The Annals of Statistics
2019, Vol. 47, No. 6, 3274–3299
https://doi.org/10.1214/18-AOS1778

SAMPLING AND ESTIMATION FOR (SPARSE) EXCHANGEABLE GRAPHS[1]

BY VICTOR VEITCH AND DANIEL M. ROY

Columbia University and University of Toronto

Sparse exchangeable graphs on $\mathbb{R}_+$, and the associated graphex framework for sparse graphs, generalize exchangeable graphs on $\mathbb{N}$, and the associated graphon framework for dense graphs. We develop the graphex framework as a tool for statistical network analysis by identifying the sampling scheme that is naturally associated with the models of the framework, formalizing two natural notions of consistent estimation of the parameter (the graphex) underlying these models, and identifying general consistent estimators in each case. The sampling scheme is a modification of independent vertex sampling that throws away vertices that are isolated in the sampled subgraph. The estimators are variants of the empirical graphon estimator, which is known to be a consistent estimator for the distribution of dense exchangeable graphs; both can be understood as graph analogues to the empirical distribution in the i.i.d. sequence setting. Our results may be viewed as a generalization of consistent estimation via the empirical graphon from the dense graph regime to also include sparse graphs.

1. Introduction. This paper is concerned with foundations for the statistical analysis of real-world networks. For densely connected networks, the graphon (dense exchangeable) framework has emerged as a powerful tool for both theory and applications in network analysis; many of the models used in practice are within the remit of this framework (e.g., [1, 18, 26–28]; see Orbanz and Roy [30] for a review). However, most real-world networks are sparsely connected; that is, as one studies larger networks, one finds that they tend to exhibit only a vanishing fraction of all possible links. The graphex (sparse exchangeable) framework has been introduced as a natural generalization of the graphon framework to include the sparse graph regime [8, 11, 32]. Sampling distribution properties of models in this framework show that they admit the rich graph structure (such as small-world connectivity and power law degree distributions) found in large real-world networks [11, 32]. Further, graphex models have already been applied to practical data modeling [11, 17, 31]. Accordingly, the graphex framework has promise for practically useful modeling that is flexible enough to capture the structure of real-world data. However, fundamental statistical questions remain open.

Received August 2017; revised October 2018.
[1]Supported by U.S. Air Force Office of Scientific Research Grant FA9550-15-1-0074.
MSC2010 subject classifications. Primary 62G05; secondary 62D05, 62G09, 62M99.
Key words and phrases. Network analysis, sampling, nonparametric estimation.

The present paper develops the graphex framework as a tool for statistical network analysis by formalizing and answering two foundational questions:

1. What is the notion of sampling naturally associated with this statistical network model? and
2. How do we use an observed dataset to consistently estimate the network model's parameters?

The answers to these questions clarify both the meaning of the modeling framework and its connection to existing frameworks.

1.1. *Background.* Statistical network models are parameterized families of probability distributions over graphs. An observed network is modeled as a realization of a random graph distributed according to some element of the model with parameters in an unknown state. The underlying properties of the network are studied by inferring the likely parameters.

The graphon framework relies on probability distributions over graphs specified by the following generative model: Each vertex $i \in \mathbb{N}$ is assigned a latent feature $x_i \in \boldsymbol{X}$, drawn independently and identically from some distribution, and, given the latent features, each edge i, j is included independently with probability $W(x_i, x_j)$, where the *graphon* $W : \boldsymbol{X} \times \boldsymbol{X} \to [0, 1]$ is a map from pairs of latent features to edge inclusion probabilities. For example, in order to model a social network, we might take the latent features to be community identities, and W to encode the probability that members of different communities interact.

Graphons generate graphs that are either empty, that is, $W = 0$ almost everywhere or dense: that is, as larger graphs are sampled, a constant fraction of all possible edges are included. It is generally thought that real-world networks do not have this property. For example, as models of social networks, graphon models posit that each individual is friends with a constant fraction of all individuals in the sample, no matter how large the sample is.

The graphex framework [8, 11, 32] resolves this pathology. The basic structure of the generative models is the same: Each vertex i is associated with a latent feature $\vartheta_i \in \boldsymbol{\vartheta}$ and, given the features, each pair of vertices i, j is connected by an edge independently with probability $W(\vartheta_i, \vartheta_j)$, where $W : \boldsymbol{\vartheta} \times \boldsymbol{\vartheta} \to [0, 1]$ is again called a graphon. The difference with the dense approach is that each vertex has a real-valued label $\theta_i \in \mathbb{R}_+$ and the latent features are no longer necessarily generated independently. Instead, the latent features are generated according to a Poisson process on the measure space $(\boldsymbol{\vartheta}, \mathcal{B}_{\boldsymbol{\vartheta}}, \nu)$. If this latent feature space has finite measure, the associated models are dense graphon models [32]. By generalizing the latent feature space to allow infinite measure, graphex models are able to produce samples with realistic network structure, including sparsity [11, 32].

In detail, let $\Pi = \{(\theta_i, \vartheta_i)\}_{i \in \mathbb{N}}$ be a Poisson (point) process on $\mathbb{R}_+ \times \boldsymbol{\vartheta}$ with intensity $\Lambda \otimes \nu$, where Λ is the Lebesgue measure. Each atom of the point process is a candidate vertex of the sampled graph; the $\{\theta_i\}$ are interpreted as (real-valued)

labels, and the $\{\vartheta_i\}$ as latent features that explain the graph structure. Each pair of points (θ_i, θ_j) with $i \leq j$ is connected independently with probability $W(\vartheta_i, \vartheta_j)$. A *size-s sample*, denoted Γ_s, is defined as the collection of connected pairs (θ_i, θ_j) such that $\theta_i, \theta_j \leq s$. We interpret Γ_s as the edge set of a graph, whose vertex set is composed of those θ_i that participate in at least one edge in Γ_s.

In general, graphex models may contain additional structure beyond the graphon. The graphon W is generalized to a parameter $\mathcal{W}$, called the *graphex*, that allows additional structure corresponding to isolated edges and stars in the sampled graphs. In full generality, graphex models are defined by a certain notion of exchangeability. Informally, this is the natural requirement that the labels of $(\Gamma_s)_{s\in\mathbb{R}_+}$ carry no information about the associated graph structure. A similar property is satisfied by dense graphon models. This is the origin of the "dense exchangeable" and "sparse exchangeable" nomenclature. For brevity, we defer the explanation of exchangeability and the associated full generative model to Section 2.

In summary, $(\Gamma_s)_{s\in\mathbb{R}_+}$ is a continuously indexed sequence of graphs with vertices labeled in $\mathbb{R}_+$, and constitutes a sample generated according to a graphex. The sequence of unlabeled graphs associated to $(\Gamma_s)_{s\in\mathbb{R}_+}$ is denoted $(G_s)_{s\in\mathbb{R}_+}$, where G_s is the graph isomorphism class of Γ_s. Statistically, we model the graph structure of an observed network as a realization of G_s for some size s.

1.2. *Sampling*. Realizations $(G_s)_{s\in\mathbb{R}_+}$ from graphex models have the property that $G_s \subset G_t$ whenever $s \leq t$. This allows us to interpret an increase in the size parameter as corresponding to the collection of additional data. This raises the question: What is the data collection mechanism that is naturally associated with the graphex framework?

We idealize a data set as a random sample from some (very large) population. In the present setting, both the population and the sample are represented as graphs, and data collection corresponds to randomly sampling a subgraph from the population graph.

The first contribution of the present paper is the identification of the sampling scheme that is naturally associated with the graphex models.

DEFINITION 1.1. For $p \in [0, 1]$, a *p-sampling* of a graph G is a random subgraph of G obtained by including each vertex of G independently with probability p, keeping only those edges connecting included vertices, and then discarding all isolated vertices in the resulting induced subgraph.

In Section 3, we prove the following key property of this sampling scheme: For $s > 0$ and $r \in [0, s]$, if G'_r is an r/s-sampling of G_s then $G'_r \stackrel{d}{=} G_r$.

This result justifies the interpretation of the parameter s as a sample size, and clarifies the sense in which a change in size parameter corresponds to collecting

additional data. Namely, if G_s represents a data set that has been collected by p-sampling some large population graph, then G_r represents a data set collected by (pr/s)-sampling from the same population. By the same token, moving from a size-r sample to a size-s sample can be interpreted as increasing the sample size by collecting additional vertices for inclusion in the induced subgraph. Note that the notion of sample size here differs from traditional approaches that take the sample size to be the number of vertices; indeed, the number of vertices of G_s is random.

1.3. *Estimation.* The connection with p-sampling provides a natural statistical interpretation of graphex models. A large graph (the population) is studied through a random subgraph (the sample) collected by p-sampling the population. Graphex models provide the statistical model for the subsample. Inference of properties of the population takes the form of inference of the underlying graphex $\mathcal{W}$. This is the analogue of the classical i.i.d. model that identifies the population of interest with the distribution that generates the sample. The fundamental statistical problem is to estimate $\mathcal{W}$ from the observed sample. Estimation has been achieved in practice for some models by restricting to special forms of the graphex [11, 17, 31]. However, key questions remain: What are natural ways to formalize consistent estimation? When, and how, can estimation be achieved?

