The Annals of Probability
2019, Vol. 47, No. 5, 2754–2800
https://doi.org/10.1214/18-AOP1320

SAMPLING PERSPECTIVES ON SPARSE EXCHANGEABLE GRAPHS

BY CHRISTIAN BORGS*, JENNIFER T. CHAYES*, HENRY COHN* AND VICTOR VEITCH†,[1]

*Microsoft Research** *and Columbia University*†

Recent work has introduced sparse exchangeable graphs and the associated graphex framework, as a generalization of dense exchangeable graphs and the associated graphon framework. The development of this subject involves the interplay between the statistical modeling of network data, the theory of large graph limits, exchangeability and network sampling. The purpose of the present paper is to clarify the relationships between these subjects by explaining each in terms of a certain natural sampling scheme associated with the graphex model. The first main technical contribution is the introduction of *sampling convergence*, a new notion of graph limit that generalizes left convergence so that it becomes meaningful for the sparse graph regime. The second main technical contribution is the demonstration that the (somewhat cryptic) notion of exchangeability underpinning the graphex framework is equivalent to a more natural probabilistic invariance expressed in terms of the sampling scheme.

1. Introduction. The present paper is concerned with the theory of graph limits, the statistical modeling of networks and the relationship between these topics 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 [6, 10, 11, 30, 31] (see [29] for a review) and provide the foundation for many of the statistical network models in current use [1, 20, 28, 32, 33] (see [34] for a review). Motivated by the importance of graphons in the dense graph setting, a recent series of papers [7, 12, 19, 23, 35–37] has developed a generalization of the graphon framework to the regime of sparse graphs, both as a tool for statistical network modeling [7, 36] and estimation [37], and as the central element of a limit theory for large graphs [7] (see also [22]). This generalization is compelling in that it preserves many of the desirable properties of the graphon framework, while simultaneously allowing much greater flexibility. However, there are some significant interpretational issues remaining. For example, it is unclear which real-world processes are appropriately modeled by the statistical network models of the new framework, or how best to

Received August 2017; revised June 2018.

[1]Work partially completed while at University of Toronto. Supported in part by U.S. Air Force Office of Scientific Research Grant #FA9550-15-1-0074 and an internship at Microsoft Research New England.

MSC2010 subject classifications. Primary 60C05; secondary 62D05, 62G09, 62G05.

Key words and phrases. Network analysis, sampling, graph limits, nonparametric estimation.

characterize the properties of large graphs that are well approximated by the new limit theory. The root of these difficulties is that the new framework is derived using a cryptic construction that represents random graphs as point processes on $\mathbb{R}_+^2$, and then formalizes the models of the generalized framework as those corresponding to point processes that are exchangeable.

In the dense setting, graphons as stochastic network models can be arrived at in at least two different ways. The first approach is simply to posit them directly. Graphon models are the class of generative models for random graphs in which each vertex i is assigned some independent latent features x_i, and conditional on these latent features, each pair of vertices i, j is connected by an edge independently with probability $W(x_i, x_j)$ determined by the latent features of i and j. 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 statistical networks literature. The second approach proceeds by identifying 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 defining 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 [2, 21].

Graphons as limit objects for dense graphs sequence also arise very 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 [6, 10, 11]. 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. Finally, exchangeable random graphs generated from a graphon can be shown to converge to the generating graphon [30], creating a first connection between graphons as models for exchangeable random graphs and as limits of sequences of sparse graphs. See [17] 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.

The key ingredient of the generalization from the dense graph setting to the sparse graph setting is a novel notion of exchangeability for random graphs. In the generalized theory, the vertices of the random graphs are labeled in $\mathbb{R}_+$, the edge sets of these graphs are represented as point processes on $\mathbb{R}_+^2$ and invariance under vertex relabeling is encoded as joint exchangeability of the point process. This rather abstruse formalization was introduced as an ad hoc solution to the problem that the more obvious notion of exchangeability implies that the corresponding

random graphs are almost surely dense. Nevertheless, the resulting models retain the essential character of the dense graphon models: each vertex i has latent feature x_i and, conditional on these latent features, each edge is included independently with a probability determined by the latent features of its endpoints. The essential difference is that the latent features are now generated as a Poisson process on a σ-finite space, rather than independently. The appeal of these models is then their close analogy to the dense graphon models, in combination with their greater flexibility.

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 [7] uses 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 paper 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 [6, 10], 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 [7] 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 [17] 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 idea in formalizing this is the following notion for sampling from a graph, introduced in [37]. Here, a vertex in a subgraph of a given graph G is called *isolated* if it is not contained in any edge (regardless of whether this edge is a loop edge or a nonloop edge) of the subgraph.

DEFINITION 1.1. A *p-sampling* $\mathsf{Smpl}(G, p)$ of a graph[2] 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. Let $e(G)$ denote the number of nonloop edges of a graph G.

DEFINITION 1.2. A sequence of graphs $G_1, G_2, \ldots$ is *sampling convergent* if, for all $r \in \mathbb{R}_+$, the random graphs $\mathsf{Smpl}(G_j, r/\sqrt{2e(G_j)})$ 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 paper.

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

[2]Throughout this paper, a graph will be a graph without multiple edges, but it may not be simple; that is, it may contain edges joining a vertex to itself. Unless explicitly mentioned, all graphs will be finite.

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

$$\mathbb{E}\big[e\big(\mathsf{Smpl}\big(G_j, r/\sqrt{2e(G_j)}\big)\big)\big] = r^2/2$$

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 a triple $\mathcal{W} = (I, S, W)$, where $I \in \mathbb{R}_+$, $S\colon \mathbb{R}_+ \to \mathbb{R}_+$ is an integrable function, and the *graphon* $W\colon \mathbb{R}_+^2 \to [0,1]$ is a symmetric integrable function. This object is the (integrable) *graphex* at the heart of the (sparse) exchangeable graph models. Each graphex defines a *graphex process* (or *Kallenberg Exchangeable Graph* in the language of [36, 37]), a family of growing random graphs $(\Gamma_s)_{s\in\mathbb{R}_+}$ with vertices labeled in $\mathbb{R}_+$. Following [7], we refer to the label of a vertex as its *birth time*, and to Γ_s as the *graphex process* at time s. For a finite labeled graph Γ_s, we denote the associated unlabeled graph by $\mathcal{G}(\Gamma_s)$. The sense in which the graphex is the natural limit object is given by Theorem 3.11: for every sampling convergent sequence $G_1, G_2, \ldots$ there is some integrable graphex $\mathcal{W}$ such that, for all $s \in \mathbb{R}_+$, $\mathsf{Smpl}(G_j, s/\sqrt{2e(G_j)}) \xrightarrow{d} \mathcal{G}(\Gamma_s)$ 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. In this case, we say that G_j is *sampling convergent to* $\mathcal{W}$.

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 certain equivalencies (Theorem 4.3), and by metrizing the convergence and characterizing the associated metric space (Theorems 6.7 and 6.8). In consequence of the former result, the (integrable) graphex process 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 processes can be characterized directly in terms of p-sampling, without appeal to either exchangeability or graph limits. The motivation in [37] for the introduction of p-sampling was the observation that a p-sampling of $\mathcal{G}(\Gamma_s)$ is equal in distribution to $\mathcal{G}(\Gamma_{ps})$; that is, this is the sampling scheme that describes the relationship between graphex process graphs at different times. We prove in Theorem 7.2 that this is in fact a defining property of the graphex process. That is, if $(G_s)_{s\in\mathbb{R}_+}$ is a family of unlabeled random graphs such that for all $s \in \mathbb{R}_+$ and all $p \in (0,1)$ the p-sampling

of G_s is equal in distribution to G_{ps}, then there is some graphex $\mathcal{W}$ such that $G_s \stackrel{d}{=} \mathcal{G}(\Gamma_s)$ 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.

1.1. *Statistical network modeling.* The major motivation in [36] for the introduction of graphex process models 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 and the exchangeable array framework, which underlie 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 graphex process.

Following [13], 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 process 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 graphex process 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, the graphex process has some properties that are highly undesirable for a model of a data generating process. For example, a graphex process 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 the 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 3.11, 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 follows:

> Finite size graphex processes 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 is 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.

In lectures and as yet unpublished work, P. Orbanz has given 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 paper is that we work out the realization of this program for p-sampling.

1.2. *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 [7], and
3. the consistent estimation of [37].

1.2.1. *Randomly labeled graphs.* The first of these is fundamental to the development of the theory in the present paper. Exchangeability is a concept of infinite size labeled random graphs, but the theory of graph limits deals with nonrandom 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 [2, 21, 27] on one hand and the independent development of the theory of dense graph limits [6, 10, 11, 30, 31] on the other. The connection between the two perspectives is explained by [4, 17], 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 each 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 nonrandom 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 $k \times k$ submatrix. 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 nonrandom 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 pairs i, j such that $\{i, j\}$ is an edge of a countable graph G (possibly containing some loops, that is, edges joining a vertex to itself) and $\theta_i \in \mathbb{R}_+$ with $\theta_i \neq \theta_j$ for $i \neq j$.

DEFINITION 1.3. Let G be a labeled or unlabeled graph and let $s > 0$. A *random labeling of G into* $[0, s)$ is a random adjacency measure obtained by labeling 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 bounded sets $A, B \subseteq \mathbb{R}_+$ such that $\max(A \cup B) \leq s_j$, the expected number of edges between vertices with labels in A and B is independent of the graph.

DEFINITION 1.4. 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 3 that the graph sequence $G_1, G_2, \dots$ is sampling convergent to $\mathcal{W}$ if and only if the canonical labelings $\mathsf{Lbl}(G_1), \mathsf{Lbl}(G_2), \dots$ 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 [3], 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 paper is that we realize this program for p-samplings of families of growing graphs.

1.2.2. *Metric convergence.* One of the important tools in the theory of dense graph limits is the cut distance between two graphs or graphons [6]. The cut metric defines a notion of distance that, essentially, captures how similar two graphs or graphons look at low resolutions; see Figure 1 below. We define cut distance formally in Section 2. One of the contributions of [7] 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, [7] 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 Figure 1. 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 under the stretched cut distance. Indeed this is so, and in Theorem 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 under the stretched cut distance implies sampling convergence.

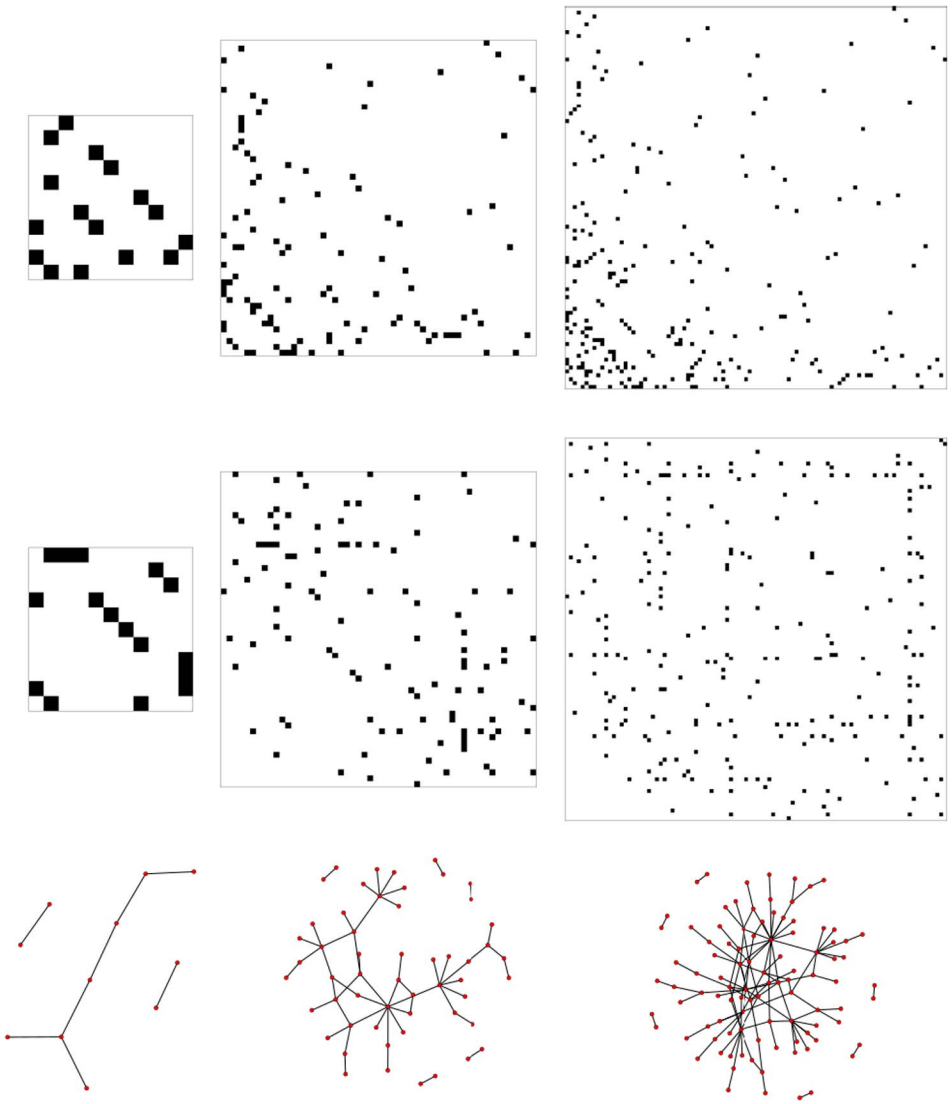

FIG. 1. *Each column shows a graph* (*bottom row*), *a corresponding stretched empirical graphon based on a random labeling of the vertices* (*middle row*), *and a corresponding stretched empirical based on an alternative labeling* (*top row*). *The three graphs are a prefix of a sequence that converges to* $(0, 0, W)$, *where* $W(x, y) = (x + 1)^{-2}(y + 1)^{-2}$. *Intuitively, the top row shows 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* $\sup_{U,V\subseteq\mathbb{R}_+} |\int_{U\times V} W_1(x, y) - W_2(x, y)\,\mathrm{d}x\,\mathrm{d}y|$, *the largest difference in any patch between the total amounts of ink in that patch.*

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.

