma(U))] \tag{3.4.8}$$

and let $A_U = \bigcap_l A_l^{(U)}$. We have shown that $\Pr(A_l^{(U)}) = 1$, and so $\Pr(A_U) = 1$ and on A_U it holds that $\lim_{k\to\infty} \Pr(\xi^k(U) \in \cdot \mid G_k) = \Pr(\Gamma(U) \in \cdot)$ weakly. Let $A = \bigcap_{U\in\mathcal{U}} A_U$, then $\Pr(A) = 1$ and on A it holds that

$$\lim_{k\to\infty} \Pr(\xi^k(U) \in \cdot \mid G_k) = \Pr(\Gamma(U) \in \cdot) \tag{3.4.9}$$

weakly for all $U \in \mathcal{U}$, completing the proof.

□

We need to do a little bit more work to show convergence in the case where the observations are taken at the jumps of the Kallenberg exchangeable graph.

Theorem 3.4.4. *Let Γ be a Kallenberg exchangeable graph generated by a non-trivial graphex $\mathcal{W}$, and let $\tau_1, \tau_2, \ldots$ be the jump times of Γ. Let $G_k = \mathcal{G}(\Gamma_{\tau_k})$ for each $k \in \mathbb{N}$. Then $\mathsf{embed}(G_k, \tau_k) \to \mathrm{GPD}(\mathcal{W}, \infty)$ weakly almost surely as $k \to \infty$.*

Proof. For each $k \in \mathbb{N}$, let ξ^k be a point process with law $\mathsf{embed}(G_k, \tau_k)$.

As in the proof of Theorem 3.4.3, to establish the claim it suffices to show that, for all bounded continuous functions f and all rectangles U, it holds that

$$\lim_{k\to\infty} \mathbb{E}[f(\xi^k(U)) \mid G_k, \tau_k] = \mathbb{E}[f(\Gamma(U))] \text{ a.s.} \tag{3.4.10}$$

Let $\mathcal{F}_{-s}$ be as in proof of Theorem 3.4.3. It is clear that $\mathcal{F}_{-\tau_k} \subset \mathcal{F}_{-\tau_{(k-1)}}$ for all k. Because $U \subset [0, r]^2$ for some finite r and $\tau_k \uparrow \infty$ a.s. as $k \to \infty$ it holds that

$$\lim_{k\to\infty} \mathbb{E}[f(\Gamma(U)) \mid \mathcal{F}_{-\tau_k}] = \lim_{k\to\infty} \mathbb{E}[f(\xi^k(U)) \mid G_k, \tau_k] \text{ a.s.} \tag{3.4.11}$$

Applying reverse martingale convergence to the l.h.s. we conclude the r.h.s. exists a.s.

It remains to identify the limit. To that end, we will define a coupling between the counts on test set U at a subsequence of the jump times and the counts on U at some deterministic sequence, which is known to converge to the desired limit. Let $s_k = \sum_{n=1}^{k} \frac{1}{n}$, let $\{\tau_{k_j}\}$ be a subsequence of the jump times defined such that at most one point in $\{\tau_{k_j}\}$ lies in $[s_l, s_{l+1})$ for all l and define s_{k_j} to be the subsequence of $\{s_k\}$ such that s_{k_j} is the largest value in $\{s_k\}$ that is smaller than τ_{k_j}. Intuitively, this gives a random subsequence of the jump times and a random subsequence of $\{s_k\}$ such that the points s_{k_j} and τ_{k_j} become arbitrarily close as $j \to \infty$. For each $j \in \mathbb{N}$, let $G_j^s = \mathcal{G}\left(\Gamma_{s_{k_j}}\right)$, and let $G_j^\tau = \mathcal{G}\left(\Gamma_{\tau_{k_j}}\right)$.

By construction, $G_j^s \subset G_j^\tau$. Label the vertices of G_j^τ as $1, \ldots, v(G_j^\tau)$ such that $1, \ldots, v(G_j^s)$ is the vertex set of G_j^s. Let $\xi^{(s,j)} = \mathsf{Lbl}_{s_{k_j}}(G_j^s)$, and let $\xi^{(\tau,j)} = \mathsf{Lbl}_{\tau_{k_j}}(G_j^\tau)$. The occupancy counts of the test may then sampled according to:

1. $V_1, \ldots, V_{\nu(G_j^\tau)} \overset{iid}{\sim} \mathrm{Uni}[0,1]$
2. $\xi^{(\tau,j)}(U) = |\{(v_i, v_j) \in e(G_j^\tau) : (V_i\tau_{k_j}, V_j\tau_{k_j}) \in U\}|$
3. $\xi^{(s,j)}(U) = |\{(v_i, v_j) \in e(G_j^s) : (V_i s_{k_j}, V_j s_{k_j}) \in U\}|$

By construction, $G_j^\tau \backslash G_j^s$ is a star; call the center of this star c. Choosing r such that $U \subset [0,r]^2$, it is clear that if $V_c\tau_{k_j} \notin [0,r]$ then $\xi^{(\tau,j)}(U) \leqslant \xi^{(s,j)}(U)$ under this coupling. The occupancy counts are the number of edges in random induced subgraphs given by including each vertex with probability $\frac{r}{\tau_{k_j}}$ and $\frac{r}{s_{k_j}}$ respectively. This perspective makes it clear that, conditional on c not being included when sampling from G_j^τ, the counts will be equal as long as no vertices of the induced subgraph of G_j^s are "forgotten" when the inclusion probability is reduced to $\frac{r}{\tau_{k_j}}$. The probability that $V_i s_{k_j} \in [0,r]$ but $V_i\tau_{k_j} \notin [0,r]$ is $\frac{r}{\tau_{k_j} s_{k_j}}(\tau_{k_j} - s_{k_j})$. Moreover, there are at most $\xi^{(s,j)}(U)$ vertices in the subgraph sampled from G_j^s so, in particular,

$$\mathbb{E}[\xi^{(s,j)}(U) - \xi^{(\tau,j)}(U) \mid E_{\bar{c}}, \Gamma] \leqslant \frac{r}{\tau_{k_j} s_{k_j}}(\tau_{k_j} - s_{k_j})\xi^{(s,j)}(U), \tag{3.4.12}$$

where $E_{\bar{c}}$ denotes the event that c is not included in the subgraph sampled from G_j^τ. Then, denoting the event $\{\xi^{(s,j)}(U) = \xi^{(\tau,j)}(U)\}$ as E_U,

$$\begin{aligned} \Pr(E_U \mid \Gamma) &\geqslant (1 - \Pr(E_{\bar{c}}))(1 - \Pr(\bar{E_U} \mid \Gamma, E_{\bar{c}})) && (3.4.13)\\ &\geqslant (1 - \frac{r}{\tau_{k_j}})(1 - \frac{r}{\tau_{k_j} s_{k_j}}(\tau_{k_j} - s_{k_j})\xi^{(s,j)}(U). && (3.4.14) \end{aligned}$$

By construction, $\tau_{k_j} - s_{k_j} \leqslant \frac{1}{k_j}$, so $\lim_{j\to\infty} \frac{\tau_{k_j} - s_{k_j}}{s_{k_j}} = 0$. In combination with $\lim_{j\to\infty} \xi^{(s,j)}(U) = \Gamma(U)$ a.s. and the fact that $\Gamma(U)$ is almost surely finite, the inequality we have just derived then implies that

$$\lim_{j\to\infty} \Pr(\xi^{(s,j)}(U) \neq \xi^{(\tau,j)}(U) \mid \Gamma) = 0 \text{ a.s.} \tag{3.4.15}$$

In view of Theorem 3.4.3, we thus have that

$$\lim_{k\to\infty} \mathbb{E}[f(\xi^k(U)) \mid G_k, \tau_k] = \mathbb{E}[f(\Gamma(U))] \text{ a.s.}, \tag{3.4.16}$$

as required.

□

Asymptotic Equivalence of Sampling Schemes

As alluded to above, a key insight for showing that $\hat{W}_{(G_k, s_k)}$ is a valid estimator is that, conditional on G_k, a graph generated according to

$\mathrm{GPD}(\hat{W}_{(G_k,s_k)}, r)$ may be viewed as a random subgraph of G_k induced by sampling $\mathrm{Poi}(\frac{r}{s_k}v(G_k))$ vertices from G_k with replacement and returning the edge set of the vertex-induced subgraph. The correctness of this scheme can be seen as follows:

1. Let Π be the latent Poisson process used to generate a sample from $\mathrm{GPD}(\hat{W}_{(G_k,s_k)}, r)$, as in Theorem 2.4.7, and let $\Pi_r = \Pi(\cdot \cap [0,r]^2)$. Because $\hat{W}_{(G_k,s_k)}$ has compact support $[0, v(G_k)/s_k]^2$, only Π_r restricted to $[0,r] \times [0, v(G_k)/s_k]$ can participate in the graph.

2. Π_r restricted to $[0,r] \times [0, v(G_k)/s_k]$ may be generated by producing $J_{s_k,r} \sim \mathrm{Poi}(rv(G_k)/s_k)$ points (θ_i, ϑ_j) where, conditional on $J_{s_k,r}$, $\theta_i \overset{iid}{\sim} \mathrm{Uni}[0,r]$ and $\vartheta_i \overset{iid}{\sim} \mathrm{Uni}[0, v(G_k)/s_k]$, also independently of each other.

3. The $\{0,1\}$-valued structure of $\hat{W}_{(G_k,s_k)}$ means that choosing latent values $\vartheta_i \overset{iid}{\sim} \mathrm{Uni}[0, v(G_k)/s_k]$ is equivalent to choosing vertices of G_k uniformly at random with replacement.

Our task is to show that the sampling scheme just described is asymptotically equivalent to r/s_k-sampling of G_k. To that end, we observe that r/s_k-sampling is the same as sampling $\mathrm{Bin}(v(G_k), r/s_k)$ vertices of G_k without replacement and returning the induced edge set. This makes it clear that there are two main distinctions between the sampling schemes: Binomial vs. Poisson number of vertices sampled, and with vs. without replacement sampling. This motivates defining three distinct random subgraphs of G_k:

1. $X_r^{(k)}$: Sample $\mathrm{Bin}(v(G_k), \frac{r}{s_k})$ vertices without replacement and return the induced edge set

2. $H_r^{(k)}$: Sample $\mathrm{Bin}(v(G_k), \frac{r}{s_k})$ vertices with replacement and return the induced edge set

3. $M_r^{(k)}$: Sample $\mathrm{Poi}(\frac{r}{s_k}v(G_k))$ vertices with replacement and return the induced edge set

The observation that, conditional on G_k, $\xi_r^k \overset{d}{=} \mathrm{Lbl}_r(X_r^{(k)})$ makes the connection with the previous subsection clear.

Our aim is to show that when r/s_k is small the different random subgraphs are all close in distribution. A natural way to encode this is the total variation distance between their distributions. However, because the distributions are themselves random (G_k measurable) variables this is rather awkward. It is instead convenient to work with couplings of the random subgraphs conditional on G_k; this gives a natural notion of conditional total variation distance. See [Hol12] for an introduction to coupling arguments.

Although we only need the sampling equivalence for sequences of graphs corresponding to a Kallenberg exchangeable graph, we state the theorems for generic random graphs where possible.

The following result, which plays a similar role in the estimation theory of graphons in the dense setting, is simply the asymptotic equivalence of sampling with and without replacement.

Lemma 3.4.5. *Let* G *be an almost surely finite random graph, with* e *edges and* v *vertices. let* X_r *be a random subgraph of* G *given by sampling* $\mathrm{Bin}(v(G), \frac{r}{s})$ *vertices without replacement and returning the induced edge set, and let* H_r *be a random subgraph of* G *given by sampling* $\mathrm{Bin}(v(G), \frac{r}{s})$ *vertices with replacement and returning the induced edge set. Then there is a coupling such that*

$$\Pr(H_r \neq X_r \mid G) \leqslant 2e\Big(\frac{r^3}{s^3} + 2\frac{r^3}{s^3v^2} + 3\frac{r^2}{s^2v} + \frac{r}{sv^2}\Big) \tag{3.4.17}$$

Moreover, specializing to the Kallenberg exchangeable graph case, with $H_r^{(k)}$ *and* $X_r^{(k)}$ *defined as above, under the same coupling,*

$$\Pr(H_r^{(k)} \neq X_r^{(k)} \mid G_k) \xrightarrow{p} 0, \tag{3.4.18}$$

as $k \to \infty$*. Further, if* $\tau_1, \tau_2, \ldots$ *are the jump times of* Γ *then taking* $s_k = \tau_k$ *for all* $k \in \mathbb{N}$*, it holds that under this coupling*

$$\Pr(H_r^{(k)} \neq X_r^{(k)} \mid G_k, \tau_k) \xrightarrow{p} 0, \tag{3.4.19}$$

as $k \to \infty$*.*

Proof. Given G, we may sample X_r according to the following scheme:

1. Sample $K_{s,r} \sim \mathrm{Bin}(v, \frac{r}{s})$
2. Sample a list $L = (L_1, L_2, \ldots, L_{K_{s,r}})$ of vertices from G without replacement
3. Return the edge set of the induced subgraph given by restricting G to L

Given G, we may sample H_r similarly, except we use a list sampled with replacement; we couple H_r and X_r by coupling with and without replacement sampling of the vertex list. The following sampling scheme for a list $\tilde{L}$ returns a list that, given G, has the distribution of a length $K_{s,r}$ list of vertices sampled with replacement from G. Given G we sample $\tilde{L}$ according to:

1. Sample L as above
2. $\tilde{L}_1 = L_1$
3. For $j = 1 \ldots K_{s,r}$, set $\tilde{L}_j = L_j$ with probability $1 - \frac{j-1}{v}$. Otherwise, sample $\tilde{L}_j$ uniformly at random from $\{L_1, \ldots, L_{j-1}\}$.

H_r is then sampled by returning the edge set of the induced subgraph given by taking $\tilde{L}$ as the vertex set.

Evidently, under this coupling, $X_r = H_r$ as long as

1. Every entry of L where $L \neq \tilde{L}$ does not participate in an edge in X_r
2. Every entry of $\tilde{L}$ where $L \neq \tilde{L}$ does not participate in an edge in X_r

Call the number of entries violating the first condition F_1 and the number of entries violating the second condition F_2, and let N be the total number of entries where $L, \tilde{L}$ differ. Observe that when $K_{s,r} > 0$, almost surely,

$$\mathbb{E}[F_1 \mid \nu(H_r), N, K_{s,r}, G] = \frac{\nu(X_r)}{K_{s,r}} N \tag{3.4.20}$$

$$\mathbb{E}[F_2 \mid \nu(H_r), N, K_{s,r}, G] = \frac{\nu(X_r)}{K_{s,r}} N. \tag{3.4.21}$$

Further observe that because the sites where the lists disagree are chosen without reference to the graph structure it holds that $\nu(X_r)$ and N are independent given G and $K_{s,r}$, so

$$\mathbb{E}[F_1 + F_2 \mid \nu(X_r), K_{s,r}, G] = 2\frac{\nu(X_r)}{K_{s,r}} \mathbb{E}[N \mid K_{s,r}, G]. \tag{3.4.22}$$

Moreover, almost surely,

$$\mathbb{E}[N \mid K_{s,r}, G] = \sum_{j=2}^{K_{s,r}} \frac{j-1}{\nu} \tag{3.4.23}$$

$$= \frac{1}{2\nu}(K_{s,r}^2 - K_{s,r}). \tag{3.4.24}$$

Using Markov's inequality along with the observation that

$$\Pr(X_r \neq H_r \mid K_{s,r} < 2) = 0, \tag{3.4.25}$$

and $K_{s,r}^2 - K_{s,r} \leqslant K_{s,r}^2$ on $K_{s,r} \geqslant 2$, Eq. (3.4.24) implies that, almost surely,

$$\Pr(X_r \neq H_r \mid \nu(X_r), K_{s,r}, G) \leqslant \mathbb{E}[F_1 + F_2 \mid \nu(X_r), K_{s,r}, G] \tag{3.4.26}$$

$$\leqslant \frac{K_{s,r}}{\nu} \nu(X_r). \tag{3.4.27}$$

To prove the first assertion of the theorem statement, we now observe that $\nu(X_r) \leqslant 2e(X_r)$ and $\mathbb{E}[e(X_r) \mid s, K_{s,r}, G] \leqslant e\frac{K_{s,r}^2}{\nu^2}$ (since each edge is included with marginal probability at most $\frac{K_{s,r}^2}{\nu^2}$), so it holds almost surely that

$$\Pr(X_r \neq H_r \mid s, G) \leqslant e\frac{2}{v^3}\mathbb{E}[K_{s,r}^3 \mid G] \tag{3.4.28}$$

$$= 2e\left(\frac{r^3}{s^3} - 3\frac{r^3}{s^3 v} + 2\frac{r^3}{s^3 v^2} + 3\frac{r^2}{s^2 v} - 3\frac{r^2}{s^2 v^2} + \frac{r}{s v^2}\right). \tag{3.4.29}$$

To prove the second assertion of the theorem statement we apply Eq. (3.4.27) to the graph G_k sampled at rate r/s_k, so

$$\Pr(X_r^{(k)} \neq H_r^{(k)} \mid \nu(X_r^{(k)}), K_{s_k,r}, G_k) \leqslant \frac{K_{s_k,r}}{\nu}\nu(X_r^{(k)}). \tag{3.4.30}$$

Markov's inequality with $\mathbb{E}[\frac{K_{s_k,r}}{\nu(G_k)} \mid G_k] = r/s_k$ implies that, given G_k, $\frac{K_{s_k,r}}{\nu(G_k)} \xrightarrow{p} 0$ as $k \to \infty$. Further, by Theorem 3.4.3 and the observation that $X_r^{(k)} \overset{d}{=} \mathcal{G}\big(\xi^k(\cdot \cap [0,r]^2)\big)$ where $\xi^k \sim \mathrm{embed}(G_k, s_k)$, it holds that $\nu(X_r^{(k)}) \xrightarrow{d} \nu(\Gamma_r)$ a.s. as $k \to \infty$. Since the integrability conditions on graphexes guarantee that $\nu(\Gamma_r)$ is almost surely finite, we have

$$\frac{K_{s_k,r}}{\nu(G_k)}\nu(X_r^{(k)}) \xrightarrow{p} 0, \tag{3.4.31}$$

as $k \to \infty$ and this implies,

$$\Pr(X_r^{(k)} \neq H_r^{(k)} \mid \nu(X_r^{(k)}), K_{s_k,r}, G_k) \xrightarrow{p} 0, \tag{3.4.32}$$

as $k \to \infty$. Now,

$$\Pr(X_r^{(k)} \neq H_r^{(k)} \mid G_k) = \mathbb{E}[\Pr(X_r^{(k)} \neq H_r^{(k)} \mid \nu(X_r^{(k)}), K_{s_k,r}, G_k) \mid G_k], \tag{3.4.33}$$

and $\Pr(X_r^{(k)} \neq H_r^{(k)} \mid G_k)$ is bounded by 1 for all k, so the second claim follows by the dominated convergence theorem for conditional expectations, [Dur10, Thm. 5.9].

The proof of the final claim goes through mutatis mutandis as the proof of the second assertion, subject to the observations that $\tau_k \uparrow \infty$ a.s., that we must condition on τ_k for each k, and that Theorem 3.4.4 should be used in place of Theorem 3.4.3. □

Remark 3.4.6. In the case that $\mathcal{W} = (0, 0, W)$ and W is integrable, it holds that $\nu(G_k) = \Omega(s_k)$ a.s. and $e(G_k) = \Theta(s_k^2)$ a.s. [BCCH16, Props. 2.18 and 5.2], in which case the rate from the first part of the above lemma is $O(r^3/s_k)$. Note that in this case, the convergence in probability may be replaced by convergence almost surely. This lemma is in fact the only component of the proof where a weakening of almost sure convergence is necessary, so (as remarked below), whenever almost sure convergence holds for the equivalence of with and without replacement sampling, almost sure convergence holds for the main estimation result.

It remains to show that the $\mathrm{Poi}(r/s_k \nu(G_k))$ and $\mathrm{Bin}(\nu(G_k), r/s_k)$ samplings are asymptotically equivalent. Note that the rate $(\nu(G_k)/s_k)$ at which the empirical graphon is dilated guarantees that the expected number of vertices sampled according to each scheme is equal; this is the reason that this rate was chosen.

