ON CONVERGENCE FOR GRAPHEXES

SVANTE JANSON

Abstract. We study four different notions of convergence for graphexes, recently introduced by Borgs, Chayes, Cohn and Holden, and by Veitch and Roy. We give some properties of them and some relations between them. We also extend results by Veitch and Roy on convergence of empirical graphons.

1. Introduction

The theory of graph limits for dense graphs and the representation of such graph limits by classical graphons has developed over the last decade and has been very succesful, see e.g. the book by Lovász [14]. Here, a classical graphon is a symmetric measurable function $W : S^2 \to [0, 1]$, where S is a probability space; without loss of generality one can take $S = [0, 1]$.

There have been many different attempts to find corresponding results for sparse graphs. One recent approach has been through random graphs defined by certain discrete exchangeable random measures on $\mathbb{R}_+^2$. Exchangeable random measures on $\mathbb{R}_+^2$ were characterized by Kallenberg [11, 13], and his general construction of such measures can be interpreted as a construction of random graphs. (Note that the classical theory of graph limits and graphons on $[0, 1]$ can be derived from the related characterizations by Aldous and Hoover of exchangeable arrays, see [1; 7], although this was not the original method or motivation.) This use of Kallenberg's construction was first done by Caron and Fox [6] in a special case, and extended by Herlau, Schmidt and Mørup [8], Veitch and Roy [15] and Borgs, Chayes, Cohn and Holden [3]. We will here use the version by Veitch and Roy [15], described in detail in Section 2.2 below. It uses a *graphex*, which is a triple (I, S, W), where the most interesting part is W which is a graphon, but a graphon in a new more general sense; W is defined on the infinite measure space $\mathbb{R}_+$ instead of $[0, 1]$.

More generally, graphons can be defined on any σ-finite measure space S. This was developed by Borgs, Chayes, Cohn and Holden [3]. However, one of their results is that it is possible to take $S = \mathbb{R}_+$ without loss of generality, and we will in the present paper only consider this case (following [15]).

Having defined graphons and graphexes, it is natural to define a topology on them, and thus a notion of convergence. For classical graphons, there are several quite different ways to define convergence, but they are all equivalent, see e.g. Borgs, Chayes, Lovász, Sós and Vesztergombi [4, 5]. (This important fact is closely related to the fact that the space of classical graphons, modulo equivalence, is compact.)

Date: 21 February, 2017; revised 23 March, 2022.

In the present, more general, context, there are also several possibilities, but, unforunately, they are not equivalent. It seems not yet clear which notion(s) of convergence that will turn out to be useful in applications, and it seems that several possibilities ought to be studied more. The present paper is a small contribution to this.

We consider in this paper four different notions of convergence for graphexes and graphons. Two of them, denoted $\to_{\mathsf{GP}}$ and $\to_{\mathsf{GS}}$, were defined by Veitch and Roy [16], based on convergence in distribution of the corresponding random graphs; we stress (which is implicit in [16]) that both convergences are metric, i.e., can be defined by (pseudo)metrics. The two other notions of convergence apply only to the special case of integrable graphons; they use the (pseudo)metrics $\delta_\square$ and $\delta_\square^{\mathsf{s}}$ defined by Borgs, Chayes, Cohn and Holden [3] (see also [10]). See Section 3 for detailed definitions. (The four metrics studied in the present are not the only possible ones. In particular, we do not consider the left convergence studied in [3].)

We show that for integrable graphons, convergence in $\delta_\square$ ($\delta_\square^{\mathsf{s}}$) implies convergence $\to_{\mathsf{GP}}$ ($\to_{\mathsf{GS}}$). We conjecture that the converses do not hold, but we leave that as an open problem.