We consider two cases distinguished by whether the sample size is considered to be part of the observed sample. In the first case, the data consists of both a size-s sample G_s from a graphex model and the sample size s itself. This case is mathematically natural, although knowing the size s is not realistic for most practical data modeling scenarios. The second case takes the observation to be only the graph, that is, the observation does not include the sample size s. Removing the sample size from the observation forces us to pinpoint the properties of the generating graphex that are identifiable from graph data alone.

Consistency is the requirement that, as more data is collected, the estimator converges in probability for some suitable notion of convergence. To formalize consistent estimation, we introduce two new notions of convergence for sequences of graphexes. Namely, a sequence of graphexes $\mathcal{W}_1, \mathcal{W}_2, \ldots$ converges to a graphex $\mathcal{W}$ if the distributions generated by $\mathcal{W}_k$ converge weakly to the distribution generated by $\mathcal{W}$. The two notions of convergence correspond to whether or not the sample size s is taken to be part of a sample generated according to $\mathcal{W}$.

In both cases, the estimator we identify is a natural analogue of the empirical distribution. The estimators we identify are related to the *empirical graphon*, the natural analogue of the empirical distribution in the dense exchangeable graph setting. For the case where sizes are part of the observation, we introduce the *dilated empirical graphon* and show that it is consistent. For the case where only the graph is observed, we build on the consistency of the dilated empirical graphon to show that the empirical graphon itself is a consistent estimator of the generating graphex.

Both estimators share the basic strengths and weaknesses of the empirical distribution. The main strength is that they are always consistent, without any restriction or assumption about the form of the generating graphex. This agnosticism to the generating graphex is also the main weakness: The estimators need not be interpretable in terms of real-world quantities of interest, and, in general, cannot be expected to yield convergence rates as fast as those arising in well-specified restricted models. Like the empirical distribution, the main statistical importance of the two estimators is foundational: The estimation results establish natural notions of consistency, and show that consistent estimation in this sense is always possible.

1.4. *Organization*. In Section 2, we give formal definitions for the basic tools of the paper. The sampling result is established in Section 3. In Section 4, we identify a consistent estimator for the setting where sizes are observed. We build on these results in Section 5 to identify a consistent estimator for the setting where the sizes are not observed. Finally, we discuss connections to related work and the interpretation of our results in Section 6.

2. Preliminaries. The edge set of a graph with labels in $\mathbb{N}$ can be represented by its *adjacency matrix* A, where $A_{ij} = 1$ indicates an edge exists between vertices i and j and $A_{ij} = 0$ indicates no edge exists. As described above, we will be interested in graphs with real-valued labels for modeling sparse graphs. The analogue of the adjacency matrix for $\mathbb{R}_+$-labeled graphs is the adjacency measure.

DEFINITION 2.1. An *adjacency measure* is a purely atomic, symmetric, locally finite measure on $\mathbb{R}_+^2$ such that every atom has unit mass.

If $\xi = \sum_i \delta_{(\theta_i', \theta_i'')}$ is a finite adjacency measure, then the *associated graph* is the graph with edge set $\{(\theta_i', \theta_i'')\}$, and vertex set $\{\theta_i'\}$. The *size-s restriction* of an adjacency measure ξ is the adjacency measure $\xi_s = \xi(\cdot \cap [0, s)^2)$. Note that the graphs $(G_s)_{s\in\mathbb{R}_+}$ associated to the size-s restrictions $(\xi_s)_{s\in\mathbb{R}_+}$ of an adjacency measure ξ satisfy $G_s \subset G_t$ for all $s \leq t$. For every finite adjacency measure ξ, the *unlabeled graph associated with* ξ, denoted $\mathcal{G}(\xi)$, is the graph isomorphism class of the associated graph.

As motivated in the Introduction, we are interested in adjacency-measure-valued stochastic process $(\Gamma_s)_{s\in\mathbb{R}_+}$ such that, for some random adjacency measure Γ on $\mathbb{R}_+^2$, each Γ_s is the size-s restriction of Γ. The defining property of the graphex model is that the underlying adjacency measure Γ is jointly exchangeable.

DEFINITION 2.2 ([21]). A random measure Γ on $\mathbb{R}_+^2$ is *jointly exchangeable* if $\Gamma \circ (\phi \otimes \phi)^{-1} \stackrel{d}{=} \Gamma$ for every measure-preserving transformation $\phi : \mathbb{R}_+ \to \mathbb{R}_+$.

Joint exchangeability implies that the labels of the vertices of the graph are uninformative about the graph structure. The relevance of jointly exchangeable random measures for modeling sparse graphs was first highlighted by Caron and Fox [11].

A representation theorem for jointly exchangeable random measures on $\mathbb{R}_+^2$ was given by Kallenberg [21, 24]. This result was translated to the setting of random graphs in [8, 32] (and was earlier quoted in [11]). Writing Λ for Lebesgue measure and letting $\mu_W(\cdot) = \int_{\mathbb{R}_+} W(x, \cdot)\,\mathrm{d}x$, the representation is expressed in terms of a graphex.

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

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

We say that a graphex is *nontrivial* if $I + \|S\|_1 + \|W\|_1 > 0$.

The next result explains how every jointly exchangeable adjacency measure on $\mathbb{R}_+^2$ arises from a possibly random graphex. The integrability conditions of Definition 2.3 are necessary and sufficient to force every size-s restriction to have a finite number of edges almost surely [24, 32]. Note that integrability of S and W suffices.

The representation theorem is the following.

THEOREM 2.4 (Theorem 4.7 [32]). *Let Γ be a random adjacency measure. Then Γ is jointly exchangeable iff there exists a (possibly random) graphex $\mathcal{W} = (I, S, W)$ such that, almost surely,*

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

for some collection of independent uniformly distributed random variables $(\zeta_{\{i,j\}})$ in $[0, 1]$ and some independent unit-rate Poisson processes $\Pi = \{(\theta_j, \vartheta_j)\}$ on $\mathbb{R}_+^2$, $\Pi_i^{\mathrm{star}} = \{(\sigma_{ij}, \chi_{ij})\}_j$, for $i \in \mathbb{N}$, on $\mathbb{R}_+^2$, and $\Pi^{\mathrm{isolate}} = \{(\rho_j, \rho'_j, \eta_j)\}$ on $\mathbb{R}_+^3$.

DEFINITION 2.5. A *Kallenberg exchangeable graph* (KEG) is a jointly exchangeable adjacency measure. A KEG Γ is an *(ergodic) KEG generated by $\mathcal{W}_0$* if

Γ satisfies the representation in Theorem 2.4 for $\mathcal{W} = \mathcal{W}_0$ a.s. Writing $\mathrm{KEG}(\mathcal{W})$ for the distribution of a KEG Γ generated by a graphex $\mathcal{W}$, a *graphex model* is a family of distributions $(\mathrm{KEG}(\mathcal{W}))$, where $\mathcal{W}$ ranges over some set of graphexes.

REMARK 2.6. We will often pass without comment from Γ to the adjacency measure-valued stochastic process $(\Gamma_s)_{s\in\mathbb{R}_+}$ defined by $\Gamma_s = \Gamma(\cdot \cap [0,s)^2)$. It will often be useful in exposition to view the restrictions Γ_s as labeled graphs, so that statements such as "the number of edges in Γ_s" are sensible.

REMARK 2.7. In the Introduction, we allowed for a general latent feature space $\boldsymbol{\vartheta}$ for the vertices. The representation theorem implies that there is no loss of generality in taking $\boldsymbol{\vartheta} = \mathbb{R}_+$, although in practice it is often natural to impose some interpretable structure on $\boldsymbol{\vartheta}$. Here, we take $\boldsymbol{\vartheta} = \mathbb{R}_+$ for mathematical convenience.

Kallenberg exchangeable graphs consist of three parts, associated with each of the three parts of the graphex. The part of a Kallenberg exchangeable graph generated by the graphon W provides the interesting graph structure; this has been described in the Introduction. The part generated by S consists of stars centered at points of of the graphon component: the terms $\delta_{\theta_j,\sigma_{jk}} + \delta_{\sigma_{jk},\theta_j}$ in the Kallenberg representation theorem correspond to edges between vertices θ_j (from the graphon component) and vertices σ_{jk} that never connect to any other vertex beyond θ_j. The part of the graph generated by I consists of isolated edges; that is, edges connecting vertices that never connect to any other part of the graph. See Figure 1 for a depiction the full generative process of the Kallenberg exchangeable graph.

3. Sampling. The restriction of a finite size-s Kallenberg exchangeable graph Γ_s to $[0,r]^2$, $r < s$, is a size-r Kallenberg exchangeable graph Γ_r. In this section, we show that this restriction has a natural relation to p-sampling: an r/s-sampling of $\mathcal{G}(\Gamma_s)$ is equal in distribution to $\mathcal{G}(\Gamma_r)$. See Section 6 for a discussion of the interpretation of this result when the larger graph is considered to be a population (i.e., as $s \to \infty$).

We require the following scheme for passing from a graph to an adjacency measure by randomly labeling the vertices independently and uniformly in some range.