Lemma 3.4.7. *Let* G *be an almost surely finite random graph with* ν *vertices. Let* H_r *be a random subgraph of* G *given by sampling* $\mathrm{Bin}(\nu, \frac{r}{s})$ *vertices with replacement and returning the induced edge set, and let* M_r *be a random subgraph of* G *given by sampling* $\mathrm{Poi}(\nu\frac{r}{s})$ *vertices with replacement and returning the induced edge set. Then there is a coupling such that*

$$\Pr(H_r \neq M_r \mid G) \leqslant \frac{r}{s} \text{ a.s.} \tag{3.4.34}$$

Proof. Conditional on G, H_r may be sampled by:

1. sample $K_{s,r} \sim \mathrm{Bin}(\nu, r/s)$ vertices with replacement from G;
2. return the edge set of the induced subgraph.

Conditional G, M_r may be sampled by:

1. sample $J_{s,r} \sim \mathrm{Poi}(r\frac{\nu}{s})$ vertices with replacement from G.
2. return the edge set of the induced subgraph.

Comparing the two sampling schemes, it is immediate that there is a coupling such that

$$\Pr(H_r \neq M_r \mid G) \leqslant \Pr(K_{s,r} \neq J_{s,r} \mid G). \tag{3.4.35}$$

Note that $\mathbb{E}[K_{s,r} \mid G] = \mathbb{E}[J_{s,r} \mid G]$. The approximation of a sum of Bernoulli random variables by a Poisson with the same expectation as the sum is well studied: if $X_1, \ldots, X_l$ are independent random variables with Bernoulli(p_i) distributions such that $\lambda = \sum_{i=1}^{l} p_i$ and $T \sim \mathrm{Poi}(\lambda)$ then there is a coupling [Hol12, Sec. 5.3] such that $\Pr(T \neq \sum_{i=1}^{l} X_i) \leqslant \frac{1}{\lambda}\sum_{i=1}^{s} p_i^2$. This implies that there is a coupling of $K_{s,r}$ and $J_{s,r}$ such that

$$\Pr(K_{s,r} \neq J_{s,r} \mid G) \leqslant \frac{r}{s}, \tag{3.4.36}$$

completing the proof. □

Estimating $\mathcal{W}$

We now combine our results to show that the law of the Kallenberg exchangeable graph generated by the empirical graphex converges to the law of a Kallenberg exchangeable graphgenerated by the underlying $\mathcal{W}$.

There is an immediate subtlety to address: Section 3.4.1 deals with convergence in distribution of point processes (i.e., labeled graphs),

and Section 3.4.2 deals with convergence in distribution of unlabeled graphs. We first give the main convergence result for the point process case. In order to state this result compactly it is convenient to metrize weak convergence. To this end, we recall that the space of boundedly finite measures may be equipped with a metric such that it is a complete separable metric space [DVJ03a, Eqn. A.2.6]. Let $d_p(\cdot,\cdot)$ be the Prokhorov metric on the space of probability measures over boundedly finite measures induced by the aforementioned metric. Then $d_p(\cdot,\cdot)$ metrizes weak convergence: i.e., for a sequence of boundedly finite random measures $\{\Pi_n\}$ it holds that $\Pi_n \xrightarrow{d} \Pi$ as $n \to \infty$ if and only if $d_p(\mathcal{L}(\Pi_n), \mathcal{L}(\Pi)) \to 0$ as $n \to \infty$.

Theorem 3.4.8. *Let Γ be a Kallenberg exchangeable graph generated by non-trivial graphex $\mathcal{W}$ and let $s_1, s_2, \ldots$ be a (possibly random) sequence in $\mathbb{R}_+$ such that $s_k \uparrow \infty$ almost surely as $k \to \infty$. Let $G_k = \mathcal{G}(\Gamma_{s_k})$ for $k \in \mathbb{N}$. Suppose that either*

1. *(s_k) is independent of Γ_k, or*
2. *$s_k = \tau_k$ for all $k \in \mathbb{N}$, where $\tau_1, \tau_2, \ldots$ are the jump times of Γ.*

Then

$$d_p(\mathrm{KEG}(\hat{W}_{(G_k,s_k)}), \mathrm{KEG}(\mathcal{W})) \xrightarrow{p} 0, \tag{3.4.37}$$

as $k \to \infty$.

Proof. For notational simplicity, we treat the deterministic index case first.

For $r \in \mathbb{R}_+$, let $\mathsf{embed}(G_k, s_k)|_r$ denote the probability measure over point processes on $[0,r]^2$ induced by generating a point process according to $\mathsf{embed}(G_k, s_k)$ and restricting to $[0,r]^2$.

By the triangle inequality,

$$\begin{aligned} d_p(\mathrm{KEG}(\hat{W}_{(G_k,s_k)}, r), \mathrm{KEG}(\mathcal{W}, r)) &\leqslant d_p(\mathrm{KEG}(\hat{W}_{(G_k,s_k)}, r), \mathsf{embed}(G_k, s_k)|_r) \qquad (3.4.38)\\ &\quad + d_p(\mathsf{embed}(G_k, s_k)|_r, \mathrm{KEG}(\mathcal{W}, r)). \qquad (3.4.39)\end{aligned}$$

Conditional on G_k and s_k, let X_r^k be an r/s_k-sampling of G_k and let M_r^k be a random subgraph of G_k given by sampling $\mathrm{Poi}(v(G_k)r/s_k)$ vertices with replacement and returning the edge set of the vertex-induced subgraph. By Lemmas 3.4.5 and 3.4.7 it holds that there is a sequence of couplings such that

$$\Pr(M_r^k \neq X_r^k \mid G_k, s_k) \xrightarrow{p} 0, \ k \to \infty. \tag{3.4.40}$$

Observe that $\Gamma_r^k \overset{d}{=} \mathsf{Lbl}_r(M_r^k, \{U_i\})$ and $\xi_r^k \overset{d}{=} \mathsf{Lbl}_r(X_r^k, \{U_i\})$, where $U_i \overset{iid}{\sim} \mathrm{Uni}[0,r]$ for $i \in \mathbb{N}$. Here ξ_r^k is a random labeling of G_k, as

in Theorem 3.4.3. Thus, the couplings of the unlabeled graphs lift to couplings of the point processes such that

$$\Pr(\Gamma_r^k \neq \xi_r^k \mid G_k, s_k) \xrightarrow{p} 0, \ k \to \infty. \tag{3.4.41}$$

The relationship between couplings and total variation distance then implies

$$\|\mathrm{KEG}(\hat{W}_{(G_k,s_k)}, r) - \mathsf{embed}(G_k, s_k)|_r\|_{\mathrm{TV}} \xrightarrow{p} 0, \ k \to \infty, \tag{3.4.42}$$

so also,

$$d_p(\mathrm{KEG}(\hat{W}_{(G_k,s_k)}, r), \mathsf{embed}(G_k, s_k)|_r) \xrightarrow{p} 0, \ k \to \infty. \tag{3.4.43}$$

Second, by Theorem 3.4.3,

$$d_p(\mathsf{embed}(G_k, s_k)|_r, \mathrm{KEG}(\mathcal{W})) \xrightarrow{p} 0, \ k \to \infty. \tag{3.4.44}$$

Thus,

$$d_p(\mathrm{KEG}(\hat{W}_{(G_k,s_k)}, r), \mathrm{KEG}(\mathcal{W}, r)) \xrightarrow{p} 0, \ k \to \infty. \tag{3.4.45}$$

By [Kalo1, Lem. 4.4], convergence in probability for each element of a sequence lifts to convergence in probability of the entire sequence:

$$(d_p(\mathrm{KEG}(\hat{W}_{(G_k,s_k)}, 1), \mathrm{KEG}(\mathcal{W}, 1)), d_p(\mathrm{KEG}(\hat{W}_{(G_k,s_k)}, 2), \mathrm{KEG}(\mathcal{W}, 2)), \cdots) \xrightarrow{p} 0, \ k \to \infty. \tag{3.4.46}$$

As the space of boundedly finite measures on $\mathbb{R}_+^2$ is homeomorphic to the space of sequences of restrictions of boundedly finite measures to $[0, r]^2$, for $r \in \mathbb{N}$, it follows that

$$d_p(\mathrm{KEG}(\hat{W}_{(G_k,s_k)}), \mathrm{KEG}(\mathcal{W})) \xrightarrow{p} 0, \ k \to \infty. \tag{3.4.47}$$

The same proof mutatis mutandis applies for convergence along the jump times. The main substitution is the use of Theorem 3.4.4 in place of Theorem 3.4.3. □

Remark 3.4.9. For graphexes such that $e(G_k)/s_k^3 \to 0$ a.s. and $v(G_k) = \Omega(s_k)$ the convergence in probability above can be replaced by almost sure convergence by replacing all the convergence in probability statements in the body of the proof by almost sure statements. This class of such graphexes includes all integrable $(0, 0, W)$.

We now turn to the analogous result for the case of unlabeled graphs generated by the dilated empirical graphon. We begin with a technical lemma that allows us to deduce convergence in distribution of unlabeled graphs from convergence in distribution of the associated adjacency measures. Note that the map taking an adjacency measure to its associated graph is measurable, but not continuous, and so this result does not follow from a naive application of the continuous mapping theorem.

Lemma 3.4.10. *Let S be a discrete space, T a metric space, $Q_1, Q_2, \ldots$ a tight sequence of probability measures on S, and K a probability kernel from S to T, such that K is injective when considered as a map from probability measures on S to probability measures on T. If $Q_1K, Q_2K, \ldots$ converge weakly to QK then $Q_1, Q_2, \ldots$ converges weakly to Q.*

Proof. Assume otherwise. Case 1: $Q_n \to Q' \neq Q$ weakly. By [Kal01, Lem. 16.24] and the discreteness of S, $Q_nK \to Q'K$ weakly. Since K is injective $Q'K \neq QK$, a contradiction.

Case 2: Q_n does not converge weakly. Since the sequence Q_n is tight it does converge subsequentially. Choose two infinite subsequences $Q_{i_1}, Q_{i_2}, \ldots$ and $Q_{j_1}, Q_{j_2}, \ldots$ with respective limits Q', Q'' with $Q' \neq Q''$. But then, by [Kal01, Lem. 16.24] and the discreteness of S, $Q'_{i_k}K \to Q'K$ and $Q''_{j_k}K \to Q''K$, hence $Q'K = QK = Q''K$, but K is injective, hence $Q' = Q = Q''$, a contradiction. □

The motivating application of this last lemma is showing that a sequence of graphs $G_1, G_2, \ldots$ converge in distribution if and only if their random labelings into $[0, s]$ for some s also converge in distribution. To parse the following theorem, note that when G is a finite random graph, and $s \in \mathbb{R}_+$, then $\Pr(G \in \cdot)\mathsf{embed}(\cdot, s) = \Pr(\mathsf{Lbl}_s(G) \in \cdot)$.

Lemma 3.4.11. *Let $K_s(\cdot) = \mathsf{embed}(\cdot, s)$ for $s \in \mathbb{R}_+$, let $Q, Q_1, Q_2, \ldots$ be probability measures on the space of almost surely finite random graphs, let $\zeta_k = Q_kK_s$ and let $\zeta = QK_s$. Then, $Q_k \to Q$ weakly as $k \to \infty$ if and only if $\zeta_k \to \zeta$ weakly as $k \to \infty$.*

Proof. The forward direction (convergence in distribution of the random graphs implies convergence in distribution of the random adjacency measures) follows immediately from the discreteness of the space of finite graphs and [Kal01, Lem. 16.24].

Conversely, suppose that $\zeta_k \to \zeta$ weakly as $k \to \infty$, and, for every $n \in \mathbb{N}$, let E_n be the set of adjacency measures ξ such that $\xi([0, s]^2) \leqslant n$, i.e., E_n is the event that the graph has fewer than n edges. Note that E_n is a ζ-continuity set by the definition of K_s, and therefore, by weak convergence, $\zeta_k(E_n) \to \zeta(E_n)$ as $k \to \infty$ for every $n \in \mathbb{N}$. Let E'_n be the set of graphs with fewer than n edges. By definition, $Q_k(E'_n) = \zeta_k(E_n)$ and $Q(E'_n) = \zeta(E_n)$, hence $Q_k(E'_n) \to Q(E'_n)$. But E'_n is a finite (hence, compact) set, hence $\{Q_k\}_{k\in\mathbb{N}}$ is tight. Noting in addition that K_s is injective, the result follows from Lemma 3.4.10. □

The following theorem is a formalization of $\hat{W}_{(G_k, s_k)} \to_{GP} \mathcal{W}$ as $k \to \infty$ in probability:

Theorem 3.4.12. *Let Γ be a Kallenberg exchangeable graph generated by non-trivial graphex $\mathcal{W}$ and let $s_1, s_2, \ldots$ be a (possibly random) sequence in $\mathbb{R}_+$ such that $s_k \uparrow \infty$ almost surely as $k \to \infty$. Let $G_k = \mathcal{G}(\Gamma_{s_k})$ for $k \in \mathbb{N}$. Suppose that either*

1. (s_k) *is independent of* Γ_k, *or*
2. $s_k = \tau_k$ *for all* $k \in \mathbb{N}$, *where* $\tau_1, \tau_2, \ldots$ *are the jump times of* Γ.

Then, for every infinite sequence $N \subseteq \mathbb{N}$, *there exists an infinite subsequence* $N' \subseteq N$, *such that*

$$\hat{W}_{(G_k, s_k)} \to_{GP} \mathcal{W} \text{ a.s.} \tag{3.4.48}$$

along N'.

Proof. We first treat the case (1) where the times (s_k) are independent of Γ.

Let $N \subseteq \mathbb{N}$ be an infinite sequence. Theorem 3.4.8 implies that there is some infinite subsequence $N' \subseteq N$ such that, for all $r \in \mathbb{R}_+$, $\mathrm{KEG}(\hat{W}_{(G_k, s_k)}, r) \to \mathrm{KEG}(\mathcal{W}, r)$ weakly almost surely along N'.

Let $r \in \mathbb{R}_+$ and $K_r(\cdot) = \mathsf{embed}(\cdot, r)$. For all $k \in N'$,

$$\mathrm{GPD}\big(\hat{W}_{(G_k, s_k)}, r\big) K_r = \mathrm{KEG}(\hat{W}_{(G_k, s_k)}, r) \text{ a.s.}, \tag{3.4.49}$$

and $\mathrm{GPD}(\mathcal{W}, r) K_r = \mathrm{KEG}(\mathcal{W}, r)$. Moreover, the graph corresponding to a size-r Kallenberg exchangeable graph is almost surely finite. Thus Lemma 3.4.11 applies and we have that $\mathrm{GPD}\big(\hat{W}_{(G_k, s_k)}, r\big) \to \mathrm{GPD}(\mathcal{W}, r)$ weakly a.s. along N'. This holds for all $r \in \mathbb{R}_+$, so we have even that $\hat{W}_{(G_k, s_k)} \to_{GP} \mathcal{W}$ a.s. along N'.

The same proof mutatis mutandis applies for convergence along the jump times. □

Remark 3.4.13. For graphexes such that $e(G_k)/s_k^3 \to 0$ a.s. and $v(G_k) = \Omega(s_k)$, Theorem 3.4.8 implies that $\hat{W}_{(G_k, s_k)} \to_{GP} \mathcal{W}$ as $k \to \infty$ almost surely and not merely in probability. The class of graphexes with these two properties includes all graphexes of the form $(0, 0, W)$ for integrable W.

ESTIMATION FOR UNKNOWN SIZES

We now turn to the case where only the graph structure of the Kallenberg exchangeable graph is observed, rather than the graph structure and the sizes of the observation.

We first show how distinct adjacency measures can give rise to the same graph sequence. For a measurable map $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ and adjacency measure ξ, define ξ^ϕ to be the measure given by $\xi^\phi(A \times B) = \xi(\phi^{-1}(A) \times \phi^{-1}(B))$, for every measurable $A, B \subseteq \mathbb{R}_+$. The graph sequences underlying an adjacency measure ξ is invariant to the action $\phi \mapsto \xi^\phi$ of every strictly monotonic and increasing function ϕ.

Proposition 3.5.1. *Let* ξ *be an adjacency measure and let* $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ *be strictly monotonic and increasing. Then* $\mathscr{G}(\xi) = \mathscr{G}(\xi^\phi)$.

Proof. Let $\{\tau_k\}$ and $\{\tau_k^\phi\}$ be the jump times of ξ and ξ^ϕ, respectively.

Since ϕ is strictly monotonic it is also invertible. From this observation it is easily seen that (θ_i, θ_j) is an atom of ξ if and only if $(\phi(\theta_i), \phi(\theta_j))$ is an atom of ξ^ϕ. It is then clear that, for all $k \in \mathbb{N}$, $\phi(\tau_k) = \tau_k^\phi$ and, moreover, the graph structure of $\{(x_i, \tau_k) : (x_i, \tau_k) \in \xi\}$ is equal to the graph structure of $\{(y_i, \tau_k^\phi) : (y_i, \tau_k^\phi) \in \xi^\phi\}$. That is, the subgraph of all edges added at the kth step is equal for both graph sequences, for all $k \in \mathbb{N}$. Moreover, the first entry of each graph sequence is (obviously) equal to the subgraph of all edges added at the first step. The proof is then completed by induction. □

If ϕ is an arbitrary strictly monotonic mapping and ξ is an exchangeable adjacency measure, it will not generally be the case that ξ^ϕ is exchangeable. One family of mappings that preserves exchangeability is $\phi(x) = cx$, for $c \in \mathbb{R}_+$. We define the c-dilation of an adjacency measure ξ to be the adjacency measure ξ^ϕ for this map. Because ξ^ϕ is exchangeable there is some graphex $\mathcal{W}'$ that generates it: the next result shows that the $\frac{1}{c}$-dilation of a Kallenberg exchangeable graph corresponds to a c-dilation of its graphex.

Lemma 3.5.2. *Let Γ be a Kallenberg exchangeable graph with graphex $\mathcal{W} = (I, S, W)$. Then the $\frac{1}{c}$-dilation of Γ is a Kallenberg exchangeable graph Γ' with generating graphex $\mathcal{W}' = (I', S', W')$ where $I' = c^2 I$, $S'(x) = cS(x/c)$, and $W'(x, y) = W(x/c, y/c)$.*

Proof. Let Γ be a Kallenberg exchangeable graph generated by W with latent Poisson processes $\Pi, \Pi_1, \Pi_2, \ldots$, and Π^i on $\mathbb{R}_+^2$ and $\mathbb{R}_+^3$, respectively. Define $f(\Pi) = \{(\frac{1}{c}\theta, c\vartheta) : (\theta, \vartheta) \in \Pi\}$, define $f(\Pi_n) = \{(\frac{1}{c}\sigma, c\chi) : (\sigma, \chi) \in \Pi_n\}$, for $n \in \mathbb{N}$, and define $f(\Pi^i) = \{(\frac{1}{c}\rho, \frac{1}{c}\rho', c^2\eta) : (\rho, \rho', \eta) \in \Pi^i\}$. Note that $f(\Pi)$ and $f(\Pi_n)$, for $n \in \mathbb{N}$, are unit-rate Poisson processes on $\mathbb{R}_+^2$, and $f(\Pi^i)$ is a unit-rate Poisson process on $\mathbb{R}_+^3$. Indeed, the joint law of $(\Pi, \Pi_1, \Pi_2, \ldots, \Pi^i)$ is the same as that of $(f(\Pi), f(\Pi_1), f(\Pi_2), \ldots, f(\Pi^i))$.

Then Γ', the $\frac{1}{c}$-dilation of Γ, is the Kallenberg exchangeable graph generated by $\mathcal{W}'$ with latent Poisson processes $f(\Pi), f(\Pi_1), f(\Pi_2), \ldots, f(\Pi^i)$ *reusing* the same i.i.d. collection $(\zeta_{\{i,j\}})$ in $[0, 1]$ as was used to generate Γ. To see this, note that Γ' includes edge $(\frac{1}{c}\theta_i, \frac{1}{c}\theta_j)$ if and only if $\zeta_{\{i,j\}} \leqslant W'(c\vartheta_i, c\vartheta_j)$ if and only if $\zeta_{\{i,j\}} \leqslant W(\vartheta_i, \vartheta_j)$ if and only if Γ includes edge (θ_i, θ_j). Similarly, Γ' includes edge $(\frac{1}{c}\theta_i, \frac{1}{c}\sigma_{ij})$ if and only if $c\chi_{ij} \leqslant S'(c\vartheta) = cS(\vartheta)$ if and only if $\chi_{ij} \leqslant S(\vartheta)$ if and only if Γ includes edge (θ_i, σ_{ij}). Finally, Γ' includes edge $(\frac{1}{c}\rho, \frac{1}{c}\rho')$ if and only if $c^2\eta \leqslant I' = c^2 I$ if and only if Γ includes edge (ρ, ρ'). Thus Γ' is a $\frac{1}{c}$-dilation of Γ, as was to be shown. □

Define the *c-dilation of a graphex* $\mathcal{W}$ to be the graphex $\mathcal{W}'$ defined in the statement of Lemma 3.5.2. We have the following consequence:

Theorem 3.5.3. *Let $\mathcal{W}$ be a graphex, let $\mathcal{W}'$ be the c-dilation of $\mathcal{W}$ for some $c > 0$, and let Γ and Γ' be Kallenberg exchangeable graphs with graphexes $\mathcal{W}$ and $\mathcal{W}'$, respectively. Then $\mathcal{G}(\Gamma) \overset{d}{=} \mathcal{G}(\Gamma')$.*

Proof. Follows immediately from Lemma 3.5.2 and Proposition 3.5.1. □

As a consequence of this result, if the observed data is the graph sequence—that is, if the size s is unknown—then the dilation of the generating graphex is not identifiable. Therefore, the notion of estimation that we used in the known-size setting is not appropriate, because it requires $\mathcal{G}_r(\mathcal{W}_n) \xrightarrow{d} \mathcal{G}_r(\mathcal{W})$ as $n \to \infty$ for all sizes $r \in \mathbb{R}_+$.