1.2.3. *Consistent estimation.* The paper [37] 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 times 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 times are included as part of the observation; both of these are closely related to the sampling convergence of the present paper.

Let $\mathrm{GPD}(\mathcal{W}, s) = \Pr(\mathcal{G}(\Gamma_s) \in \cdot \mid \mathcal{W})$ denote the probability distribution over unlabeled time s graphs generated by $\mathcal{W}$, where GPD stands for graphex process distribution. In the setting where the times 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 $\mathcal{W}$.

For a graph G, define $\widehat{W}_{(G,s)}\colon [0, v(G)/s]^2 \to \{0, 1\}$, the dilated empirical graphon of G with dilation s, to be the function given by representing the adjacency matrix[3] of G as a step function where each pixel has size $1/s \times 1/s$; see Figure 1. The estimator used by [37] in the setting where the times 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: $\widehat{W}_{(G_j,s_j)} \to_{\mathrm{GP}} \mathcal{W}$ is equivalent to $\mathsf{Smpl}(G_j, \frac{r}{s_j}) \xrightarrow{d} \mathcal{G}(\Gamma_r)$ for all $r \in \mathbb{R}_+$. To explain this connection, we recall a pair of ideas from [37] (themselves adapted from

[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. All our subsequent notions do not 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.

[25]). First, generating a sample from $\mathrm{GPD}(\widehat{W}_{(G_j,s_j)}, r)$ is equivalent to sampling a subgraph from G_j by selecting $\mathrm{Poi}(\frac{r}{s_j} v(G_j))$ 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 of the two notions of convergence 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 $\mathcal{W} = (0, 0, W)$, [7] proves that $e(G_j)/s_j^2 \to \frac{1}{2}\|W\|_1$ a.s. as $j \to \infty$, and 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 the dilated empirical graphon of G with dilation $\sqrt{2e(G)}$. The salient fact, spelled out in Lemma 5.4, is that $G_1, G_2, \ldots$ is sampling convergent to $\mathcal{W}$ if and only if $W^{G_j,s} \to_{\mathrm{GP}} \mathcal{W}$ 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 $\|\mathcal{W}\|_1 = 1$.

Veitch and Roy (2016) follow a different approach. In the case where the sample times 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 $\mathcal{W}_j \to_{\mathrm{GS}} \mathcal{W}$ 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 $s_1, s_2, \ldots$.

Because the empirical graphon relies only on the graph (and not the latent observation time), 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 graphex process), and say that the sequence is *GS convergent* to $\mathcal{W}$, written $G_j \to_{\mathrm{GS}} \mathcal{W}$ as $j \to \infty$, whenever $W^{G_j} \to_{\mathrm{GS}} \mathcal{W}$ as $j \to \infty$. [37], Lemma 5.6, shows that as long as $\mathcal{W} \neq 0$, $\mathcal{W}_j \to_{\mathrm{GP}} \mathcal{W}$ as $j \to \infty$ implies also $\mathcal{W}_j \to_{\mathrm{GS}} \mathcal{W}$ as $j \to \infty$, from which it follows that sampling convergence implies GS convergence. The converse is not true: the consistent estimation results of [37] establish that graph sequences generated by nonintegrable $\mathcal{W}$ are GS convergent to $\mathcal{W}$, 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 process theory.

We note that [23] includes a discussion of the relationship between various notions of convergence of graphexes, and is closely related to the development in this section.

1.3. *Organization.* We give formal definitions and recall some important results in Section 2. The basic results for sampling convergence—most importantly, the limits are graphexes—are given in Section 3. In Section 4, we prove that a graph sequence generated by integrable graphex $\mathcal{W}$ is almost surely sampling convergent to a canonical dilation of $\mathcal{W}$; this has the particular consequence that (a canonical representative of) every integrable graphex arises as the sampling limit of some graph sequence. In Section 5, we relate convergence in distribution of graphex sequences generated by $\mathcal{W}_1, \mathcal{W}_2, \ldots$ to the metric convergence of [7]. In Section 6, we metrize sampling convergence and show that the metric completion of the space of finite unlabeled loopless graphs is compact (a less elegant statement is required for loops). In Section 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 $\mathcal{W}$ such that $G_s = \mathcal{G}(\Gamma_s)$ for some $(\Gamma_s)_{s\in\mathbb{R}_+}$ generated by $\mathcal{W}$.

2. Preliminaries. As usual, we denote the set of edges and vertices of a graph G by $E(G)$ and $V(G)$, respectively. In general, $E(G)$ will consist of both loop and nonloop edges; we denote the number of nonloop edges by $e(G)$ and the number of loop edges by $\ell(G)$.

Some of the basic objects of interest in this paper are locally finite point processes on $\mathbb{R}_+^2$, interpreted as the edge sets of random graphs with vertices labeled in $\mathbb{R}_+$. Here, as usual, a *locally finite point process* on $\mathbb{R}_+^2$ is a random element ξ of the set $\mathcal{N} = \mathcal{N}(\mathbb{R}_+^2)$ of locally finite counting measure on $\mathbb{R}_+^2$ (i.e., the set of integer valued measures ξ such that $\xi(A) < \infty$ for all bounded Borel sets $A \subset \mathbb{R}_+^2$), equipped with the Borel σ-algebra inherited from the vague topology, defined as the coarsest topology for which the maps $\mu \mapsto \int f \, d\mu$ are continuous for all continuous functions with bounded support. As shown in, for example, [14], this topology can be metrized in such a way that $\mathcal{N}$ becomes a complete, separable metric space. Convergence in distribution for locally finite point processes is defined as weak convergence with respect to this topology, so that $\xi_n \xrightarrow{d} \xi$ is defined by the condition that $\mathbb{E}[F(\xi_n)] \to \mathbb{E}[F(\xi)]$ for all continuous, bounded functions F, with continuity defined with respect to the vague topology on $\mathcal{N}$.

DEFINITION 2.1. An *adjacency measure* is a purely atomic, symmetric locally finite counting measure on $\mathbb{R}_+^2$ for which all atoms have weight 1. A *random adjacency measure* is a locally finite point process ξ on $\mathbb{R}_+^2$ such that ξ is almost surely an adjacency measure.

We associated a graph with labels in $\mathbb{R}_+$ to an adjacency measure ξ by writing it as $\xi = \sum_{i,j} \delta_{(\theta_i,\theta_j)}$, defining the set $\{(\theta_i, \theta_j)\}$ with $\theta_i \le \theta_j$ as its edge set, and defining the set of points θ_i that participate in at least one edge as its vertex set. Most of the time, we will not distinguish between the countable graph associated with ξ and the adjacency measure ξ itself.

The defining property of graphex processes is that, intuitively speaking, the labels of the vertices of the graphs are uninformative about their structure. This is formalized by requiring the associated adjacency measure to be jointly exchangeable.

DEFINITION 2.2. A random adjacency measure ξ is *jointly exchangeable* if $\xi \circ (\phi \otimes \phi) \stackrel{d}{=} \xi$ for every measure-preserving transformation $\phi\colon \mathbb{R}_+ \to \mathbb{R}_+$. It is called an *extremal exchangeable adjacency measure* if its distribution cannot be written as a nontrivial superposition of distributions over jointly exchangeable adjacency measures, that is, if a representation of the distribution as $\alpha P_1 + (1 - \alpha) P_2$ for some $\alpha \in (0, 1)$ implies that $P_1 = P_2$ a.e.

A representation theorem for jointly exchangeable random measures on $\mathbb{R}_+^2$ was given by Kallenberg [24, 27]. This result was translated to the setting of random graphs in [7, 36]. Writing Λ for Lebesgue measure and $\mu_W(\cdot) = \int_{\mathbb{R}_+} W(x, \cdot)\,\mathrm{d}x$, the defining object of the representation theorem is as follows.

DEFINITION 2.3. A *graphex* is a triple (I, S, W), where $I \ge 0$ is a nonnegative real, $S\colon \mathbb{R}_+ \to \mathbb{R}_+$ is a measurable function such that $\min(S, 1)$ is integrable, and the *graphon* $W\colon \mathbb{R}_+^2 \to [0, 1]$ is a symmetric, measurable function that satisfies:

1. $\Lambda\{\mu_W = \infty\} = 0$ and $\Lambda\{\mu_W > 1\} < \infty$,
2. $\int_{\mathbb{R}_+^2} W(x, y) 1[\mu_W(x) \le 1] 1[\mu_W(y) \le 1]\,\mathrm{d}x\,\mathrm{d}y < \infty$, and
3. $\int_{\mathbb{R}_+} W(x, x)\,\mathrm{d}x < \infty$.

REMARK 2.4. Integrability of W (and its diagonal) is a sufficient but not necessary condition for it to be a graphon. If the graphon, its diagonal $W_D(x) = W(x, x)$, and the function S are integrable, then we say that $\mathcal{W}$ is an *integrable graphex*. We set

$$\|\mathcal{W}\|_1 = \|W\|_1 + 2\|S\|_1 + 2I.$$

Integrability plays a fundamental role in sampling convergence.

Each graphex gives rise to a random adjacency measure, which in turn leads to a graph-valued stochastic process:

DEFINITION 2.5. Given a graphex $\mathcal{W} = (I, S, W)$, let ξ be the random adjacency measure

$$\begin{aligned}\xi = &\sum_{i,j} 1\big[\zeta_{\{i,j\}} \le W(\vartheta_i, \vartheta_j)\big]\delta_{\theta_i,\theta_j} \\ &+ \sum_{j,k} 1\big[\chi_{jk} \le S(\vartheta_j)\big](\delta_{\theta_j,\sigma_{jk}} + \delta_{\sigma_{jk},\theta_j}) \\ &+ \sum_{k} 1[\eta_k \le I](\delta_{\rho_k,\rho'_k} + \delta_{\rho'_k,\rho_k}),\end{aligned} \tag{2.1}$$

where $(\zeta_{\{i,j\}})$ is a collection of independent uniformly distributed random variables in $[0, 1]$, $\{(\theta_j, \vartheta_j)\}$ and $\{(\sigma_{ij}, \chi_{ij})\}_j$, for $i \in \mathbb{N}$, are independent unit rate Poisson processes on $\mathbb{R}_+^2$, and $\{(\rho_j, \rho'_j, \eta_j)\}$ are independent unit rate Poisson processes on $\mathbb{R}_+^3$, all of them independent of each other.[4]

Let Γ be the (in general countably infinite) graph corresponding to the adjacency measure ξ defined in (2.1), and let Γ_s be the (a.s. finite) graph corresponding to the adjacency measure $\xi_s(\cdot) = \xi(\cdot \cap [0, s]^2)$. The *graphex process* associated with graphex (I, S, W) is the family $(\Gamma_s)_{s\in\mathbb{R}_+}$.

REMARK 2.6. One might be tempted to identify graphexes that are equal almost everywhere. While this is possible, one must pay attention to details here, since changing a graphon on the diagonal is only a change on a set of measure zero, but it changes the graphex process associated to the graphex. This problem can be easily addressed by introducing the function $W_D(x) = W(x, x)$, identifying a graphex (I, S, W) with the quadruple (I, S, W, W_D) and considering the latter as an element of $\mathbb{R}_+ \times L^0(\mathbb{R}_+, \Lambda) \times L^0(\mathbb{R}_+^2, \Lambda^2) \times L^0(\mathbb{R}_+, \Lambda)$.

REMARK 2.7. In [7], a nominally more general definition of a graphon (and the associated graphon process) is used. There the domain of W is allowed to be an arbitrary σ-finite measure space modeling a space of latent features. The associated process is then defined by labeling vertices with a pair of labels, namely their birth time and their feature. In the above definition of $(\Gamma_s)_{s\in\mathbb{R}_+}$, the feature space is assumed to be $\mathbb{R}_+$, and vertices are just labeled by their birth times, not a pair of labels. By Theorem 2.8 below, or the explicit measure-preserving mappings constructed in [7], every such model is equivalent to one with latent feature space $\mathbb{R}_+$, so there is no loss of generality in our definition. The motivation for the more general notion is that in many situations there is a natural choice for the space of latent features, and strong-arming the feature space to $\mathbb{R}_+$ may obfuscate the conceptual

[4]By the results of [24], the integrability conditions from Definition 2.3 imply that the above sums are a.s. convergent in the vague topology, which in turn implies that ξ is a.s. locally finite. It is furthermore not hard to show that a.s., ξ is simple, implying that ξ is an adjacency measure.

underpinnings of the model or destroy certain nice theoretical properties (such as continuity of the graphon). In the present paper, this is not a concern, so we prefer the simpler definition with graphons defined over $\mathbb{R}_+$. We also label vertices in the graphex process $(\Gamma_s)_{s\in\mathbb{R}_+}$ by just their birth time, since in this paper, the latent feature of a vertex is usually not important. Indeed, as we will see below, we often remove even the birth time label of our vertices, leading to processes of unlabeled graphs.

Given Definitions 2.3 and 2.5, we can now state the Kallenberg representation theorem.

THEOREM 2.8. *Let ξ be a random adjacency measure. Then ξ is jointly exchangeable iff there exists a (possibly random) graphex $\mathcal{W}$ such that ξ is of the form (2.1). The graphex $\mathcal{W}$ can be chosen to be nonrandom if and only if ξ is extremal.*

REMARK 2.9. By a random measurable function $f\colon X \to \mathbb{R}$, we mean a measurable function $f'\colon [0,1]\times X \to \mathbb{R}$ and a randomization variable $\alpha \sim \mathrm{Uni}[0,1]$ such that $f(x) = f'(\alpha, x)$; see, for example, [18], Chapter 4. By a random graphex, we mean a quadruple (I, S, W, W_D) (see Remark 2.6 above) such that each component is an appropriate random measurable function all sharing a common randomization variable α, and such that the graphex integrability requirements are almost surely satisfied; by conditioning on a graphex $\mathcal{W}$ we mean conditioning on the randomization parameter α. We separate out the diagonal of the graphon so that two graphexes that are equal a.e. generate the same distribution over adjacency measures; this sidesteps some measurability technicalities.