DEFINITION 3.1. Let G be a graph with edge set E, and let $s > 0$. A *random labeling of G into* $[0,s]$ is a random adjacency measure with the same distribution as $\mathsf{Lbl}_s(G,\{U_i\}) = \sum_{(i,j)\in E}(\delta_{(U_i,U_j)} + \delta_{(U_j,U_i)})$, where $U_i \overset{\text{i.i.d.}}{\sim} \mathrm{Uni}[0,s]$, for $i \in \mathbb{N}$. If G is an unlabeled graph, we define the random labeling of G to be a random labeling of any representative of the graph isomorphism class. Where there is no risk of confusion, we will write $\mathsf{Lbl}_s(G)$ for $\mathsf{Lbl}_s(G,\{U_i\})$ where $U_i \overset{\text{i.i.d.}}{\sim} \mathrm{Uni}[0,s]$, for $i \in \mathbb{N}$, independently of everything else.

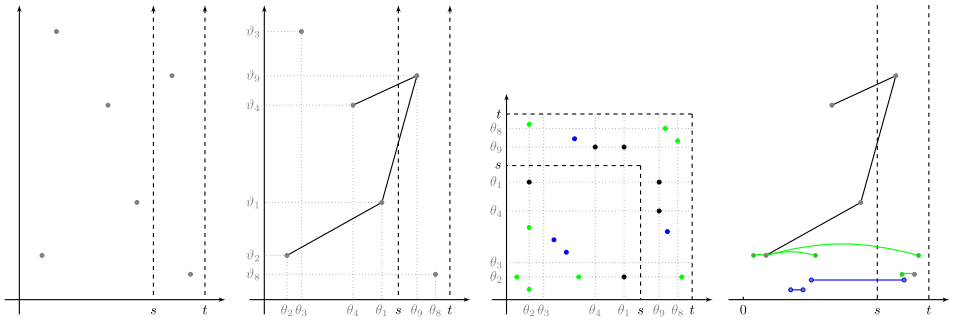

FIG. 1. *Generative process of a Kallenberg exchangeable graph with graphex* $\mathcal{W} = (I, S, W)$. *First panel*: *a realization of the latent Poisson process* $\Pi_t = \{(\theta_i, \vartheta_i) : \theta_i \le t\}$. *Second panel*: *Edges due to the graphon component* W *are sampled by connecting each distinct pair of points* $(\theta_i, \vartheta_i), (\theta_j, \vartheta_j) \in \Pi_t$ *independently with probability* $W(\vartheta_i, \vartheta_j)$. *Only a finite number of edges appear, despite there being an infinite number of points in* Π_t. *Third panel*: *The edge set* Γ_t *represented as an adjacency measure on* $[0, t]^2$. *Edges in the graphon component appear as* (*symmetric pairs of*) *black dots*; *edges from the star component* (S) *appear in green*; *isolated edges* (*from the* I *component*) *appear in blue*. *At size* s, *only the edges in* $[0, s]^2$ (*inner dashed black line*) *appear in the graph*. *The graphon component contains only a single edge at size* s. *The final panel shows the graphs corresponding to the adjacency measure at sizes* s *and* t.

The first result we need is that random labelings preserve the law of exchangeable adjacency measures. This is an intuitive consequence of exchangeability, although we rely on the representation theorem equation (2.1) for the proof.

LEMMA 3.2. *Let* $s > 0$ *and let* Γ_s *be a size-*s *Kallenberg exchangeable graph generated by* $\mathcal{W}$. *Then* $\mathsf{Lbl}_s(\mathcal{G}(\Gamma_s)) \overset{d}{=} \Gamma_s$.

PROOF. Recall the generative model for Γ_s given in equation (2.1). For simplicity of exposition, suppose that the generating graphex is $(0, 0, W)$, and the associated latent Poisson process is $\Pi_s = \{(\theta_i, \vartheta_i)\}$. Let $\{\theta'_i\}_{i\in\mathbb{N}} \overset{\text{i.i.d.}}{\sim} \text{Uni}[0, s]$, and let $\Pi'_s = \{(\theta'_i, \vartheta_i) : (\theta_i, \vartheta_i) \in \Pi_s\}$. By a property of the Poisson process, $\Pi'_s \overset{d}{=} \Pi_s$. Let Γ'_s be a size-s Kallenberg exchangeable graph generated using the same latent variables as Γ_s, but with Π'_s replacing Π_s. Then, by construction, $\Gamma'_s \overset{d}{=} \mathsf{Lbl}_s(\mathcal{G}(\Gamma_s))$. Moreover, Γ'_s is a size-s Kallenberg exchangeable graph, so $\Gamma'_s \overset{d}{=} \Gamma_s$.

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

Recall that the definition of p-sampling is given in Definition 1.1. The main sampling result is the following.

THEOREM 3.3. *Let* $\mathcal{W}$ *be a graphex*, *let* $s > 0$ *and* $r \in [0, s]$, *let* Γ *be generated by* $\mathcal{W}$ *and let* G_r *be an* r/s*-sampling of* $\mathcal{G}(\Gamma_s)$. *Then* $G_r \overset{d}{=} \mathcal{G}(\Gamma_r)$.

PROOF. Let $\xi_s = \mathsf{Lbl}_s(G_s)$. By Lemma 3.2, $\xi_s \stackrel{d}{=} \Gamma_s$. Let ξ_r be the restriction of ξ_s to $[0, r]^2$, so $\xi_r \stackrel{d}{=} \Gamma_r$. Each vertex of ξ_s has a label in $[0, r]$ independently with probability r/s; thus, $\mathcal{G}(\xi_r) \stackrel{d}{=} G_r$. □

4. Estimation with known sizes. This section describes a canonical estimator for the graphex $\mathcal{W}$ generating a Kallenberg exchangeable graph $(\Gamma_s)_{s\in\mathbb{R}_+}$ when one observes a sequence

$$(G_1, s_1), (G_2, s_2), \ldots,$$

where $s_1, s_2, \ldots \in \mathbb{R}_+$, $s_k \uparrow \infty$, and

$$G_k = \mathcal{G}(\Gamma_{s_k})$$

is the observed graph structure of the size-s_k restriction.

To formalize estimation, we require notation for the distribution corresponding to a graphex $\mathcal{W}$. Recall that $\mathrm{KEG}(\mathcal{W})$ denotes the distribution of a Kallenberg exchangeable graph Γ generated by $\mathcal{W}$.

DEFINITION 4.1. Let $\mathcal{W}$ be a graphex, let Γ be a Kallenberg exchangeable graph, and let $(\Gamma_s)_{s\in\mathbb{R}_+}$ be its restrictions to finite sizes. The *finite Kallenberg exchangeable graph distribution* with parameters $\mathcal{W}$ and s, denoted $\mathrm{KEG}(\mathcal{W}, s)$, is the distribution of Γ_s. The *finite unlabeled Kallenberg exchangeable graph distribution* with parameters $\mathcal{W}$ and s, denoted $\mathrm{uKEG}(\mathcal{W}, s)$ is the distribution of $\mathcal{G}(\Gamma_s)$.

We formalize estimation in terms of the following notion of convergence.

DEFINITION 4.2. Let $\mathcal{W}_1, \mathcal{W}_2, \ldots$ be a sequence of graphexes. Write $\mathcal{W}_n \to_{\mathrm{GP}} \mathcal{W}$ as $n \to \infty$ when, for all $s \in \mathbb{R}_+$, it holds that $\mathrm{uKEG}(\mathcal{W}_n, s) \to \mathrm{uKEG}(\mathcal{W}, s)$ weakly as $n \to \infty$.

For each k, we will construct a random *graphon* $\hat{W}_k$, measurable with respect to (G_k, s_k), such that, in probability, as $k \to \infty$,

$$(0, 0, \hat{W}_k) \to_{\mathrm{GP}} \mathcal{W}.$$

We now describe our estimator. Let $v(G)$ denote the number of vertices of a graph G. Our results build on results for estimating (dense) exchangeable arrays [23], presented here in the special case of graphs, as in [30].

DEFINITION 4.3 (Empirical graphon). Let G be some nonempty graph over the vertex set $\{1, \ldots, n\}$, and write $A \in \{0, 1\}^{n\times n}$ for its adjacency matrix. The *empirical graphon* $\tilde{W}_G : (0, 1]^2 \to \{0, 1\}$ of G is the function $\tilde{W}_G(x, y) = A_{\lceil nx\rceil, \lceil ny\rceil}$,

that is, the function produced by partitioning $(0, 1]^2$ into an $n \times n$ grid and mapping every input in the (i, j)-cell to the value 1 if edge (i, j) is included in G, and 0 otherwise. When G is empty, the empirical graphon is defined by $\tilde{W}_G = 0$. Two empirical graphons are *equivalent* when their underlying graphs are isomorphic as graphs. We define the empirical graphon of a finite unlabeled graph G as the equivalence class of the empirical graphon of an arbitrary labeling of G, with vertex set $[v(G)]$.

The estimator we identify is a *dilation* of the empirical graphon.

DEFINITION 4.4. Let G be a finite graph and let $s > 0$. The *dilated empirical graphon* of (G, s) is the function $\hat{W}_{(G,s)} : [0, v(G)/s)^2 \to \{0, 1\}$ defined by

$$\hat{W}_{(G,s)}(x, y) = \begin{cases} \tilde{W}_G\left(\dfrac{x}{v(G)/s}, \dfrac{y}{v(G)/s}\right) & x, y \le v(G)/s, \\ 0 & \text{otherwise.} \end{cases}$$

Intuitively, when the generating graphex is $\mathcal{W} = (0, 0, W)$, the sequence of dilated empirical graphon estimates are higher and higher resolution pixel pictures of the generating graphex $\mathcal{W}$. See Figure 2.