The appropriate notion of estimation in this setting is then:

Definition 3.5.4. Let $\mathcal{W}, \mathcal{W}_1, \mathcal{W}_2, \ldots$ be a sequence of graphexes, and let $\Gamma, \Gamma^1, \Gamma^2, \ldots$ be Kallenberg exchangeable graphs generated by each graphex. Write $\mathcal{W}_k \to_{GS} \mathcal{W}$ as $k \to \infty$ when $\mathcal{G}(\Gamma^k) \xrightarrow{d} \mathcal{G}(\Gamma)$ as $k \to \infty$.

Note that this is equivalent to requiring convergence in distribution of the length-l prefixes of the graph sequences, for all $l \in \mathbb{N}$. Intuitively, a length-l graph sequence generated by the estimator is close in distribution to a length-l graph sequence generated by the true graphex, provided the observed graph is large enough. This perspective explains how a sequence of compactly supported graphexes can estimate a graphex that is not itself compactly supported. The following is immediate from Theorem 3.5.3.

Corollary 3.5.5. *Let $\mathcal{W}, \mathcal{W}_1, \mathcal{W}_2, \ldots$ be a sequence of graphexes, let $c, c_1, c_2, \cdots > 0$, and let $\mathcal{W}^c, \mathcal{W}_1^{c_1}, \mathcal{W}_2^{c_2}, \ldots$ be the corresponding dilations. Then $\mathcal{W}_k \to_{GS} \mathcal{W}$ as $k \to \infty$ if and only if $\mathcal{W}_k^{c_k} \to_{GS} \mathcal{W}^c$ as $k \to \infty$.*

Intuitively speaking, $\mathcal{W}_k \to_{GS} \mathcal{W}$ as $k \to \infty$ demands less than $\mathcal{W}_k \to_{GP} \mathcal{W}$ as $k \to \infty$, because in the former case we don't need to find a correct rate of dilation for the graphex. The intuition that convergence in distribution of the graph sequence is weaker than convergence in distribution of $(\mathcal{G}(\Gamma_s))_{s \in \mathbb{R}_+}$ is borne out by the next lemma:

Lemma 3.5.6. *Let $\mathcal{W}, \mathcal{W}_1, \mathcal{W}_2, \ldots$ be graphexes where $\mathcal{W}$ is non-trivial and $\mathcal{W}_k \to_{GP} \mathcal{W}$ as $k \to \infty$. Then $\mathcal{W}_k \to_{GS} \mathcal{W}$ as $k \to \infty$.*

Proof. Let Γ^k be Kallenberg exchangeable graphs generated by $\mathcal{W}_k$, and let Γ be generated by $\mathcal{W}$. For $n \in \mathbb{N}$, let $G_n^k = \mathcal{G}(\Gamma_n^k)$, and let $G_n = \mathcal{G}(\Gamma_n)$.

Consider the sequence $H_n^k = (\mathcal{G}(\Gamma_1^k), \mathcal{G}(\Gamma_2^k), \ldots, \mathcal{G}(\Gamma_n^k))$, where each entry is itself an a.s. finite graph sequence and entry j is a prefix of entry $j+1$. Let $\eta_n^k = \Pr(H_n^k \in \cdot)$, and let $\eta_n = \Pr((\mathcal{G}(\Gamma_1), \mathcal{G}(\Gamma_2), \ldots, \mathcal{G}(\Gamma_n)) \in \cdot)$. Intuitively speaking, we are breaking up the graph sequence of the entire Kallenberg exchangeable graph into the graph sequences up

to size $1, 2, \dots$ and η_n is the joint distribution of the first n of these partial graph sequences. Our short term goal is to show that $\eta_n^k \to \eta_n$ weakly as $k \to \infty$.

To that end, let G be a finite graph and consider the random variable

$$L_n(G) = (\mathscr{G}(\mathsf{Lbl}_n(G)([0,j)^2 \cap \cdot)))_{j=1,\dots,n}. \tag{3.5.1}$$

This is a nested sequence of graph sequences given by mapping G to an adjacency measure on $[0,n)^2$ and then returning the sequence of graph sequences corresponding to this adjacency matrix at sizes $1, \dots, n$. The significance of this construction is that we may use it to define a probability kernel,

$$K_n(G, \cdot) = \Pr(L_n(G) \in \cdot), \tag{3.5.2}$$

such that that $\Pr(G_n^k \in \cdot)K_n = \mathbb{E}K_n(G_n^k, \cdot) = \eta_n^k$ and $\Pr(G_n \in \cdot)K_n = \mathbb{E}K_n(G_n, \cdot) = \eta_n$. By assumption, we have $\mathcal{W}_k \to_{GP} \mathcal{W}$ as $k \to \infty$, whence $G_n^k \xrightarrow{d} G_n$ as $k \to \infty$. By the discreteness of the space of finite graphs and [Kal01, Lem. 16.24] it then holds that,

$$\Pr(G_n^k \in \cdot)K_n \to \Pr(G_n \in \cdot)K_n, \tag{3.5.3}$$

weakly as $k \to \infty$. It thus holds by the construction of K_n that

$$\eta_n^k \to \eta_n, \tag{3.5.4}$$

weakly as $k \to \infty$.

We now have that an arbitrary length prefix of the graph sequence converges in distribution, when the notion of length is given by the latent sizes. It remains to argue that this convergence holds for arbitrary prefixes in the usual sequence sense. To that end, we observe that because Eq. (3.5.4) holds for all $n \in \mathbb{N}$, by [Kal01, Thm. 4.29] it further holds that

$$(\mathscr{G}(\Gamma_1^k), \mathscr{G}(\Gamma_2^k), \dots) \xrightarrow{d} (\mathscr{G}(\Gamma_1), \mathscr{G}(\Gamma_2), \dots), \ k \to \infty. \tag{3.5.5}$$

There is function f such that, for every locally finite but infinite adjacency measure ξ on $\mathbb{R}_+^2$, with restrictions ξ_j to $[0,j)^2$,

$$f(\mathscr{G}(\xi_1), \mathscr{G}(\xi_2), \dots) = \mathscr{G}(\xi) \tag{3.5.6}$$

and f is even continuous because every finite prefix of $\mathscr{G}(\xi)$ is determined by some finite prefix of the left hand side. Extend f to the space of all nested graph sequences arbitrarily. Note that f is continuous at Γ a.s. because Γ is a.s. locally finite and $\|W\|_1 > 0$ implies Γ is a.s. infinite. Hence, the result follows by the continuous mapping theorem [Kal01, Thm. 4.27]. □

We now turn to establishing the main estimation result for the setting where the sizes are not included as part of the observation. In this setting, the observations are increasing sequences of graphs $G_1, G_2, \ldots$. There are two natural models for the observations: In one model, $G_k = \mathcal{G}(\Gamma_{s_k})$ for some Kallenberg exchangeable graph Γ and (possibly random, independent, and a.s.) increasing and diverging sequence of sizes $s_1, s_2, \ldots$. Alternatively, in the other model, the sequence $G_1, G_2, \ldots$ is the graph sequence $\mathscr{G}(\Gamma)$ of some Kallenberg exchangeable graph Γ.

A natural estimator is the empirical graphon, $\tilde{W}_{G_k}$, reflecting the intuition that the dilation necessary in the previous section for convergence of the generated Kallenberg exchangeable graph is irrelevant for convergence in distribution of the associated graph sequence. Somewhat more precisely, we view the empirical graphon as the canonical representative of the equivalence class of graphons given by equating graphons that induce the same distribution on graph sequences. The main result of this section is that $\tilde{W}_{G_k} \to_{GS} \mathcal{W}$ as $k \to \infty$ in probability, for either of the natural models for the observed sequence $G_1, G_2, \ldots$.

Theorem 3.5.7. *Let Γ be a Kallenberg exchangeable graph generated by some non-trivial graphex $\mathcal{W}$ and let $G_1, G_2, \ldots$ be some sequence of graphs such that either*

1. *There is some random sequence (s_k), independent from Γ, such that $s_k \uparrow \infty$ a.s. and $G_k = \mathcal{G}(\Gamma_{s_k})$ for all $k \in \mathbb{N}$, or*
2. *$(G_1, G_2, \ldots) = \mathscr{G}(\Gamma)$.*

Then, for every infinite sequence $N \subseteq \mathbb{N}$, there is an infinite subsequence $N' \subseteq N$, such that

$$\tilde{W}_{G_k} \to_{GS} \mathcal{W} \text{ a.s.}, \tag{3.5.7}$$

along N'.

Proof. We prove case (1). Case (2) follows mutatis mutandis, substituting τ_k for s_k.

Let $\hat{W}_{(G_k, s_k)}$ denote the dilated empirical graphon of G_k with observation size s_k. By Theorem 3.4.12, for every sequence $N \subseteq \mathbb{N}$, there is an infinite subsequence $N' \subseteq N$, such that $\hat{W}_{(G_k, s_k)} \to_{GP} \mathcal{W}$ along N'. By Lemma 3.5.6 and $\mathcal{W}$ being non-trivial, this implies that $\hat{W}_{(G_k, s_k)} \to_{GS} \mathcal{W}$ along N'. For every k, $\tilde{W}_{G_k}$ is some dilation of $\hat{W}_{(G_k, s_k)}$, hence, the result follows by Corollary 3.5.5.

□

4

SAMPLING PERSPECTIVES ON SPARSE EXCHANGEABLE GRAPHS

INTRODUCTION

This chapter is concerned with the connections between the theory of graph limits, the statistical modeling of networks, and exchangeability. In the setting of dense graphs, these topics meet in the theory of graphons, which are fundamental in the study of graph limits [BCLS+06; LS06; LS07; BCLS+08b; BCLS+12] (see [Lov13] for a review) as well as providing the foundation for many of the statistical network models in current use, e.g., [NS01; HRH02; ABFX08; MJG09; LOGR12] (see [OR15] for a review). Our goal in this chapter is to explain the connection between these subjects in the sparse graph setting. This will lead us naturally to the resolution of some conceptual sticking points of the graphex models.

Graphons as limit objects for dense graphs sequences arise naturally in the dense setting: many natural notions of similarity, such as left convergence motivated by extremal graph theory, right convergence motivated by studying statistical physics (or, equivalently, graphical) models on graphs, as well as quotient convergence motivated by combinatorial optimization, all lead to graphons over probability spaces as the completion of the space of dense graphs [BCLS+06; BCLS+08b; BCLS+12]. These notions of convergence turn out to all be equivalent, and can be metrized by the *cut metric* (discussed below), making the theory of graph convergence a well rounded mathematical theory. Exchangeable random graphs generated from a graphon can be shown to converge to the generating graphon [LS06], creating a first connection between graphons as models for exchangeable random graphs and as limits of sequences of sparse graphs. See [DJ08] for a systematic overview of the relationship between the theory of graph convergence and the theory of exchangeable random graphs in the dense graph setting.

Following this graph limits perspective, [BCCH16] arrived at the graphex framework independently of [VR16]. A major contribution of their paper is a generalization of convergence in cut distance to the sparse graph regime; the natural limit objects of the resultant theory are graphexes of the form $(0, 0, W)$. The graphex framework then offers a generalization of the dense graph theory to the sparse graph regime that is compelling in that it preserves the essential character of the traditional graphon models, while simultaneously allowing much

greater flexibility. This applies both for statistical network modeling, and for graph limits.

However, this picture is somewhat superficial, since it leaves many questions unanswered. Why do we represent graphs as point processes? Why does the corresponding notion of exchangeability give a much broader class of models than the adjacency matrix exchangeability? Why should the points in the latent feature space be distributed according to a Poisson process? What motivates the particular way of embedding graphs into the space of graphons over $\mathbb{R}_+$ that [BCCH16] use to translate convergence in the cut metric for graphons into a notion of convergence in metric for graphs? Why are graph limits and statistical network modeling so closely tied together? The contribution of the present chapter is to resolve these conceptual difficulties by relating the core ideas—graph limits, statistical network modeling and exchangeability—to a certain natural scheme for sampling random subgraphs from larger graphs.

Our first main contribution is the introduction and development of *sampling convergence*, a new notion of graph limit that generalizes left convergence, [BCLS+06; BCLS+08b] a core concept in the graphon theory of limits of dense graphs, to a notion that is also meaningful for sparse graphs. We show that sampling convergence both generalizes the metric convergence of [BCCH16], and allows us to formalize the notion of sampling a data set from an infinite size population network; it thereby connects graph limits and statistical network modeling. Our second main contribution is that the ad hoc assumption of exchangeability may be replaced by a more natural equivalent invariance given in terms of the sampling scheme. This symmetry makes no reference to the point process representation of random graphs or to the associated notion of exchangeability; this allows us to understand these ideas as mathematical artifices rather than conceptual cornerstones of the theory.

We begin by explaining our limit theory as a natural generalization of the dense graph limit theory. In the setting of dense graphs, one of the core limit notions is left convergence, the convergence of subgraph densities. In the course of explaining the connection between exchangeability and graph limits in the dense graph setting, Diaconis and Janson [DJ08] present the following perspective on left convergence. Given a graph G_j, for each $k \in \mathbb{N}$ we draw a random subgraph $H_{j,k}$ of G_j by selecting k vertices independently at random and returning the induced subgraph; a sequence $G_1, G_2, \ldots$ is left convergent when, for all $k \in \mathbb{N}$, the random graphs $H_{j,k}$ converge in distribution as $j \to \infty$. Intuitively speaking, this notion of convergence encodes the idea that two large graphs are similar when it is difficult to tell them apart by randomly sampling small subgraphs from each.

It is straightforward to see why left convergence is informative only for dense graph sequences: if the graph sequence $G_1, G_2, \ldots$ is sparse then the probability that a random k vertex subgraph of G_j contains even a single edge goes to 0 as j becomes large. The resolution we propose here is, intuitively speaking, to generalize this sampling scheme in a way that fixes the target number of *edges* in the randomly sampled subgraph, instead of the number of vertices.

The first key tool for formalizing this is p-sampling, introduced in the previous chapter.

Definition 4.1.1. A p-*sampling* $\mathsf{Smpl}(G, p)$ of a graph[1] G is a random subgraph of G given by including each vertex of G independently with probability $\min(p, 1)$, then discarding all isolated vertices in the resulting induced subgraph, and finally returning the unlabeled graph corresponding to this subgraph.

The critical property that distinguishes p-sampling from independent vertex sampling is that vertices that do not participate in any edges in the vertex induced subgraph are thrown away. Note that by definition, $\mathsf{Smpl}(G, p)$ is always unlabeled, whether G is labeled or not.

We may now define our notion of graph limit:

Definition 4.1.2. Let $e(G)$ denote the number of non-loop edges of a graph G.

Definition 4.1.3. A sequence of graphs $G_1, G_2, \ldots$ is *sampling convergent* if, for all $r \in \mathbb{R}_+$, the random graphs $\mathsf{Smpl}\big(G_j, r/\sqrt{2e(G_j)}\big)$ induced by $r/\sqrt{2e(G_j)}$-sampling of G_j converge in distribution as $j \to \infty$.

For the remainder of the introduction, we will restrict our attention to sequences of simple graphs; loops are treated in the body of the chapter.

Sampling convergence can be understood as a modification of left convergence as follows: we draw an increasing number of vertices as $j \to \infty$ because if we drew only a fixed number k then the induced graph would be empty in the limit. Since the number of sampled vertices diverges, we instead fix the target number of sampled edges. Because we are selecting vertices at random, the number of edges in the vertex induced subgraph must be random, so a natural way to fix the size of the sampled subgraph as $j \to \infty$ is to require the expected number of edges to be constant. This requirement dictates that each vertex is included with probability proportional to $1/\sqrt{e(G_j)}$; the convention we choose for the proportionality constant gives

$$\mathbb{E}\left[e\left(\mathsf{Smpl}\left(G_j, r/\sqrt{2e(G_j)}\right)\right)\right] = r^2/2 \tag{4.1.1}$$

1 Throughout this chapter, a graph will be a graph without multiple edges, but it may not be simple, i.e., it may contain edges joining a vertex to itself; unless explicitly mentioned, all graphs will be finite.

for all $j \in \mathbb{N}$. Because the number of sampled vertices goes to infinity as $j \to \infty$, it is not possible to have convergence in distribution of the vertex sampled subgraphs. This problem is solved by using p-sampling instead of independent vertex sampling; that is, we simply throw away the vertices that are isolated in the sampled subgraph.

Our first main result is that the natural limit object of a sampling convergent sequence is an integrable graphex $\mathcal{W} = (I, S, W)$.[2] The sense in which the graphex is the natural limit object is given by Theorem 4.3.10: for every sampling convergent sequence $G_1, G_2, \ldots$ there is some integrable graphex $\mathcal{W}$ such that, for all $s \in \mathbb{R}_+$, $\mathsf{Smpl}\big(G_j, s/\sqrt{2e(G_j)}\big) \xrightarrow{d} \mathcal{G}(\Gamma_s)$ as $j \to \infty$, where $(\Gamma_s)_{s\in\mathbb{R}_+}$ is generated by $\mathcal{W}$. That is, the limiting distribution of the sampled subgraph is characterized by the graphex that is the sampling convergent limit.

We complete the limit theory by showing that every integrable graphex arises as the sampling convergent limit of some graph sequence, at least up to some equivalencies (Theorem 4.4.3), and by metrizing the convergence and characterizing the associated metric space (Theorems 4.6.5 and 4.6.6). In consequence of the former result, the (integrable) graphex models can be understood conceptually as originating as the limit objects of sampling convergence, without any direct appeal to exchangeability (although in fact our technical arguments lean heavily on exchangeability and the associated machinery).

This last observation raises the question of whether the graphex models can be characterized directly in terms of p-sampling, without appeal to either exchangeability or graph limits. The motivation in [VR16] for the introduction of p-sampling was the observation that a p-sampling of $\mathcal{G}(\Gamma_s)$ is equal in distribution to $\mathcal{G}(\Gamma_{ps})$; i.e., this is the sampling scheme that describes the relationship between Kallenberg exchangeable graphs at different sizes. We prove in Theorem 4.7.2 that this is in fact a defining property of the graphex models. That is, if $(G_s)_{s\in\mathbb{R}_+}$ is a family of unlabeled random graphs such that for any $s \in \mathbb{R}_+$ and any $p \in (0,1)$ the p-sampling of G_s is equal in distribution to G_{ps}, then there is some graphex $\mathcal{W}$ such that $G_s \overset{d}{=} \mathcal{G}(\Gamma_s)$ for all $s \in \mathbb{R}_+$, where $(\Gamma_s)_{s\in\mathbb{R}_+}$ is generated by $\mathcal{W}$. This gives a formal sense in which this sampling invariance is equivalent to the notion of exchangeability originally used to define exchangeable random graphs.

We now turn to explaining the connection between our results and statistical network modeling, and the relationship to other notions of graph limits.

2 We say a graphex is integrable if each of its components is integrable.

Statistical Network Modeling

The major motivation in [VR15] for the introduction of the graphex framework was as a tool for the statistical analysis of network-valued data sets. These models are attractive for this purpose because they offer a sparse graph generalization of the graphon model of the exchangeable arrays framework, which underlies many popular models. In this setting, the conceptual challenge brought on by exchangeability is that because it is unclear what the symmetry means in practical terms it is also unclear what the practical applicability of the models is. In particular, we would like a clear articulation of the circumstances under which it is appropriate to model a data set by a Kallenberg exchangeable graph.