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

DEFINITION 2.10. Let ξ be a finite adjacency measure. The *unlabeled graph associated with* ξ is $\mathcal{G}(\xi)$.

Similarly, we will often want to move from unlabeled graphs to adjacency measures. To do so, we must invent labels for the vertices; a simple scheme is to produce labels independently and uniformly in some range:

DEFINITION 2.11. Let G be a graph with edge set E, and let $s > 0$. A *random labeling of G into* $[0,s)$, denoted $\mathsf{Lbl}_s(G, \{U_i\})$, is a random adjacency measure $\mathsf{Lbl}_s(G, \{U_i\}) = \sum_{(i,j)\in E} \delta_{(U_i,U_j)}$, where the sum contains both orientations of each nonloop edge and $U_i \overset{\text{iid}}{\sim} \mathrm{Uni}[0,s)$ for each vertex i in G. Where there is no risk of confusion, we will write $\mathsf{Lbl}_s(G)$ for $\mathsf{Lbl}_s(G, \{U_i\})$ where $U_i \overset{\text{iid}}{\sim} \mathrm{Uni}[0,s)$ for all vertices i, independently of everything else. The random labeling is called a *canonical labeling of G* and denoted by $\mathsf{Lbl}(G)$ if $s = \sqrt{2e(G)}$.

Note that for an unlabeled graph, we need to fix a labeling of the vertices of G to define $\mathsf{Lbl}_s(G, \{U_i\})$; since the distribution of $\mathsf{Lbl}_s(G, \{U_i\})$ is independent of what labeling we chosen for G, the choice of this labeling is irrelevant.

2.1. *Graph limits.* We now recall some important definitions and results on the metric convergence of [7], 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 $\psi : \mathbb{R}_+ \to \mathbb{R}_+$, we let $W^\psi(x, y) = W(\psi(x), \psi(y))$.

DEFINITION 2.12. 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 $\psi_j : \mathbb{R}_+ \to \mathbb{R}_+$ 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 as follows.

DEFINITION 2.13. The *cut distance* between two integrable graphons W_1, W_2 is

$$\delta_\square(W_1, W_2) = \inf_{\psi_1,\psi_2} \sup_{U,V\subseteq\mathbb{R}_+} \left| \int_{U\times V} W_1^{\psi_1}(x, y) - W_2^{\psi_2}(x, y)\,\mathrm{d}x\,\mathrm{d}y \right|,$$

where the infimum is over all measure-preserving transformations $\psi_j : \mathbb{R}_+ \to \mathbb{R}_+$ for $j = 1, 2$ and the supremum is over Borel sets $U, V \subseteq \mathbb{R}_+$.

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 Figure 1.

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 2.14. The *empirical graphon* $W^G : [0, 1]^2 \to \{0, 1\}$ 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 2.15. The *stretched canonical graphon* $W^{G,s}\colon \mathbb{R}_+^2 \to \{0,1\}$ of a graph G is defined to be

$$W^{G,s}(x,y) = W^G\big(\|W^G\|_1^{1/2}x, \|W^G\|_1^{1/2}y\big)$$

if $x, y \in [0, \|W^G\|_1^{-1/2})$ and $W^{G,s}(x,y) = 0$ otherwise.

See Figure 1. 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 $\mathbb{E}[e(H_r)] = r^2/2$. 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 2.16. 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 [7], Theorem 28, 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 paper, 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 2.17. 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 $\{W^{G_j,s}\}_{j\in\mathbb{N}}$ has uniformly regular tails.

The main results about sequences with uniformly regular tails are that any such sequence has a further subsequence that converges in cut distance—that is, any such sequence is relatively compact in cut distance—and that any sequence that is convergent in cut distance also has uniformly regular tails (see [7], Corollary 17). 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.

2.2. *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 graphex processes. It is most convenient to express this by introducing notation for the distributions of these graphs.

DEFINITION 2.18. The *canonical sampling distribution* with parameters s and G is $\mathrm{SmplD}(G,s)(\cdot) = \Pr(\mathsf{Smpl}(G, s/\sqrt{2e(G)}) \in \cdot | G)$.

DEFINITION 2.19. Let $(\Gamma_s)_{s\in\mathbb{R}_+}$ be a graphex process generated by $\mathcal{W}$, with $\mathcal{W}$ possibly random. The *unlabeled graphex process 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 $\mathsf{Smpl}(G_j, s/\sqrt{2e(G_j)}) \xrightarrow{d} \mathcal{G}(\Gamma_s)$ as $j \to \infty$, we may now equivalently write $\mathrm{SmplD}(G_j,s) \to \mathrm{GPD}(\mathcal{W},s)$ weakly as $j \to \infty$. This has the advantages that it makes the limit object $\mathcal{W}$ explicit, it does not introduce extraneous randomness (nonrandom graphs are mapped to nonrandom probability measures), and it allows us to deal easily with cases where the graph sequence or $\mathcal{W}$ is random.

DEFINITION 2.20. Let $\mathcal{W}$ be a graphex and let $G_1, G_2, \dots$ be a sequence of graphs. We say that $G_1, G_2, \dots$ is *sampling convergent* if $\mathrm{SmplD}(G_j,s)$ converges weakly as $j \to \infty$ for every s. We say that the sequence is *sampling convergent to* $\mathcal{W}$ or *sampling convergent with limit* $\mathcal{W}$ if $\mathrm{SmplD}(G_j,s) \to \mathrm{GPD}(\mathcal{W},s)$ weakly as $j \to \infty$ for every s.

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 2.21. A *with-replacement p-sampling* $\mathsf{SmplWR}(G,p)$ of a graph G 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 replacement, 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 time-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

$$\mathrm{GPD}(W^{G,s}, r) = \Pr\left(\mathsf{SmplWR}\left(G, \frac{r}{\sqrt{2e(G)}}\right) \in \cdot | G\right).$$

This observation (essentially) originates in [37], in the context of the study of the empirical graphons of $G_1, G_2, \dots$ generated by $\mathcal{W}$ at times $s_1, s_2, \dots$, 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.

Recall that $\ell(G)$ denotes the number of loops of a graph G. Asymptotic equivalence of with and without replacement sampling translates to the following lemma.

LEMMA 2.22. *Let G be a random graph with e edges and ℓ loops, and let $p \le 1$. Then* $\mathsf{Smpl}(G, p)$ *and* $\mathsf{SmplWR}(G, p)$ *can be coupled in such a way that a.s.,*

$$\Pr\bigl(\mathsf{Smpl}(G, p) \neq \mathsf{SmplWR}(G, p) \mid G\bigr) \le 4p^3 e + 2p^2 \ell.$$

PROOF. Note that $\mathbb{E}[v(\mathsf{Smpl}(G, p)) \mid G] \le 2p^2 e + p\ell$. [23], Lemma 5.2, establishes that there exists a coupling such that, almost surely,

$$\Pr\bigl(\mathsf{Smpl}(G, p) \neq \mathsf{SmplWR}(G, p) \mid \mathsf{Smpl}(G, p), G\bigr) \le 2pv\bigl(\mathsf{Smpl}(G, p)\bigr).$$

The result follows immediately.

We note that [23], Lemma 5.2, does not explicitly treat graphs with loops, but the proof given there applies verbatim to this case. □

LEMMA 2.23. *Let $G_1, G_2, \ldots$ be a sequence of (possibly random) graphs such that a.s., G_j is finite, $e(G_j) \to \infty$, and $\ell(G_j) = O(\sqrt{e(G_j)})$ as $j \to \infty$. Then a.s. with respect to the randomness of the sequence $G_1, G_2, \ldots$, we have that* $\mathsf{Smpl}(G_j, r/\sqrt{2e(G_j)}) \xrightarrow{d} H$ *for some finite random graph H if and only if* $\mathsf{SmplWR}(G_j, r/\sqrt{2e(G_j)}) \xrightarrow{d} H$.

PROOF. The proof follows immediately from the previous lemma by setting $p = r/\sqrt{2e(G_j)}$, $e = e(G_j)$, and $\ell = \ell(G_j)$. □

2.3. *Coupling.* Much of this paper involves convergence of probability measures. We will often make use of coupling techniques in order to establish these results; see [16] for an overview. A coupling of probability measures P and P', both on the measurable space $(E, \mathcal{E})$, is a probability measure $\widehat{P}$ on $(E \times E, \sigma(\mathcal{E} \times \mathcal{E}))$ with marginals P and P'. Such a coupling $\widehat{P}$ bounds the total variation distance $\|P - P'\|_{\mathrm{TV}}$ between P and P' by

$$\|P - P'\|_{\mathrm{TV}} \le \widehat{P}(X \neq X'),$$

where X and X' are random variables on E with distributions P and P' (which we then view as functions of the two coordinates on $E \times E$). Moreover, if E is a Polish space, then there exists some coupling that saturates this bound.

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 bound $\Pr(X \neq X')$ under this algorithm.

2.4. *Distributional convergence of point processes.* 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 [14, 15] for a friendly introduction and [26], Chapter 16, for a very general treatment.

For our purposes, the main result needed to understand distributional convergence of point processes is the following theorem.

THEOREM 2.24 ([14], Theorem 11.1.VII). *Let $\xi, \xi_1, \xi_2, \dots$ be locally finite point processes on $\mathbb{R}_+^2$. Then $\xi_j \xrightarrow{d} \xi$ as $j \to \infty$ if and only if*

$$\big(\xi_j(B_1), \dots, \xi_j(B_n)\big) \xrightarrow{d} \big(\xi(B_1), \dots, \xi(B_n)\big)$$

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. There are generally consistency requirements between the counts on different test sets, and in consequence it actually suffices to check convergence on a smaller collection.

3. 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.

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 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

$$\mathrm{embed}(G)(\cdot) = \Pr\big(\mathsf{Lbl}(G) \in \cdot \mid G\big).$$

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 3.2. *Let $G_1, G_2, \ldots$ be a graph sequence with $e(G_j) \to \infty$ as $j \to \infty$. The graph sequence is sampling convergent if and only if the sequence* $\mathrm{embed}(G_1), \mathrm{embed}(G_2), \ldots$ *converges weakly, that is, if and only if the random labelings converge in distribution. Further, denoting the limiting distributions of* $\mathrm{SmplD}(G_j, r)$ *and* $\mathrm{embed}(G_j)$ *by η_r and ζ, respectively, if $H_r \sim \eta_r$ and $\xi \sim \zeta$ then $\mathsf{Lbl}_r(H_r) \stackrel{d}{=} \xi([0,r)^2 \cap \cdot)$.*

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 $U[0, r)$ i.i.d. label. Denote this restriction by $\mathsf{Lbl}(G_j)|_r$. We have just shown that $\mathsf{Lbl}(G_j)|_r \stackrel{d}{=} \mathsf{Lbl}_r(\mathsf{Smpl}(G_j, r/\sqrt{2e(G_j)}))$. [37], Lemma 4.13, shows that if $G', G'_1, G'_2, \ldots$ are unlabeled random graphs then $G'_j \stackrel{d}{\to} G'$ as $j \to \infty$ if and only if $\mathsf{Lbl}_r(G'_j) \stackrel{d}{\to} \mathsf{Lbl}_r(G')$ 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. To do so, we let $B_1, \ldots, B_n \subseteq \mathbb{R}_+^2$ be bounded Borel sets, choose r such that $B_1, \ldots, B_n \subseteq [0, r)^2$, and demand that

$$\big\{\xi(B_1), \ldots, \xi(B_n)\big\} \stackrel{d}{=} \lim_{j \to \infty} \big\{\mathsf{Lbl}(G_j)|_r(B_1), \ldots, \mathsf{Lbl}(G_j)|_r(B_n)\big\}.$$

To see that the right-hand side is well-defined (i.e., independent of the choice of r) notice that for $r < r'$, $(\mathsf{Lbl}(G_j)|_{r'})([0, r)^2 \cap \cdot) \stackrel{d}{=} \mathsf{Lbl}(G_j)|_r$. 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 distributions are counts with respect to $\lim_{j\to\infty} \mathsf{Lbl}(G_j)|_r$. By the Kolmogorov existence theorem for point processes (see [14], Theorem 9.2.X), this suffices to show that ξ exists and has a well-defined distribution.

It is immediate that $\mathsf{Lbl}(G_j) \stackrel{d}{\to} \xi$ as $j \to \infty$ because, by construction,

$$\big\{\mathsf{Lbl}(G_j)(B_1), \ldots, \mathsf{Lbl}(G_j)(B_n)\big\} \stackrel{d}{\to} \big\{\xi(B_1), \ldots, \xi(B_n)\big\} \qquad \text{as } j \to \infty,$$

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 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 $Q_r \mid \mathcal{W} = \mathrm{GPD}(\mathcal{W}, r)$.*

PROOF. Notice that $\ell(G_j) = O(\sqrt{e(G_j)})$ for any sampling convergent sequence, since otherwise the number of vertices in the random subgraph diverges. By Lemma 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 $[0, r)$, $\xi \circ (\phi \otimes \phi) \stackrel{d}{=} \xi$. In particular then, for any dyadic partitioning of $\mathbb{R}_+$ and any transposition τ of this dyadic partitioning, $\xi \circ (\tau \otimes \tau) \stackrel{d}{=} \xi$, and by [27], 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 3.2 then establishes the result. □

We now turn to establishing that the limiting graphex $\mathcal{W}$ in Lemma 3.3 is nonrandom and integrable.

The next lemma gives a tractable criterion for determining when an exchangeable point process is ergodic, that is, when $\mathcal{W}$ is nonrandom. 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 lemma is an analogue of [27], Lemma 7.35, attributed there to David Aldous.

LEMMA 3.4. *Let Γ be an exchangeable adjacency measure on $\mathbb{R}_+^2$. Then 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 2.8) immediately implies the result.