Each graphex defines a random graph process $(G_s(\mathcal{W}))_{s\geqslant0}$, see Section 2.2 below. Borgs, Chayes, Cohn and Holden [3, Theorem 2.23] show that for any integrable graphon W, the empirical graphon defined by the random graph $G_s(W)$ a.s. converges to W in the metric $\delta_\square^{\mathsf{s}}$ as $s\to\infty$. Similarly, Veitch and Roy [16] show that for any graphex $\mathcal{W}$, the empirical graphon defined by $G_s(\mathcal{W})$ converges to the graphex $\mathcal{W}$ in $\to_{\mathsf{GP}}$ (after suitable stretching) and in $\to_{\mathsf{GS}}$ as $s\to\infty$; however, they prove this only for a sequence $s_k\to\infty$, and in general only with convergence of probability. We extend their theorems to convergence for the full family G_s with a continuous parameter $s\to\infty$; moreover, we show a.s. convergence. (See Theorems 5.1 and 5.3.) Furthermore, in order to prove this result, Veitch and Roy [16] first show a related convergence result for a randomly relabelled version of the random graph, again for sequences $s_k\to\infty$. Again we improve their result to convergence for the continuous parameter $s\to\infty$ (Theorem 4.1). (In both cases, we make only some minor technical improvements in the proof; the proofs are thus essentially due to [16].)

2. Notation and preliminaries

Much of the notation follows Veitch and Roy [15, 16], but there are various modifications and additions for our purposes.

λ denotes Lebesgue measure (in one or several dimensions).

$\mathbb{R}_+:=[0,\infty)$, the set of non-negative real numbers.

If S is a measurable space, then $\mathcal{P}(S)$ is the set of probability measures on S. If X is a random variable in some measurable space S, then $\mathcal{L}(X)$ denotes the distribution of X; thus $\mathcal{L}(X)\in\mathcal{P}(S)$. If S is a metric (or metrizable) space; we equip $\mathcal{P}(S)$ with the usual weak topology, see e.g. [2] or [12, Chapter 4]. Note that if S is a Polish space (i.e., it can be given a complete and separable metric), then $\mathcal{P}(S)$ is Polish too, see [2, Appendix III]. We denote convergence in distribution of random variables in S by $\xrightarrow{\mathrm{d}}$; recall that $X_n\xrightarrow{\mathrm{d}}X$ means $\mathcal{L}(X_n)\to\mathcal{L}(X)$ in $\mathcal{P}(S)$.

If X and Y are random variables (defined on the same probability space and, for simplicity, with values in some Polish spaces $\mathcal{S}_X$ and $\mathcal{S}_Y$), then $\mathcal{L}(X \mid Y)$ denotes the conditional distribution of X given Y; note that this is a random probability measure on $\mathcal{S}_X$ that can be regarded as a function of Y. We use also $(X \mid Y)$ for a random variable with this conditional distribution.

If furthermore S is a locally compact Polish space (= locally compact second countable Hausdorff space), then $\mathcal{M}(S)$ is the set of locally finite Borel measures on S. (We will only use $S = \mathbb{R}_+^2$ and subsets thereof.) We equip $\mathcal{M}(S)$ with the vague topology, which makes $\mathcal{M}(S)$ into a Polish space, see [12, Appendix A.2 and Theorem A.2.3]. Furthermore, $\mathcal{N}(S)$ is the subset of integer-valued measures in $\mathcal{M}(S)$, i.e., the set of all locally finite sums of unit point masses δ_x, and we let $\mathcal{N}_{\mathsf{s}}(S)$ be the subset of simple integer-valued measures, i.e., locally finite sums of distinct unit point masses. It is easily seen that $\mathcal{N}(S)$ and $\mathcal{N}_{\mathsf{s}}(S)$ are measurable subsets of $\mathcal{M}(S)$.

2.1. **Graphs and adjacency measures.** We consider both unlabelled and labelled graphs; in the labelled case, each vertex is labelled with a real number in $\mathbb{R}_+$, and these labels are supposed to be distinct. The graphs may be finite or countably infinite, but we always assume that there are no isolated vertices; thus a graph G is determined by its edge set $E(G)$. We furthermore consider only graphs that are simple in the sense that there are no multiple edges; in general we allow loops (but in many applications we do not have any).

We denote the vertex set and edge set of a graph G by $V(G)$ and $E(G)$, and let $v(G) := |V(G)|$ and $e(G) := |E(G)|$ be the numbers of vertices and edges.