Following [CD15], a statistical model can be understood as consisting of two parts: a data generating process and a sampling scheme for collecting a data set from a realization of this process. In the network setting, this is envisioned as some real world process that generates a large population graph from which the data set is then somehow sampled. In order to assess the applicability of a statistical network model we should articulate the associated data generation mechanism and sampling scheme.

The most obvious sampling scheme to associate with the graphex model is p-sampling. Having assumed p-sampling, the question of what data generating mechanism gives rise to the population is subtle. One obvious guiding principle is that we ought to be able to make meaningful inferences about the population on the basis of the sample. For example, if the data generating process is itself a Kallenberg exchangeable graph with graphex $\mathcal{W}$ then the sample will be distributed as finite graph generated by $\mathcal{W}$; inferences about the population then take the form of inferences about $\mathcal{W}$. However, KEGs have some properties that are highly undesirable for a model of a data generating process. For example, KEGs can only grow and, moreover, can grow only by adding edges connecting to vertices that have never been seen before. As a model for a social network this would mean that two people who are friends may never stop being friends, and two people who are not yet friends may never form a link in the future.

In classical statistics, data sets are often envisioned as being drawn independently from some very large population, often idealized as infinite. In our setting, the analogous thing is to envision a particular (fixed size) observation as a draw from a very large population network where each vertex is included independently with small probability. To formalize the infinite-size population idealization, consider the limit where the size of the population, created according to data generating mechanism, becomes infinite while the vertex inclusion probability goes to 0 at a rate that keeps the size of the observed data

set constant. That is, we imagine $e(G_j) \to \infty$ and the inclusion probability $p_j = \Theta(1/\sqrt{e(G_j)})$. In this case, a minimal requirement for the sampled data set to be informative about the limiting population is that the distribution of the sample should converge. We have thus been led to the following precept: the data generating mechanism should give rise to a sequence of population graphs that is sampling convergent. This is as far as we need go: by Theorem 4.3.10 the requirement of sampling convergence already implies that the observation is distributed according to some integrable graphex $\mathcal{W}$.

The preceding can be summarized as:

> Finite size KEGs approximate statistical network models that arise from vertex sampling of a population that is generated according to some sampling convergent data generating process. In the infinite population limit this approximation becomes exact.

It's worth emphasizing that this is much broader than it may appear at first glance. For example, this perspective may even be appropriate in situations where we observe the entire available network, as long as the physical mechanism generating the network is sampling convergent and the process that restricts to a finite size observation can be modeled approximately as an independent sampling of the vertices.

[Orb17] gives a treatment of the broad idea of defining schemes for statistical network modeling by way of defining a sampling scheme and studying the models compatible with the symmetries thereby induced. One perspective on the present chapter is that we work out the realization of this program for p-sampling.

Graph Limits

Sampling convergence gives a notion of graph limit for deterministic sequences of unlabeled graphs. We now explain the connection to several other notions of large graph limit, namely

1. the convergence of sequences of randomly labeled graphs,
2. the metric convergence of [BCCH16], and
3. the consistent estimation of [VR16].

Randomly Labeled Graphs

The first of these is fundamental to the development of the theory in the present chapter. Exchangeability is a concept of infinite size labeled random graphs, but the theory of graph limits deals with non-random sequences of graphs. It is then somewhat mysterious

why there should be such a close connection between graph limits and exchangeable random graphs.

In the dense graph setting, this manifested as the development of the theory of exchangeable arrays [Ald81; Hoo79; Kal05] on one hand and the independent development of the theory of dense graph limits [BCLS+06; LS06; LS07; BCLS+08b; BCLS+12] on the other. The connection between the two perspectives is explained by [DJ08; Aus08], the development of which is roughly as follows. In the dense graph setting the popular notions of graph limits are all equivalent to left convergence, which says that a growing sequence of graphs G_j converges if, for any fixed graph F, the proportion of copies of F in G_j converges. The first key insight is that this can be phrased in probabilistic language by viewing left convergence as requiring convergence in distribution of random subgraphs $H_{j,k}$ drawn by selecting k vertices independently from G_j, for all $k \in \mathbb{N}$. The second key insight is that we may pass from non-random sequences of graphs $(G_j)_{j\in\mathbb{N}}$ to sequences of random adjacency matrices $(A(G_j))_{j\in\mathbb{N}}$ by randomly labeling the vertices of each G_j by $\{1, \dots, v(G_j)\}$; this gives a construction such that for each fixed j the random adjacency matrix is exchangeable. We then observe that convergence in distribution of randomly sampled k vertex subgraphs is equivalent to convergence in distribution of the random adjacency matrices given by restricting $A(G_j)$ to its upper left $k \times k$ sub-matrix. Now, using standard probability theory machinery, distributional convergence of all size k prefixes is enough for even distributional convergence of $A(G_j)$ as $j \to \infty$. As one might expect, the limit of $A(G_1), A(G_2), \dots$ is an infinite exchangeable array. By the Aldous–Hoover theorem there is then some graphon W that characterizes the distribution of this array. This graphon is the same as the left convergent limit of the graph sequence $G_1, G_2, \dots$

In the present context, the relationship between non-random graph sequences and sequences of randomly labeled objects is captured as a correspondence between edge sets and point processes. The point processes will be given in terms of *adjacency measures*, defined as locally finite measures of the form $\xi = \sum_{i,j} \delta_{(\theta_i,\theta_j)}$, where the sum goes over all ordered paris i, j such that $\{i, j\}$ is an edge of a countable graph G (possibly containing some loops, i.e., edges joining a vertex to itself) and $\theta_i \in \mathbb{R}_+$ with $\theta_i \neq \theta_j$ for $i \neq j$.

Definition 4.1.4. Let G be labeled or unlabeled graph and let $s > 0$. A *random labeling of* G *into* $[0, s)$ is a random adjacency measure obtained by labelling the vertices randomly with i.i.d. labels in $[0, s)$.

For a graph sequence $G_1, G_2, \dots$ it may not be immediately obvious what the ranges $[0, s_1), [0, s_2), \dots$ of the random labelings should be. Our choice here is $s_j = \sqrt{2e(G_j)}$, which has the virtue that for all measurable sets $A, B \subseteq \mathbb{R}_+$, the expected number of edges between vertices with labels in A and B is independent of the graph.

Definition 4.1.5. We define the *canonical labeling* $\mathsf{Lbl}(G)$ of a graph G to be the random labeling of G into $[0, \sqrt{2e(G)})$.

The relationship between sampling convergence of a graph sequence and the distributional convergence of the canonical labelings is closely analogous to the relationship between left convergence of a graph sequence and the distributional convergence of the associated random adjacency matrices. We show in Section 4.3 that the graph sequence $G_1, G_2, \ldots$ is sampling convergent to $\mathcal{W}$ if and only if the canonical labelings $\mathsf{Lbl}(G_1), \mathsf{Lbl}(G_2), \ldots$ converge in distribution to an infinite exchangeable point process characterized by $\mathcal{W}$. Indeed, the machinery of distributional convergence of point processes is core to many of our main results.

In [Ald09] a broad program for studying the limits of complex structures of increasing size is outlined. The basic idea is to define a notion of sampling on these structures such that for each complex object C_j we may sample some substructure $D_j^{(k)}$ of size k; convergence is then defined as convergence in distribution of $\{D_j^{(k)}\}$ as $j \to \infty$ for all sizes k. The natural limit is then the joint distribution of the limiting object for all sizes k. This object will have some symmetries imposed by the sampling scheme, and so might admit some more compact representation, which would then be the natural limit object. One perspective on the present chapter is that we realize this program for p-samplings of families of growing graphs.

Metric Convergence

One of the important tools in the theory of dense graph limits is the cut distance between two graphs or graphons first introduced in [BCLS+06]. The cut metric defines a notion of distance that, essentially, captures how similar two graphs or graphons look at low resolutions; see Fig. 9. We define cut distance formally in Section 4.2. One of the contributions of [BCCH16] was to generalize the cut distance to graphons supported on general σ-finite spaces, and in particular for graphons $W\colon \mathbb{R}_+^2 \to [0,1]$, and to use this notion to compare two graphs via an embedding of the space of graphs into the space of graphons $W\colon \mathbb{R}_+^2 \to [0,1]$, mapping a graph G into what they called the stretched canonical graphon $W^{G,s}$ of G. Using this embedding, they then introduced the "stretched cut distance" between two graphs as the cut distance between the stretched canonical graphons of these graphs. That paper developed a theory of graph limits based on convergence in this stretched cut distance, where the essential idea is to transform a sequence of graphs into a sequence of stretched canonical graphons and ask for cut metric convergence of this sequence; see Fig. 9. This turns out to generalize the dense graph cut metric convergence, and the generalized limit objects are the same generalized graphons that arise as limits in sampling convergence.

In the dense graph setting, convergence in cut distance is equivalent to left convergence. Given that sampling convergence is an analogue of left convergence, it is natural to expect that there should be some connection with convergence in the stretched cut distance. Indeed this is so, and in Theorem 4.5.5 we show that the two notions of convergence coincide for any graph sequence that is subsequentially convergent with respect to the stretched cut metric. Thus, in particular, convergence in the stretched cut distance implies sampling convergence.

Our main motivation for the introduction of sampling convergence is conceptual clarity. However, it is also worth noting that sampling convergence (and the associated move from graphons to graphexes) has some pleasant mathematical properties that stretched cut convergence does not. In particular, every graph sequence is subsequentially sampling convergent, but this is not true for stretched cut metric convergence.

Consistent Estimation

[VR16] (equivalently, Chapter 3) deals with the problem of estimating $\mathcal{W}$ from a growing sequence of unlabeled graphs $G_1, G_2, \dots$ generated from $\mathcal{W}$. Simplifying somewhat, the data set is modeled as $G_j = \mathcal{G}(\Gamma_{s_j})$ for some sequence $s_1, s_2, \dots$ of observation sizes with $s_j \uparrow \infty$ and $(\Gamma_s)_{s\in\mathbb{R}_+}$ generated by $\mathcal{W}$. The basic goal of estimation is to produce a sequence of graphexes $\mathcal{W}_{G_1}, \mathcal{W}_{G_2}, \dots$ such that $\mathcal{W}_{G_j} \to \mathcal{W}$ as $j \to \infty$, for some notion of convergence that formalizes the idea that the distribution defined by the estimated graphex should be asymptotically the same as the distribution defined by the true underlying graphex. In the graphex setting, there are two natural distinct notions of estimation depending on whether the observation sizes are included as part of the observation; both of these are closely related to the sampling convergence of the present chapter.

Let $\mathrm{GPD}(\mathcal{W}, s) = \Pr(\mathcal{G}(\Gamma_s) \in \cdot \mid \mathcal{W})$ denote the probability distribution over unlabeled size s graphs generated by $\mathcal{W}$. In the setting where the sizes are known, estimation is formalized by defining $\mathcal{W}_j \to_{\mathrm{GP}} \mathcal{W}$ as $j \to \infty$ to mean $\mathrm{GPD}(\mathcal{W}_j, s) \to \mathrm{GPD}(\mathcal{W}, s)$ weakly as $j \to \infty$, for all $s \in \mathbb{R}_+$. That is, $\mathcal{W}_1, \mathcal{W}_2, \dots$ estimates $\mathcal{W}$ if the random graphs generated by the estimators converge in distribution to the random graphs generated by $\mathcal{W}$. Here, GP stands for graphon process, and GPD stands for graphon process distribution.

For a graph G, define $\hat{W}_{(G,s)} \colon [0, v(G)/s]^2 \to \{0, 1\}$, the dilated empirical graphon of G with dilation s, to be the function given by representing the adjacency matrix of G as a step function where each pixel has size $1/s \times 1/s$; see Fig. 9[3]. The estimator used by [VR16]

3 Implicitly, this notion requires us to order the vertices of G, since otherwise it is not clear which interval of length $1/s$ should be mapped to a given vertex; we will choose an arbitrary, fixed ordering for each unlabeled, finite graph G, but note that

in the setting where the sizes s_j are included as part of the observation are dilated empirical graphons of G_j with dilation s_j. The basic structure of estimation—map a sequence of graphs to a sequence of graphons and define a notion of convergence on the graphons—looks very similar to the development of (stretched) cut metric convergence, and, as with stretched cut convergence, there is a close connection to sampling convergence: $\hat{W}_{(G_j,s_j)} \to_{GP} \mathcal{W}$ is equivalent to $\mathsf{Smpl}\left(G_j, \frac{r}{s_j}\right) \xrightarrow{d} \mathcal{G}(\Gamma_r)$ for all $r \in \mathbb{R}_+$. To explain this connection, we recall a pair of ideas from [VR16] (themselves adapted from [Kal99]). First, generating a sample from $\mathrm{GPD}\left(\hat{W}_{(G_j,s_j)}, s\right)$ is equivalent to sampling a subgraph from G_j by selecting $\mathrm{Poi}\left(\frac{r}{s_j}v(G_j)\right)$ vertices with replacement, and returning the vertex induced subgraph without its isolated vertices. Second, this with-replacement sampling scheme is asymptotically equivalent to r/s_j-sampling (without replacement). The equivalence follows immediately.

If $s_1, s_2, \ldots$ are not included as part of the observation, then we require a different approach to estimation. For graphexes of the form $\mathcal{W} = (0, 0, W)$, [BCCH16] proves that $e(G_j)/s_j^2 \to \frac{1}{2}\|W\|_1$ a.s. as $j \to \infty$, but it is not hard to extend this result to general integrable graphexes, showing that $e(G_j)/s_j^2 \to \frac{1}{2}\|\mathcal{W}\|_1$ a.s. as $j \to \infty$, where we define the L_1-norm of a graphex $\mathcal{W} = (I, S, W)$ as $\|\mathcal{W}\|_1 = \|W\|_1 + \frac{1}{2}\|S\|_1 + \frac{1}{2}I$. This suggests making a canonical choice of $\|W\|_1 = 1$ and defining the stretched canonical graphon $W^{G,s}$ of a graph G as dilated empirical graphon of G with dilation $1/\sqrt{2e(G)}$. The salient fact, spelled out in Lemma 4.5.4, is that $G_1, G_2, \ldots$ is sampling convergent to $\mathcal{W}$ if and only if $W^{G_j,s} \to_{GP} \mathcal{W}$ as $j \to \infty$. In conjunction with our result that graph sequences generated by $\mathcal{W}$ are sampling convergent to $\mathcal{W}$, this establishes that the stretched canonical graphon is a consistent estimator for $\mathcal{W}$ if $\|\mathcal{W}\|_1 = 1$.

[VR16] follows a different approach. In the case where the sample sizes are not included as part of the observation, the most general observation is the sequence of all distinct (unlabeled) graph structures taken on by $(\mathcal{G}(\Gamma_s))_{s\in\mathbb{R}_+}$; call this collection $\mathscr{G}(\Gamma)$, the graph sequence of Γ. Intuitively, this is the structure that remains when the labels are stripped from $(\Gamma_s)_{s\in\mathbb{R}_+}$. The natural notion of estimation for graph sequences is then to say that $\mathcal{W}_j \to_{GS} \mathcal{W}$ as $j \to \infty$ whenever $\mathscr{G}(\Gamma^j) \xrightarrow{d} \mathscr{G}(\Gamma)$, where Γ^j is generated by $\mathcal{W}_j$; that is, $\mathcal{W}_1, \mathcal{W}_2, \ldots$ estimates $\mathcal{W}$ if the distribution over unlabeled structures generated by $\mathcal{W}_j$ is asymptotically equal to the distribution over unlabeled structures generated by $\mathcal{W}$. It turns out that the empirical graphon (without any dilation) is a consistent estimator for $\mathcal{W}$ in the graph sequence sense; so indeed estimation is possible without any knowledge of $s_1, s_2, \ldots$.

all our subsequent notions don't depend on the particular ordering, and hence are well defined for unlabeled graphs, as well as graphs with vertices labeled by labels in an unordered set.

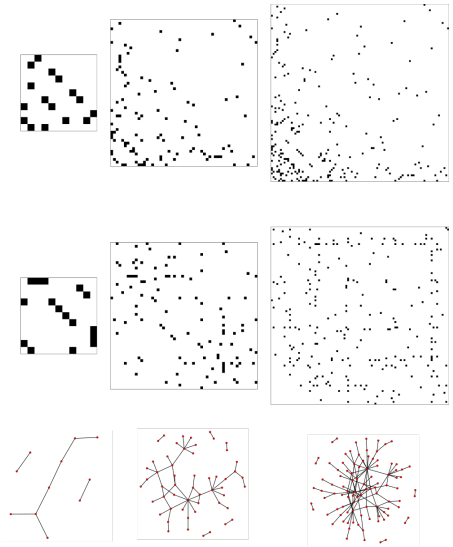

Figure 9: Each column shows a graph (bottom row), a corresponding stretched empirical graphon based on a random labeling of the vertices (middle row), and a corresponding stretched empirical based on an alternative labeling (top row). The three graphs are a prefix of a sequence that converges to $(0, 0, W)$ where $W(x, y) = (x+1)^{-2}(y+1)^{-2}$. Intuitively, the top row are pixel picture approximations to the limiting graphon. The cut metric formalizes this intuition: the graphons are aligned according to some optimal measure preserving transformation, and the distance between them is then $\sup_{U,V\subseteq\mathbb{R}_+^2} \left|\int_{U\times V} W_1(x, y) - W_2(x, y)\,dx\,dy\right|$, the largest difference in any patch between the total amounts of ink in that patch.

Because the empirical graphon relies only on the graph (and not the latent observation size), it can be used to define a notion of graph limit. Let $G_1, G_2, \ldots$ be a sequence of graphs (not necessarily corresponding to a KEG), and say that the sequence is GS convergent to W, written $G_j \to_{GS} W$ as $j \to \infty$, whenever $W^{G_j} \to_{GS} W$ as $j \to \infty$. [VR16, Lemma 5.6] shows that $W_j \to_{GP} W \neq 0$ as $j \to \infty$ implies also $W_j \to_{GS} W$ as $j \to \infty$, from which it follows that sampling convergence implies GS convergence. The converse is not true: the consistent estimation results of [VR16] establish that graph sequences generated by non-integrable W are GS convergent to W, but sampling convergent limits are always integrable. Thus GS convergence provides an even more general notion of graph limit. However, it is unclear whether GS convergence has any interpretation or motivation outside the graphex theory.

Organization

We give formal definitions and recall some important results in Section 4.2. The basic results for sampling convergence—most importantly, the limits are graphexes—are given in Section 4.3. In Section 4.4 we prove that a graph sequence generated by integrable graphex W is almost surely sampling convergent to a canonical dilation of W; this has the particular consequence that (a canonical representative of) every integrable graphex arises as the sampling limit of some graph sequence. In Section 4.5 we relate convergence in distribution of graphex sequences generated by $W_1, W_2, \ldots$ to the metric convergence of [BCCH16]. In Section 4.6 we metrize sampling convergence and show that the metric completion of the space of finite unlabeled graphs is compact. In Section 4.7 we prove that if a graph-valued stochastic process $(G_s)_{s\in\mathbb{R}_+}$ has the property that, for all $p \in (0,1)$ and all $s \in \mathbb{R}_+$, a p-sampling of G_s is equal in distribution to G_{ps}, then there is some graphex W such that $G_s = \mathcal{G}(\Gamma_s)$ for some $(\Gamma_s)_{s\in\mathbb{R}_+}$ generated by W.

PRELIMINARIES

Graph Limits

We now recall some important definitions and results on the metric convergence of [BCCH16], specializing to the case of graphons defined over $\mathbb{R}_+^2$ and sequences of simple graphs.

There are two main notions of distance between integrable graphons that we will need. The first is a modification of the L^1 distance that accounts for the fact that graphons have a natural equivalence under measure preserving transformations. For $\psi\colon \mathbb{R}_+ \to \mathbb{R}_+$, we let $W^\psi(x,y) = W(\psi(x),\psi(y))$.

Definition 4.2.1. The *invariant* L^1 *distance* between integrable graphons W_1, W_2 is $\delta_1(W_1, W_2) = \inf_{\psi_1,\psi_2} \|W_1^{\psi_1} - W_2^{\psi_1}\|_1$, where the infimum is over all measure preserving transformations $\psi_j \colon \mathbb{R}_+ \to \mathbb{R}_+$ for $j = 1, 2$.

Intuitively, the invariant L^1 distance lines up the two graphons as closely as possible and then takes the L^1 distance between them.