To prove the converse direction, we use the following notation from [37]. 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 (the possibly random) $\mathcal{W}$.

Suppose that Γ is not extremal. By a consistent estimation result [37], Theorem 4.8, for any sequence $s_1, s_2, \ldots$ such that $s_j \uparrow \infty$,

$$\lim_{j \to \infty} \Pr\bigl(\Gamma\bigl([0, r)^2 \cap \cdot\bigr) \in \cdot \mid \Gamma\bigl([r, s_j)^2 \cap \cdot\bigr)\bigr) = \Pr\bigl(\Gamma\bigl([0, r)^2 \cap \cdot\bigr) \in \cdot \mid \mathrm{KEG}(\mathcal{W})\bigr)$$

(i.e., $\mathcal{W}$ can be estimated from an infinite size sample). Since, by nonextremity, $\Pr(\Gamma([0,r)^2 \cap \cdot) \in \cdot \mid \mathrm{KEG}(\mathcal{W})) \neq \Pr(\Gamma([0,r)^2 \cap \cdot) \in \cdot)$, this means that there is some $r' \in \mathbb{R}_+$ such that

$$\Pr\big(\Gamma\big([0,r)^2 \cap \cdot\big) \in \cdot \mid \Gamma\big([r,r')^2 \cap \cdot\big)\big) \neq \Pr\big(\Gamma\big([0,r)^2 \cap \cdot\big) \in \cdot\big),$$

as required. □

LEMMA 3.5. *The limiting graphex $\mathcal{W}$ in Lemma 3.3 is nonrandom.*

PROOF. As in the proof of Lemma 3.3, $\ell(G_j) = O(\sqrt{e(G_j)})$ for any sampling convergent sequence. We then make use of Lemma 2.23, the asymptotic equivalence of $\mathsf{Smpl}(G_j, r/\sqrt{2e(G_j)})$ and $\mathsf{SmplWR}(G_j, r/\sqrt{2e(G_j)})$. 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 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 $\mathrm{Poi}(rv(G_j)/\sqrt{2e(G_j)})$ and $\mathrm{Poi}((r'-r)v(G_j)/\sqrt{2e(G_j)})$ 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 3.4 implies that ξ is ergodic, or, equivalently, that $\mathcal{W}$ is nonrandom. □

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 3.6. *Let $(\Gamma_s)_{s\in\mathbb{R}_+}$ be generated by $\mathcal{W} = (I, S, W)$. Then*

$$\mathbb{E}\big[e(\Gamma_s)\big] = \frac{s^2}{2}\|\mathcal{W}\|_1 \quad \textit{and} \quad \mathbb{E}\big[\ell(\Gamma_s)\big] = s\int W(x,x)\,\mathrm{d}x.$$

PROOF. Let e_s^I, e_s^S and e_s^W be the number of nonloop edges generated by the I, S and W components of the graphex, noting that the edge sets generated by the different components are disjoint.

The equation $\mathbb{E}[e_s^I] = Is^2$ is immediate from Campbell's theorem.

By [36], Theorem 5.3, $\mathbb{E}[e_s^W] = s^2 \frac{1}{2}\|W\|_1$ and $\mathbb{E}[\ell(\Gamma_s)] = s \int W(x,x)\,\mathrm{d}x$.

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 $(t_i, x_i) \in \Pi_s$ let $M(x_i)$ be the number of rays that (t_i, x_i) has due to the star component of the graphex. By viewing $M(x_i)$ as a marking of Π_s, and recalling that $M(x_i) \sim \mathrm{Poi}(sS(x_i))$, we have from Campbell's theorem that $\mathbb{E}[e_s^S] = s^2\|S\|_1$. $\square$

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 nonempty 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 $\mathbb{E}[e(H_r)] = r^2/2$ for $H_r \sim \mathrm{GPD}(\mathcal{W}, r)$.

DEFINITION 3.7. 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,

$$\frac{1}{e(G_j)} \sum_{i=1}^{v(G_j)} d_{j,i} 1\big[d_{j,i} > k\sqrt{e(G_j)}\big] < \epsilon,$$

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: a sequence that is not uniformly sampling regular is also not uniformly tail regular (see Remark 3.8 below), but for example, graph sequences that consist of only isolated edges are uniformly sampling regular but not uniformly tail regular.

REMARK 3.8. For a sequence of graphs, the sets U_j in Definition 2.17 can without loss of generality be assumed to correspond to the high degree vertices in G_j. Formulated differently, a sequence of graphs $G_1, G_2, \ldots$ has uniformly regular tails iff for each $\varepsilon > 0$ we can find an $M < \infty$ such that when vertices are ordered from highest to lowest degrees, then

$$\frac{1}{2e(G_j)} \sum_{i > M\sqrt{e(G_j)}} d_i(G_j) \le \varepsilon,$$

for all j, where $d_i(G_j)$ denotes the degree of vertex i in G_j; see [7], Remark 18. While this is a statement about the negligible contribution of the low degree tail of the degree distribution, it interestingly also implies that vertices of large degrees only have a negligible contribution; that is, it implies that the sequence $G_1, G_2, \ldots$ is uniformly sampling regular.

PROOF. For $M < k$, the degree of a vertex of degree at least $k\sqrt{e(G)}$ clearly does not change by more than a factor of $(1-M/k)$ if we remove at most $M\sqrt{e(G)}$ of its neighbors from the graph. As a consequence,

$$\begin{aligned}
&\sum_i d_i(G)1\bigl[d_i > k\sqrt{e(G)}\bigr] \\
&\quad\le \sum_i 1\bigl[d_i > k\sqrt{e(G)}\bigr]\Biggl(\frac{1}{1-M/k}\sum_{\ell > M\sqrt{e(G)}} 1\bigl[(i,\ell)\in E(G)\bigr]\Biggr) \\
&\quad\le \frac{1}{1-M/k}\sum_{\ell > M\sqrt{e(G)}} d_\ell.
\end{aligned}$$

With the help of this bound, the proof is straightforward. □

LEMMA 3.9. *Let $G_1, G_2, \ldots$ be a graph sequence with $e_j = e(G_j) \to \infty$. Then $e(\mathsf{Smpl}(G_j, r/\sqrt{2e_j}))$ is uniformly integrable for every r if and only if $G_1, G_2, \ldots$ is uniformly sampling regular.*

PROOF. Let $e_j^r \stackrel{d}{=} e(\mathsf{Smpl}(G_j, r/\sqrt{2e_j}))$. Uniform integrability is the statement that for each $\varepsilon > 0$ we can find an $M < \infty$ such that

$$\limsup_{j\to\infty} \mathbb{E}\bigl[e_j^r 1\bigl[e_j^r > M\bigr]\bigr] \le \varepsilon.$$

Let $d_{j,i}$ denote the degree of vertex i in G_j (ignoring loops, as usual), and 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. Then $e_j^r = \frac{1}{2}\sum_i D_{j,i}^r$. As we will see, the contributions to this sum that determine whether e_j^r is uniformly integrable come from the high-degree vertices in G_j, specifically from the vertices in a set of the form $H_j = \{i \in V(G_j) : d_{j,i} > k\sqrt{e_j}\}$ for a suitable $k > 1$.

To show that uniform sampling regularity is necessary for uniform integrability, we observe that $D_{j,i}^r$ is given by

$$D_{j,i}^r = X_{j,i}^r B_{j,i}^r,$$

where $X_{j,i}^r$ and $B_{j,i}^r$ are independent random variables with

$$X_{j,i}^r \sim \mathrm{Bern}\Biggl(\frac{r}{\sqrt{2e_j}}\Biggr) \quad\text{and}\quad B_{j,i}^r \sim \mathrm{Bin}\Biggl(d_{j,i}, \frac{r}{\sqrt{2e_j}}\Biggr).$$

(Specifically, $B^r_{j,i}$ is the number of neighbors of vertex i that are sampled, and $X^r_{j,i}$ is the indicator function for whether i is sampled itself.) In particular, we can rewrite the sum from Definition 3.7 as

$$\frac{1}{e(G_j)}\sum_{i=1}^{v(G_j)} d_{j,i}1\big[d_{j,i} > k\sqrt{e(G_j)}\big] = \frac{2}{r^2}\sum_{i\in H_j}\mathbb{E}\big[D^r_{j,i}\big].$$

Assume for a moment that for k large and $i \in H_j$,

$$\mathbb{E}\big[D^r_{j,i}\big] \le 4\mathbb{E}\big[D^r_{j,i}1\big[D^r_{j,i} > kr/4\big]\big]. \tag{3.1}$$

This would allow us to bound our sum by

$$\begin{aligned}\frac{2}{r^2}\sum_{i\in H_j}\mathbb{E}\big[D^r_{j,i}\big] &\le \frac{8}{r^2}\sum_{i\in H_j}\mathbb{E}\big[D^r_{j,i}1\big[D^r_{j,i} > kr/4\big]\big]\\ &\le \frac{8}{r^2}\sum_{i\in V(G_j)}\mathbb{E}\big[D^r_{j,i}1\big[e^r_j > kr/4\big]\big]\\ &= \frac{8}{r^2}\mathbb{E}\big[e^r_j 1\big[e^r_j > kr/4\big]\big].\end{aligned}$$

If we assume uniform integrability, the right-hand side can be made arbitrarily small by choosing k large enough, showing that uniform sampling regularity is necessary for uniform integrability, once we establish the bound (3.1).

To prove (3.1), we observe that $\Pr(B^r_{j,i} \le \frac{1}{2}\mathbb{E}[B^r_{j,i}]) \le \exp(-\frac{1}{8}\mathbb{E}[B^r_{j,i}])$ by the multiplicative Chernoff bound. For $i \in H_j$, we have that $\mathbb{E}[B^r_{j,i}] = \frac{rd_{j,i}}{\sqrt{2e_j}} \ge \frac{kr}{\sqrt{2}}$, so for $k \ge \frac{8}{r}$, we have that $\Pr(B^r_{j,i} \le \frac{1}{2}\mathbb{E}[B^r_{j,i}]) \le \exp(-1/\sqrt{2}) \le \frac{1}{2}$. Combined with the fact that $kr/4 < \frac{1}{2}\mathbb{E}[B^r_{j,i}]$ if $i \in H_j$, this allows us to bound

$$\begin{aligned}\mathbb{E}\big[D^r_{j,i}1\big[D^r_{j,i} > kr/4\big]\big] &\ge \mathbb{E}\left[D^r_{j,i}1\left[D^r_{j,i} > \frac{1}{2}\mathbb{E}\big[B^r_{j,i}\big]\right]\right]\\ &= \frac{r}{\sqrt{2e_j}}\mathbb{E}\left[D^r_{j,i}1\left[D^r_{j,i} > \frac{1}{2}\mathbb{E}\big[B^r_{j,i}\big]\right]|X^r_{j,i} = 1\right]\\ &= \frac{r}{\sqrt{2e_j}}\mathbb{E}\left[B^r_{j,i}1\left[B^r_{j,i} > \frac{1}{2}\mathbb{E}\big[B^r_{j,i}\big]\right]\right]\\ &\ge \frac{r}{\sqrt{2e_j}}\frac{1}{2}\mathbb{E}\big[B^r_{j,i}\big]\Pr\left(B^r_{j,i} > \frac{1}{2}\mathbb{E}\big[B^r_{j,i}\big]\right)\\ &\ge \frac{r}{\sqrt{2e_j}}\frac{1}{4}\mathbb{E}\big[B^r_{j,i}\big] = \frac{1}{4}\mathbb{E}\big[D^r_{j,i}\big],\end{aligned}$$

proving (3.1), and hence the necessity of uniform sampling regularity.

To prove that uniform sampling regularity is sufficient to for uniform integrability, we will need to control various other terms, but it turns out that the contribution of the vertices in H_j is the only one that requires uniform sampling regularity. The details are tedious, and are given in the rest of this proof.

Consider first the number of isolated edges, $e_j^{r,I}$, that is, the number of edges $\{i,i'\} \in E(G_j)$ such that $D_{j,i}^r = D_{j,i'}^r = 1$. The probability that a given edge in $E(G_j)$ is an isolated edge is then bounded by r^2/e_j, and the probability that two edges $b, b' \in E(G_j)$ are both isolated is at most r^4/e_j^2, that is, the probability that all four termini are selected by the sampling, except for the case that $b = b'$, in which case we only have the upper bound r^2/e_j. As a consequence, the expectation of $(e_j^{r,I})^2$ is bounded by $e_j(e_j-1)r^4/e_j^2 + e_j r^2/e_j \le r^4 + r^2$. Thus $e_j^{r,I}$ is square integrable uniformly in j, and hence uniformly integrable.