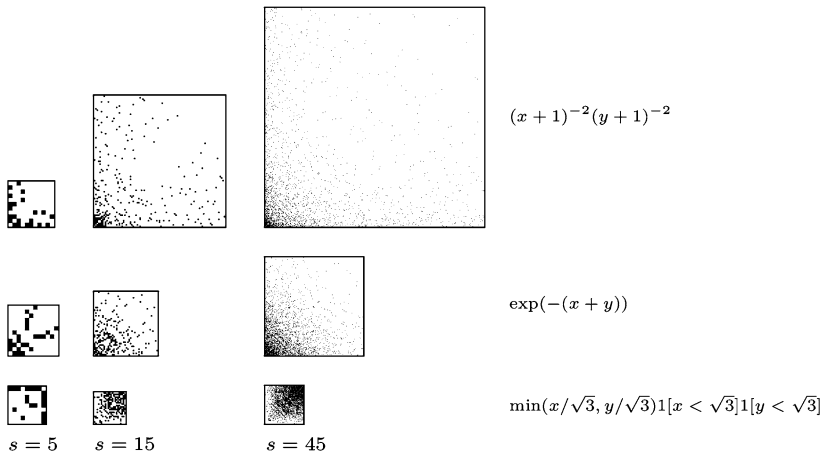

FIG. 2. *Realizations of dilated empirical graphons of KEGs generated by* $(0, 0, W)$ *for* W *given in the rightmost column, at sizes given in the bottom row. Note that the ordering of the vertices used to define the estimator is arbitrary. Here, we have suggestively ordered the vertices according to the latent values from the process simulations; with this ordering the dilated empirical graphons are approximate pixel pictures of the generating graphon where the resolution becomes finer as the size of the observed sample grows. All three graphons satisfy* $\|W\|_1 = 1$, *and thus the expected number of edges (black pixels) at each size* s *is* $\frac{1}{2}s^2$ *in each column. Note that the rate of dilation is faster for sparser graphs; as established in* [32], *the topmost KEG used for this example is sparser than the middle KEG, and the graphon generating the bottom KEG corresponds to a dense graph.*

The main result of this section is that the dilated empirical graphon of (G_k, s_k) is a consistent estimator for the $\mathcal{W}$, that is, $\hat{W}_{(G_k,s_k)} \to_{\mathrm{GP}} \mathcal{W}$ in probability as $k \to \infty$.

REMARK 4.5. Building on an earlier version of the present paper, Janson [20] has provided various technical improvements to the proofs of this section. Particularly, Janson extends convergence through the continuous sample size parameter s, and extends our result from convergence in probability to almost sure convergence.

4.1. *Convergence in distribution of randomly labeled graphs.* This section is devoted to proving that, with probability one, as $k \to \infty$, the conditional distribution of $\mathsf{Lbl}_{s_k}(G_k)$ given G_k converges weakly to $\mathrm{KEG}(\mathcal{W})$. This establishes that the distribution generated by $\mathcal{W}$ can be recovered from observed data.

THEOREM 4.6. *Let Γ be a KEG generated by a nontrivial graphex $\mathcal{W}$, let $s_1, s_2, \ldots$ be a sequence in $\mathbb{R}_+$ such that $s_k \uparrow \infty$ as $k \to \infty$ and let $G_k = \mathcal{G}(\Gamma_{s_k})$ for all k. Then $\mathrm{P}(\mathsf{Lbl}_{s_k}(G_k) \in \cdot \mid G_k) \to \mathrm{KEG}(\mathcal{W})$ weakly almost surely.*

SKETCH. For brevity, we sketch the main idea and defer the technical details to Section 2 of the Supplementary Material, where we provide a full proof [33].

Given G_k, let $\xi^k = \mathsf{Lbl}_{s_k}(G_k)$. Let $U \subset [0, r)^2$ be a rectangle with rational coordinates, for some $r \in \mathbb{R}_+$. We sketch an argument showing that $\mathrm{P}(\xi^k(U) \in \cdot \mid G_k) \to \mathrm{P}(\Gamma(U) \in \cdot)$ weakly as $k \to \infty$. It can be shown that this suffices to establish the claim by invoking results from the theory of distributional convergence of point processes.

It suffices to show that $\lim_{k\to\infty} \mathbb{E}[f(\xi^k(U)) \mid G_k] \to \mathbb{E}[f(\Gamma(U))]$ almost surely for arbitrary bounded functions f. Let Γ_{-s} be the partially labeled graph derived from Γ by forgetting the labels of all nodes with label $\theta_i < s$, and let $\mathcal{F}_{-s}$ be the σ-algebra generated by Γ_{-s}. The first key observation is that, for $s_k > r$,

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

In words: randomly relabeling the graph is the same as deleting the label information.

Next, let $U_t = U + (t, t)$ and define

$$X_s^{(r)} = \frac{1}{s-r} \int_0^{s-r} f\big(\Gamma(U_t)\big)\, \mathrm{d}t.$$

The second key observation is that, for all $s > r$ and all $t < s - r$, the joint exchangeability of Γ implies $\mathbb{E}[f(\Gamma(U_t)) \mid \mathcal{F}_{-s}] = \mathbb{E}[f(\Gamma(U)) \mid \mathcal{F}_{-s}]$, and thus also $\mathbb{E}[X_s^{(r)} \mid \mathcal{F}_{-s}] = \mathbb{E}[f(\Gamma(U)) \mid \mathcal{F}_{-s}]$.

Finally, $X_s^{(r)} \to \mathbb{E}[f(\Gamma(U))]$ almost surely by the ergodic theorem, and the required result then follows by a reverse martingale convergence argument. □

It is natural to take the sequence $G_1, G_2, \ldots$ to be all distinct graph structures that occur in $(\Gamma_s)_{s\in\mathbb{R}_+}$. To formalize this, let ξ be an adjacency measure and define $E : \mathbb{R}_+ \to \mathbb{N}$ by

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

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

DEFINITION 4.7. Let ξ be an adjacency measure. The *jump sizes* of ξ, written as $\tau(\xi)$, is the sequence $\tau_1, \tau_2, \ldots$ of jumps of E in order of appearance as s increases.

$\tau_1, \tau_2, \ldots$ are the sample sizes at which edges are added to the unlabeled graph as more data is collected.

The convergence also holds in the case where the observed samples are taken at the jumps of the Kallenberg exchangeable graph.

THEOREM 4.8. *Let Γ be a Kallenberg exchangeable graph generated by a nontrivial graphex $\mathcal{W}$, and let $\tau_1, \tau_2, \ldots$ be the jump sizes of Γ. Let $G_k = \mathcal{G}(\Gamma_{\tau_k})$ for each $k \in \mathbb{N}$. Then $\mathrm{P}(\mathsf{Lbl}_{\tau_k}(G_k) \in \cdot \mid G_k, \tau_k) \to \mathrm{KEG}(\mathcal{W})$ weakly almost surely as $k \to \infty$.*

SKETCH. For brevity, we sketch the main idea and defer the technical details to Section 2 of the Supplementary Material, where we provide a full proof [33].

As in the proof of Theorem 4.6, we consider convergence of $\mathsf{Lbl}(G_k, \tau_k)(U)$. We couple this random variable with $\mathsf{Lbl}(G_k, s_k)(U)$ for a sequence $s_1, s_2, \ldots$ chosen so that $|\tau_k - s_k| \to 0$ almost surely as $k \to \infty$. The required convergence then follows from Theorem 4.6. □

4.2. *Asymptotic equivalence of sampling schemes.* Let $\mathrm{Poi}(\lambda)$ denote the Poisson distribution with expectation λ, and let $\mathrm{Bin}(n, p)$ denote the binomial distribution with parameters n and p.

An r/s_k-sample from G_k can be drawn by restricting $\mathsf{Lbl}_{s_k}(G_k)$ to vertices with label less than r, and then dropping the labels of the induced subgraph. The results of the previous subsection then suggest that r/s_k-samples from G_k will converge in distribution to a sample from $\mathrm{uKEG}(\mathcal{W}, r)$ as $k \to \infty$. Indeed, the results of the next subsection show that this holds. Given this observation, it is natural to seek a coupling of $\mathrm{uKEG}(\hat{W}_{(G_k,s_k)}, r)$ and r/s_k-sampling of G_k. Such a coupling will allow us to use the results of Section 4.1 to show that $\hat{W}_{(G_k,s_k)}$ is a consistent estimator.

The key insight is: conditional on G_k, a graph generated according to $\mathrm{uKEG}(\hat{W}_{(G_k,s_k)}, r)$ may be viewed as a random subgraph of G_k induced by sampling $\mathrm{Poi}(\frac{r}{s_k}v(G_k))$ vertices from G_k with replacement and returning the edge set

of the vertex-induced subgraph. The correctness of this scheme can be seen as follows:

1. Let Π be the latent Poisson process used to generate a sample from $\mathrm{uKEG}(\hat{W}_{(G_k,s_k)}, r)$, as in Theorem 2.4, and let $\Pi_r = \Pi(\cdot \cap [0,r]^2)$. Because $\hat{W}_{(G_k,s_k)}$ has support $[0, v(G_k)/s_k]^2$, only atoms of Π_r in $[0,r] \times [0, v(G_k)/s_k]$ can ever participate in any edge.