The invariant L^1 distance is too stringent of a notion for many cases of interest. In particular, it is obviously impossible to approximate a general graphon by a $\{0,1\}$-valued graphon under that notion of distance. The weakened distance we use is:

Definition 4.2.2. The *cut distance* between two integrable graphons W_1, W_2 is

$$\delta_\square(W_1, W_2) = \inf_{\psi_1,\psi_2} \sup_{U,V\subseteq\mathbb{R}_+^2} \left| \int_{U\times V} W_1^{\psi_1}(x,y) - W_2^{\psi_2}(x,y)\,dx\,dy \right|, \tag{4.2.1}$$

where the infimum is over all measure preserving transformations $\psi_j \colon \mathbb{R}_+ \to \mathbb{R}_+$ for $j = 1, 2$ and the supremum is over Borel sets $U, V \subseteq \mathbb{R}_+^2$.

Intuitively, the cut distance lines up two graphons as closely as possible, then "smears them out" so that they are close in the cut sense if their mass on every rectangular region is close. This allows a $\{0,1\}$-valued graphon to approximate an arbitrary graphon as a pixel-picture approximation to a grayscale image; see Fig. 9.

The cut metric defines a form of convergence for sequences of integrable graphons. To lift this to convergence of sequences of graphs we need a canonical way to map graphs to graphons.

Definition 4.2.3. The *empirical graphon* $W^G \colon [0,1]^2 \to \{0,1\}$ of a graph G is the function produced by partitioning $[0,1]^2$ into a $v(G) \times v(G)$ grid and setting square (i,j) to take value 1 if edge (i,j) is included in G, and 0 otherwise.

The empirical graphon is the "right" mapping in the dense graph setting, but it needs a modification in the sparse graph setting:

Definition 4.2.4. The *stretched canonical graphon* $W^{G,s} \colon \mathbb{R}_+^2 \to \{0,1\}$ of a graph G is defined to be

$$W^{G,s}(x,y) = W^G\left(\frac{x}{\sqrt{2e(G)}}, \frac{y}{\sqrt{2e(G)}}\right) \tag{4.2.2}$$

if $x, y \in \left[0, 1/\sqrt{2e(G_j)}\right)$ and $W^{G,s}(x,y) = 0$ otherwise.

See Fig. 9. The basic intuition for this definition is that $\|W^{G,s}\|_1 = 1$, so that if $H_r \sim \mathrm{GPD}(W^{G,s}, r)$ then $\mathbb{E}[H_r] = r^2/2$. That is, the

canonical stretched graphon is stretched such that the corresponding graphon process has a fixed "growth rate" irrespective of the graph used as input.

We now have an obvious notion for convergence of graph sequences.

Definition 4.2.5. A graph sequence $G_1, G_2, \ldots$ converges in stretched cut distance to W if $\delta_\square(W^{G_j,s}, W) \to 0$ as $j \to \infty$.

A key property of stretched cut convergence is, by [BCCH16, Theorem 2.23], if $(G_s)_{s\in\mathbb{R}_+}$ is a graphon process generated by W such that $\|W\|_1 = 1$ then, almost surely, $\delta_\square(W^{G_s,s}, W) \to 0$ as $s \to \infty$. In this chapter we will establish the analogous result for sampling convergence.

The space of graphons equipped with the cut metric is not relatively compact, so a further restriction is needed for subsequential convergence.

Definition 4.2.6. A set of graphons $\{W_j\}_{j\in\mathbb{N}}$ has *uniformly regular tails* if for every $\epsilon > 0$ there is some $M > 0$ such that for each j there is some $U_j \subseteq \mathbb{R}_+$ with $|U_j| < M$ and $\|W_j - W_j 1_{U_j\times U_j}\|_1 < \epsilon$ for all j. A set of graphs $\{G_j\}_{j\in\mathbb{N}}$ is said to have uniformly regular tails if $\{W^{G_j,s}\}$ has uniformly regular tails.

The main results about sequences with uniformly regular tails is that any such sequence has a further subsequence that converges in cut distance, and that any sequence that is convergent in cut distance also has uniformly regular tails. Intuitively speaking, the uniformly regular tail condition requires the graphs to have "dense cores" where a constant fraction of all edges of G_j occur between only $\Theta(\sqrt{e(G_j)})$ vertices.

Sampling

Sampling convergence requires subgraphs sampled from $G_1, G_2, \ldots$ to converge in distribution to finite size random graphs given by dropping the labels from finite size KEGs. It is most convenient to express this by introducing notation for the distributions of these graphs.

Definition 4.2.7. The *canonical sampling distribution* with parameters s and G is $\mathrm{SmplD}(G, s)(\cdot) = \Pr\Big(\mathsf{Smpl}\Big(G, s/\sqrt{2e(G)}\Big) \in \cdot \mid G\Big)$.

Definition 4.2.8. Let $(\Gamma_s)_{s\in\mathbb{R}_+}$ be a Kallenberg exchangeable graph generated by $\mathcal{W}$, with $\mathcal{W}$ possibly random. The *unlabeled KEG distribution* with parameters $\mathcal{W}$ and s is $\mathrm{GPD}(\mathcal{W}, s)(\cdot) = \Pr(\mathcal{G}(\Gamma_s) \in \cdot \mid \mathcal{W})$.

Instead of $\mathsf{Smpl}\big(G_j, s/\sqrt{2e(G_j)}\big) \xrightarrow{d} \mathcal{G}(\Gamma_s)$ as $j \to \infty$, we may now equivalently write $\mathrm{SmplD}\big(G_j, s\big) \to \mathrm{GPD}(\mathcal{W}, s)$ weakly as $j \to \infty$. This has the advantages that it makes the limit object $\mathcal{W}$ explicit, it doesn't

introduce extraneous randomness (non-random graphs are mapped to non-random probability measures), and it allows us to deal easily with cases where the graph sequence or $\mathcal{W}$ is random.

We will make use of another sampling scheme that is asymptotically equivalent to p-sampling with $p = r/\sqrt{2e(G_j)}$. The alternative sampling scheme will again be defined for labeled or unlabeled input graphs, and as in the case of p-sampling, outputs an unlabeled graph, whether the input graph is labeled or not.

Definition 4.2.9. A *with-replacement* p*-sampling* of a graph G, $\mathsf{SmplWR}(G, p)$, is an unlabeled graph obtained by sampling $\mathrm{Poi}(pv(G))$ vertices from G with replacement and returning the vertex-induced "subgraph" without its isolated vertices. Explicitly, if $x_1, \dots, x_k$ are the vertices of G chosen by sampling with replacements, we first form a graph on $[k]$ by joining $i, j \in [k]$ by an edge whenever $x_i x_j$ is an edge in G (whether that edge was a loop or an edge between two different vertices), then deleting isolated vertices, and then returning the resulting graph without its labels.

The motivation for this definition is the observation that generating a size-r graph according to the canonical stretched empirical graphon of G is equivalent to a with replacement $r/\sqrt{2e(G)}$-sampling of G, in the sense that

$$\mathrm{GPD}\big(W^{G,s}, r\big) = \Pr\left(\mathsf{SmplWR}\left(G, \frac{r}{\sqrt{2e(G)}}\right) \in \cdot \mid G\right). \tag{4.2.3}$$

This observation (essentially) originates in [VR16], in the context of the study of the empirical graphons of $G_1, G_2, \dots$ generated by $\mathcal{W}$ at sizes $s_1, s_2, \dots$, and stretched out by a factor of s_j at each stage (instead of $\sqrt{2e(G_j)}$). In our setting there is a small additional complication arising from possible loops in G.

Definition 4.2.10. Let $l(G)$ denote the number of loops of graph G.

Asymptotic equivalence of with and without replacement sampling translates to:

Lemma 4.2.11. *Let* $G_1, G_2, \dots$ *be some sequence of graphs such that* $e(G_j) \to \infty$ *as* $j \to \infty$ *and* $l(G_j) = O(\sqrt{e(G_j)})$. *Then* $\mathsf{Smpl}\big(G_j, r/\sqrt{2e(G_j)}\big) \xrightarrow{d} H$ *for some finite random graph* H *if and only if* $\mathsf{SmplWR}\big(G_j, r/\sqrt{2e(G_j)}\big) \xrightarrow{d} H$.

Proof. Immediate from [VR16, Lemmas 4.7 and 4.9]. □

Coupling

Much of this chapter involves convergence of probability measures. We will often make use of coupling techniques in order to establish

these results, see [Hol12] for an overview. A coupling of probability measures P and P', both on measurable space $(E, \mathcal{E})$, is a probability measure $\hat{P}$ on $(E \times E, \sigma(\mathcal{E} \times \mathcal{E}))$ with marginals P and P'. The sense in which a coupling gives a notion of distance between two probability measures is, for any coupling $\hat{P}$ of P and P', given X, X' with distributions P, P', it holds that $\|P - P'\|_{TV} \leqslant \hat{P}(X \neq X')$; moreover, if E is Polish, there exists some coupling that saturates the bound. Here, $\|P - P'\|_{TV}$ denotes the total variation distance.

It is often convenient to describe a coupling as a scheme for jointly sampling X and X'. In this case we may refer to the coupling as a coupling of the random variables. In this case, the basic proof technique is to describe an algorithm for jointly sampling X and X', and then bounding $\Pr(X \neq X')$ under this algorithm.

Distributional Convergence of Point Process

Our technical development relies on techniques from point process theory, particularly the theory of distributional convergence of point processes viewed as random measures. Good references include [DVJ03a; DVJ03b] for a friendly introduction, and [Kal01, Chapter 16] for a very general treatment.

For our purposes, the main result needed to understand distributional convergence of point processes is, for example, [DVJ03b, Theorem 11.1.VII]:

Theorem 4.2.12. *Let $\xi, \xi_1, \xi_2, \ldots$ be locally finite point processes on $\mathbb{R}_+^2$. Then $\xi_j \xrightarrow{d} \xi$ as $j \to \infty$ if and only if $(\xi_j(B_1), \ldots, \xi_j(B_n)) \xrightarrow{d} (\xi(B_1), \ldots, \xi(B_n))$ as $j \to \infty$, where $B_i \subseteq \mathbb{R}_+^2$ are bounded Borel sets such that $\Pr(\xi(\partial B_i) = 0) = 1$.*

That is, convergence in distribution of point processes is just convergence in distribution of the counts on arbitrary collections of test sets. Generally, there are consistency requirements between the counts on different test sets, and in consequence it actually suffices to check convergence on a smaller collection.

SAMPLING LIMITS OF GRAPH SEQUENCES

In this section we show that for graph sequences with size going to infinity the limits of sampling convergence are graphexes and every such sequence is subsequentially sampling convergent.

The main technical idea is to use the canonical labeling to introduce a map from graphs to probability distributions over point processes, and then establish the claimed results by way of tools from the theory of distributional convergence of point processes. Recall that the canonical labeling of a graph G is a random adjacency measure corresponding to independently randomly labeling each vertex of G

uniformly in $[0, \sqrt{2e(G)})$. We introduce notation for the probability distribution of the random labeling:

Definition 4.3.1. The *embedding* of a (possibly random) graph G is a probability distribution over point processes on $[0, \sqrt{2e(G)})^2$ given by $\mathrm{embed}(G)(\cdot) = \Pr(\mathsf{Lbl}(G) \in \cdot \mid G)$.

Our first lemma relates distributional convergence of the point processes given by the canonical random labelings of $G_1, G_2, \dots$ to sampling convergence of the graph sequence. Intuitively, sampling convergence is equivalent to distributional convergence of the point processes, and the limiting random graph of $r/\sqrt{2e(G_j)}$-sampling is isomorphic to the graph given by restricting the limiting adjacency measure to vertices with label less than r. To parse the lemma statement, note that sampling convergence may be written as, for all $r \in \mathbb{R}_+$, $\mathrm{SmplD}(G_j, r)$ converges weakly as $j \to \infty$. It may also be helpful to note that part of our goal in this section is to establish that the limit η_r below is equal to $\mathrm{GPD}(\mathcal{W}, r)$ for some integrable graphex $\mathcal{W}$.

Lemma 4.3.2. *Let $G_1, G_2, \dots$ be a graph sequence with $e(G_j) \to \infty$ as $j \to \infty$. The graph sequence is sampling convergent if and only if $\mathrm{embed}(G_1), \mathrm{embed}(G_2), \dots$ converge weakly; that is, if and only if the random labelings converge in distribution. Further, denoting the limiting distributions as $\mathrm{SmplD}(G_j, r) \to \eta_r$ as $j \to \infty$ and $\mathrm{embed}(G_j) \to \zeta$ as $j \to \infty$, if $H_r \sim \eta_r$ and $\xi \sim \zeta$ then $\mathsf{Lbl}_r(H_r) \overset{d}{=} \xi([0, r)^2 \cap \cdot)$.*

Proof. Suppose first that the sequence is sampling convergent. Fix r and notice that, for $\sqrt{2e(G_j)} > r$, under the canonical labelings of G_j each vertex has a label in $[0, r)$ independently with probability $r/\sqrt{2e(G_j)}$. Moreover, restricted to $[0, r)$, each vertex has a $\mathrm{U}[0, r)$ i.i.d. label. Denote this restriction by $\mathsf{Lbl}(G_j)|_r$. We have just shown that $\mathsf{Lbl}(G_j)|_r \overset{d}{=} \mathsf{Lbl}_r(\mathrm{Smpl}(G_j, r/\sqrt{2e(G_j)}))$. [VR16, Lemma 4.13] shows that if $G', G'_1, G'_2, \dots$ are unlabeled random graphs then $G'_j \xrightarrow{d} G'$ as $j \to \infty$ if and only if $\mathsf{Lbl}_r(G'_j) \xrightarrow{d} \mathsf{Lbl}_r(G')$ as $j \to \infty$. Hence, by the assumption of sampling convergence, $\mathsf{Lbl}(G_j)|_r$ converges in distribution as $j \to \infty$.

Next we lift this convergence on arbitrary prefixes $\mathsf{Lbl}(G_j)|_r$ to convergence of the entire point process. We first identify the limiting point process ξ. Let $B_1, \dots, B_n \subseteq \mathbb{R}_+^2$ be bounded Borel sets, choose r such that $B_1, \dots, B_n \subseteq [0, r)^2$ and demand

$$\{\xi(B_1), \dots, \xi(B_n)\} \overset{d}{=} \lim_{j \to \infty} \{\mathsf{Lbl}(G_j)|_r(B_1), \dots, \mathsf{Lbl}(G_j)|_r(B_n)\}. \tag{4.3.1}$$

To see that the right hand side is well defined (that is, independent of the choice of r) notice that for $r < r'$, $(\mathsf{Lbl}(G_j)|_{r'})([0, r)^2 \cap \cdot) \overset{d}{=} \mathsf{Lbl}(G_j)|_r$. The right hand side converges in distribution because $\mathsf{Lbl}(G_j)|_r$

converges in distribution. Moreover, the consistency conditions necessary for the right hand side to be counts with respect to some point process are satisfied, because the limiting joint distribution are counts with respect to $\lim_{j\to\infty} \mathsf{Lbl}(G_j)|_r$. By the Kolmogorov existence theorem for point processes (see [DVJ03b, Theorem 9.2.X]) this suffices to show that ξ exists and has a well-defined distribution.

It is immediate that $\mathsf{Lbl}(G_j) \xrightarrow{d} \xi$ as $j \to \infty$ because, by construction,

$$\{\mathsf{Lbl}(G_j)(B_1), \ldots, \mathsf{Lbl}(G_j)(B_n)\} \xrightarrow{d} \{\xi(B_1), \ldots, \xi(B_n)\}, \quad j \to \infty, \tag{4.3.2}$$

for all bounded Borel sets $B_1, \ldots, B_n \subseteq \mathbb{R}_+^2$.

The reverse direction follows similarly. □

The next result establishes that graphexes are the natural limit objects of sampling convergent sequences.

Lemma 4.3.3. *Let $G_1, G_2, \ldots$ be a sampling convergent graph sequence with $e(G_j) \to \infty$ as $j \to \infty$. Then the limit is a graphex, in the sense that there is some (possibly random) $\mathcal{W}$ such that if $\mathrm{SmplD}(G_j, r) \to Q_r$ then $Q_r \mid W = \mathrm{GPD}(\mathcal{W}, r)$.*

Proof. Notice that $l(G_j) = O(\sqrt{e(G_j)})$ for any sampling convergent sequence, since otherwise the number of vertices in the random subgraph diverges. By Lemma 4.3.2, the canonical random labelings of G_j are convergent to some point process ξ on $\mathbb{R}_+^2$. Observe that for any r and any measure preserving transformation ϕ on $[0, r)$, $\xi \circ (\phi \otimes \phi) \overset{d}{=} \xi$. In particular then, for any dyadic partitioning of $\mathbb{R}_+$ and any transposition τ of this dyadic partitioning, $\xi \circ (\tau \otimes \tau) \overset{d}{=} \xi$, and by [Kal05, Proposition 9.1] this implies that ξ is exchangeable. Then by the Kallenberg representation theorem there is some (possibly random) graphex $\mathcal{W}$ that generates ξ. That is, $\mathrm{embed}(G_j)$ converges weakly to the distribution over point processes defined by (marginalizing over) $\mathcal{W}$. Lemma 4.3.2 then establishes the result. □

We now turn to establishing that the limiting graphex $\mathcal{W}$ in Lemma 4.3.3 is non-random and integrable.

The next lemma gives a tractable criterion for determining when an exchangeable point process is ergodic; i.e. when $\mathcal{W}$ is non-random. Basically, an adjacency measure is ergodic if for all $r, r' \in \mathbb{R}_+$, with $r < r'$, the induced subgraph with vertex labels less than r gives no information about the induced subgraph with vertex labels between r and r'. This is an analogue of [Kal05, Lemma 7.35], attributed there to David Aldous.

It will be convenient to have an additional definition from [VR16] for our argument:

Definition 4.3.4. Let Γ be generated by $\mathcal{W}$ and let $\mathrm{KEG}(\mathcal{W}) = \Pr(\Gamma \in \cdot \mid \mathcal{W})$, the (possibly random) probability measure over adjacency measures induced by (possibly random) $\mathcal{W}$.

Lemma 4.3.5. *Let Γ be an exchangeable adjacency measure on $\mathbb{R}_+^2$. Γ is extremal if and only if for all $r < r' \in \mathbb{R}_+$, $\Gamma([0,r)^2 \cap \cdot)$ and $\Gamma([r,r')^2 \cap \cdot)$ are independent.*

Proof. If the point process is extremal, the Kallenberg representation theorem (Theorem 2.4.7) immediately implies the result.

Conversely, suppose that Γ is not extremal. By a consistent estimation result [VR16, Theorem 4.8], for any sequence $s_1, s_2, \ldots$ such that $s_j \uparrow \infty$,

$$\lim_{j\to\infty} \Pr(\Gamma([0,r)^2 \cap \cdot) \in \cdot \mid \Gamma([r,s_j)^2 \cap \cdot)) = \Pr(\Gamma([0,r)^2 \cap \cdot) \in \cdot \mid \mathrm{KEG}(\mathcal{W})). \tag{4.3.3}$$

(That is, $\mathcal{W}$ can be estimated from an infinite size sample). Since, by non-extremity, $\Pr(\Gamma([0,r)^2 \cap \cdot) \in \cdot \mid \mathrm{KEG}(\mathcal{W})) \neq \Pr(\Gamma([0,r)^2 \cap \cdot) \in \cdot)$, this means that there is some $r' \in \mathbb{R}_+$ such that

$$\Pr(\Gamma([0,r)^2 \cap \cdot) \in \cdot \mid \Gamma([r,r')^2 \cap \cdot)) \neq \Pr(\Gamma([0,r)^2 \cap \cdot) \in \cdot), \tag{4.3.4}$$

as required. □

Lemma 4.3.6. *The limiting graphex $\mathcal{W}$ in Lemma 4.3.3 is non-random.*

Proof. Notice that $l(G_j) = O(\sqrt{e(G_j)})$ for any sampling convergent sequence, since otherwise the number of vertices in the random subgraph diverges. We then make use of Lemma 4.2.11, the asymptotic equivalence of $\mathsf{Smpl}\big(G_j, r/\sqrt{2e(G_j)}\big)$ and $\mathsf{SmplWR}\big(G_j, r/\sqrt{2e(G_j)}\big)$. Let $r' \in \mathbb{R}_+$ and produce a sequence of adjacency measures $\xi_{j,r'}$ by, for each $j \in \mathbb{N}$, sampling a subgraph from G_j according to the with replacement scheme (with probability $r'/\sqrt{2e_j}$) and then randomly labeling this subgraph in $[0,r')$. By the asymptotic equivalence of the sampling schemes and Lemma 4.3.2, $\xi_{j,r'} \xrightarrow{d} \xi([0,r')^2 \cap \cdot)$, where ξ is an adjacency measure generated by $\mathcal{W}$.