Next, given $k > 1$, we partition the vertices of G_j into three sets:

$$\begin{aligned} H_j &= \{i \in V(G_j) : d_{j,i} > k\sqrt{e_j}\}, \\ M_j &= \{i \in V(G_j) : d_{j,i} \in [\sqrt{e_j}/k, k\sqrt{e_j}]\}, \\ L_j &= \{i \in V(G_j) : d_{j,i} < \sqrt{e_j}/k\}. \end{aligned}$$

We then partition the set of edges contributing to $e_j^r - e_j^{r,I}$ into several classes, starting with the edges which have one endpoint of degree 1 in L_j and one endpoint of degree at least 2 in M_j. Denote the number of these edges by $e_j^{L,r,1}$, and consider the expectation of $(e_j^{L,r,1})^2$. We then bound $e_j^{L,r,1}$ by $\sum_{i\in M_j}\sum_{u\in L_j} X_{iu}$, where $X_{iu} = 1[\{i,u\} \in E(G_j)]1[D_{j,i}^r \ge 2]1[D_{j,u}^r = 1]$. Observe that $\mathbb{E}[X_{iu}X_{i'u'}] \le r^4/4e_j^2$ if $i \ne i'$, because each of i, i', u, u' must be selected by the sampling. As a consequence,

$$\begin{aligned} \mathbb{E}\big[\big(e_j^{L,r,1}\big)^2\big] &\le \sum_{\substack{i,i'\in M_j \\ i\ne i' \\ u,u'\in L_j \\ \{i,u\},\{i',u'\}\in E(G_j)}} \mathbb{E}[X_{iu}X_{i'u'}] + \sum_{\substack{i\in M_j \\ u,u'\in L_j}} \mathbb{E}[X_{iu}X_{iu'}] \\ &\le \sum_{\{i,u\},\{i',u'\}\in E(G_j)} \frac{r^4}{4e_j^2} + \sum_{i\in M_j} \mathbb{E}\Big[\Big(\sum_{u\in L_j} X_{iu}\Big)^2\Big] \\ &\le \frac{r^4}{4} + \sum_{i\in M_j} \mathbb{E}\big[\big(D_{j,i}^r\big)^2\big] \\ &\le \frac{r^4}{4} + \frac{r}{\sqrt{2e_j}} \sum_{i\in M_j} \mathbb{E}\big[\big(B_{j,i}^r\big)^2\big] \end{aligned}$$

$$\leq \frac{r^4}{4} + \frac{r}{\sqrt{2e_j}} \sum_{i \in M_j} \left(\frac{r^2}{2e_j} d_{j,i}^2 + \frac{r}{\sqrt{2e_j}} d_{j,i} \right)$$

$$\leq \frac{r^4}{4} + \frac{r}{\sqrt{2e_j}} \sum_{i \in M_j} \left(\frac{r^2 k}{2\sqrt{e_j}} d_{j,i} + \frac{r}{\sqrt{2e_j}} d_{j,i} \right),$$

where in the last step we used that $d_{j,i} \leq k\sqrt{e_j}$ when $i \in M_j$. Since $\sum_{i \in M_j} d_{j,i} \leq 2e_j$, we see that for each k, the right-hand side is bounded uniformly in j, as required.

The remaining contribution to e_j^r will be bounded by

$$e_j^{H,r} + e_j^{M,r} + e_j^{L,r,\geq 2},$$

where $e_j^{H,r} = \sum_{i \in H_j} D_{j,i}^r$, $e_j^{M,r} = \sum_{i \in M_j} D_{j,i}^{M,r}$, $D_{j,i}^{M,r}$ is the degree of i of edges in subgraph of the sampled graph $\mathsf{Smpl}(G_j, r/\sqrt{2e_j})$ induced by restricting to vertices that belong to M_j in G_j, and

$$e_j^{L,r,\geq 2} = \sum_{i \in L_j} D_{j,i}^r 1[D_{j,i}^r \geq 2].$$

Let $M_j^r \subseteq M_j$ be defined by keeping each vertex in M_j independently with probability $r/\sqrt{2e_j}$. Then $e_j^{M,r} \leq |M_j^r|^2$. Observing that are at most $2k\sqrt{e_j}$ vertices in M_j, since otherwise there would be too many edges, we stochastically bound $|M_j^r|$ by $v_j^{M,r} \sim \mathrm{Bin}(2k\sqrt{e_j}, r/\sqrt{2e_j})$. Since the expectation of $(v_j^{M,r})^4$ is bounded uniformly in j, this proves that for each k, $e_j^{M,r}$ is square integrable uniformly in j, and hence uniformly integrable.

By the assumption of uniform sampling regularity and the fact that

$$\mathbb{E}[e_j^{H,r}] = \frac{r^2}{2e_j} \sum_{i=1}^{v_j} d_{j,i} 1[d_{j,i} > k\sqrt{e_j}],$$

we may uniformly force $\mathbb{E}[e_j^{H,r}]$ to be arbitrarily small by choosing k sufficiently large.

Finally, direct computation gives that

$$\mathbb{E}[D_{j,i}^r 1[D_{j,i}^r \geq 2]] = \frac{r}{\sqrt{2e_j}} (\mathbb{E}[B_{j,i}^r] - \Pr(B_{j,i}^r = 1)) \leq (r/\sqrt{2e_j})^3 d_{j,i}^2,$$

whereby

$$\mathbb{E}[e_j^{L,r,\geq 2}] \leq \left(\frac{r}{\sqrt{2}} \right)^3 \sum_{i \in L_j} \frac{d_{j,i}}{\sqrt{e_j}} \frac{d_{j,i}}{e_j} \leq \left(\frac{r}{\sqrt{2}} \right)^3 \frac{1}{k} \sum_{i \in L_j} \frac{d_{j,i}}{e_j} \leq \frac{r^3}{\sqrt{2}k},$$

where the second line has used that $d_{j,i}/\sqrt{e_j} < 1/k$ for every vertex i in L_j.

Now, for any constant $c' > 0$,

$$\begin{aligned}\mathbb{E}[e_j^r 1[e_j^r > c']] &\le \mathbb{E}[((e_j^{r,I} + e_j^{r,L,1} + e_j^{M,r})1[((e_j^{r,I} + e_j^{r,L,1} + e_j^{M,r}) > c'/2]] \\ &\quad + \mathbb{E}[((e_j^{r,I} + e_j^{r,L,1} + e_j^{M,r})1[e_j^{L,r,\ge 2} + e_j^{H,r} > c'/2]] \\ &\quad + \mathbb{E}[e_j^{L,r,\ge 2}] + \mathbb{E}[e_j^{H,r}].\end{aligned}$$

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 $\lim_{c'\to\infty} \Pr(e_j^{L,r,\ge 2} + e_j^{H,r} > c'/2) = 0$. [26], 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 $e_j^{r,I} + e_j^{r,L,1} + e_j^{M,r}$, we may choose c' (depending on k) large enough such that the second 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

$$\mathbb{E}[e_j^r 1[e_j^r > c']] < \epsilon$$

uniformly, as required. □

COROLLARY 3.10. *The limiting graphex $\mathcal{W} = (I, S, W)$ in Lemma 3.3 is integrable, with $\|\mathcal{W}\|_1 \le 1$ and $\int W(x,x)\,dx = \lim_{j\to\infty} \ell(G_j)/\sqrt{2e(G_j)}$. Further, the bound is saturated if and only if the graph sequence is uniformly sampling regular.*

PROOF. Let ξ be the limiting point process as in the proof of Lemma 3.3, let $\Gamma_r = \xi|_{[0,r]^2}$, let $\ell_r = \ell(\Gamma_r)$ and let e^r and e_j^r be defined by $e^r = e(\xi(\cdot \cap [0,r)^2))$ and $e_j^r = e(\mathsf{Lbl}(G_j)(\cdot \cap [0,r)^2))$.

Observe that $\ell(\mathsf{Smpl}(G_j, r/\sqrt{2e_j})) \sim \mathrm{Bin}(\ell(G_j), r/\sqrt{2e_j})$, and that loops in the sampled subgraph can only occur by selecting loops in the original graph. It then follows that $\mathbb{E}[\ell_r] = r \lim_{j\to\infty} \ell(G_j)/\sqrt{2e(G_j)}$ for all $r \in \mathbb{R}_+$. Comparing this expression with Lemma 3.6 establishes the claim about the diagonal part of W.

We have $\mathbb{E}[e^r] \le \lim_{j\to\infty} \mathbb{E}[e_j^r] = \frac{1}{2}r^2$ by a version of Fatou's lemma [26], Lemma 4.11. Comparing with Lemma 3.6 establishes that $\|\mathcal{W}\|_1 \le 1$.

The second claim follows from the observation that $e(\mathsf{Lbl}(G_j)(\cdot \cap [0,r)^2)) \overset{d}{=} e(\mathsf{Smpl}(G_j, r/\sqrt{2e_j}))$, Lemma 3.9, and the fact that a sequence of nonnegative random variables $X_1, X_2, \ldots$ that converges in distribution to X also satisfies $\mathbb{E}[X_j] \to \mathbb{E}[X]$ if and only if it is uniformly integrable. □

We now have the ingredients of the main result characterizing the limits of sampling convergent sequences.

THEOREM 3.11. *Let $G_1, G_2, \dots$ be a sampling convergent graph sequence such that $e(G_j) \to \infty$ as $j \to \infty$. Then the limit is a nonrandom graphex $\mathcal{W}$ such that $\|\mathcal{W}\|_1 \le 1$, in the sense that* $\mathrm{SmplD}(G_j, r) \to \mathrm{GPD}(\mathcal{W}, r)$ *weakly as $j \to \infty$ for all $r \in \mathbb{R}_+$. The bound on $\|\mathcal{W}\|_1$ is saturated if and only if the sequence is uniformly sampling regular.*

PROOF. Immediate from Lemmas 3.3 and 3.5 and Corollary 3.10. □

In some other sparse graph limit theories [7–9], only graph sequences satisfying certain constraints are subsequentially convergent. We prove a compactness result in Section 6 that has the following corollary.

THEOREM 3.12. *Every sequence of graphs $G_1, G_2, \dots$ satisfying $\ell(G_j) = O(\sqrt{e(G_j)})$ is subsequentially sampling convergent.*

PROOF. This will be immediate from Theorem 6.8. □

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 sparse graph sequences with relatively homogeneous degrees. In particular, the next result applies to sequences of bounded degree graphs, for which there is already a well developed limit theory [5]. It also applies to the random graph $G_{n,p}$ as long as $p \to 0$ and $n^2 p \to \infty$ as $n \to \infty$, or more generally, to inhomogeneous random graphs obtained by first choosing a dense random graph sequence generated by a bounded graphon and then subsampling it so that it becomes sparse, again as long as it is dense enough to guarantee that the number of edges goes to infinity a.s.

To state the theorem, we define the average degree and square average degree of a graph G as $\overline{d}(G) = \frac{1}{v(G)} \sum_i d_i(G)$ and $\overline{d^2}(G) = \frac{1}{v(G)} \sum_i (d_i(G))^2$, where $d_i(G)$ is the degree of vertex i not counting loops. We also recall that the edge density of G is defined as $\rho(G) = 2e(G)/(v(G))^2 = \overline{d}(G)/v(G)$.

THEOREM 3.13. *Let $G_1, G_2, \dots$ be a sampling convergent graph sequence with $e(G_j) \to \infty$ as $j \to \infty$. Suppose that the maximal degree of G_j is $o(\sqrt{e(G_j)})$ or, more generally, that*

$$\frac{\overline{d^2}(G)}{(\overline{d}(G))^2} \sqrt{\rho(G_j)} = o(1). \tag{3.2}$$

Then $G_1, G_2, \dots$ is sampling convergent to a graphex of the form $(1/2, 0, W)$, where the graphon W is zero except on the diagonal.

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 $D^r_{j,i}$ 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.

We first prove that the assumption (3.2) implies uniform sampling regularity. To this end, we bound

$$\begin{aligned}\frac{1}{e(G_j)} \sum_{i=1}^{v(G_j)} d_{j,i} 1\big[d_{j,i} > k\sqrt{e(G_j)}\big] &\le \frac{1}{k(e(G_j))^{3/2}} \sum_{i=1}^{v(G_j)} (d_{j,i})^2 \\ &= 2\frac{\sqrt{2}}{k} \frac{\overline{d^2}(G_j)\sqrt{\rho(G_j)}}{(\overline{d}(G_j))^2},\end{aligned}$$

from which the claim follows.

Next, we recall that

$$D^r_{j,i} \mid B_{j,i} \sim (1-p_j)\delta_0 + p_j \delta_{B_{j,i}} \qquad \text{where } B_{j,i} \sim \mathrm{Bin}(d_{j,i}, p_j),$$

so in particular

$$\Pr\big(D^r_{j,i} \ge 2\big) = p_j\big(1 - \big[(1-p_j)^{d_{j,i}} + d_{j,i}\, p_j (1-p_j)^{d_{j,i}-1}\big]\big) \le p_j^3 d_{j,i}^2,$$

using Bernoulli's inequality. Let N_j be the number of vertices with degree greater than 1 in the sampled subgraph. Then

$$\mathbb{E}[N_j] \le \sum_{i \le v_j} \Pr\big(D^r_{j,i} \ge 2\big) \le p_j^3 \sum_{i \le v_j} d_{j,i}^2 = r^3 \frac{\overline{d^2}(G)}{(\overline{d}(G))^2} \sqrt{\rho(G_j)} = o(1).$$

Markov's inequality then implies that $N_j \xrightarrow{p} 0$ as $j \to \infty$. Since r was arbitrary, this implies convergence to a graphex of the claimed form. □

As a corollary of the theorem, the limit of a sequence of preferential attachment graphs is the pure edge graphex. More generally, we have the following corollary.

COROLLARY 3.14. *Let $G_1, G_2, \ldots$ be a random sequence of simple graphs such that almost surely* (a) *the empirical degree distribution converges to a distribution with finite, positive mean*, (b) *the average degree converges to the mean of the limiting degree distribution and* (c) $\limsup_{j\to\infty} \frac{\max_i d_i(G_j)}{\sqrt{v(G_j)}} < \infty$ *and* $\lim_{j\to\infty} e(G_j) = \infty$. *Then a.s., $G_1, G_2, \ldots$ is sampling convergent to the graphex* $(1/2, 0, 0)$.