If Γ is a labelled graph, then the corresponding unlabelled graph, obtained by ignoring the labels, is denoted $\mathcal{G}(\Gamma)$. Conversely, if G is an unlabelled graph and $s > 0$, then $\mathsf{Lbl}_s(G)$ is the (random) labelled graph obtained by labelling the vertices by random i.i.d. labels that are $U(0, s)$, i.e., uniformly distributed in $(0, s)$. (Note that this yields distinct labels a.s., so we may assume that the labels are distinct as required above.) If G is a labelled graph, we define $\mathsf{Lbl}_s(G)$ in the same way; thus relabelling the vertices randomly (regardless of their original labels). In other words, $\mathsf{Lbl}_s(G) := \mathsf{Lbl}_s(\mathcal{G}(G))$. When G is a random graph (labelled or not), $\mathsf{Lbl}_s(G)$ is defined by taking the labelling independent of G.

If Γ is a labelled graph, we represent the edge set $E(\Gamma)$ of Γ (and thus the graph Γ itself) by the measure

$$\xi = \xi(\Gamma) := \sum_{x,y\in\mathbb{R}_+:(x,y)\in E(\Gamma)} \delta_{(x,y)} \tag{2.1}$$

on $\mathbb{R}_+^2$; an edge between two distinct vertices labelled x and y is thus represented by the two point masses $\delta_{(x,y)} + \delta_{(y,x)}$, while a loop (if such exist) at a vertex labelled x is represented by $\delta_{(x,x)}$. (We consider undirected graphs, and thus the endpoints of an edge have to be treated symmetrically.) Note that Γ is determined (as a labelled graph) by ξ.

If Γ is a labelled graph, and ξ the corresponding measure, then for $r \geqslant 0$, $\Gamma|_r$ denotes the induced subgraph of Γ obtained by first eliminating all

vertices with labels $> r$, and any edges incident to such a vertex, and then also removing all remaining vertices that have become isolated. In other words, we keep the edges whose endpoints both have labels $\leqslant r$, and the endpoints of these edges. We let $\xi|_r$ denote the corresponding measure on $\mathbb{R}_+^2$, and note that this is just the restriction of ξ to $[0,r]^2$.

A labelled graph Γ is *locally finite* if $\Gamma|_r$ is finite for each $r < \infty$; equivalently, $\xi(\Gamma)$ is a locally finite measure. We consider only locally finite graphs. We say that a measure ξ on $\mathbb{R}_+^2$ is an *adjacency measure* if it is given by (2.1) for some locally finite labelled graph Γ. Hence, a measure is an adjacency measure if and only if it is a symmetric measure in $\mathcal{N}_{\mathsf{s}}(\mathbb{R}_+^2)$. We denote the set of adjacency measures by $\mathcal{N}_{\mathsf{s},\mathsf{s}}(\mathbb{R}_+^2)$, and let $\widehat{\mathfrak{G}}$ be the set of locally finite labelled graphs.

Thus, (2.1) defines a 1–1 correspondence $\Gamma \leftrightarrow \xi(\Gamma)$ between the sets $\widehat{\mathfrak{G}}$ and $\mathcal{N}_{\mathsf{s},\mathsf{s}}(\mathbb{R}_+^2)$ of locally finite labelled graphs and adjacency measures. We give the set of adjacency measures $\mathcal{N}_{\mathsf{s},\mathsf{s}}(\mathbb{R}_+^2)$ the subspace topology as a subset of $\mathcal{M}(\mathbb{R}_+^2)$, and give $\widehat{\mathfrak{G}}$ the corresponding topology induced by the correspondence (2.1). Thus $\mathcal{N}_{\mathsf{s},\mathsf{s}}(\mathbb{R}_+^2)$ and $\widehat{\mathfrak{G}}$ are metric spaces, and a sequence $\Gamma_n \to \Gamma$ in $\widehat{\mathfrak{G}}$ if and only if $\xi(\Gamma_n) \to \xi(\Gamma)$ in $\mathcal{M}(\mathbb{R}_+^2)$, i.e., in the vague topology.

Returning to unlabelled graphs, we let $\mathfrak{G}$ be the set of finite or countably infinite unlabelled graphs, and $\mathfrak{G}_{\mathsf{f}}$ the subset of finite unlabelled graphs. Then $\mathfrak{G}_{\mathsf{f}}$ is countable, and we give it the discrete topology. (We do not define a topology on $\mathfrak{G}$.)