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

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

Our aim is to show that the sampling scheme just described is asymptotically equivalent to r/s_k-sampling of G_k. Observe that r/s_k-sampling is the same as sampling $\mathrm{Bin}(v(G_k), r/s_k)$ vertices of G_k without replacement and returning the induced edge set. Thus, there are two main distinctions between the sampling schemes: Binomial versus Poisson number of vertices sampled, and with versus without replacement sampling.

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

LEMMA 4.9. *Let G be an almost surely finite random graph with v vertices. Let H_r be a random subgraph of G given by sampling* $\mathrm{Bin}(v, \frac{r}{s})$ *vertices without replacement and returning the induced edge set, and let N_r be a random subgraph of G given by sampling* $\mathrm{Poi}(v\frac{r}{s})$ *vertices without replacement and returning the induced edge set. Then there is a coupling such that*

$$\mathrm{P}(H_r \neq N_r \mid G) \leq \frac{r}{s} \qquad a.s.$$

PROOF. If the number of vertices chosen by each sampling scheme is equal, then we may couple the graphs by choosing the same set of vertices. Hence, it suffices to couple the number of vertices, say $K_{s,r} \sim \mathrm{Bin}(v, r/s)$ and $J_{s,r} \sim \mathrm{Poi}(r\frac{v}{s})$.

Note that $\mathbb{E}[K_{s,r} \mid G] = \mathbb{E}[J_{s,r} \mid G]$. The approximation of a sum of Bernoulli random variables by a Poisson with the same expectation as the sum is well studied: if $X_1, \ldots, X_l$ are independent random variables with $\mathrm{Bern}(p_i)$ distributions such that $\lambda = \sum_{i=1}^{\ell} p_i$ and $T \sim \mathrm{Poi}(\lambda)$ then there is a coupling [15], Section 5.3,

such that $\mathrm{P}(T \neq \sum_{i=1}^{\ell} X_i) \leq \frac{1}{\lambda} \sum_{i=1}^{\ell} p_i^2$. This implies that there is a coupling of $K_{s,r}$ and $J_{s,r}$ such that

$$\mathrm{P}(K_{s,r} \neq J_{s,r} \mid G) \leq \frac{r}{s},$$

completing the proof. □

LEMMA 4.10. *Let G be an almost surely finite random graph with v vertices. Let N_r be a random subgraph of G given by sampling* $\mathrm{Poi}(v\frac{r}{s})$ *vertices without replacement and returning the induced edge set, and let M_r be a random subgraph of G given by sampling* $\mathrm{Poi}(v\frac{r}{s})$ *vertices with replacement and returning the induced edge set. Then there is a coupling such that*

$$\mathrm{P}\big(M_r \neq N_r \mid G, v(N_r)\big) \leq v(N_r)\frac{1}{1+\exp(-\frac{r}{s})s/2r} \qquad a.s.$$

PROOF. By coupling the number of vertices selected by each sampling scheme, we can choose a coupling such that the unique vertices of N_r contain the unique vertices of M_r. Label the vertices of the graph by $1, \ldots, v(N_r)$ such that $1, \ldots, v(M_r)$ are the unique vertices selected by M_r.

Note that M_r can be sampled by including O_i copies of each vertex independently, with $O_i \sim \mathrm{Poi}(\frac{r}{s})$. The sampled graphs will be equal if $O_i \leq 1$ for all i. Let F be the number of vertices sampled more than once. Under the coupling described in the first paragraph,

$$F = \sum_i 1[O_i \geq 2] = \sum_{i \leq v(M_r)} 1[O_i \geq 2] \leq \sum_{i \leq v(N_r)} 1[O_i \geq 2].$$

Then

$$\mathbb{E}\big[F \mid v(N_r)\big] \leq v(N_r)\mathrm{P}(O_i \geq 2 \mid O_i \geq 1) \tag{4.1}$$

$$= v(N_r)\frac{1}{1+\mathrm{P}(O_i = 1)/\mathrm{P}(O_i \geq 2)} \tag{4.2}$$

$$\leq v(N_r)\frac{1}{1+\exp(-\frac{r}{s})\frac{r}{s}/\frac{r^2}{s^2}}. \tag{4.3}$$

The final line follows from Chebyshev's inequality. □

LEMMA 4.11. *Let Γ be generated by graphex $\mathcal{W}$ and fix $r \in \mathbb{R}_+$. For all $s > r \in \mathbb{R}_+$, let H_r^s be a r/s-sampling of $\mathcal{G}(\Gamma_s)$ and let $M_r^s \sim \mathrm{uKEG}(\hat{W}_{(\mathcal{G}(\Gamma_s),s)}, r)$. Then, for any sequence $s_1, s_2 \uparrow \infty$, there is a sequence of couplings such that*

$$\mathrm{P}\big(H_r^{s_k} \neq M_r^{s_k} \mid \mathcal{G}(\Gamma_{s_k})\big) \to 0$$

in probability as $k \to \infty$. Further, letting $\tau_1, \tau_2, \ldots$ be the jump sizes of Γ,

$$\mathrm{P}\big(H_r^{\tau_k} \neq M_r^{\tau_k} \mid \mathcal{G}(\Gamma_{\tau_k})\big) \to 0$$

in probability as $k \to \infty$.

PROOF. Let N_r^s be a random subgraph of $\mathcal{G}(\Gamma_s)$ given by sampling $\mathrm{Poi}(v\frac{r}{s})$ vertices without replacement and returning the induced edge set. Recall that a $M_r^s \sim \mathrm{uKEG}(\hat{W}_{(\mathcal{G}(\Gamma_s),s)}, r)$ may be sampled by sampling $\mathrm{Poi}(v\frac{r}{s})$ vertices with replacement and returning the induced edge set. Thus, by Lemma 4.10,

$$\mathrm{P}\big(H_r^s \neq M_r^s \mid \mathcal{G}(\Gamma_s)\big) \leq v\big(H_r^s\big)\frac{1}{1+\exp(-\frac{r}{s})s/2r} + \mathrm{P}\big(H_r^s \neq N_r^s \mid \mathcal{G}(\Gamma_s)\big).$$

Then, by Lemma 4.9,

$$\mathrm{P}\big(H_r^s \neq M_r^s \mid \mathcal{G}(\Gamma_s)\big) \leq v\big(H_r^s\big)\frac{1}{1+s/2r} + \frac{r}{s}.$$

Finally, note that $v(H_r^s) \leq 2e(H_r^s)$, where $e(\cdot)$ denotes number of edges. Using the fact that r/s-sampling is equivalent to random relabeling in $[0, s]$ and restricting to vertices with label $\leq r$, Theorem 4.6 shows that $e(H_r^{s_k})$ converges in distribution to a bounded random variable. This establishes the first part of the claim.

The second part follows mutatis mutandis by substituting Theorem 4.8 for Theorem 4.6. □

4.3. *Estimating* $\mathcal{W}$. We now combine our results to show that the law of the Kallenberg exchangeable graph generated by the dilated empirical graphon converges to the law of a Kallenberg exchangeable graph generated by the underlying $\mathcal{W}$.

In Section 4.1, we showed that r/s-samples from the data set are asymptotically distributed as samples from the true generating graphex. In Section 4.2, we showed that samples generated according to the dilated empirical graphon are asymptotically equivalent subgraphs drawn by r/s-sampling the data set. Together, these results imply that, asymptotically, samples generated according to the dilated empirical graphon have the same distribution as samples from the true underlying graphex; this is the required estimation result.

There is an immediate subtlety to address: Section 4.1 deals with convergence in distribution of point processes (i.e., labeled graphs), and Section 4.2 deals with convergence in distribution of unlabeled graphs. We first give the main convergence result for the point process case.

To prove this result, it is convenient to metrize weak convergence. The space of boundedly finite measures may be equipped with a metric such that it is a complete separable metric space ([14], equation (A.2.6)). Let $d_{\mathrm{p}}(\cdot,\cdot)$ be the Prokhorov metric on the space of probability measures over boundedly finite measures induced by the aforementioned metric. Then $d_{\mathrm{p}}(\cdot,\cdot)$ metrizes weak convergence.

THEOREM 4.12. *Let* Γ *be a Kallenberg exchangeable graph generated by nontrivial graphex* $\mathcal{W}$ *and let* $s_1, s_2, \ldots$ *be a (possibly random) sequence in* $\mathbb{R}_+$ *such that* $s_k \uparrow \infty$ *almost surely as* $k \to \infty$. *Let* $G_k = \mathcal{G}(\Gamma_{s_k})$ *for* $k \in \mathbb{N}$. *Suppose that either*:

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

Then

$$\mathrm{KEG}(\hat{W}_{(G_k,s_k)}) \to \mathrm{KEG}(\mathcal{W}),$$

weakly in probability, as $k \to \infty$.

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

For $r \in \mathbb{R}_+$, let $\mathrm{Q}_{s_k}(G_k, r) = \mathrm{P}(\mathsf{Lbl}_{s_k}(G_k)|_r \in \cdot \mid G_k, s_k)$, where $\mathsf{Lbl}_{s_k}(G_k)|_r = \mathsf{Lbl}_{s_k}(G_k)(\cdot \cap [0, r)^2)$. That is, $\mathrm{Q}_{s_k}(G_k, r)$ is the distribution induced by randomly labeling the observed graph and restricting to vertices with labels less than r.