PROOF. Let P_d be the limit of the probability that a random vertex in G_j has degree d, let $\overline{d}$ be the mean of P and let $d_{j,i}$ and v_j be as in the last proof. Then

$$\lim_{j\to\infty} \frac{1}{v_j} \sum_i d_{j,i} 1[d_{j,i} \geq k] \overset{\text{a.s.}}{=} \overline{d} - \lim_{j\to\infty} \frac{1}{v_j} \sum_i d_{j,i} 1[d_{j,i} < k]$$
$$\overset{\text{a.s.}}{=} \overline{d} - \sum_{d<k} d P_d = \sum_{d\geq k} d P_d.$$

Given $\varepsilon > 0$, let k be a (possibly random) finite constant such that the right-hand side is at most $\varepsilon/2$, and let $J < \infty$ be such that for $j \geq J$,

$$\frac{1}{v_j} \sum_i d_{j,i} 1[d_{j,i} \geq k] \leq \varepsilon.$$

Defining $C_j = \frac{1}{\sqrt{v_j}} \max_i d_{j,i}$, we then have that

$$\frac{1}{v_j} \sum_i d_{j,i}^2 \leq \frac{C_j}{\sqrt{v_j}} \sum_i d_{j,i} 1[d_{j,i} \geq k] + \frac{k}{v_j} \sum_i d_{j,i} \leq \varepsilon C_j \sqrt{v_j} + k\overline{d}(G_j).$$

Using that $\rho(G_j) = \overline{d}(G_j)/v_j$, this shows that

$$\frac{\sqrt{\rho(G_j)}}{(\overline{d}(G_j))^2} \overline{d^2}(G_j) \leq \overline{d}(G_j)^{-3/2} \varepsilon C_j + \frac{k}{\sqrt{v_j \overline{d}(G_j)}}.$$

Recalling that $v_j \overline{d}(G_j) = 2e(G_j)$, we can now first take the limit superior over j and then the limit $\varepsilon \to 0$ to see that the condition (3.2) is a.s. satisfied. To complete the proof, we use that every sequence of loopless graphs $G_1, G_2, \ldots$ with $e(G_j) \to \infty$ has a convergent subsequence. □

4. Graphex processes are sampling convergent. We now turn to characterizing the sampling limits of sequences of graphs generated by a graphex process. 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 [37], although we will appeal to the technically stronger versions of [23]. We need the following (implicit) result from those papers.

LEMMA 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 $\widehat{W}_{(G_s,s)}$ be the empirical graphon of G_s stretched so that each pixel is $1/s \times 1/s$. [23], Theorem 5.1, shows that $\mathrm{GPD}(\widehat{W}_{(G,s)}, r) \to \mathrm{GPD}(\mathcal{W}, r)$ weakly almost surely. As noted earlier, $\mathrm{GPD}(\widehat{W}_{(G_s,s)}, r) = \Pr(\mathsf{SmplWR}(G_s, r/s) \in \cdot \mid G_s)$, and so the result follows from the asymptotic equivalence of with and without replacement sampling, Lemma 2.22. Indeed, for each fixed r, we have that a.s. $e(G_s)(r/s)^3 \to 0$ and $\ell(G_s)(r/s)^2 \to 0$ as $s \to \infty$ (by, e.g., Lemma 4.2 below). Lemma 2.22 then implies that conditioned on $(\Gamma_s)_{s\in\mathbb{R}_+}$, the total variation distance between the with and without replacement distributions goes to zero a.s. as $s \to \infty$. □

To drop the latent times, we will need an extension of a result of [7] relating $e(G_j)$ and s_j. It will be convenient to partition each Γ_s into three components, which correspond to the three terms in (2.1). We will use the notation of the Kallenberg representation theorem (Theorem 2.8). Let Π be the latent Poisson process used in the Kallenberg representation construction, and let Π_s be the restriction of Π to $[0, s) \times \mathbb{R}_+$. We partition Γ_s into the following three pieces, corresponding to the three terms in the representation theorem:

1. Γ_s^W: the edge induced subgraph given by restricting to edges between vertices that belong to the underlying Poisson process Π_s; 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}; 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; this is the part of the graph generated by $(I, 0, 0)$.

LEMMA 4.2. *Let $(\Gamma_s)_{s\in\mathbb{R}_+}$ be a graphex process 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^W, Γ_s^S and Γ_s^I, respectively. Then, almost surely,*

$$\lim_{s\to\infty} e_s^W/s^2 = \frac{1}{2}\|W\|_1, \qquad \lim_{s\to\infty} e_s^S/s^2 = \|S\|_1, \qquad \lim_{s\to\infty} e_s^I/s^2 = I, \quad \text{and}$$

$$\lim_{s\to\infty} \ell(\Gamma_s)/s = \int W(x, x)\,\mathrm{d}x.$$

PROOF. First, $e_s^W/s^2 \to \frac{1}{2}\|W\|_1$ a.s. by [7], Proposition 30.

The case $\|S\|_1 = 0$ is trivial. Assume $\|S\|_1 > 0$. The star component of the graphex process can be understood as assigning a $\mathrm{Poi}(s\,S(x_i))$ number of rays to each point of the underlying point process $(t_i, x_i) \in \Pi_s$, independent of everything else. By the additive property of independent Poisson distributions, we then have $e_s^S \mid \Pi_s \sim \mathrm{Poi}(s \sum_{(t_i,x_i)\in\Pi_s} S(x_i))$. Since $s \sum_{(t_i,x_i)\in\Pi_s} S(x_i) \uparrow \infty$ a.s. as $s \to \infty$, the law of large numbers implies $e_s^S/(s \sum_{(t_i,x_i)\in\Pi_s} S(x_i)) \to 1$ a.s. as $j \to \infty$. The

law of large numbers for Poisson processes gives $\sum_{(t_i,x_i)\in\Pi_s} S(x_i)/s \to \|S\|_1$ a.s. as $j \to \infty$, whereby $e_s^S/s^2 \to \|S\|_1$ a.s. as $j \to \infty$.

We have $e_s^I/s^2 \to I$ 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 (t_i, x_i) included with probability $W(x_i, x_i)$. The fact that $\lim_{s\to\infty} \ell(\Gamma_s)/s = \int W(x,x)\,\mathrm{d}x$ then follows by the law of large numbers for Poisson processes. $\square$

By the two previous lemmas, the limiting distribution of $\mathsf{Smpl}(G_s, r/s)$ is generated by $\mathcal{W}$, and (temporarily simplifying to the case $\|W\|_1 + 2\|S\|_1 + 2I = 1$) we have $s \approx \sqrt{2e(G_s)}$ when s is large. Thus, to prove our main result we would like to couple $\mathsf{Smpl}(G_s, r/s)$ and $\mathsf{Smpl}(G_s, r/\sqrt{2e(G_s)})$.

THEOREM 4.3. *Let $(\Gamma_s)_{s\in\mathbb{R}_+}$ be a graphex process generated by an integrable graphex $\mathcal{W} = (I, S, W)$ such that $\|\mathcal{W}\|_1 > 0$, and let $G_s = \mathcal{G}(\Gamma_s)$ for all $s \in \mathbb{R}_+$. Then $(G_s)_{s\in\mathbb{R}_+}$ is sampling convergent to $\mathcal{W}'$, that is,* $\mathrm{SmplD}(G_s, r) \to \mathrm{GPD}(\mathcal{W}', r)$ *weakly almost surely, where $\mathcal{W}' = (I', S', W')$ is defined by*

$$I' = I/\|\mathcal{W}\|_1, \qquad S'(x) = \big(\|\mathcal{W}\|_1\big)^{-1/2} S\big(x\|\mathcal{W}\|_1^{1/2}\big), \quad \textit{and}$$
$$W'(x,y) = W\big(x\|\mathcal{W}\|_1^{1/2}, y\|\mathcal{W}\|_1^{1/2}\big).$$

PROOF. First, for any graph G and any $q, p \in [0,1]$ such that $q < p$, there is a coupling such that

$$\Pr\big(\mathsf{Smpl}(G,p) \neq \mathsf{Smpl}(G,q)\big) \le \big(2p^2 e(G) + p\ell(G)\big)(1 - q/p).$$

Explicitly, we sample $\mathsf{Smpl}(G,p)$ as usual, and we sample $\mathsf{Smpl}(G,q)$ as $\mathsf{Smpl}(\mathsf{Smpl}(G,p), q/p)$. Then the expected number of vertices included in $\mathsf{Smpl}(G,p)$ that are not selected as candidates for $\mathsf{Smpl}(G,q)$ is

$$\begin{aligned}\mathbb{E}\big[v\big(\mathsf{Smpl}(G,p)\big)\big](1-q/p) &\le \big(2\mathbb{E}\big[e\big(\mathsf{Smpl}(G,p)\big)\big] \\ &\quad + \mathbb{E}\big[\ell\big(\mathsf{Smpl}(G,p)\big)\big]\big)(1-q/p) \\ &= \big(2p^2 e(G) + p\ell(G)\big)(1-q/p),\end{aligned}$$

and the claimed inequality follows by Markov's inequality and the observation that $\mathsf{Smpl}(G,p) = \mathsf{Smpl}(G,q)$ if every vertex of $\mathsf{Smpl}(G,p)$ is included as a candidate for $\mathsf{Smpl}(G,q)$.

Let $c = \|\mathcal{W}\|_1^{-1/2}$. Under the above coupling,

$$\Pr\left(\mathsf{Smpl}\left(G_s, \frac{r}{\sqrt{2e(G_s)}}\right) \neq \mathsf{Smpl}\left(G_s, \frac{rc}{s}\right)\right) \leq \begin{cases} \left(2r^2c^2\dfrac{e(G_s)}{s^2} + rc\dfrac{\ell(G_s)}{s}\right)\left(1 - \dfrac{s/c}{\sqrt{2e(G_s)}}\right) \\ \qquad \text{for } s/c < \sqrt{2e(G_s)}, \text{ and} \\ \left(r^2 + r\dfrac{\ell(G_s)}{\sqrt{2e(G_s)}}\right)\left(1 - \dfrac{\sqrt{2e(G_s)}}{s/c}\right) \\ \qquad \text{for } s/c \geq \sqrt{2e(G_s)}. \end{cases}$$

By Lemma 4.2, the right-hand side goes to 0 almost surely as $s \to \infty$. The theorem statement then follows by Lemma 4.1 and [37], Lemma 5.2, which implies that $\mathrm{GPD}(\mathcal{W}, rc) = \mathrm{GPD}(\mathcal{W}^c, r)$, where $\mathcal{W}^c = (c^2 I, cS(\cdot/c), W(\cdot/c, \cdot/c))$. □

COROLLARY 4.4. *For any integrable graphex $\mathcal{W}$ such that $\|\mathcal{W}\|_1 \leq 1$ 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 4.3.

Next, suppose that $0 < \|\mathcal{W}\|_1 < 1$, and as above, let $G_j = \mathcal{G}(\Gamma_{s_j})$, with Γ generated by $\mathcal{W}$, and let $S_1, S_2, \ldots$ be a sequence of stars such that $e(S_j) \to \infty$ as $j \to \infty$ and $\lim_{j\to\infty} e(G_j)/(e(G_j) + e(S_j)) = \frac{1}{2}\|\mathcal{W}\|_1$. Under the obvious coupling,

$$\lim_{j\to\infty} \mathsf{Smpl}\left(G_j \cup S_j, \frac{r}{\sqrt{e(G_j \cup S_j)}}\right) \overset{\text{a.s.}}{=} \lim_{j\to\infty} \mathsf{Smpl}\left(G_j, \frac{r}{\sqrt{e(G_j \cup S_j)}}\right),$$

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 tends to 0. By Lemma 4.2, $e(G_j)/s_j^2 \to \frac{1}{2}\|W\|_1$ a.s. as $j \to \infty$, implying that $e(G_j \cup S_j)/s_j^2 \to 1$ a.s. as $j \to \infty$. By essentially the same coupling argument used in the proof of Theorem 4.3, $\mathrm{SmplD}(G_j \cup S_j, r) \to \mathrm{GPD}(\mathcal{W}, r)$ weakly as $j \to \infty$, showing that $G_1 \cup S_1, G_2 \cup S_2, \ldots$ is sampling convergent to $\mathcal{W}$.

Next, consider a sequence $G_1, G_2, \ldots$ generated by a graphon W that is 0 except on the diagonal, and take $e(S_j) = \lceil (\ell(G_j)/\int W(x,x)\,\mathrm{d}x)^2 \rceil$. By Lemma 4.2 and the fact that $e(G_j) = 0$ a.s., we see that $e(G_j \cup S_j)/s_j^2 \to 1$ a.s., showing that $G_1 \cup S_1, G_2 \cup S_2, \ldots$ is sampling convergent to $(0, 0, W)$.

Finally, the sampling limit of $S_1, S_2, \ldots$ with $e(S_j) = j$ is $(0, 0, 0)$, completing the proof. $\square$

5. Graphon metrics and sampling distributions. In this section, we relate sampling convergence to the metric convergence of [7]. 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. Similarly, we assume the graphons are integrable, since otherwise the metrics are not defined.

LEMMA 5.1. *Let W and W' be integrable graphons with vanishing diagonals, and let $H_r^{(1)} \sim \mathrm{GPD}(W_1, r)$ and $H_r^{(2)} \sim \mathrm{GPD}(W_2, r)$. Then there is a coupling of $H_r^{(1)}$ and $H_r^{(2)}$ such that under this coupling*

$$\Pr\bigl(H_r^{(1)} \neq H_r^{(2)}\bigr) \leq \frac{1}{2} r^2 \delta_1(W_1, W_2).$$

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 (t_i, t_j) in graph $H_r^{(k)}$ if and only if $W_k(x_i, x_j) > U_{ij}$.
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 (t_i, x_i) and (t_j, x_j) in Π the probability, conditional on Π, that (t_i, t_j) is an edge in one graph and not an edge in the other is $|W_1(x_i, x_j) - W_2(x_i, x_j)|$. The expected number of edges that disagree between the two graphs is then

$$\frac{1}{2}\mathbb{E}\Biggl[\sum_{x_i, x_j \in \Pi} \bigl|W_1(x_i, x_j) - W_2(x_i, x_j)\bigr|\Biggr] = \frac{r^2}{2}\|W_1 - W_2\|_1,$$

where the expectation is computed by an application of the Slivnyak–Mecke theorem.

The graphs are equal if there are no edges that disagree, so Markov's inequality then gives

$$\Pr\bigl(H_r^{(1)} \neq H_r^{(2)}\bigr) \leq \frac{r^2}{2}\|W_1 - W_2\|_1.$$

For any measure-preserving transformation ϕ of $\mathbb{R}_+$, $\mathrm{GPD}(W \circ (\phi \otimes \phi), r) = \mathrm{GPD}(W, r)$. It then follows that

$$\Pr\bigl(H_r^{(1)} \neq H_r^{(2)}\bigr) \leq \min_{\phi_1,\phi_2} \frac{r^2}{2} \bigl\| W_1 \circ (\phi_1 \otimes \phi_1) - W_2 \circ (\phi_2 \otimes \phi_2) \bigr\|_1,$$

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 will need a translation of the corresponding result ([10], Theorem 3.7a) from the theory of dense graph convergence.