2.2. **Graphons, graphexes and random graphs.** A *graphon* is (in the present context) a symmetric, measurable function $W : \mathbb{R}_+^2 \to [0,1]$ that satisfies the integrability conditions, where $\mu_W(x) := \int_0^\infty W(x,y)\,\mathrm{d}y$:

(i) $\mu_W(x) < \infty$ for a.e. x and $\lambda\{x : \mu_W(x) > 1\} < \infty$;
(ii) $\int_{\mathbb{R}_+^2} W(x,y)\mathbf{1}\{\mu_W(x) \leqslant 1\}\mathbf{1}\{\mu_W(y) \leqslant 1\}\,\mathrm{d}x\,\mathrm{d}y < \infty$;
(iii) $\int_{\mathbb{R}_+} W(x,x) < \infty$.

Note that these conditions are satisfied if W is integrable, but they are also satisfied for some non-integrable W.

A *graphex* is a triple $\mathcal{W} = (I, S, W)$, where $I \geqslant 0$ is a non-negative real number, $S : \mathbb{R}_+ \to \mathbb{R}_+$ is measurable with $S \wedge 1$ integrable and $\mathcal{W}$ is a graphon. Let $\mathfrak{W}$ be the set of all graphexes.

Each graphex $\mathcal{W} \in \mathfrak{W}$ defines a random adjacency measure $\xi = \xi(\mathcal{W})$, and thus a corresponding random labelled graph $\Gamma = \Gamma(\mathcal{W})$, by the following construction; see further Veitch and Roy [15, 16] and Kallenberg [13]: Take realizations of independent unit-rate Poisson processes $\Xi = \{(\theta_j, \vartheta_j)\}_j$ on $\mathbb{R}_+^2$, $\Xi'_i = \{(\sigma_{ij}, \chi_{ij})\}_j$ on $\mathbb{R}_+^2$ for $i \in \mathbb{N}$, and $\Xi'' = \{(\rho_j, \rho'_j, \eta_j)\}_j$ on $\mathbb{R}_+^3$. We regard θ_i, σ_{ij}, ρ_i and ρ'_j as potential vertex labels, while ϑ_j, χ_{ij} and η_j can be regarded as types of the corresponding labels. Given $\mathcal{W} = (I, S, W)$ and these realizations, and a family of i.i.d. random variables $\zeta_{i,j} \sim U(0,1)$

independent of them, define the adjacency measure

$$\begin{aligned}\xi(\mathcal{W}) = \sum_{i,j} \mathbf{1}\{\zeta_{i,j} \leqslant W(\vartheta_i,\vartheta_j)\}\delta_{\theta_i,\theta_j} \\ + \sum_{j,k} \mathbf{1}\{\chi_{jk} \leqslant S(\vartheta_j)\}\big(\delta_{\theta_j,\sigma_{jk}} + \delta_{\sigma_{jk},\theta_j}\big) \\ + \sum_{k} \mathbf{1}\{\eta_k \leqslant I\}\big(\delta_{\rho_k,\rho'_k} + \delta_{\rho'_k,\rho_k}\big). \end{aligned} \tag{2.2}$$

In other words, the corresponding random labelled graph $\Gamma(\mathcal{W})$ is defined to have the following edges, with all random choices independent:

(G1) (θ_i,θ_j) with probability $W(\vartheta_i,\vartheta_j)$ for each pair (i,j) with $i \leqslant j$,
(G2) (θ_j,σ_{jk}) for each j and k with $\chi_{jk} \leqslant S(\vartheta_j)$
(G3) (ρ_k,ρ'_k) for each k such that $\eta_k \leqslant I$.

Equivalently, we can define Γ by starting with the Poisson process $\Xi = (\theta_j,\vartheta_j)$ and first define the edges in (G1), and then add for each j a star with centre in θ_j and peripheral vertices labelled by a Poisson process $\{\sigma_{jk}\}_k$ on $\mathbb{R}_+$ with intensity $S(\vartheta_j)$, and finally add edges (ρ_k,ρ'_k) according to a Poisson process with intensity $2I$ in $\{(x,x') \in \mathbb{R}_+^2 : x < x'\}$. (Again all random choices are independent.)