By the triangle inequality, $d_{\mathrm{p}}(\mathrm{KEG}(\hat{W}_{(G_k,s_k)}, r), \mathrm{KEG}(\mathcal{W}, r))$ is bounded by $d_{\mathrm{p}}(\mathrm{KEG}(\hat{W}_{(G_k,s_k)}, r), \mathrm{Q}_{s_k}(G_k, r)) + d_{\mathrm{p}}(\mathrm{Q}_{s_k}(G_k, r), \mathrm{KEG}(\mathcal{W}, r))$.

Conditional on G_k and s_k, let H_r^k be an r/s_k-sampling of G_k and let $M_r^k \sim \mathrm{uKEG}(\hat{W}_{(G_k,s_k)}, r)$. Note that the couplings of Lemma 4.11 lift to a sequence coupling such that

$$\mathrm{P}\big(\mathsf{Lbl}_r\big(M_r^k\big) \neq \mathsf{Lbl}_r\big(H_r^k\big) \mid G_k, s_k\big) \xrightarrow{p} 0, \qquad k \to \infty.$$

This is simply by using the same random labels for both graphs.

Observe that $\mathsf{Lbl}_r(M_r^k)$ is distributed as a sample from $\mathrm{KEG}(\hat{W}_{(G_k,s_k)}, r)$ and $\mathsf{Lbl}_r(H_r^k)$ is distributed as a sample from $\mathrm{Q}_{s_k}(G_k, r)$. The relationship between couplings and total variation distance then implies

$$\big\|\mathrm{KEG}(\hat{W}_{(G_k,s_k)}, r) - \mathrm{Q}_{s_k}(G_k, r)\big\|_{\mathrm{TV}} \xrightarrow{p} 0, \qquad k \to \infty,$$

so also,

$$d_{\mathrm{p}}\big(\mathrm{KEG}(\hat{W}_{(G_k,s_k)}, r), \mathrm{Q}_{s_k}(G_k, r)\big) \xrightarrow{p} 0, \qquad k \to \infty.$$

Second, by Theorem 4.6,

$$d_{\mathrm{p}}\big(\mathrm{Q}_{s_k}(G_k, r), \mathrm{KEG}(\mathcal{W})\big) \xrightarrow{p} 0, \qquad k \to \infty.$$

Thus,

$$d_{\mathrm{p}}\big(\mathrm{KEG}(\hat{W}_{(G_k,s_k)}, r), \mathrm{KEG}(\mathcal{W}, r)\big) \xrightarrow{p} 0, \qquad k \to \infty. \tag{4.4}$$

It remains to extend convergence for each r to convergence of the entire point process. Let $\Gamma^k \sim \mathrm{KEG}(\hat{W}_{(G_k,s_k)})$ for each k, and let $\Gamma \sim \mathrm{KEG}(\mathcal{W})$. Let B be some bounded set, and fix r such that $B \subset [0, r]^2$. By equation (4.4), there is some subsequence $k_1, k_2, \ldots$ such that, almost surely, $\Gamma^{k_j}(B) \xrightarrow{d} \Gamma(B)$. By taking further subsequences, this argument extends to show that for any countable collection of bounded sets $\{B_n\}_n$ there is a further subsequence $k_1', k_2', \ldots$ of $k_1, k_2, \ldots$ such

that, almost surely, $\Gamma^{k_j}(B_n) \xrightarrow{d} \Gamma(B_n)$ for each B_n. Since Γ is a simple point process, [22], Lemma 16.16, establishes that $\mathrm{KEG}(\hat{W}_{(G'_k,s'_k)}) \to \mathrm{KEG}(\mathcal{W})$ weakly a.s. along $k'_1, k'_2, \ldots$.

The same argument shows that if we first choose any infinite sequence $N \subset \mathbb{N}$ then there is a further subsequence $N' \subset N$ such that $\mathrm{KEG}(\hat{W}_{(G'_k,s'_k)}) \to \mathrm{KEG}(\mathcal{W})$ weakly a.s. along N'. Hence, $\mathrm{KEG}(\hat{W}_{(G_k,s_k)}) \to \mathrm{KEG}(\mathcal{W})$ weakly in probability, as required.

The same proof mutatis mutandis applies for convergence along the jump sizes. The main substitution is the use of Theorem 4.8 in place of Theorem 4.6. □

The following theorem is a formalization of $\hat{W}_{(G_k,s_k)} \to_{\mathrm{GP}} \mathcal{W}$ as $k \to \infty$ in probability, and is the main estimation result for the case where the sizes are includes as part of the observation.

THEOREM 4.13. *Let Γ be a Kallenberg exchangeable graph generated by nontrivial graphex $\mathcal{W}$ and let $s_1, s_2, \ldots$ be a (possibly random) sequence in $\mathbb{R}_+$ such that $s_k \uparrow \infty$ almost surely as $k \to \infty$. Let $G_k = \mathcal{G}(\Gamma_{s_k})$ for $k \in \mathbb{N}$. Suppose that either*:

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

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

$$\hat{W}_{(G_k,s_k)} \to_{\mathrm{GP}} \mathcal{W} \quad a.s.$$

along N'.

The proof is deferred to Section 3 of the Supplementary Material [33]. The basic strategy is to use the convergence in distribution of the adjacency measures established in Theorem 4.12. The main difficulty is that the map from adjacency measures to graph structures is measurable but not continuous, so the result does not follow immediately from the continuous mapping theorem.

5. Estimation for unknown sizes. We now establish estimation results for the case where only the graph structure of the Kallenberg exchangeable graph is observed, rather than the graph structure and the sizes of the observation. We first characterize what is possible to estimate in principle—namely, the graphex up to dilation—and we then adapt the known sizes estimation results to define an estimator for this setting—namely, the empirical graphon up to dilation—and to prove consistency.

In this section, we are interested in the distinct unlabeled graphs that occur in $(\Gamma_s)_{s \in \mathbb{R}_+}$.

DEFINITION 5.1. Let ξ be an adjacency measure, with jump sizes $\tau_1, \tau_2, \ldots$. The *graph sequence of* ξ, written $\mathcal{G}(\xi)$, is the sequence, $\mathcal{G}(\xi(\cdot \cap [0, \tau_1]^2)), \mathcal{G}(\xi(\cdot \cap [0, \tau_2]^2)), \ldots$, consisting of all distinct graph unlabeled graphs in $(\xi_s)_{s \in \mathbb{R}_+}$.

The next result shows that distinct adjacency measures may give rise to the same graph sequence. For a measurable map $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ and adjacency measure ξ, define ξ^ϕ to be the adjacency measure given by $\xi^\phi(A \times B) = \xi(\phi^{-1}(A) \times \phi^{-1}(B))$, for every measurable $A, B \subseteq \mathbb{R}_+$.

LEMMA 5.2. *Let ξ be an adjacency measure and let $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ be strictly monotonic. Then $\mathcal{G}(\xi) = \mathcal{G}(\xi^\phi)$.*

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

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

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

LEMMA 5.3. *Let Γ be a Kallenberg exchangeable graph with graphex $\mathcal{W} = (I, S, W)$. Then the $\frac{1}{c}$-dilation of Γ is a Kallenberg exchangeable graph Γ' with generating graphex $\mathcal{W}' = (I', S', W')$ where $I' = c^2 I$, $S'(x) = cS(x/c)$, and $W'(x, y) = W(x/c, y/c)$.*

PROOF. For simplicity of exposition, we prove the result for $\mathcal{W} = (0, 0, W)$. The same argument extends to general graphexes by considering dilations of the latent Poisson processes used to generate the star and isolated edges components.

Recall the generative model for Γ given in equation (2.1). Let the Π be the latent Poisson processes used to generate Γ, and let $(\zeta_{\{i,j\}})$ be the i.i.d. uniform random variables.

Define $f(\Pi) = \{(\frac{1}{c}\theta, c\vartheta) : (\theta, \vartheta) \in \Pi\}$. Note that $f(\Pi)$ is a unit-rate Poisson process on $\mathbb{R}_+^2$, so that a Γ' generated by $\mathcal{W}' = (0, 0, W')$ may be generated by using latent Poisson process $f(\Pi)$ and by reusing $(\zeta_{\{i,j\}})$.

Note that Γ' includes edge $(\frac{1}{c}\theta_i, \frac{1}{c}\theta_j)$ if and only if $\zeta_{\{i,j\}} \leq W'(c\vartheta_i, c\vartheta_j)$ which occurs if and only if $\zeta_{\{i,j\}} \leq W(\vartheta_i, \vartheta_j)$ which itself occurs if and only if Γ includes edge (θ_i, θ_j). Thus, Γ' is the $\frac{1}{c}$-dilation of Γ, as was to be shown. □

Define the *c-dilation of a graphex* $\mathcal{W}$ to be the graphex $\mathcal{W}'$ defined in the statement of Lemma 5.3.

THEOREM 5.4. *Let $\mathcal{W}$ be a graphex, let $\mathcal{W}'$ be the c-dilation of $\mathcal{W}$ for some $c > 0$, and let Γ and Γ' be Kallenberg exchangeable graphs with graphexes $\mathcal{W}$ and $\mathcal{W}'$, respectively. Then $\mathcal{G}(\Gamma) \overset{d}{=} \mathcal{G}(\Gamma')$.*

PROOF. The proof follows immediately from Lemmas 5.3 and 5.2. □

As a consequence of this result, when the observed data is a graph sequence—that is, the size s is unknown—then the dilation of the generating graphex is not identifiable. We introduce a weaker notion of convergence to accommodate this.