LEMMA 5.2. *Let $W_1, W_2, \ldots$ be a sequence of integrable graphons with vanishing 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 $\mathcal{W}$, 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,*

$$\lim_{j\to\infty} \Pr\bigl(H_r^{(j)} \neq H_r\bigr) = 0 \qquad a.s.$$

PROOF. Because C is compact, $C \subseteq [0, c]^2$ for some $c \in \mathbb{R}_+$. We only require a C such that $\mathrm{supp}(W_j) \subseteq C$, and hence we may assume without loss of generality that $C = [0, c]^2$.

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 $i = 1, \ldots, N_r$ sample features $x_k \overset{\text{iid}}{\sim} U[0, c]$.
3. Include each edge (k, l) independently with probability $W(x_k, x_l)$.
4. Drop the labels in $[N_r]$ and 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. [10], Theorem 3.7a, shows that if $\delta_\square(W_j, W) \to 0$ as $j \to \infty$, then for each fixed graph F,

$$\lim_{j\to\infty} \bigl|\Pr\bigl(H_r^{(j)} = F \mid N_r\bigr) - \Pr(H_r = F \mid N_r)\bigr| = 0.$$

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. $\square$

The next result extends this to the case of arbitrary cut convergent graphon sequences. The same result has recently been independently proved as [23], Theorem 3.4.

LEMMA 5.3. *Let $W_1, W_2, \ldots$ be a sequence of integrable graphons with vanishing diagonals such that $\delta_\square(W_j, W) \to 0$ a.s. as $j \to \infty$ for some integrable graphon W with vanishing diagonal. 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,*

$$\lim_{j\to\infty} \Pr\bigl(H_r^{(j)} \neq H_r\bigr) = 0.$$

PROOF. If the sequence is compactly supported then the result follows from Lemma 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 $j > j'$,

$$\Pr\bigl(H_r^{(j)} \neq H_r \mid W_j\bigr) \leq \epsilon.$$

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 5.2. We control the error introduced by restricting to the dense cores by Lemma 5.1.

Because $\delta_\square(W_j, W) \to 0$ a.s. as $j \to \infty$, we can find a sequence of measure-preserving maps $\phi_j : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\|W_j^{\phi_j} - W\|_\square \to 0$. Replacing W_j by $W_j^{\phi_j}$, we may therefore assume without loss of generality that $\|W_j - W\|_\square \to 0$. Since W is integrable, we can find a constant $M_{r,\epsilon}$ such that $\|W - W1_{[0,M_{r,\epsilon}]^2}\|_1 \leq \varepsilon r^{-2}$. Next, we observe that

$$\begin{aligned}
&\|W_j - W_j 1_{[0,M_{r,\epsilon}]^2}\|_1 \\
&\quad = \int (W_j - W_j 1_{[0,M_{r,\epsilon}]^2}) \\
&\quad = \int (W - W1_{[0,M_{r,\epsilon}]^2}) + \int (W_j - W) + \int (W_j - W) 1_{[0,M_{r,\epsilon}]^2} \\
&\quad \leq \|W - W1_{[0,M_{r,\epsilon}]^2}\| + 2\|W_j - W\|_\square,
\end{aligned}$$

showing that for j large enough, $\|W_j - W_j 1_{[0,M_{r,\epsilon}]^2}\|_1 \leq \varepsilon r^{-2}/2$.

We will construct a series of couplings of $H_r^{(j)}$ and H_r by first coupling $G_r^{(j)} \sim \mathrm{GPD}(W_j 1_{[0,M_{r,\epsilon}]^2}, r)$ and $G_r \sim \mathrm{GPD}(W 1_{[0,M_{r,\epsilon}]^2}, r)$ in such a way that $\Pr(G_r^{(j)} \neq G_r) \leq \epsilon/4$ for all sufficiently large j. To see that such couplings exists, we first note that if we define $\widetilde{W}_j = W_j 1_{[0,M_{r,\epsilon}]^2}$ and $\widetilde{W} = W 1_{[0,M_{r,\epsilon}]^2}$, then $\|\widetilde{W}_j - \widetilde{W}\|_\square \leq \|W - W_j\|_\square \to 0$ as $j \to \infty$. We can therefore use Lemma 5.2 to get a sequence of couplings of G_r and $G_r^{(j)}$ such that for j sufficiently large, $\Pr(G_r^{(j)} \neq G_r) \leq \epsilon/4$.

We now observe that given $G_r^{(j)}$, we may sample $H_r^{(j)}$ according the following scheme:

1. 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 $W_j 1_{[0,M_{r,\epsilon}]^2}$ using Π. Draw $\Pi \mid G_r^{(j)} \sim \Pr(\Pi, G_r(W_j 1_{[0,M_{r,\epsilon}]^2})) \in \cdot \mid G_r(W_j 1_{[0,M_{r,\epsilon}]^2}) = G_r^{(j)})$.

2. Generate a graph $E_r^{(j)}$ according to $W_j 1_{\mathbb{R}_+^2 \setminus ([0,M_{e,\varepsilon}]^2}$ using Π.

3. 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).

We define 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 $W_j 1_{[0,M_\varepsilon]^2}$ using Π. This makes it clear that the sampling scheme reproduces the distribution given by the Kallenberg representation construction, that is, $H_r^{(j)} \sim \mathrm{GPD}(W_j, r)$. Also note that $E_r^{(j)} \sim \mathrm{GPD}(W_j 1_{\mathbb{R}_+^2 \setminus [0,M_\varepsilon]^2}, r)$, and $E_r \sim \mathrm{GPD}(W 1_{\mathbb{R}_+^2 \setminus [0,M_\varepsilon]^2}, r)$ (marginalizing $G_r^{(j)}$ and G_r).

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

$$\begin{aligned}
\Pr\bigl(H_r^{(j)} \neq H_r\bigr) &\leq \Pr\bigl(G_r^{(j)} \neq G_r \text{ or } e\bigl(E_r^{(j)}\bigr) > 0 \text{ or } e(E_r) > 0\bigr) \\
&\leq \Pr\bigl(G_r^{(j)} \neq G_r\bigr) + \Pr\bigl(e\bigl(E_r^{(j)}\bigr) > 0\bigr) + \Pr\bigl(e(E_r) > 0\bigr) \qquad (5.1) \\
&\leq \varepsilon/4 + \Pr\bigl(e\bigl(E_r^{(j)}\bigr) > 0\bigr) + \Pr\bigl(e(E_r) > 0\bigr). \qquad (5.2)
\end{aligned}$$

By Lemma 5.1, the last two terms of (5.2) are each at most $\epsilon/2$ and $\varepsilon/4$, respectively, proving the claim. □

We now turn from the convergence of graphons to convergence of graphs.

LEMMA 5.4. *Let $G_1, G_2, \ldots$ be a sequence of graphs such that $e(G_j) \to \infty$ as $j \to \infty$. The following are equivalent*:

1. *The sequence is sampling convergent to $\mathcal{W}$.*

2. *The graphon process corresponding to the stretched empirical graphon converges to* $\mathcal{W}$, *in the sense that, for all* $r \in \mathbb{R}_+$, $\mathrm{GPD}(W^{G_j,s}, r) \to \mathrm{GPD}(\mathcal{W}, r)$ *weakly as* $j \to \infty$.

PROOF. Note that $\ell(G_j) = O(\sqrt{e(G_j)})$ is a necessary condition for convergence in either sense. If $H_{j,r} \sim \mathrm{GPD}(W^{G_j,s}, r)$ then $H_{r,j}$ may be generated by first sampling $\mathrm{Poi}(\frac{r}{\sqrt{2e(G_j)}} v(G_j))$ vertices with replacement from G_j and then returning the edge set of the vertex induced subgraph. The claim is then simply Lemma 2.23, the asymptotic equivalence of this with replacement sampling scheme and $r/\sqrt{2e(G_j)}$-sampling. □

THEOREM 5.5. *Let* $G_1, G_2, \ldots$ *be a uniformly tail regular sequence of simple graphs and let* W *be some nonrandom 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* $r \in \mathbb{R}_+$, $\mathrm{GPD}(W^{G_j,s}, r) \to \mathrm{GPD}(W, r)$ *weakly as* $j \to \infty$.

PROOF. The equivalence of (2) and (3) is a special case of Lemma 5.4.

By Lemma 5.3 the convergence in stretched cut distance implies that, almost surely,

$$\mathrm{GPD}\big(W^{G_j,s}, r\big) \to \mathrm{GPD}(W, r),$$

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 [7], Theorem 15. 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. By [7], Theorem 27, graphexes with stretched cut distance not equal to 0 generate distinct distributions. Distinct subsequential limits thus contradict the assumption of sampling convergence, and so (2) implies (1). □

REMARK 5.6. Stretched cut convergent graph sequences are always tail regular, so convergence in stretched cut distance implies sampling convergence without any need to explicitly check tail regularity.

6. Metrization. We now translate our main limit result to the language of metric convergence and give a compactness result.

Recall that a sequence of graphexes $\mathcal{W}_1, \mathcal{W}_2, \dots$ converges in GP to $\mathcal{W}$ if for all $r \in \mathbb{R}_+$, $\mathrm{GPD}(\mathcal{W}_j, r) \to \mathrm{GPD}(\mathcal{W}, r)$ weakly as $j \to \infty$. Let δ_{GP} be a pseudometric on graphexes that metrizes convergence in GP [23]. Then δ_{GP} is a proper metric on the space of equivalence classes of graphexes under the relation that identifies graphexes that generate the same probability distribution. We will slightly abuse notation in the case where $\mathcal{W}_j = (W_j, 0, 0)$ and write $\delta_{\mathrm{GP}}(W_1, W_2) = \delta_{\mathrm{GP}}(\mathcal{W}_1, \mathcal{W}_2)$.

DEFINITION 6.1. Given two finite unlabeled graphs G, H, we define $\delta_{\mathrm{GP}}(G, H) = \delta_{\mathrm{GP}}(W^{G,s}, W^{H,s}) + |1/e(G) - 1/e(H)|$.

The metric δ_{GP} on graphs metrizes sampling convergence: For sequences such that $e(G_j) \uparrow \infty$ (so the limit is a graphex), this is Lemma 5.4. For sequences such that $e(G_j) < k$ for some $k < \infty$ for all j, this is trivial because such a sequence is sampling convergent (and δ_{GP} convergent) if and only if there is some finite graph H such that, for all j sufficiently large, G_j is isomorphic to H after excluding isolated vertices. A sequence that satisfies neither condition fails to be sampling convergent and fails to be δ_{GP} convergent.

The term $|1/e(G) - 1/e(H)|$ ensures that $\delta_{\mathrm{GP}}(G, H) = 0$ only if G and H are isomorphic after removing isolated vertices; without this term we would identify complete bipartite symmetric graphs $K_{n,n}$ for all n.

For completeness, we also define a natural metric between graphs and graphexes, although we do not make explicit use of it.

DEFINITION 6.2. Given a finite unlabeled graph G and a graphex $\mathcal{W}$, we define $\delta_{\mathrm{GP}}(G, \mathcal{W}) = \delta_{\mathrm{GP}}(W^{G,s}, \mathcal{W}) + 1/e(G)$.

DEFINITION 6.3. Let $\mathscr{G}$ be the metric space of all edge sets of finite graphs equipped with δ_{GP} (identifying G and H whenever $\delta_{\mathrm{GP}}(G, H) = 0$). Also, let $\mathscr{G}_0 \subset \mathscr{G}$ be the metric space of all simple graphs in $\mathscr{G}$.

DEFINITION 6.4. Let $\mathscr{G}^*$ and $\mathscr{G}_0^*$ be the metric completions of $\mathscr{G}$ and $\mathscr{G}_0$, respectively.

Our aim is to identify $\mathscr{G}^*$ with a graphex space.

DEFINITION 6.5. Let $\mathscr{W}^k$ be the space of equivalence classes of stretched empirical graphons of k edge graphs, under the equivalence relation $\sim$ defined by $W_1 \sim W_2$ if and only if $\delta_{\mathrm{GP}}(W_1, W_2) = 0$.

Let $\mathscr{W}^\infty$ be the space of equivalence classes of graphexes $\mathcal{W}$ satisfying $\|\mathcal{W}\|_1 \le 1$, under the equivalence relation $\sim$ defined by $\mathcal{W}_1 \sim \mathcal{W}_2$ if and only if $\delta_{\mathrm{GP}}(\mathcal{W}_1, \mathcal{W}_2) = 0$.

Let $\mathscr{W} = (\mathscr{W}^\infty \times \{0\}) \cup (\bigcup_{k=1}^\infty \mathscr{W}^k \times \{1/k\})$, equipped with the metric δ_{GP} defined by $\delta_{\mathrm{GP}}((\mathcal{W}_1, p), (\mathcal{W}_2, q)) = \delta_{\mathrm{GP}}(\mathcal{W}_1, \mathcal{W}_2) + |p - q|$.

The space $\mathscr{W}$ is the natural set of limit points of sampling convergent graph sequences. Splitting the empirical graphons according to the number of edges of the corresponding graphs allows for an identification with $\mathscr{G}$.

It is also convenient to define a version of $\mathscr{W}$ that excludes loops.

DEFINITION 6.6. Let $\mathscr{W}_0 \subset \mathscr{W}$ be the subspace where the graphons have an a.e. vanishing diagonal (i.e., $W(x,x) = 0$ for almost all $x \in \mathbb{R}_+$).

The next theorem encapsulates two of our results: limits of sampling convergent sequences are graphexes, and (up to natural equivalencies) all integrable graphexes arise in this way.