It follows from more general results by Kallenberg [11], [13, Theorem 9.24] that this construction yields all jointly exchangeable random adjacency measures, provided we allow the graphex $\mathcal{W}$ to be random; see Veitch and Roy [15] for the present context. (See also [3, Theorem 2.21] for a related result.) The conditions (i)–(iii) are precisely the conditions needed to guarantee that the constructed measure ξ a.s. is locally finite (and thus an adjacency measure), see [15] and [13, Proposition 9.25].

Note that the edges of type (G3) are independent of everything else and a.s. isolated; they form a dust of little interest. Also the stars produced by (G2) are of minor interest. One therefore often takes $S = 0$ and $I = 0$. A graphex $(0,0,W)$ can be identified with the graphon W, and we write $\xi(W) = \xi(0,0,W)$ and $\Gamma(W) = \Gamma(0,0,W)$. In this case, the construction uses only the Poisson process (θ_j,ϑ_j) and gives the edges in (G1); see also [3] (at least when W is integrable).

We write for convenience $\Gamma_s(\mathcal{W}) = \Gamma(\mathcal{W})|_s$, and we are particularly interested in the graph valued process $(\Gamma_s(\mathcal{W}))_{s\geqslant 0}$. Note that this is an increasing process of finite labelled graphs, where Γ_0 is empty and Γ_r is an induced subgraph of Γ_s whenever $0 \leqslant r \leqslant s$. Furthermore, $\Gamma = \bigcup_{s\geqslant 0}\Gamma_s$, so the (typically infinite) graph Γ (or the measure $\xi(\Gamma)$) and the process $(\Gamma_s)_{s\geqslant 0}$ determine each other.

We consider also the corresponding processes $\xi_s(\mathcal{W}) := \xi(\Gamma_s(\mathcal{W})) = \xi(\mathcal{W})|_s$ of finite adjacency measures and $G_s(\mathcal{W}) := \mathcal{G}(\Gamma_s(\mathcal{W}))$ of finite unlabelled graphs. It follows from the construction above, that for every fixed $s > 0$, given the unlabelled graph $G_s(\mathcal{W})$, the vertex labels on the labelled graph $\Gamma_s(\mathcal{W})$ are i.i.d. with the distribution $U(0,s)$; in other words, conditioned on $G_s(\mathcal{W})$, and therefore also unconditionally,

$$\Gamma_s(\mathcal{W}) \stackrel{\mathrm{d}}{=} \mathsf{Lbl}_s(G_s(\mathcal{W})). \tag{2.3}$$

2.3. **The classical case of graphons on $[0,1]$.** As said above, the classical theory of graphons considers graphons $W:[0,1]^2\to[0,1]$ defined on $[0,1]^2$; we can identify them with graphons defined on $\mathbb{R}_+^2$ that vanish outside $[0,1]^2$. As said above, we furthermore identify the graphon W and the graphex $(0,0,W)$. (There is no dust and no added stars in the classical theory.)

In the classical theory, one defines for each $n\in\mathbb{N}$ a random graph $\overline{G}_n(W)$ with n vertices by taking n i.i.d. random numbers $\vartheta_i\sim U(0,1)$ and conditionally on these variables, letting there be an edge ij with probability $W(\vartheta_i,\vartheta_j)$ for each pair ij. (Furthermore, there are no loops. We may impose this by assuming that W vanishes on the diagonal $\Delta:=\{(x,x)\}_{x\in\mathbb{R}_+}$. For convenience, we tacitly assume this; note that redefining a graphon W to be 0 on Δ is equivalent to ignoring loops.)

On the other hand, in the construction above of $\Gamma_s(W)$, (G2) and (G3) do not appear, and for (G1) we can ignore every (θ_i,ϑ_i) with $\vartheta_i>1$ (since then $W(\vartheta_i,\vartheta_j)=0$ for every j). In the construction of $\Gamma_s(W)$ and $G_s(W)$, we thus consider points $(\theta_i,\vartheta_i)\in[0,s]\times[0,1]$. Let the number of these points be $N_s:=\Xi([0,s]\times[0,1])\sim\mathrm{Po}(s)$. It follows that conditioned on N_s, the random graph $G_s(W)$ constructed above equals $\overline{G}_{N_s}(W)$, with all isolated vertices deleted.

3. Graphex equivalence and convergence

In this section, we define the four different types of convergence that we consider in the present paper.