As a consequence of the with replacement sampling scheme, for all $j \in \mathbb{N}$, $\xi_{j,r'}([0,r)^2 \cap \cdot)$ is independent of $\xi_{j,r'}([r,r')^2 \cap \cdot)$, for any $r < r'$. To see this, note first that each sampled vertex has a label in $[0,r)$ independently with probability r/r', so that, by a property of the Poisson distribution, the number of vertices in $[0,r)$ and in $[r,r')$ have independent $\mathrm{Poi}\big(rv(G_j)/\sqrt{2e(G_j)}\big)$ and $\mathrm{Poi}\big((r'-r)v(G_j)/\sqrt{2e(G_j)}\big)$ distributions. Second, because the vertex sampling is with replacement, the structure of the graph with labels in $[0,r)$ contains no information about the structure of the graph with labels in $[r,r')$.

The independence of $\xi_{j,r'}([0,r)^2 \cap \cdot)$ and $\xi_{j,r'}([r,r')^2 \cap \cdot)$ for all $j \in \mathbb{N}$ implies that $\xi([0,r)^2 \cap \cdot)$ is independent of $\xi([r,r')^2 \cap \cdot)$. Because r, r' were arbitrary, Lemma 4.3.5 implies that ξ is ergodic, or, equivalently, that $\mathcal{W}$ is non-random. □

Next we show that the limiting $\mathcal{W}$ is integrable, we bound the integral, and we give a condition for when the bound is saturated. We will need the following lemma:

Lemma 4.3.7. *Let* $(\Gamma_s)_{s\in\mathbb{R}_+}$ *be generated by* $\mathcal{W} = (I, S, W)$. *Then,* $\mathbb{E}[e(\Gamma_s)] = s^2\big(I + \|S\|_1 + \frac{1}{2}\|W\|_1\big)$, *and* $\mathbb{E}[l(\Gamma_s)] = s\int W(x,x)\,dx$.

Proof. Let e_s^I, e_s^S, and e_s^W be the number of non-loop edges generated by the I, S, and W components of the graphex, noting that the edge sets generated by the different components are disjoint.

$\mathbb{E}[e_s^I] = Is^2$ is immediate from Campbell's theorem.

By [VR15, Theorem 5.3], $\mathbb{E}[e_s^W] = s^2\frac{1}{2}\|W\|_1$, and $\mathbb{E}[l(\Gamma_s)] = s\int W(x,x)\,dx$.

To treat the star component, let Π_s be the latent Poisson process, restricted to $[0,s)\times\mathbb{R}_+$, used to generate Γ and for each $(\theta_i,\vartheta_i)\in\Pi_s$ let $M(\vartheta_i)$ be the number of rays that (θ_i,ϑ_i) has due to the star component of the graphex. By viewing $M(\vartheta_i)$ as a marking of Π_s, and recalling that $M(\vartheta_i) \sim \mathrm{Poi}(sS(\vartheta_i))$, we have from Campbell's theorem that $\mathbb{E}[e_s^S] = s^2\|S\|_1$. □

By construction, $\mathbb{E}\big[e\big(\mathsf{Smpl}\big(G_j, r/\sqrt{2e(G_j)}\big)\big)\big] = r^2/2$ for any simple graph G_j. However, it is not necessarily true that the expected number of edges of the limiting graph is $r^2/2$. For example, consider the case where G_j is a star with j rays. In this case, the sampled subgraph is non-empty only if the center of the star is selected by the vertex sampling. The probability that this happens goes to 0 as $j\to\infty$, so the limiting graph is the empty graph. The following property characterizes when the limiting graphex $\mathcal{W}$ satisfies $\mathbb{E}[H_r] = r^2/2$ for $H_r \sim \mathrm{GPD}(\mathcal{W}, r)$:

Definition 4.3.8. A sequence of graphs $G_1, G_2, \ldots$ is *uniformly sampling regular* if for all $\epsilon > 0$ there is some $k > 0$ such that, uniformly for all j,

$$\frac{1}{e(G_j)}\sum_{i=1}^{v(G_j)} d_{j,i}\,1\left[d_{j,i} > k\sqrt{e(G_j)}\right] < \epsilon, \tag{4.3.5}$$

where $d_{j,i}$ is the degree of vertex i in G_j ignoring loops.

Intuitively, this property is the requirement that, asymptotically, only a vanishing fraction of the edges of the graph are due to vertices with exceptionally high degree. This is a weakening of the condition of uniform tail regularity: it's easy to see that a sequence that is not uniformly sampling regular is also not uniformly tail regular, but, for example, graph sequences that consist of only isolated edges are uniformly sampling regular but not uniformly tail regular.

Lemma 4.3.9. *The limiting graphex* $\mathcal{W} = (I, S, W)$ *in Lemma 4.3.3 is integrable, with* $\frac{1}{2}\|W\|_1 + \|S\|_1 + I \leqslant 1/2$ *and* $\int W(x,x)\,dx = \lim_{j\to\infty} l(G_j)/\sqrt{2e(G_j)}$. *Further, the bound is saturated if and only if the graph sequence is uniformly sampling regular.*

Proof. We first treat the diagonal. Let $l(P)$ be the number of loops in the limiting graph. Observe that $l(\mathsf{Smpl}(G_j, r/\sqrt{2e_j})) \sim \mathrm{Bin}(l(G_j), r/\sqrt{2e_j})$, and that loops in the sampled subgraph can only occur by selecting loops in the original graph. It then follows that $\mathbb{E}[l(P)] = r \lim_{j\to\infty} l(G_j)/\sqrt{2e(G_j)}$ for all $r \in \mathbb{R}_+$. Comparing this expression with Lemma 4.3.7 establishes the claim.

Let P be the limiting point process as in the proof of Lemma 4.3.3. By Fatou's lemma, $\mathbb{E}[e(P(\cdot \cap [0,r)^2))] \leqslant \lim_{j\to\infty} \mathbb{E}[e(\mathsf{Lbl}(G_j)(\cdot \cap [0,r)^2))] = r^2/2$. By Lemma 4.3.7, $\mathbb{E}[e(P(\cdot \cap [0,r)^2))] = \left(\frac{1}{2}\|W\|_1 + \|S\|_1 + I\right) r^2$. This proves the first claim.

For notational brevity, let $e_j = e(G_j)$ and let $e_j^r = e(\mathsf{Lbl}(G_j)(\cdot \cap [0,r)^2))$. A sequence of random variables that converges in distribution also converges in L_1 if and only if it is uniformly integrable. Thus, to prove the second claim, we need to establish that uniform integrability of $\{e_j^r\}_{j\in\mathbb{N}}$ is equivalent to uniform sampling regularity.

Observe that $e_j^r \stackrel{d}{=} e(\mathsf{Smpl}(G_j, r/\sqrt{2e_j}))$. Suppose that the graph sequence is not uniformly sampling regular. Let $D_{j,i}^r$ be the degree in the sampled subgraph of vertex i in G_j, where $D_{j,i}^r = 0$ if vertex i is not selected. Let $d_{j,i}$ denote the degree of vertex i in G_j. Then, the distribution of $D_{j,i}^r$ is:

$$\begin{aligned} B_{j,i}^r &\sim \mathrm{Bin}\left(d_{j,i}, \frac{r}{\sqrt{2e_j}}\right) \\ D_{j,i}^r \mid B_{j,i}^r &\sim \left(1 - \frac{r}{\sqrt{2e_j}}\right)\delta_0 + \frac{r}{\sqrt{2e_j}}\delta_{B_{j,i}^r}. \end{aligned}$$

Note that $\mathbb{E}[D_{j,i}^r] = \frac{r^2}{2e_j} d_{j,i}$, and that if $d_{j,i} > k\sqrt{e_j}$ then there is some constant $c > 0$ such that $\liminf_{j\to\infty} \Pr(D_{j,i}^r > kr/2 \mid D_{j,i}^r > 0) > c$. Accordingly,

$$\begin{aligned} \limsup_{j\to\infty} \mathbb{E}[e_j^r 1[e_j^r > kr/2]] &\geqslant \limsup_{j\to\infty} \frac{1}{2} \sum_i \mathbb{E}[D_{j,i}^r 1[D_{j,i}^r > kr/2]] \\ &\geqslant \limsup_{j\to\infty} \frac{1}{2} \sum_{i\,:\, d_{j,i} > k\sqrt{e_j}} \mathbb{E}[D_{j,i}^r] \Pr(D_{j,i}^r > kr/2 \mid D_{j,i}^r > 0) \\ &\geqslant \limsup_{j\to\infty} \frac{c}{2} \frac{r^2}{2e_j} \sum_{i=1}^{v_j} d_{j,i} 1\left[d_{j,i} > k\sqrt{e(G_j)}\right]. \end{aligned}$$

Because k is arbitrary, the failure of uniform sampling regularity thus also implies the failure of uniform integrability. Intuitively speaking, this argument is that if there are too many edges belonging to ultra-high degree vertices then whenever an ultra-high degree vertex is selected, the sampled subgraph will have "too many" edges.

Conversely, suppose that the graph sequence is uniformly sampling regular. For $k > 1$, we partition the vertices of G_j into three sets:

$$\begin{aligned} H_j &= \{i \in v_j \,:\, d_{j,i} > k\sqrt{e_j}\} \\ M_j &= \{i \in v_j \,:\, d_{j,i} \in [\sqrt{e_j}/k, k\sqrt{e_j}]\} \\ L_j &= \{i \in v_j \,:\, d_{j,i} < \sqrt{e_j}/k\}, \end{aligned}$$

and define $e_j^{H,r} = \sum_{i\in H_j} D_{j,i}^r$, $e_j^{L,r} = \sum_{i\in L_j} D_{j,i}^r$, and $e_j^{M,r} = |\{(i,j) \in M_j \times M_j \,:\, (i,j) \in \mathsf{Smpl}\big(G_j, r/\sqrt{2e_j}\big)\}|$, the number of edges in the subgraph of the sampled graph induced by restricting to vertices that belong to M_j in G_j. Note that $e_j^r \leqslant e_j^{H,r} + e_j^{M,r} + e_j^{L,r}$.

There are at most $2k\sqrt{e_j}$ vertices in M_j, since otherwise there would be too many edges. Let $v_j^{M,r} \sim \mathrm{Bin}(2k\sqrt{e_j}, r/\sqrt{2e_j})$ be an upper bound for the number of candidate vertices selected from M_j by the vertex sampling, and note that $e_j^{M,r} < (v_j^{M,r})^2$. But $(v_j^{M,r})^2$ is uniformly integrable, so, by dominated convergence, $e_j^{M,r}$ is uniformly integrable for any fixed k.

Next, let $e_j^{L',r} = \sum_{i\in L_j} D_{j,i}^r 1[D_{j,i}^r \geqslant 2]$, and let e_r^I be the number of isolated edges the sampled subgraph. Note that $e_j^{L,r} \leqslant e_j^{L',r} + e_r^I$, because every edge is either isolated or connects to a vertex of degree at least 2. Direct computation gives that $\mathbb{E}[D_{j,i}^r 1[D_{j,i}^r \geqslant 2]] = \mathbb{E}[D_{j,i}^r] - \Pr(D_{j,i}^r = 1) \leqslant \big(r/\sqrt{2e_j}\big)^3 d_{j,i}^2$, whereby

$$\begin{aligned} \mathbb{E}[e_j^{L',r}] &\leqslant \Big(\frac{r}{\sqrt{2}}\Big)^3 \sum_{i\in L_j} \frac{d_{j,i}}{\sqrt{e_j}} \frac{d_{j,i}}{e_j} \\ &\leqslant \Big(\frac{r}{\sqrt{2}}\Big)^3 \frac{1}{k} \sum_{i\in L_j} \frac{d_{j,i}}{e_j} \\ &\leqslant \frac{r^3}{\sqrt{2}k}, \end{aligned}$$

where the second line has used that $d_{j,i}/\sqrt{e_j} < 1/k$ for every vertex i in L_j.

We further note that each edge in G_j is an isolated edge in the sampled subgraph with probability at most $r^2/(2e_j)$, so e_r^I is bounded by a $\mathrm{Bin}(e_j, r^2/(2e_j))$ random variable, and is thus uniformly integrable.

The last intermediate step is to note that

$$\mathbb{E}[e_j^{H,r}] = \frac{r^2}{2e_j} \sum_{i=1}^{v_j} d_{j,i} 1\big[d_{j,i} > k\sqrt{e_j}\big], \tag{4.3.6}$$

so, by the assumption of uniform sampling regularity, we may uniformly force $\mathbb{E}[e_j^{H,r}]$ to be arbitrarily small by choosing k sufficiently large.

Now, for any constant $c' > 0$,

$$\begin{aligned}\mathbb{E}[e_j^r 1[e_j^r > c']] \leqslant \mathbb{E}[(e_r^I + e_j^{M,r})1[(e_r^I + e_j^{M,r}) > c'/2]] \\ + \mathbb{E}[(e_r^I + e_j^{M,r})1[e_j^{L',r} + e_j^{H,r} > c'/2]] + \mathbb{E}[e_j^{L',r}] + \mathbb{E}[e_j^{H,r}].\end{aligned}$$

For $\epsilon > 0$, we may guarantee that the last two terms are each at most $\epsilon/4$ by choosing k sufficiently large. For any fixed k, Markov's inequality shows that $\lim_{c'\to\infty} \Pr(e_j^{L',r} + e_j^{H,r} > c'/2) = 0$. [Kal01, Lemma 4.10] shows that for any uniformly integrable family $\{X_j\}$ and sequence of events $A_1, A_2, \ldots$ such that $\lim_{k\to\infty} \Pr(A_k) = 0$, we have $\lim_{k\to\infty} \sup_j \mathbb{E}[X_j 1[A_k]] = 0$; accordingly, invoking the uniform integrability of $e_r^I + e_j^{M,r}$, we may choose c' (depending on k) large enough such that the third term is at most $\epsilon/4$. Similarly, by uniform integrability, we may choose c' large enough such that the first term is at most $\epsilon/4$. Thus, for any ϵ there is a $c' > 0$ such that,

$$\mathbb{E}[e_j^r 1[e_j^r > c']] < \epsilon, \tag{4.3.7}$$

uniformly, as required. □

We now have the ingredients of the main result characterizing the limits of sampling convergent sequences:

Theorem 4.3.10. *Let $G_1, G_2, \ldots$ be a sampling convergent graph sequence such that $e(G_j) \to \infty$ as $j \to \infty$. Then the limit is a non-random graphex $\mathcal{W}$ such that $\|\mathcal{W}\|_1 \leqslant 1$, in the sense that $\mathrm{SmplD}(G_j, r) \to \mathrm{GPD}(\mathcal{W}, r)$ weakly as $j \to \infty$ for all $r \in \mathbb{R}_+$.*

Proof. Immediate from Lemmas 4.3.3, 4.3.6 and 4.3.9. □

In some other sparse graph limit theories [BCCZ14a; BCCZ14b; BCCH16], only graph sequences satisfying certain constraints are subsequentially convergent. In Theorem 4.6.6 we show that sampling convergence has no such caveat. On the basis of this result, one might hope that sampling convergent limits are informative about all sparse graph sequences, or at least all uniformly sampling regular sequences. The next result helps clarify that there are further limitations. Intuitively speaking, it shows that the sampling limit is degenerate for graph sequences that are very sparse. In particular, the next result applies to sequences of bounded degree graphs, for which there is already a well developed limit theory [BS01].

Theorem 4.3.11. *Let $G_1, G_2, \ldots$ be a sampling convergent graph sequence with $e(G_j) \to \infty$ as $j \to \infty$, suppose that the maximum degree of any vertex in G_j is $o(\sqrt{e(G_j)})$. Then the sampling limit is the isolated edge graphex $(\frac{1}{2}, 0, 0)$.*

Proof. Let $r \in \mathbb{R}_+$. For brevity, let $v_j = v(G_j)$, $e_j = e(G_j)$ and $p_j = r/\sqrt{2e_j}$. Let $d_{j,i}$ be the degree of vertex i in G_j and let $D_{j,i}^r$ be the

degree of this vertex in a p_j-sampled subgraph, where $D^r_{j,i} = 0$ is understood to mean that the vertex is not included in the subgraph. Then,

$$\begin{aligned} B_{j,i} &\sim \mathrm{Bin}(d_{j,i}, p_j) \\ D^r_{j,i} \mid B_{j,i} &\sim (1-p_j)\delta_0 + p_j \delta_{B_{j,i}}. \end{aligned}$$

In particular,

$$\begin{aligned} \Pr(D^r_{j,i} \geqslant 2) &= p_j(1 - [(1-p_j)^{d_{j,i}} + d_{j,i} p_j (1-p_j)^{d_{j,i}-1}]) \\ &\leqslant p_j^3 d_{j,i}^2, \end{aligned}$$

using Bernoulli's inequality.

Let N_j be the number of vertices with degree greater than 1 in the sampled subgraph. Then,

$$\begin{aligned} \mathbb{E}[N_j] &\leqslant \sum_{i \in v_j} \Pr(D^r_{j,i} \geqslant 2) \\ &\leqslant p_j^3 \sum_{i \in v_j} d_{j,i}^2 \\ &= o\left(p_j^2 \sum_{i \in v_j} d_{j,i} \right) \\ &= o(1). \end{aligned}$$

Markov's inequality then implies that $N_j \xrightarrow{p} 0$ as $j \to \infty$, which implies the claimed result since r was arbitrary. □

KALLENBERG EXCHANGEABLE GRAPHS ARE SAMPLING CONVERGENT

We now turn to characterizing the sampling limits of sequences of graphs generated by a graphex. Let $s_1, s_2, \ldots$ be some sequence such that $s_j \uparrow \infty$ as $j \to \infty$ and let $G_j = \mathcal{G}(\Gamma_{s_j})$, where Γ is generated by an integrable graphex $\mathcal{W}$. Intuitively speaking, our aim is to show that the sampling limit of $G_1, G_2, \ldots$ is $\mathcal{W}$.

The basic strategy makes use of the consistent estimation results first established in [VR16], although we will appeal to the technically stronger versions of [Jan17]. We need the following (implicit) result from those chapters:

Lemma 4.4.1. *Let $G_s = \mathcal{G}(\Gamma_s)$, where $(\Gamma_s)_{s \in \mathbb{R}_+}$ is generated by an integrable graphex $\mathcal{W}$, then $\Pr(\mathsf{Smpl}(G_s, r/s) \in \cdot \mid G_s) \to \mathrm{GPD}(\mathcal{W}, r)$ weakly almost surely as $s \to \infty$, for all $r \in \mathbb{R}_+$.*

Proof. Let $\hat{W}_{(G_s,s)}$ be the empirical graphon of G_s stretched so that each pixel is $1/s \times 1/s$. [Jan17, Theorem 5.1] shows that $\mathrm{GPD}\big(\hat{W}_{(G,s)}, r\big) \to \mathrm{GPD}(\mathcal{W}, r)$ weakly almost surely. As remarked earlier, $\mathrm{GPD}\big(\hat{W}_{(G_s,s)}, r\big) =$

$\Pr(\mathsf{SmplWR}(G_s, r/s) \in \cdot \mid G_s)$, and so the result follows from the asymptotic equivalence of with and without replacement sampling [VR16, Lemmas 4.7 and 4.9]. □

To drop the latent sizes, we will need an extension of a result of [BCCH16] relating $e(G_j)$ and s_j. It will be convenient to partition each Γ_s into three components. Using the notation of the Kallenberg representation theorem (Theorem 2.4.7):

1. Γ_s^W: the edge induced subgraph given by restricting to edges between vertices that belong to the underlying Poisson process Π; intuitively, this is the part of the graph generated by $(0, 0, W)$.
2. Γ_s^S: the edge induced subgraph given by restricting to edges where one vertex belongs to any latent star Poisson process σ_{jk}; intuitively, this is the part of the graph generated by $(0, S, 0)$.
3. Γ_s^I: the induced subgraph given by restricting to the remaining edges; intuitively, this is the part of the graph generated by $(I, 0, 0)$.

Then,

Lemma 4.4.2. *Let $(\Gamma_s)_{s\in\mathbb{R}_+}$ be a KEG generated by graphex $\mathcal{W} = (I, S, W)$, and let e_s^W, e_s^S, and e_s^I be the number of edges of Γ_s^I, Γ_s^S, and Γ_s^W respectively. Then, almost surely,*

$$\begin{aligned}
\lim_{s\to\infty} e_s^W/s^2 &= \frac{1}{2}\|W\|_1 \\
\lim_{s\to\infty} e_s^S/s^2 &= \|S\|_1 \\
\lim_{s\to\infty} e_s^I/s^2 &= I \\
\lim_{s\to\infty} l(\Gamma_s)/s &= \int W(x,x)\,dx
\end{aligned}$$

Proof. First, $e_s^W/s^2 \to \frac{1}{2}\|W\|_1$ a.s. by [BCCH16, Proposition 2.25].