THEOREM 6.7. *$\mathscr{G}^*$ and $\mathscr{G}_0^*$ are isometric to $\mathscr{W}$ and $\mathscr{W}_0$, respectively.*

PROOF. Let $G_1, G_2, \dots$ be a Cauchy sequence in $\mathscr{G}$. If $G_j = H$ for some graph H and all sufficiently large j, then we identify the sequence with $(W^{H,s}, 1/e(H))$. If $e(G_j) \to \infty$ as $j \to \infty$, Theorem 3.11 shows that the sampling convergent limit is identified with some $\mathcal{W} \in \mathscr{W}^\infty$. We then identify the sequence with $(\mathcal{W}, 0)$. We have thus defined a map from $\mathscr{G}^*$ into $\mathscr{W}$.

Suppose $G_1, G_2, \dots$ maps to $(\mathcal{W}_1, p)$ and that $H_1, H_2, \dots$ is a second Cauchy sequence that maps to $(\mathcal{W}_2, q)$. Then

$$\begin{aligned}\lim_n \delta_{\mathrm{GP}}(G_n, H_n) &= \lim_n \delta_{\mathrm{GP}}\big((W^{G_n,s}, 1/e(G_n)), (W^{H_n,s}, 1/e(H_n))\big)\\ &= \delta_{\mathrm{GP}}\big((\mathcal{W}_1, p), (\mathcal{W}_2, q)\big),\end{aligned}$$

where the first equality is by definition and the second is by Lemma 5.4 and the observation that δ_{GP} metrizes sampling convergence. The map is thus an isometry.

Finally, the map is surjective: It follows from Corollary 4.4 that for each $(\mathcal{W}, 0) \in \mathscr{W}$ there is some graph sequence with $\mathcal{W}$ as the sampling convergent limit. The analogous statement for $(W, 1/k) \in \mathscr{W}$ with $k < \infty$ is immediate from the definition of $\mathscr{W}$.

The fact that under this isometry, $\mathscr{G}_0^*$ gets mapped into $\mathscr{W}_0$ is trivial. □

THEOREM 6.8. *If $G_1, G_2, \dots$ in $\mathscr{G}$ is an infinite sequence such that $\ell(G_j) = O(\sqrt{e(G_j)})$, then it has a subsequence that is convergent in $\mathscr{G}^*$. In particular, the metric completion $\mathscr{G}_0^*$ of the space of simple graphs equipped with δ_{GP} is compact.*

PROOF. Let $G_1, G_2, \dots$ be some sequence in $\mathscr{G}$. If there is some $k \in \mathbb{N}$ such that $\sup_j e(G_j) + \ell(G_j) < k$, 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 $\ell(G_j) = O(\sqrt{e(G_j)})$ is sequentially compact. By Lemma 3.2, it is equivalent to show that the canonical embeddings of the graph sequence are sequentially

compact in the topology of weak convergence. [14], 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 $G_1, G_2, \ldots$, the canonical labelings have the property that $\mathbb{E}[\mathsf{Lbl}(G_j)([0,r]^2)] \le r^2 + r\ell(G_j)/\sqrt{2e(G_j)}$ (with equality whenever $\sqrt{2e(G_j)} > r$), from which the uniform tightness condition follows trivially. The result then follows by Prokhorov's theorem. □

7. Sampling defines exchangeable random graphs. The time parameter of a graphex process is related to p-sampling by the observation that if $G \sim \mathrm{GPD}(\mathcal{W}, s)$ then $\mathsf{Smpl}(G, p) \sim \mathrm{GPD}(\mathcal{W}, ps)$. 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 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 $s \le t$, $G_s \subseteq G_t$ in the sense that there is some subgraph of G_t that is isomorphic to G_s.

THEOREM 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 $p \in (0,1)$,*

$$\mathsf{Smpl}(G_s, p) \overset{d}{=} G_{ps},$$

then there is some (possibly random, possibly nonintegrable) almost surely nonzero 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 $\mathsf{Lbl}_s(G_s) \overset{d}{=} \Gamma_s$, where $(\Gamma_s)_{s\in\mathbb{R}_+}$ is a graphex process generated by $\mathcal{W}$.

Let $r, s \in \mathbb{R}_+$ be such that $r < s$. Then

$$\mathsf{Lbl}_s(G_s)\big([0,r)^2 \cap \cdot\big) \overset{d}{=} \mathsf{Lbl}_r\left(\mathsf{Smpl}\left(G_s, \frac{r}{s}\right)\right) \overset{d}{=} \mathsf{Lbl}_r(G_r). \tag{7.1}$$

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 $B_1, \ldots, B_n \subseteq \mathbb{R}_+^2$,

$$\big\{\xi(B_1), \ldots, \xi(B_n)\big\} \overset{d}{=} \lim_{s\to\infty} \big\{\mathsf{Lbl}_s(G_s)(B_1), \ldots, \mathsf{Lbl}_s(G_s)(B_n)\big\}.$$

Equation (7.1) makes it clear that the limiting distribution on the right-hand side is well-defined. Moreover, using the fact that the joint distribution is defined as counts of the random labeling point process, the consistency conditions necessary for $\lim_{s\to\infty}\{\mathsf{Lbl}_s(G_s)(B_1),\dots,\mathsf{Lbl}_s(G_s)(B_n)\}$ to be counts with respect to some point process can easily be seen to be satisfied. By the Kolmogorov existence theorem for point processes (see [14], 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 (7.1) and the definition of ξ it holds that, for all $r \in \mathbb{R}_+$,

$$\mathsf{Lbl}_r(G_r) \stackrel{d}{=} \xi\big([0,r)^2 \cap \cdot\big). \tag{7.2}$$

In consequence, for any measure-preserving transformation ϕ on $[0,r)$, $\xi \circ (\phi \otimes \phi) \stackrel{d}{=} \xi$. 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 $\xi \circ (\tau \otimes \tau) \stackrel{d}{=} \xi$. By [27], 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 2.8, there is some graphex $\mathcal{W}$ such that ξ is generated by $\mathcal{W}$. The proof is then completed by again invoking (7.2). □

Acknowledgments. The authors thank Daniel Roy for many helpful discussions, Svante Janson for many valuable comments and suggestions on an earlier draft and the anonymous referee for detailed and helpful comments.

REFERENCES

[1] AIROLDI, E. M., BLEI, D. M., FEINBERG, S. E. and XING, E. P. (2008). Mixed-membership stochastic blockmodels. *J. Mach. Learn. Res.* **9** 1981–2014.

[2] ALDOUS, D. J. (1981). Representations for partially exchangeable arrays of random variables. *J. Multivariate Anal.* **11** 581–598. MR0637937

[3] ALDOUS, D. J. (2009). More uses of exchangeability: Representations of complex random structures. Preprint. Available at arXiv:0909.4339.

[4] AUSTIN, T. (2008). On exchangeable random variables and the statistics of large graphs and hypergraphs. *Probab. Surv.* **5** 80–145. MR2426176

[5] BENJAMINI, I. and SCHRAMM, O. (2001). Recurrence of distributional limits of finite planar graphs. *Electron. J. Probab.* **6** 23. MR1873300

[6] BORGS, C., CHAYES, J., LOVÁSZ, L., SÓS, V. T. and VESZTERGOMBI, K. (2006). Counting graph homomorphisms. In *Topics in Discrete Mathematics. Algorithms Combin.* **26** 315–371. Springer, Berlin. MR2249277

[7] BORGS, C., CHAYES, J. T., COHN, H. and HOLDEN, N. (2017). Sparse exchangeable graphs and their limits via graphon processes. *J. Mach. Learn. Res.* **18** 210. MR3827098

[8] BORGS, C., CHAYES, J. T., COHN, H. and ZHAO, Y. (2018). An L^p theory of sparse graph convergence I: Limits, sparse random graph models, and power law distributions. *Trans. Amer. Math. Soc.* To appear.

[9] BORGS, C., CHAYES, J. T., COHN, H. and ZHAO, Y. (2018). An L^p theory of sparse graph convergence II: LD convergence, quotients and right convergence. *Ann. Probab.* **46** 337–396. MR3758733

[10] BORGS, C., CHAYES, J. T., LOVÁSZ, L., SÓS, V. T. and VESZTERGOMBI, K. (2008). Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing. *Adv. Math.* **219** 1801–1851. MR2455626

[11] BORGS, C., CHAYES, J. T., LOVÁSZ, L., SÓS, V. T. and VESZTERGOMBI, K. (2012). Convergent sequences of dense graphs II. Multiway cuts and statistical physics. *Ann. of Math.* (2) **176** 151–219. MR2925382

[12] CARON, F. and FOX, E. B. (2017). Sparse graphs using exchangeable random measures. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 1295–1366. MR3731666

[13] CRANE, H. and DEMPSEY, W. (2015). A framework for statistical network modeling. Preprint. Available at arXiv:1509.08185.

[14] DALEY, D. J. and VERE-JONES, D. (2003). *An Introduction to the Theory of Point Processes. Vol. II: General Theory and Structure*, 2nd ed. *Probability and Its Applications* (*New York*). Springer, New York. MR1950431

[15] DALEY, D. J. and VERE-JONES, D. (2003). *An Introduction to the Theory of Point Processes. Vol. I: Elementary Theory and Methods*, 2nd ed. *Probability and Its Applications* (*New York*). Springer, New York. MR1950431

[16] DEN HOLLANDER, F. (2012). Probability Theory: The Coupling Method. Available at http://websites.math.leidenuniv.nl/probability/lecturenotes/CouplingLectures.pdf.

[17] DIACONIS, P. and JANSON, S. (2008). Graph limits and exchangeable random graphs. *Rend. Mat. Appl.* (7) **28** 33–61. MR2463439

[18] GIKHMAN, I. I. and SKOROKHOD, A. V. (1969). *Introduction to the Theory of Random Processes*. Dover, Mineola, NY.

[19] HERLAU, T., SCHMIDT, M. N. and MØRUP, M. (2016). Completely random measures for modelling block-structured sparse networks. In *Advances in Neural Information Processing Systems* 29 (D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon and R. Garnett, eds.) 4260–4268. Curran Associates, Red Hook, NY.

[20] HOFF, P. D., RAFTERY, A. E. and HANDCOCK, M. S. (2002). Latent space approaches to social network analysis. *J. Amer. Statist. Assoc.* **97** 1090–1098. MR1951262

[21] HOOVER, D. N. (1979). Relations on probability spaces and arrays of random variables. Technical report, Institute of Advanced Study, Princeton, NJ.

[22] JANSON, S. (2016). Graphons and cut metric on sigma-finite measure spaces. Preprint. Available at arXiv:1608.01833.

[23] JANSON, S. (2017). On convergence for graphexes. Preprint. Available at arXiv:1702.06389.

[24] KALLENBERG, O. (1990). Exchangeable random measures in the plane. *J. Theoret. Probab.* **3** 81–136. MR1031426

[25] KALLENBERG, O. (1999). Multivariate sampling and the estimation problem for exchangeable arrays. *J. Theoret. Probab.* **12** 859–883. MR1702867

[26] KALLENBERG, O. (2002). *Foundations of Modern Probability*, 2nd ed. *Probability and Its Applications* (*New York*). Springer, New York. MR1876169

[27] KALLENBERG, O. (2005). *Probabilistic Symmetries and Invariance Principles. Probability and Its Applications* (*New York*). Springer, New York. MR2161313

[28] LLOYD, J., ORBANZ, P., GHAHRAMANI, Z. and ROY, D. M. (2012). Random function priors for exchangeable arrays with applications to graphs and relational data. In *Advances in Neural Information Processing Systems* 25 (F. Pereira, C. J. C. Burges, L. Bottou and K. Q. Weinberger, eds.) 998–1006. Curran Associates, Red Hook, NY.

[29] LOVÁSZ, L. (2012). *Large Networks and Graph Limits. American Mathematical Society Colloquium Publications* **60**. Amer. Math. Soc., Providence, RI. MR3012035

[30] LOVÁSZ, L. and SZEGEDY, B. (2006). Limits of dense graph sequences. *J. Combin. Theory Ser. B* **96** 933–957. MR2274085

[31] LOVÁSZ, L. and SZEGEDY, B. (2007). Szemerédi's lemma for the analyst. *Geom. Funct. Anal.* **17** 252–270. MR2306658

[32] MILLER, K., JORDAN, M. I. and GRIFFITHS, T. L. (2009). Nonparametric latent feature models for link prediction. In *Advances in Neural Information Processing Systems* 22 (Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams and A. Culotta, eds.) 1276–1284. Curran Associates, Red Hook, NY.

[33] NOWICKI, K. and SNIJDERS, T. A. B. (2001). Estimation and prediction for stochastic blockstructures. *J. Amer. Statist. Assoc.* **96** 1077–1087. MR1947255

[34] ORBANZ, P. and ROY, D. M. (2015). Bayesian models of graphs, arrays and other exchangeable random structures. *IEEE Trans. Pattern Anal. Mach. Intell.* **37** 437–461.

[35] TODESCHINI, A. and CARON, F. (2016). Exchangeable random measures for sparse and modular graphs with overlapping communities. Preprint. Available at arXiv:1602.02114.

[36] VEITCH, V. and ROY, D. M. (2015). The class of random graphs arising from exchangeable random measures. Preprint. Available at arXiv:1512.03099.

[37] VEITCH, V. and ROY, D. M. (2016). Sampling and estimation for (sparse) exchangeable graphs. Preprint. Available at arXiv:1611.00843.

C. BORGS
J. T. CHAYES
H. COHN
MICROSOFT RESEARCH
ONE MEMORIAL DRIVE
CAMBRIDGE, MASSACHUSETTS 02142
USA
E-MAIL: borgs@microsoft.com
jchayes@microsoft.com
cohn@microsoft.com

V. VEITCH
DEPARTMENT OF STATISTICS
COLUMBIA UNIVERSITY
1255 AMSTERDAM AVE
NEW YORK, NEW YORK 10027
USA
E-MAIL: vcv2109@columbia.edu