3.1. **Convergence of corresponding random graphs: $\to_{\mathsf{GP}}$.** The construction in Section 2 defines a random labelled graph $\Gamma(\mathcal{W})$ and a corresponding random adjacency measure $\xi(\mathcal{W})$ for every graphex $\mathcal{W}$. This defines a map $\Psi:\mathfrak{W}\to\mathcal{P}(\mathcal{M}(\mathbb{R}_+^2))$ by $\Psi(\mathcal{W}):=\mathcal{L}(\xi(\mathcal{W}))$, the distribution of $\xi(\mathcal{W})$. Unfortunately, Ψ is not injective, i.e., a graphex $\mathcal{W}$ is not uniquely determined by the distribution of $\xi(\mathcal{W})$. We say that two graphexes $\mathcal{W}$ and $\mathcal{W}'$ are *equivalent* if $\mathcal{L}(\xi(\mathcal{W}))=\mathcal{L}(\xi(\mathcal{W}'))$; we denote this by $\mathcal{W}\cong\mathcal{W}'$. Let $\widetilde{\mathfrak{W}}:=\mathfrak{W}/\cong$, the set of equivalence classes; then Ψ can be regarded as an injection $\widetilde{\mathfrak{W}}\to\mathcal{P}(\mathcal{M}(\mathbb{R}_+^2))$, and we can identify $\widetilde{\mathfrak{W}}$ with its image and equip $\widetilde{\mathfrak{W}}$ with the subspace topology inherited from $\mathcal{P}(\mathcal{M}(\mathbb{R}_+^2))$. Since $\mathcal{M}(\mathbb{R}_+^2)$ is a Polish space, $\mathcal{P}(\mathcal{M}(\mathbb{R}_+^2))$ is metrizable, and thus the topology of $\widetilde{\mathfrak{W}}$ can be defined by a metric d_{GP}. (There are many possible choices d_{GP} but we assume that one is chosen; we will not distinguish an explicit choice.) Moreover, we can also regard d_{GP} as defined on $\mathfrak{W}$; this makes $\mathfrak{W}$ into a pseudometric space, with

$$\mathcal{W}\cong\mathcal{W}' \iff d_{\mathsf{GP}}(\mathcal{W},\mathcal{W}')=0. \tag{3.1}$$

Convergence in the pseudometric d_{GP} is denoted $\to_{\mathsf{GP}}$ by Veitch and Roy [16]. (They actually use (vi) in Theorem 3.1 below as the definition.) Convergence $\to_{\mathsf{GP}}$ can be characterized as follows, which is at least implicit in [16] but stated explicitly here for easy reference. Note that since the topology is metric, it suffices to consider convergence of sequences; the theorem extends immediately to, e.g., convergence of families with a continuous parameter.

Theorem 3.1. *Let $\mathcal{W}_n$, $n \geqslant 1$, and $\mathcal{W}$ be graphexes. Then the following are equivalent, as $n \to \infty$.*

(i) $\mathcal{W}_n \to_{\mathsf{GP}} \mathcal{W}$.
(ii) $d_{\mathsf{GP}}(\mathcal{W}_n, \mathcal{W}) \to 0$.
(iii) $\mathcal{L}(\xi(\mathcal{W}_n)) \to \mathcal{L}(\xi(\mathcal{W}))$ *in* $\mathcal{P}(\mathcal{M}(\mathbb{R}_+^2))$.
(iv) $\xi(\mathcal{W}_n) \xrightarrow{\mathrm{d}} \xi(\mathcal{W})$ *in* $\mathcal{M}(\mathbb{R}_+^2)$.
(v) $\xi_s(\mathcal{W}_n) \xrightarrow{\mathrm{d}} \xi_s(\mathcal{W})$ *in* $\mathcal{M}(\mathbb{R}_+^2)$ *for every* $s < \infty$.
(vi) $G_s(\mathcal{W}_n) \xrightarrow{\mathrm{d}} G_s(\mathcal{W})$ *in* $\mathfrak{G}_{\mathsf{f}}$, *for every* $s < \infty$.

In (v)–(vi), *it suffices to consider s in a given unbounded subset of $\mathbb{R}_+$, for example $s \in \mathbb{N}$.*

Proof. (i) $\iff$ (ii) $\iff$ (iii) holds by the definitions above.