DEFINITION 5.5. Let $\mathcal{W}, \mathcal{W}_1, \mathcal{W}_2, \ldots$ be a sequence of graphexes, and let $\Gamma, \Gamma^1, \Gamma^2, \ldots$ be Kallenberg exchangeable graphs generated by each graphex. Write $\mathcal{W}_k \to_{\mathrm{GS}} \mathcal{W}$ as $k \to \infty$ when $\mathcal{G}(\Gamma^k) \xrightarrow{d} \mathcal{G}(\Gamma)$ as $k \to \infty$.

Note that this is equivalent to requiring convergence in distribution of the length-l prefixes of the graph sequences, for all $l \in \mathbb{N}$. A length-l graph sequence generated by the estimator is close in distribution to a length-l graph sequence generated by the true graphex, provided the observed graph is large enough. This perspective explains how a sequence of graphexes with support on sets of finite measure can estimate a graphex that has support on a set of infinite measure.

The following is immediate from Theorem 5.4.

COROLLARY 5.6. *Let $\mathcal{W}, \mathcal{W}_1, \mathcal{W}_2, \ldots$ be a sequence of graphexes, let $c, c_1, c_2, \ldots > 0$ and let $W^c, W_1^{c_1}, W_2^{c_2}, \ldots$ be the corresponding dilations. Then $\mathcal{W}_k \to_{\mathrm{GS}} \mathcal{W}$ as $k \to \infty$ if and only if $\mathcal{W}_k^{c_k} \to_{\mathrm{GS}} \mathcal{W}^c$ as $k \to \infty$.*

$\mathcal{W}_k \to_{\mathrm{GS}} \mathcal{W}$ as $k \to \infty$ demands less than $\mathcal{W}_k \to_{\mathrm{GP}} \mathcal{W}$ as $k \to \infty$ because in the former case we do not need to find a correct rate of dilation for the graphex. This is borne out by the next lemma.

LEMMA 5.7. *Let $\mathcal{W}, \mathcal{W}_1, \mathcal{W}_2, \ldots$ be graphexes where $\mathcal{W}$ is nontrivial and $\mathcal{W}_k \to_{\mathrm{GP}} \mathcal{W}$ as $k \to \infty$. Then $\mathcal{W}_k \to_{\mathrm{GS}} \mathcal{W}$ as $k \to \infty$.*

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

Consider the sequence $H_n^k = (\mathscr{G}(\Gamma_1^k), \mathscr{G}(\Gamma_2^k), \ldots, \mathscr{G}(\Gamma_n^k))$, where each entry is itself an a.s. finite graph sequence and entry j is a prefix of entry $j+1$. Let $\eta_n^k = \mathrm{P}(H_n^k \in \cdot)$, and let $\eta_n = \mathrm{P}((\mathscr{G}(\Gamma_1), \mathscr{G}(\Gamma_2), \ldots, \mathscr{G}(\Gamma_n)) \in \cdot)$. We are breaking up the graph sequence of the entire Kallenberg exchangeable graph into the graph sequences up to size $1, 2, \ldots$, and η_n is the joint distribution of the first n of these partial graph sequences. Our short term goal is to show that $\eta_n^k \to \eta_n$ weakly as $k \to \infty$.

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

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

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

$$K_n(G, \cdot) = \mathrm{P}\big(L_n(G) \in \cdot\big),$$

such that $\mathrm{P}(G_n^k \in \cdot)K_n = \mathbb{E}K_n(G_n^k, \cdot) = \eta_n^k$ and $\mathrm{P}(G_n \in \cdot)K_n = \mathbb{E}K_n(G_n, \cdot) = \eta_n$. By assumption, we have $\mathcal{W}_k \to_{\mathrm{GP}} \mathcal{W}$ as $k \to \infty$, whence $G_n^k \xrightarrow{d} G_n$ as $k \to \infty$. By the discreteness of the space of finite graphs and [22], Lemma 16.24, it then holds that

$$\mathrm{P}\big(G_n^k \in \cdot\big)K_n \to \mathrm{P}(G_n \in \cdot)K_n,$$

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

$$\eta_n^k \to \eta_n, \tag{5.1}$$

weakly as $k \to \infty$.

We now have that an arbitrary length prefix of the graph sequence converges in distribution, when the notion of length is given by the latent sizes. It remains to argue that this convergence holds for arbitrary prefixes in the usual sequence sense. To that end, we observe that because equation (5.1) holds for all $n \in \mathbb{N}$, by [22], Theorem 4.29, it further holds that

$$\big(\mathscr{G}\big(\Gamma_1^k\big), \mathscr{G}\big(\Gamma_2^k\big), \ldots\big) \xrightarrow{d} \big(\mathscr{G}(\Gamma_1), \mathscr{G}(\Gamma_2), \ldots\big), \qquad k \to \infty. \tag{5.2}$$

Now, observe that there exists some function f that maps prefixes of graph sequences to the associated graph sequence. That is, for every locally finite measure ξ on $\mathbb{R}_+^2$ with restrictions ξ_j to $[0, j)^2$,

$$f\big(\mathscr{G}(\xi_1), \mathscr{G}(\xi_2), \ldots\big) = \mathscr{G}(\xi).$$

Moreover, f is even continuous because every finite prefix of $\mathscr{G}(\xi)$ is determined by some finite prefix of the left-hand side. Hence, the claim follows by the continuous mapping theorem [22], Theorem 4.27, and equation (5.2). □

We now establish the main estimation result of the paper. The observed data is an increasing sequences of graphs $G_1 \subset G_2 \subset \cdots$. We consider two models for this sequence: In one model, $G_k = \mathcal{G}(\Gamma_{s_k})$ for some Kallenberg exchangeable graph Γ and increasing and diverging sequence of sizes $s_1, s_2, \ldots$. In the other model, the sequence $G_1, G_2, \ldots$ is the graph sequence $\mathscr{G}(\Gamma)$ of some Kallenberg exchangeable graph Γ.

Our estimator is the empirical graphon, $\tilde{W}_{G_k}$, defined in Definition 4.3. This choice reflects the intuition that, because the dilation of the generating graphex is not identifiable, we need not dilate the estimator. Somewhat more precisely, we view the empirical graphon as the canonical representative of the equivalence class of graphons given by equating graphons that induce the same distribution on graph sequences. The main estimation result is the following.

THEOREM 5.8. *Let Γ be a Kallenberg exchangeable graph generated by some nontrivial graphex $\mathcal{W}$ and let $G_1, G_2, \ldots$ be some sequence of graphs such that either*:

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

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

$$\tilde{W}_{G_k} \to_{\mathrm{GS}} \mathcal{W} \qquad a.s.,$$

along N'.

PROOF. We prove case (1). Case (2) follows mutatis mutandis, substituting τ_k for s_k.

Let $\hat{W}_{(G_k,s_k)}$ denote the dilated empirical graphon of G_k with observation size s_k. By Theorem 4.13, for every sequence $N \subseteq \mathbb{N}$, there is an infinite subsequence $N' \subseteq N$, such that $\hat{W}_{(G_k,s_k)} \to_{\mathrm{GP}} \mathcal{W}$ along N' a.s. By Lemma 5.7 and $\mathcal{W}$ being nontrivial, this implies that $\hat{W}_{(G_k,s_k)} \to_{\mathrm{GS}} \mathcal{W}$ along N' a.s. For every k, $\tilde{W}_{G_k}$ is some dilation of $\hat{W}_{(G_k,s_k)}$, hence the result follows by Corollary 5.6. □

6. Discussion and related work. We end by explaining some connections to related work, and remarking on the interpretation of the results developed here.

Percolated graphon models. The motivation for graphex models is that they generalize classical graphon models to allow for sparse graphs. A different approach to resolving the dense graph pathology is to sparsify samples from graphon models by randomly deleting edges [3, 4, 9, 10]. We call these percolated graphon models. This approach has been particularly impactful in the study of graph limits, and has also provided the setting for significant statistical theory.

To enforce sparsity, the probability of edge deletion must grow as the size of the graph increases. Accordingly, $G_n \subsetneq G_{n+1}$ for typical samples from percolated graphon models. That is, as the size of the sampled graphs increases, edges that are present in smaller samples are deleted. This means that network growth in these models cannot be interpreted as the collection of additional data. This is fundamental distinction with the graphex models we consider. For instance, there is no natural percolated graphon analogue of Theorem 3.3 interpreting the size parameter in terms of some sampling scheme. This considerably complicates the statistical interpretation and application of these models in the case where the network is understood as a sample from a larger population.

Sampling. We idealize observed data as a random sample from some (very large) population graph. If the population graph G_t is itself a size-t sample from a graphex process, and the observation is selected by s/t-sampling then Theorem 3.3 shows that the observation is marginally distributed as a size-s sample from a graphex process. If t is very large—as assumed for a population—then Theorem 4.13 shows that the observed data is approximately distributed as a size-s sample from a graphex process even conditional on the population. The approximation becomes exact as $t \to \infty$.