The star component of the KEG can be understood as assigning a $\mathrm{Poi}(sS(\vartheta_i))$ number of rays to each point of the underlying point process $(\theta_i, \vartheta_i) \in \Pi_s$, independent of everything else. By the additive property of independent Poisson distributions we then have $e_s^S \mid \Pi_s \sim \mathrm{Poi}(s\sum_{\vartheta_i\in\Pi_s} S(\vartheta_i))$. Since $s\sum_{\vartheta_i\in\Pi_s} S(\vartheta_i) \uparrow \infty$ a.s. as $s \to \infty$ the law of large numbers implies $e_s^S/(s\sum_{\vartheta_i\in\Pi_s} S(\vartheta_i)) \to 1$ a.s. as $j \to \infty$. The law of large numbers for Poisson processes gives $\sum_{\vartheta_i\in\Pi_s} S(\vartheta_i)/s \to \|S\|_1$ a.s. as $j \to \infty$, whereby $e_s^S/s^2 \to \|S\|_1$ a.s. as $j \to \infty$.

$e_s^I/s^2 \to I$ a.s. as $s \to \infty$ by the law of large numbers for Poisson processes.

Finally, we may view the loops as an independent marking of the latent Poisson process, with a loop on (θ_i, ϑ_i) included with probability $W(\vartheta_i, \vartheta_i)$. $\lim_{s\to\infty} l(\Gamma_s)/s = \int W(x,x)\,dx$ then follows by the law of large numbers for Poisson processes. □

The two previous lemmas show that the limiting distribution of $\mathsf{Smpl}(G_s, r/s)$ is generated by $\mathcal{W}$, and that (temporarily simplifying to the case $\|W\|_1 + 2\|S\|_1 + 2I = 1$), $s \approx \sqrt{2e(G_s)}$ when s is large. Thus, to prove our main result we would like to couple $\mathsf{Smpl}(G_s, r/s)$ and $\mathsf{Smpl}\left(G_s, r/\sqrt{2e(G_s)}\right)$.

Theorem 4.4.3. *Let $(\Gamma_s)_{s\in\mathbb{R}_+}$ be a KEG generated by graphex $\mathcal{W} = (I, S, W)$ such that $\|\mathcal{W}\|_1 = \|W\|_1 + 2\|S\|_1 + 2I > 0$, and let $G_s = \mathcal{G}(\Gamma_s)$ for all $s \in \mathbb{R}_+$. Then $\mathsf{SmplD}\left(G_s, r/\sqrt{2e(G_s)}\right) \to \mathrm{GPD}(\mathcal{W}', r)$ weakly almost surely, where, $\mathcal{W}' = (I', S', W')$ is defined by*

$$\begin{aligned} I' &= I/\|\mathcal{W}\|_1 \\ S'(x) &= (\|\mathcal{W}\|_1)^{-1/2} S(x\|\mathcal{W}\|_1^{1/2}) \\ W'(x,y) &= W(x\|\mathcal{W}\|_1^{1/2}, y\|\mathcal{W}\|_1^{1/2}) \end{aligned}$$

Proof. Let $\mathcal{W}^c = (c^2 I, cS(\cdot/c), W(\cdot/c, \cdot/c))$. By [VR16, Lemma 5.2], $\mathrm{GPD}(\mathcal{W}, s) = \mathrm{GPD}(\mathcal{W}^c, s/c)$ for all $s, c \in \mathbb{R}_+$. As a particular consequence, we may take $\|W\|_1 + 2\|S\|_1 + 2I = 1$ without loss of generality.

First, for any graph G and any $q < p \in [0,1]$ there is a coupling such that

$$\Pr(\mathsf{Smpl}(G,p) \neq \mathsf{Smpl}(G,q)) \leqslant 2\mathbb{E}[e(\mathsf{Smpl}(G,p))]\left(1 - \frac{q}{p}\right). \tag{4.4.1}$$

Namely, we sample $\mathsf{Smpl}(G,p)$ as usual, and sample $\mathsf{Smpl}(G,q)$ as $\mathsf{Smpl}(\mathsf{Smpl}(G,p), q/p)$. The expected number of vertices included in $\mathsf{Smpl}(G,p)$ but not in $\mathsf{Smpl}(G,q)$ is bounded by $\mathbb{E}[v(\mathsf{Smpl}(G,p))(1 - q/p)]$, and the claimed inequality follows by $v(\mathsf{Smpl}(G,p)) \leqslant 2e(\mathsf{Smpl}(G,p))$ and Markov's inequality.

Under this coupling,

$$\begin{aligned} &\Pr\left(\mathsf{Smpl}\left(G_s, \frac{r}{\sqrt{2e(G_s)}}\right) \neq \mathsf{Smpl}\left(G_s, \frac{r}{s}\right)\right) \\ &\qquad \leqslant r^2 \frac{\sqrt{2e(G_s)}}{s} \max\left(\frac{\sqrt{2e(G_s)}}{s} - 1, \frac{s}{\sqrt{2e(G_s)}} - 1\right). \end{aligned}$$

By Lemma 4.4.2, the right hand side goes to 0 almost surely as $s \to \infty$. The theorem statement then follows by Lemma 4.4.1. □

Corollary 4.4.4. *For any integrable graphex $\mathcal{W}$ such that $\|\mathcal{W}\|_1 \leqslant 1$ there is some graph sequence that is sampling convergent to $\mathcal{W}$.*

Proof. Suppose $\|\mathcal{W}\|_1 = 1$, and let $s_1, s_2, \ldots$ be some sequence such that $s_j \uparrow \infty$ as $j \to \infty$ and let $G_j = \mathcal{G}(\Gamma_{s_j})$, where Γ is generated by $\mathcal{W}$; the sequence $G_1, G_2, \ldots$ is almost surely sampling convergent to $\mathcal{W}$ by Theorem 4.4.3.

Next, let $S_1, S_2, \dots$ be a sequence of stars such that $e(S_j) \to \infty$ as $j \to \infty$.

$$\lim_{j\to\infty} \mathrm{Smpl}\left(G_j \cup S_j, \frac{r}{\sqrt{e(G_j \cup S_j)}}\right) \overset{\text{a.s.}}{=} \lim_{j\to\infty} \mathrm{Smpl}\left(G_j, \frac{r}{\sqrt{e(G_j \cup S_j)}}\right), \tag{4.4.2}$$

because the probability of seeing even a single edge sampled from S_j is bounded by the probability of selecting the center of the star as a candidate vertex, which goes to 0. Further,

$$\mathrm{Smpl}\left(G_j, \frac{r}{\sqrt{e(G_j \cup S_j)}}\right) \overset{\text{d}}{=} \mathrm{Smpl}\left(\mathrm{Smpl}\left(G_j, \frac{r}{\sqrt{2e(G_j)}}\right), \frac{\sqrt{e(G_j)}}{\sqrt{e(G_j)+e(S_j)}}\right). \tag{4.4.3}$$

Let $c \in (0,1]$ and choose $S_1, S_2, \dots$ such that $\lim_{j\to\infty} \sqrt{e(G_j)}/\sqrt{e(G_j)+e(S_j)} = c$. By essentially the same coupling argument used in the proof of Theorem 4.4.3,

$$\mathrm{SmplD}\big(G_j \cup S_j, r\big) \to \mathrm{GPD}(\mathcal{W}, cr) \tag{4.4.4}$$

weakly as $j \to \infty$. By [VR16, Lemma 5.2], $\mathrm{GPD}(\mathcal{W}, rc) = \mathrm{GPD}(\mathcal{W}^c, r)$ where $\mathcal{W}^c = (c^2 I, cS(\cdot/c), W(\cdot/c, \cdot/c))$. Thus $G_1 \cup S_1, G_2 \cup S_2, \dots$ is sampling convergent to $\mathcal{W}^c$; because c is arbitrary this shows the claimed result for any non-zero graphex.

Next, consider a sequence $G_1, G_2, \dots$ generated by a graphon W that is 0 except on the diagonal, and take $e(S_j) = l(G_j)$. By repeating the argument of Theorem 4.4.3, using that $\lim_{s\to\infty} l(\Gamma_s)/s = \int W(x,x)\,dx$ by Lemma 4.4.2, we have that the sequence $G_1 \cup S_1, G_2 \cup S_2, \dots$ is sampling convergent to $(0, 0, W)$.

The sampling limit of $S_1, S_2, \dots$ is $(0,0,0)$, completing the proof. □

GRAPHON METRICS AND SAMPLING DISTRIBUTIONS

In this section we relate sampling convergence to the metric convergence of [BCCH16]. Intuitively, the basic idea is to show that if $\delta_1(W_1, W_2)$ or $\delta_\square(W_1, W_2)$ is small then we can construct a coupling of $\mathrm{GPD}(W_1, r)$ and $\mathrm{GPD}(W_2, r)$ such that $\Pr(G_r^1 \neq G_r^2)$ is also small, where $G_r^k \sim \mathrm{GPD}(W_k, r)$ marginally. Note that we require the diagonals to be 0 throughout because the graphon metrics do not control distance between diagonals.

Lemma 4.5.1. *Let W and W' be graphons with 0 diagonals. There is a coupling of $\mathrm{GPD}(W_1, r)$ and $\mathrm{GPD}(W_2, r)$ such that, for $H_r^{(1)}$ and $H_r^{(2)}$ with distributions $\mathrm{GPD}(W_1, r)$ and $\mathrm{GPD}(W_2, r)$ respectively,*

$$\Pr(H_r^{(1)} \neq H_r^{(2)}) \leqslant \frac{1}{2} r^2 \delta_1(W_1, W_2), \tag{4.5.1}$$

under the coupling.

Proof. We couple $H_r^{(1)}$ and $H_r^{(2)}$ according to the following generative scheme:

1. Draw $\Pi \sim \mathrm{PP}([0,r) \times \mathbb{R}_+, \lambda \otimes \lambda)$
2. Draw U-array $\{U_{ij}\}$
3. Include edge (θ_i, θ_j) in graph $H_r^{(k)}$ if and only if $W_k(\vartheta_i, \vartheta_j) > U_{ij}$
4. Drop the labels of the graphs

That is, we generate both graphon processes using the same latent Poisson process and U-array. Marginally, this is just the standard graphon process scheme and so the coupling is obviously valid.

Under this coupling, for each pair of points (θ_i, ϑ_i) and (θ_j, ϑ_j) in Π the probability, conditional on Π, that (θ_i, θ_j) is an edge in one graph and not an edge in the other is $|W_1(\vartheta_i, \vartheta_j) - W_2(\vartheta_i, \vartheta_j)|$. The expected number of edges that disagree between the two graphs is then

$$\frac{1}{2}\mathbb{E}[\sum_{\vartheta_i, \vartheta_j \in \Pi} |W_1(\vartheta_i, \vartheta_j) - W_2(\vartheta_i, \vartheta_j)|] = \frac{r^2}{2}\|W_1 - W_2\|_1, \quad (4.5.2)$$

where the expectation is computed by an application of Slivnyak-Mecke.

The graphs are equal if there are no edges that disagree, so Markov's inequality then gives

$$\Pr(H_r^{(1)} \neq H_r^{(2)}) \leqslant \frac{r^2}{2}\|W_1 - W_2\|_1. \quad (4.5.3)$$

For any measure preserving transformations ϕ of $\mathbb{R}_+$, $\mathrm{GPD}(W \circ (\phi \otimes \phi), r) = \mathrm{GPD}(W, r)$. It then follows that

$$\Pr(H_r^{(1)} \neq H_r^{(2)}) \leqslant \min_{\phi_1, \phi_2} \frac{r^2}{2}\|W_1 \circ (\phi_1 \otimes \phi_1) - W_2 \circ (\phi_2 \otimes \phi_2)\|_1, \quad (4.5.4)$$

where the minimization is over all pairs of measure preserving transformations. □

To show that convergence in stretched cut distance implies convergence of the laws of the graphs generated by the graphons we'll need a translation of the corresponding result ([BCLS+08b, Theorem 3.7a]) from the theory of dense graph convergence:

Lemma 4.5.2. *Let $W_1, W_2, \ldots$ be a sequence of graphons with 0 diagonals. Suppose that there is some compact set C such that $\mathrm{supp}(W_j) \subseteq C$ for all $j \in \mathbb{N}$. If $\lim_{j\to\infty} \delta_\square(W_j, W) = 0$ for some graphon W, then there is a sequence of couplings of $\mathrm{GPD}(W_j, r)$ and $\mathrm{GPD}(W, r)$ such that, for $H_r^{(j)}$ and H_r distributed according to $\mathrm{GPD}(W_j, r)$ and $\mathrm{GPD}(W, r)$ respectively,*

$$\lim_{j\to\infty} \Pr(H_r^{(j)} \neq H_r) = 0 \text{ a.s.} \quad (4.5.5)$$

Proof. We may assume without loss of generality that $C = [0, c)^2$ for some $c \in \mathbb{R}_+$, as there is always some measure preserving transformation ϕ taking C to $[0, c)$, and $\mathrm{GPD}(W \circ (\phi \otimes \phi), r) = \mathrm{GPD}(W, r)$.

The first ingredient of the coupling is the observation that a sample from $\mathrm{GPD}(W, r)$ may be generated according to the following scheme:

1. Sample $N_r \sim \mathrm{Poi}(cr)$
2. For $i = 1 \dots N_r$ sample features $x_k \overset{iid}{\sim} U[0, c]$
3. Include each edge (k, l) independently with probability $W(x_k, x_l)$
4. Return the edge set

That is, in the compactly supported graphon case, the edges are sampled independently conditional on the number of candidate vertices. This is essentially the same generative model as is used in the dense graph theory, with the distinction that the number of vertices is now random and that vertices that do not connect to any edges are not included in the graph. Our aim is to build a sequence of couplings that exploits this observation along with the equivalence of left convergence and cut convergence in the dense graph setting.

Using a common N_r for sampling from each W_j allows us to use results from the dense graph setting. [BCLS+08b, Theorem 3.7a] shows that if $\delta_\square(W_j, W) \to 0$ as $j \to \infty$ then for any fixed graph F,

$$\lim_{j\to\infty} \left| \Pr(H_r^{(j)} = F \mid N_r) - \Pr(H_r = F \mid N_r) \right| = 0. \tag{4.5.6}$$

It is immediate that the limit is also 0 unconditionally; that is, $\mathrm{GPD}(W_j, r) \to \mathrm{GPD}(W, r)$ weakly as $j \to \infty$. Since the space of graphs is discrete, weak convergence also implies convergence in total variation. This implies the existence of the sequence of couplings in the lemma statement. □

The next result extends this to the case of arbitrary cut convergent graphon sequences. The same result has recently been independently proved as [Jan17, Theorem 3.4].

Lemma 4.5.3. *Let $W_1, W_2, \dots$ be a sequence of graphons with 0 diagonals such that $\delta_\square(W_j, W) \to 0$ a.s. as $j \to \infty$ for some W. Then there is a sequence of couplings such that, given $H_r^{(j)}$ and H_r distributed according to $\mathrm{GPD}(W_j, r)$ and $\mathrm{GPD}(W, r)$ respectively,*

$$\lim_{j\to\infty} \Pr(H_r^{(j)} \neq H_r) = 0 \text{ a.s.} \tag{4.5.7}$$

Proof. If the sequence is compactly supported then the result follows from Lemma 4.5.2, so assume otherwise.

It suffices to show that for all $\epsilon' > 0$ there is a sequence of couplings (indexed by j) such that there is some j' such that for all $j > j'$,

$$\Pr(H_r^{(j)} \neq H_r \mid W_j) \leqslant \epsilon'. \tag{4.5.8}$$

The basic structure of our couplings is to pick out compactly supported "dense cores" of W and W_j such that, with high probability, every edge of $H_r^{(j)}$ and H_r is due to the dense cores, and then couple these cores by Lemma 4.5.2. We control the error introduced by restricting to the dense cores by Lemma 4.5.1.

Because the sequence $W_1, W_2, \ldots$ is convergent in the cut norm it is in particular uniformly tail regular, by [BCCH16, Theorem 2.12]. This means that there is some $M_{r,\epsilon} > 0$ such that for each j there is a set $S_{r,\epsilon}^j$ such that $\Lambda(S_{r,\epsilon}^j) \leqslant M_{r,\epsilon}$ and $\|W_j - W_j 1_{S_{r,\epsilon}^j \times S_{r,\epsilon}^j}\|_1 < \frac{\epsilon}{2r^2}$. We may assume without loss of generality that $\Lambda(S_{r,\epsilon}^j) = M_{r,\epsilon}$.

We may even assume that $S_{r,\epsilon}^j = U_{r,\epsilon}$ for all $j \in \mathbb{N}$, where $U_{r,\epsilon} = [0, M_{r,\epsilon})$. To see that this is true, we first observe that, as a straightforward consequence of the fact that there is a Borel isomorphism mapping any probability space to $[0,1]$ equipped with Lebesgue measure, see e.g. [Ker14, Theorem A.20], for each $j \in \mathbb{N}$, there is a measure preserving transformation $\phi_j \colon \mathbb{R}_+ \to \mathbb{R}_+$ such that $\phi_j(U_{r,\epsilon}) = S_{r,\epsilon}^j$. Now,

$$\|W_j \circ (\phi_j \otimes \phi_j) 1_{\mathbb{R}_+^2 \setminus (U_{r,\epsilon} \times U_{r,\epsilon})}\|_1 = \|W_j 1_{\mathbb{R}_+^2 \setminus S_{r,\epsilon}^j \times S_{r,\epsilon}^j}\|_1. \tag{4.5.9}$$

Because $\mathrm{GPD}\big(W \circ (\phi_j \otimes \phi_j), r\big) = \mathrm{GPD}(W, r)$, this justifies our claim that we may take $S_{r,\epsilon}^j = U_{r,\epsilon}$ independent of j.

We now observe that $H_r^{(j)}$ may be sampled according the following scheme:

1. Sample $G_r^{(j)} \sim \mathrm{GPD}\big(W_j 1_{U_{r,\epsilon} \times U_{r,\epsilon}}, r\big)$
2. Let $(\Pi, G_r(W_j 1_{U_{r,\epsilon} \times U_{r,\epsilon}}))$ be the tuple of the latent point process used to generate a graph, and the graph generated by $W_j 1_{U_{r,\epsilon} \times U_{r,\epsilon}}$ using Π. Draw $\Pi \mid G_r^{(j)} \sim \Pr((\Pi, G_r(W_j 1_{U_{r,\epsilon} \times U_{r,\epsilon}})) \in \cdot \mid G_r(W_j 1_{U_{r,\epsilon} \times U_{r,\epsilon}}) = G_r^{(j)})$
3. Generate a graph $E_r^{(j)}$ according to $W_j 1_{\mathbb{R}_+^2 \setminus (U_{r,\epsilon} \times U_{r,\epsilon})}$ using Π
4. Return the edge set of the graph union of $E_r^{(j)}$ and $G_r^{(j)}$ (taking the common vertex set to be Π, and dropping the labels)

Replacing W_j by W makes this a sampling scheme for H_r; we define G_r and E_r corresponding to W in the obvious way.

Notice that, by construction, the joint distribution of $(\Pi, G_r^{(j)})$ is the same as the distribution given by drawing Π as a unit rate Poisson process and then generating $G_r^{(j)}$ according to $W_j 1_{U_{r,\epsilon} \times U_{r,\epsilon}}$ using Π.

This makes it clear that the sampling scheme reproduces the distribution given by the Kallenberg representation construction, i.e. $H_r^{(j)} \sim \mathrm{GPD}(W_j, r)$. Also note that $E_r^{(j)} \sim \mathrm{GPD}\left(W_j 1_{\mathbb{R}_+^2 \setminus (U_{r,\epsilon} \times U_{r,\epsilon})}, r\right)$, and $E_r^{(j)} \sim \mathrm{GPD}\left(W 1_{\mathbb{R}_+^2 \setminus (U_{r,\epsilon} \times U_{r,\epsilon})}, r\right)$ (marginalizing $G_r^{(j)}$ and Π').