This provides an interpretation of graphex models in terms of p-sampling. However, the requirement that the population is generated by some graphex process is onerous. This requirement has been significantly weakened by Borgs et al. [5], building on an earlier version of the present paper. In brief, they study the case where the population graph G_t generated according to some arbitrary mechanism, and the observed data is collected from the population by $s/\sqrt{e(G_t)}$-sampling, where $e(G_t)$ is the number of edges of G_t. They show that there is some graphex that characterizes the distribution of the sample in the limit $e(G_t) \to \infty$. Accordingly, the graphex encapsulates the information about the population that can be learned by p-sampling. This enforces an identification of the graphex and the population. This idea has been further generalized to other sampling schemes by Orbanz [29]. See also Crane and Dempsey [13] for a general discussion of the role of sampling design in statistical network modeling.

It is not clear when realistic data collection schemes are well approximated by p-sampling, and so it may be unclear when graphex models should be used in practice. We note that the analogous sampling scheme for classical graphon (dense exchangeable) models is independent vertex sampling. It is also unclear when, if ever, independent vertex sampling is a good model for realistic data collection

procedures. Nevertheless, graphon models are in widespread use and have been successfully applied in practice. Because p-sampling is a simple modification of independent vertex sampling that allows for sparsity, it seems appropriate to use graphex models in situations where a graphon model would have been used, but where the graph is believed to be sparse. Alternatively, comparing p-sampling and independent vertex sampling reveals that the key difference between graphon models and graphex models is that the latter disallow isolated vertices; thus, it is appropriate to use graphex models in situations where isolates are excluded from the observed data. This is common in practice.

Estimation. Our results connect with the literature on (nonparametric) graphon estimation [2, 6, 7, 12, 16, 25, 34, 35]. Typically, these results allow for sparsity of the observation by making use of the percolated graphon model. The real-world problems motivating this literature are the same as the problems motivating graphon estimation in the graphex framework. However, the fact that the percolated graphon models do not admit a sampling interpretation makes the statistical interpretation of the models very different. A clear articulation of the statistical connection between the two approaches is an open problem. A particularly interesting question is whether, and how, the estimation ideas from the percolated graphon setting can be imported to the graphex setting; the proof techniques and development of the present paper are largely disjoint from the approach of the percolated graphon development.

Our estimation results are inspired by Kallenberg's development of the theory of estimation for exchangeable arrays [23]. Restricted to the graph setting (i.e., 2-dimensional arrays interpreted as adjacency matrices), and translated into modern language, that paper introduced the empirical graphon (although not named as such) and formalized consistency in a fashion analogous to our approach. Our estimation results may be seen as generalizations of Kallenberg [23] to the sparse graph regime. Additionally, our proof technique for estimator consistency in the known-sizes setting is partially inspired by Kallenberg's development, although our technical development is wholly disjoint.

Graph limits. The present paper is also closely related to the recent paper [8]. Building on the work of Caron and Fox [11], they generalize notions of graph limits from the dense graph regime to the sparse graph regime. The key idea is to map graphs to graphons, and define a notion of convergence for sequences of graphons (thereby also defining a notion of convergence for sequences of graphs). An estimator is also a mapping from graphs to graphons. In fact, the mapping that they use is nearly identical to the dilated empirical graphon. The difference is that they replace the size s by $\sqrt{e(G_s)}$.

This suggests an alternative approach to consistent estimation with unknown sizes: namely, use the known-sizes estimator of the present paper with the substitution $s \to \sqrt{e(G_s)}$. Indeed, building on an earlier version of the present paper, this approach to estimation has now been realized [5].

The connection between graph limits and estimation is also manifest in Janson [19, 20]. Janson [20] builds on an earlier version of the present paper, studying the convergence notions we introduce. In particular, that paper metrizes the convergence, and strengthens some of our technical results—notably, proving consistency almost surely for general graphexes, while the present paper establishes only convergence in probability.

Acknowledgments. The authors would like to thank Christian Borgs, Jennifer Chayes and Henry Cohn for helpful discussions.

SUPPLEMENTARY MATERIAL

Supplemental Material for "Sampling and estimation for (sparse) exchangeable graphs" (DOI: 10.1214/18-AOS1778SUPP; .pdf). Proofs of several results, including several technical lemmas.

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] AIROLDI, E. M., COSTA, T. B. and CHAN, S. H. (2013). Stochastic blockmodel approximation of a graphon: Theory and consistent estimation. In *Advances in Neural Information Processing Systems* 26 692–700.

[3] BOLLOBÁS, B., JANSON, S. and RIORDAN, O. (2007). The phase transition in inhomogeneous random graphs. *Random Structures Algorithms* **31** 3–122. MR2337396

[4] BOLLOBÁS, B. and RIORDAN, O. (2007). Metrics for sparse graphs.

[5] BORGS, C., CHAYES, J., COHN, H. and VEITCH, V. (2017). Sampling perspectives on sparse exchangeable graphs. *Ann. Probab.* To appear.

[6] BORGS, C., CHAYES, J. and SMITH, A. (2015). Private graphon estimation for sparse graphs. In *Advances in Neural Information Processing Systems* **28** (C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama and R. Garnett, eds.) 1369–1377. Curran Associates, Red Hook, NY.

[7] BORGS, C., CHAYES, J. T., COHN, H. and GANGULY, S. (2015). Consistent nonparametric estimation for heavy-tailed sparse graphs. Preprint. Available at arXiv:1508.06675.

[8] 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

[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., COHN, H. and ZHAO, Y. (2019). An L^p theory of sparse graph convergence I: Limits, sparse random graph models, and power law distributions. *Trans. Amer. Math. Soc.* **372** 3019–3062. MR3988601

[11] 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

[12] CHAN, S. and AIROLDI, E. (2014). A consistent histogram estimator for exchangeable graph models. In *Proceedings of the* 31*st International Conference on Machine Learning* (E. P. Xing and T. Jebara, eds.). *Proceedings of Machine Learning Research* **32** 208–216. PMLR, Bejing.

[13] CRANE, H. and DEMPSEY, W. (2015). A framework for statistical network modeling. Preprint. Available at arXiv:1509.08185.
[14] DALEY, J. and VERE-JONES, (2003). *An Introduction to the Theory of Point Processes*: *Volume I*: *Elementary Theory and Methods*, 2nd ed. Springer, New York.
[15] DEN HOLLANDER, F. (2012). *Probability Theory*: *The Coupling Method*.
[16] GAO, C., LU, Y. and ZHOU, H. H. (2015). Rate-optimal graphon estimation. *Ann. Statist.* **43** 2624–2652. MR3405606
[17] 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** 4260–4268.
[18] 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
[19] JANSON, S. (2016). Graphons and cut metric on sigma-finite measure spaces. Preprint. Available at arXiv:1608.01833.
[20] JANSON, S. (2017). On convergence for graphexes. Preprint. Available at arXiv:1702.06389.
[21] KALLENBERG, O. (1990). Exchangeable random measures in the plane. *J. Theoret. Probab.* **3** 81–136. MR1031426
[22] KALLENBERG, O. (1997). *Foundations of Modern Probability*. *Probability and Its Applications* (*New York*). Springer, New York. MR1464694
[23] KALLENBERG, O. (1999). Multivariate sampling and the estimation problem for exchangeable arrays. *J. Theoret. Probab.* **12** 859–883. MR1702867
[24] KALLENBERG, O. (2005). *Probabilistic Symmetries and Invariance Principles*. *Probability and Its Applications* (*New York*). Springer, New York. MR2161313
[25] LATOUCHE, P. and ROBIN, S. (2013). Variational Bayes model averaging for graphon functions and motif frequencies inference in W-graph models. Availabe at arxiv.org/abs/1310.6150.
[26] LLOYD, J. R., ORBANZ, P., GHAHRAMANI, Z. and ROY, D. M. (2012). Random function priors for exchangeable arrays. In *Advances in Neural Information Processing Systems* **25** 1007–1015.
[27] 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** 1276–1284.
[28] NOWICKI, K. and SNIJDERS, T. A. B. (2001). Estimation and prediction for stochastic blockstructures. *J. Amer. Statist. Assoc.* **96** 1077–1087. MR1947255
[29] ORBANZ, P. (2017). Subsampling large graphs and invariance in networks. Preprint. Available at arXiv:1710.04217.
[30] 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. DOI:10.1109/TPAMI.2014.2334607.
[31] TODESCHINI, A., MISCOURIDOU, X. and CARON, F. (2016). Exchangeable random measures for sparse and modular graphs with overlapping communities. Preprint. Available at arXiv:1602.02114.
[32] VEITCH, V. and ROY, D. M. (2015). The class of random graphs arising from exchangeable random measures. Preprint. Available at arXiv:1512.03099.
[33] VEITCH, V. and ROY, D. M (2019). Supplement to "Sampling and estimation for (sparse) exchangeable graphs." DOI:10.1214/18-AOS1778SUPP.
[34] WOLFE, P. J. and OLHEDE, S. C. (2013). Nonparametric graphon estimation. Preprint. Available at arXiv:1309.5936.
[35] YANG, J., OHAN, C. and AIROLDI, E. (2014). Nonparametric estimation and testing of exchangeable graph models. In *Proceedings of the Seventeenth International Conference*

on Artificial Intelligence and Statistics (S. Kaski and J. Corander, eds.). *Proceedings of Machine Learning Research* **33** 1060–1067. PMLR, Reykjavik.

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

DEPARTMENT OF STATISTICAL SCIENCES
SIDNEY SMITH HALL
100 ST GEORGE STREET
TORONTO, ONTARIO, M5S 3G3
CANADA
E-MAIL: droy@utstat.toronto.edu