The point of this sampling scheme is that now a coupling of $G_r^{(j)}$ and G_r immediately lifts to a coupling of $H_r^{(j)}$ and H_r such that

$$\begin{aligned}\Pr(H_r^{(j)} \neq H_r) &\leqslant \Pr(G_r^{(j)} \neq G_r \text{ or } e(E_r^{(j)}) > 0 \text{ or } e(E_r) > 0) \\ &\leqslant \Pr(G_r^{(j)} \neq G_r) + \mathbb{E}[E_r^{(j)}] + \mathbb{E}[E_r]. \end{aligned} \tag{4.5.10}$$

By Lemma 4.5.2 there is some sequence of couplings such that, for any $\epsilon' > 0$, there is some $\delta > 0$ such that $\delta_\square(W_j 1_{U_{r,\epsilon} \times U_{r,\epsilon}}, W 1_{U_{r,\epsilon} \times U_{r,\epsilon}}) < \delta$ implies $\Pr(G_r^{(j)} \neq G_r) \leqslant \epsilon'/2$. By the triangle inequality,

$$\begin{aligned}\delta_\square(W_j 1_{U_{r,\epsilon} \times U_{r,\epsilon}}, W 1_{U_{r,\epsilon} \times U_{r,\epsilon}}) &\leqslant \delta_\square(W_j, W) + \delta_\square(W_j 1_{U_{r,\epsilon} \times U_{r,\epsilon}}, W_j) + \delta_\square(W 1_{U_{r,\epsilon} \times U_{r,\epsilon}}, W) \\ &\leqslant \delta_\square(W_j, W) + \epsilon/r^2,\end{aligned}$$

where the second line follows by the construction of $U_{r,\epsilon}$. By the assumption of cut convergence, there is some j' such that the first term is less than $\delta/2$ for all $j > j'$. Taking $\epsilon < r^2\delta/2$ then guarantees that the first term in Eq. (4.5.10) is less than $\epsilon'/2$ for sufficiently large j.

By Lemma 4.5.1, the last two terms of Eq. (4.5.10) are each at most $\epsilon/4$. Thus, choosing $\epsilon < \epsilon'$ guarantees that, for any $\epsilon' > 0$, there is some j' such that for all $j > j'$,

$$\Pr(H_r^{(j)} \neq H_r) \leqslant \epsilon', \tag{4.5.11}$$

as required. □

We now turn from the convergence of graphons to convergence of graphs.

Lemma 4.5.4. *Let $G_1, G_2, \ldots$ be a sequence of simple graphs. The following are equivalent:*

1. *The sequence is sampling convergent to W*
2. *The graphon process corresponding to the stretched empirical graphon converges to W, in the sense that, for all $r \in \mathbb{R}_+$, $\mathrm{GPD}(W^{G_j,s}, r) \to \mathrm{GPD}(W, r)$ weakly as $j \to \infty$.*

Proof. If $H_{j,r} \sim \mathrm{GPD}(W^{G_j,s}, r)$ then H_r may be generated by sampling $\mathrm{Poi}\left(\frac{r}{\sqrt{2e(G_j)}} v(G_j)\right)$ vertices with replacement from G_j and returning the edge set of the vertex induced subgraph. The claim is then simply Lemma 4.2.11, the asymptotic equivalence of this with replacement sampling scheme and $r/\sqrt{2e(G_j)}$-sampling. □

Theorem 4.5.5. *Let $G_1, G_2, \ldots$ be a uniformly tail regular sequence of simple graphs and let W be some non-random graphon. The following are equivalent:*

1. *The sequence converges in stretched cut distance to W*
2. *The sequence is sampling convergent to W*
3. *The graphon process corresponding to the stretched empirical graphon converges to W, in the sense that, for all $r \in \mathbb{R}_+$, $\mathrm{GPD}\big(W^{G_j,s}, r\big) \to \mathrm{GPD}(W, r)$ weakly as $j \to \infty$.*

Proof. The equivalence of 2 and 3 is a special case of Lemma 4.5.4.

By Lemma 4.5.3 the convergence in stretched cut distance implies that, almost surely,

$$\mathrm{GPD}\big(W^{G_j,s}, r\big) \to \mathrm{GPD}(W, r), \tag{4.5.12}$$

weakly as $j \to \infty$, for all $r \in \mathbb{R}_+$. Thus 1 implies 3.

Assume the sequence is sampling convergent. Because the sequence is assumed to be tail regular, it is subsequentially convergent in the stretched cut distance, by [BCCH16, Theorem 2.12]. If there are two subsequences with distinct limits then, because 1 implies 2, each of these subsequences will be sampling convergent with the laws of the sampled graphs given by distinct graphexes. This contradicts the assumption of sampling convergence, so we have that 2 implies 1. □

Remark 4.5.6. Because stretched cut convergent graph sequences are always tail regular, the previous lemma also shows that convergence in stretched cut distance implies sampling convergence.

METRIZATION

We now translate our main limit result in the language of metric convergence. We also show that the metric completion of the space of finite graphs equipped with a natural metrization of sampling convergence is compact.

Definition 4.6.1. Given two finite unlabeled graphs G, H, we define $\delta_{\mathrm{smpl}}(G, H) = \sum_k 2^{-k} \|\mathrm{SmplD}(G, k) - \mathrm{SmplD}(H, k)\|_{\mathrm{TV}}$

The point of introducing this metric is that it metrizes sampling convergence: it is immediate that a sampling convergent sequence is δ_{smpl}-Cauchy, the converse follows from the observation that $\|\mathrm{SmplD}(G, r) - \mathrm{SmplD}(H, r)\|_{\mathrm{TV}} \leqslant \|\mathrm{SmplD}(G, k) - \mathrm{SmplD}(H, k)\|_{\mathrm{TV}}$ whenever $r < k$, which is in turn clear from $\mathsf{Smpl}\Big(G, r/\sqrt{2e(G)}\Big) \overset{\mathrm{d}}{=} \mathsf{Smpl}\Big(\mathsf{Smpl}\Big(G, k/\sqrt{2e(G)}\Big), r/k\Big)$.

Remark 4.6.2. We could also have defined $\delta_{\mathrm{GP}}(G, H) = \delta_{\mathrm{GP}}(W^{G,s}, W^{H,s})$ where δ_{GP} on graphons is the metric that metrizes the topology of GP convergence [Jan17]. This is also a metrization of sampling convergence by Lemma 4.5.4.

Definition 4.6.3. Let $\mathscr{G}$ be the metric space of all edge sets of finite graphs equipped with δ_{smpl}, and let $\mathscr{G}^*$ be the metric completion.

Notice that finite graphs G and H and identified in $\mathscr{G}$ if and only if they are isomorphic after removing all isolated vertices.

Definition 4.6.4. Let $\mathscr{W}$ be the set of equivalence classes of graphexes satisfying $\|\mathcal{W}\|_1 \leqslant 1$ under the relation $\mathcal{W}_1 \sim W_2$ if and only if $\mathrm{GPD}(\mathcal{W}_1, r) = \mathrm{GPD}(\mathcal{W}_2, r)$ for all $r \in \mathbb{R}_+$.

$\mathscr{W}$ is the natural set of limit points of sampling convergent graph sequences of diverging size.

Theorem 4.6.5. *$\mathscr{G}^*$ is isomorphic to $\mathscr{G} \cup \mathscr{W}$.*

Proof. For any sequence of graphs $G_1, G_2, \dots$ that is Cauchy in $\mathscr{G}$ either $G_j = G$ for some finite graph G and all sufficiently large j, or $e(G_j) \to \infty$ as $j \to \infty$. In the former case, the sequence is identified with $G \in \mathscr{G}$. In the later case, we may identify the sequence with some element of $\mathscr{W}$ by Theorem 4.3.10.

Conversely, for every element for every element $G \in \mathscr{G}$ there is obviously some Cauchy sequence identified with it (namely, $G_j = G$ for all j). By Corollary 4.4.4 there is also a Cauchy sequence identified with every element of $\mathscr{W}$. □

Theorem 4.6.6. *Every infinite sequence $G_1, G_2, \dots$ in $\mathscr{G}$ such that $l(G_j) = O(\sqrt{e(G_j)})$ has a subsequence that is convergent in $\mathscr{G}^*$.*

Proof. Let $G_1, G_2, \dots$ be some sequence in $\mathscr{G}$. If there is some $k \in \mathbb{N}$ such that $\sup_j e(G_j) + l(G_j) < k$ then the existence of a convergent subsequence is obvious.

It now suffices to show that the closure in $\mathscr{G}^*$ of sequences such that $e(G_j) \to \infty$ and $l(G_j) = O(\sqrt{e(G_j)})$ is sequentially compact. By Lemma 4.3.2 it is equivalent to show that the canonical embeddings of the graph sequence are sequentially compact in the topology of weak convergence. [DVJ03b, Proposition 11.1.VI] shows that a sufficient condition for uniform tightness of a family of probability measures on the space of boundedly finite random measures on $\mathbb{R}_+^2$, say $(\Pr(\xi_s \in \cdot))_{s \in \mathcal{I}}$, is that for any bounded Borel set B and any $\epsilon > 0$ there is some $M \in \mathbb{R}_+$ such that $\Pr(\xi_s(B) > M) < \epsilon$ for all $s \in \mathcal{I}$. For a graph sequence $G_1, G_2, \dots$ the canonical labelings have the property that $\mathbb{E}[\mathrm{Lbl}(G_j)([0, r]^2)] \leqslant r^2 + r l(G_j)/\sqrt{2e(G_j)}$ (with equality whenever $\sqrt{2e(G_j)} > r$), from which the uniform tightness condition follows trivially. The result then follows by Prokhorov's theorem. □

SAMPLING DEFINES EXCHANGEABLE RANDOM GRAPHS

The size parameter of a KEG is related to p-sampling by the observation that if $G \sim \mathrm{GPD}(\mathcal{W}, s)$ then $\mathrm{Smpl}(G, p) \sim \mathrm{GPD}(\mathcal{W}, ps)$. That

is, the relationship between graphs at different times is captured by p-sampling. In this section we show that this is in fact a defining property of sparse exchangeable random graphs.

Definition 4.7.1. Call $(G_s)_{s\in\mathbb{R}_+}$ an *unlabeled random graph process* indexed by $\mathbb{R}_+$ if, for all s, G_s is a finite unlabeled graph, and, for all $s \leqslant t$, $G_s \subseteq G_t$ in the sense that there is some subgraph of G_t that is isomorphic to G_s.

Theorem 4.7.2. *Let $(G_s)_{s\in\mathbb{R}_+}$ be an unlabeled random graph process such that $e(G_s) \uparrow \infty$ a.s. as $s \to \infty$. For each $s \in \mathbb{R}_+$ and $p \in (0,1)$, let $\mathsf{Smpl}(G_s, p)$ be a p-sampling of G_s. If for all $s \in \mathbb{R}_+$ and $p \in (0,1)$,*

$$\mathsf{Smpl}(G_s, p) \stackrel{d}{=} G_{ps}, \tag{4.7.1}$$

then there is some (possibly random, possibly non-integrable) almost surely non-zero graphex $\mathcal{W}$ such that, for all $s \in \mathbb{R}_+$, $G_s \mid \mathcal{W} \sim \mathrm{GPD}(\mathcal{W}, s)$.

Proof. To establish the claimed result, it obviously suffices to show that there is some $\mathcal{W}$ such that $\mathsf{Lbl}_s(G_s) \stackrel{d}{=} \Gamma_s$, where $(\Gamma_s)_{s\in\mathbb{R}_+}$ is a KEG generated by $\mathcal{W}$.

Let $r, s \in \mathbb{R}_+$ be such that $r < s$. Then,

$$\mathsf{Lbl}_s(G_s)([0,r)^2 \cap \cdot) \stackrel{d}{=} \mathsf{Lbl}_r(\mathsf{Smpl}\left(G_s, \frac{r}{s}\right)) \stackrel{d}{=} \mathsf{Lbl}_r(G_r). \tag{4.7.2}$$

The first equality follows by the observation that each vertex of $\mathsf{Lbl}_s(G_s)$ has label in $[0,r)$ independently with probability r/s, so that $\mathsf{Lbl}_s(G_s)$ restricted to $[0,r)^2$ has the same distribution as $\mathsf{Lbl}_r(\mathsf{Smpl}(G_s, r/s))$. The second equality is by hypothesis.

Let ξ be a point process with distribution defined by, for any bounded Borel sets $B_1, \ldots, B_n \subseteq \mathbb{R}_+^2$,

$$\{\xi(B_1), \ldots, \xi(B_n)\} \stackrel{d}{=} \lim_{s\to\infty} \{\mathsf{Lbl}_s(G_s)(B_1), \ldots, \mathsf{Lbl}_s(G_s)(B_n)\}. \tag{4.7.3}$$

Eq. (4.7.2) makes it clear that the limiting distribution on the right hand side is well defined. Moreover, the consistency conditions necessary for $\lim_{s\to\infty}\{\mathsf{Lbl}_s(G_s)(B_1), \ldots, \mathsf{Lbl}_s(G_s)(B_n)\}$ to be counts with respect to some point process are obviously satisfied, because the joint distribution is defined as counts of the random labeling point process. By the Kolmogorov existence theorem for point processes (see [DVJ03b, Theorem 9.2.X]), this suffices to show that ξ exists and has a well-defined distribution. Also note that ξ is purely atomic by construction.

Observe that by Eq. (4.7.2) and the definition of ξ, it holds that, for all $r \in \mathbb{R}_+$,

$$\mathsf{Lbl}_r(G_r) \stackrel{d}{=} \xi([0,r)^2 \cap \cdot). \tag{4.7.4}$$

In consequence, for any measure preserving transformation ϕ on $[0, r)$, $\xi \circ (\phi \otimes \phi) \overset{d}{=} \xi$. In particular then, for any dyadic partitioning of $\mathbb{R}_+$ and any transposition τ of this dyadic partitioning we may take r large enough such that the transposition acts only in $[0, r)$ and thus $\xi \circ (\tau \otimes \tau) \overset{d}{=} \xi$. By [Kal05, Proposition 9.1] this implies that ξ is exchangeable.

We now have that ξ is a purely atomic exchangeable point process, so by the Kallenberg representation theorem, Theorem 2.4.7, there is some graphex $\mathcal{W}$ such that ξ is generated by $\mathcal{W}$. The proof is then completed by again invoking Eq. (4.7.4). □

BIBLIOGRAPHY

[ABFX08] E. M. Airoldi, D. M. Blei, S. E. Feinberg, and E. P. Xing. "Mixed-membership stochastic blockmodels". In: *J. Machine Learn. Res.* 9 (2008), pp. 1981–2014.

[AG89] R. Arratia and L. Gordon. "Tutorial on large deviations for the binomial distribution". English. In: *Bulletin of Mathematical Biology* 51.1 (1989), pp. 125–131. ISSN: 0092-8240.

[Ald09] D. J. Aldous. "More Uses of Exchangeability: Representations of Complex Random Structures". In: *ArXiv e-prints* (Sept. 2009). arXiv: 0909.4339 [math.PR].

[Ald81] David J. Aldous. "Representations for partially exchangeable arrays of random variables". In: *J. Multivariate Anal.* 11.4 (1981), pp. 581–598. ISSN: 0047-259X.

[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.

[Aus08] Tim Austin. "On exchangeable random variables and the statistics of large graphs and hypergraphs". In: *Probab. Surv* 5.80-145 (2008).

[BA99] Albert-László Barabási and Réka Albert. "Emergence of Scaling in Random Networks". In: *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". In: *ArXiv e-prints* (Jan. 2014). arXiv: 1401.2792 [math.PR].

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

[BCCH16] C. Borgs, J. T. Chayes, H. Cohn, and N. Holden. "Sparse exchangeable graphs and their limits via graphon processes". In: *ArXiv e-prints* (Jan. 2016). arXiv: 1601.07134 [math.PR].

[BCCV17] C. Borgs, J. Chayes, H. Cohn, and V. Veitch. *Sampling perspectives on sparse exchangeable graphs*. Preprint. 2017.

[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". In: *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". In: *ArXiv e-prints* (Aug. 2014). arXiv: 1408.0744 [math.CO].

[BCLS+06] C. Borgs, J. Chayes, L. Lovászsz, V. T. Sós, and K. Vesztergombi. "Counting graph homomorphisms". In: *Topics in Discrete Mathematics*. Ed. by M. Klazar, J. Kratochvil, M. Loebl, J. Matousek, R. Thomas, and P. Valtr. Springer, 2006, pp. 315–371.

[BCLS+08a] C. Borgs, J. T. Chayes, L. Lovasz, V. T. Sós, and K. Vesztergombi. "Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing". In: *Advances in Mathematics* 219.6 (2008), pp. 1801 –1851. ISSN: 0001-8708.

[BCLS+08b] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi. "Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing". In: *Adv. Math.* 219.6 (2008), pp. 1801–1851. ISSN: 0001-8708.

[BCLS+12] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi. "Convergent sequences of dense graphs. II. Multiway cuts and statistical physics". In: *Ann. Math.* 176 (2012), pp. 151–219.

[BCS15] C. Borgs, J. T. Chayes, and A. Smith. "Private Graphon Estimation for Sparse Graphs". In: *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". In: *Random Struct. Alg.* 31.1 (2007), pp. 3–122. ISSN: 1042-9832.

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

[BS01] Itai Benjamini and Oded Schramm. "Recurrence of distributional limits of finite planar graphs". In: *Electron. J. Probab.* 6 (2001), no. 23, 13 pp. (electronic). ISSN: 1083-6489.

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

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

[CD15] H. Crane and W. Dempsey. "A framework for statistical network modeling". In: *ArXiv e-prints* (Sept. 2015). arXiv: 1509.08185 [math.ST].

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

[CSKM13] Sung Nok Chiu, Dietrich Stoyan, Wilfrid S. Kendall, and Joseph Mecke. *Stochastic Geometry and Its Applications*. 3rd ed. Wiley, 2013.

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

[DVJ03a] Daryl J Daley and David Vere-Jones. *An introduction to the theory of point processes: volume I: elementary theory and methods*. Second. Springer Science & Business Media, 2003.

[DVJ03b] Daryl J Daley and David Vere-Jones. *An introduction to the theory of point processes: volume II: general theory and structure*. Second. Springer Science & Business Media, 2003.

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

[Dur10] R. Durrett. *Probability: Theory and Examples*. 4th Edition. Cambridge U Press, 2010.

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

[Fin37] Bruno de Finetti. "La prévision: ses lois logiques, ses sources subjectives". In: *Ann. Inst. H. Poincaré* 7.1 (1937), pp. 1–68. ISSN: 0365-320X.

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

[HRH02] Peter D Hoff, Adrian E Raftery, and Mark S Handcock. "Latent space approaches to social network analysis". In: *Journal of the American Statistical Association* 97.460 (2002), pp. 1090–1098.

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

[Hol12] F. den Hollander. *Probability Theory: The Coupling Method*. 2012.

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

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

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

[Kal90] Olav Kallenberg. "Exchangeable random measures in the plane". English. In: *Journal of Theoretical Probability* 3.1 (1990), pp. 81–136. ISSN: 0894-9840.

[Kal99] Olav Kallenberg. "Multivariate sampling and the estimation problem for exchangeable arrays". In: *J. Theoret. Probab.* 12.3 (1999), pp. 859–883. ISSN: 0894-9840.

[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.

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

[LS07] László Lovász and Balázs Szegedy. "Szemerédi's Lemma for the Analyst". In: *Geom. Func. Anal.* 17.1 (2007), pp. 252–270.

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

[MJG09] Kurt Miller, Michael I Jordan, and Thomas L Griffiths. "Nonparametric latent feature models for link prediction". In: *Advances in neural information processing systems*. 2009, pp. 1276–1284.

[NS01] Krzysztof Nowicki and Tom A. B. Snijders. "Estimation and prediction for stochastic blockstructures". In: *J. Amer. Statist. Assoc.* 96.455 (2001), pp. 1077–1087. ISSN: 0162-1459.

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

[OR15] P. Orbanz and D.M. Roy. "Bayesian Models of Graphs, Arrays and Other Exchangeable Random Structures". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 37.2 (Feb. 2015), pp. 437–461. ISSN: 0162-8828.

[Orb17] P. Orbanz. *Subsampling and invariance in networks*. Preprint. 2017.

[VR15] V. Veitch and D. M. Roy. "The Class of Random Graphs Arising from Exchangeable Random Measures". In: *ArXiv e-prints* (Dec. 2015). arXiv: 1512.03099 [math.ST].

[VR16] V. Veitch and D. M. Roy. "Sampling and Estimation for (Sparse) Exchangeable Graphs". In: *ArXiv e-prints* (Nov. 2016). arXiv: 1611.00843 [math.ST].

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

[Jan17] S. Janson. "On convergence for graphexes". In: *ArXiv e-prints* (Feb. 2017). arXiv: 1702.06389 [math.PR].