(iii) $\iff$ (iv) holds by the definition of convergence in distribution.

(iv) $\implies$ (v). Let, $\hat{\mathcal{S}}$ be the class of bounded measurable subsets of $\mathbb{R}_+^2$, and for a random measure ξ, let $\hat{\mathcal{S}}_\xi := \{B \in \hat{\mathcal{S}} : \xi(\partial B) = 0 \text{ a.s.}\}$. By [12, Theorem 16.16], a sequence of random measures $\xi_n \xrightarrow{\mathrm{d}} \xi$ if and only if, for every finite sequence $B_1, \dots, B_k \in \hat{\mathcal{S}}$,

$$\big(\xi_n(B_1), \dots, \xi_n(B_k)\big) \xrightarrow{\mathrm{d}} \big(\xi(B_1), \dots, \xi(B_k)\big). \tag{3.2}$$

Let $Q_s := [0, s]^2$. The random graph $\Gamma(\mathcal{W})$ has a.s. no label s, and thus the random measure $\xi(\mathcal{W})$ has a.s. no mass at ∂Q_s, i.e., $Q_s \in \hat{\mathcal{S}}_{\xi(\mathcal{W})}$. Hence, if $B \in \hat{\mathcal{S}}_{\xi_s(\mathcal{W})}$, then $B \cap Q_s \in \hat{\mathcal{S}}_{\xi(\mathcal{W})}$. It follows, see (3.2), that if $\xi(\mathcal{W}_n) \xrightarrow{\mathrm{d}} \xi(\mathcal{W})$ and $B_1, \dots, B_k \in \hat{\mathcal{S}}_{\xi_s(\mathcal{W})}$, then

$$\begin{aligned}\big(\xi_s(\mathcal{W}_n)(B_j)\big)_{j=1}^k &= \big(\xi(\mathcal{W}_n)(B_j \cap Q_s)\big)_{j=1}^k \\ &\xrightarrow{\mathrm{d}} \big(\xi(\mathcal{W})(B_j \cap Q_s)\big)_{j=1}^k = \big(\xi_s(\mathcal{W})(B_j)\big)_{j=1}^k.\end{aligned} \tag{3.3}$$

Consequently, $\xi_s(\mathcal{W}_n) \xrightarrow{\mathrm{d}} \xi_s(\mathcal{W})$ by [12, Theorem 16.16] again. (Alternatively, this implication follows by the continuous mapping theorem since a similar argument shows that the mapping $\xi \to \xi|_s$ from $\mathcal{M}(\mathbb{R}_+^2)$ to itself is continuous at every ξ with $\xi(\partial Q_s) = 0$.)

(v) $\implies$ (iv). Let $B_1, \dots, B_k \in \hat{\mathcal{S}}_{\xi(\mathcal{W})}$, and let s be so large that $\bigcup_j B_j \subset Q_s$. Then $\xi_s(B_j) = \xi(B_j)$ and thus $\xi_s(\mathcal{W}_n) \xrightarrow{\mathrm{d}} \xi_s(\mathcal{W})$ implies, again using [12, Theorem 16.16], $\big(\xi(\mathcal{W}_n)(B_j)\big)_{j=1}^k \xrightarrow{\mathrm{d}} \big(\xi(\mathcal{W})(B_j)\big)_{j=1}^k$. Hence, $\xi(\mathcal{W}_n) \xrightarrow{\mathrm{d}} \xi(\mathcal{W})$.

(v) $\iff$ (vi). Fix $s > 0$. By (2.3), $\xi_s(\mathcal{W}) \stackrel{\mathrm{d}}{=} \xi(\mathsf{Lbl}_s(G_s(\mathcal{W})))$, and similarly for $\mathcal{W}_n$. Hence the equivalence follows from Lemma 3.2 below, taken from [16]. □

Lemma 3.2 ([16, Lemma 4.11])**.** *If G_n, $n \geqslant 1$, and G are any random graphs in $\mathfrak{G}_{\mathsf{f}}$, and $s > 0$, then $G_n \xrightarrow{\mathrm{d}} G$ in $\mathfrak{G}_{\mathsf{f}}$ if and only if $\mathsf{Lbl}_s(G_n) \xrightarrow{\mathrm{d}} \mathsf{Lbl}_s(G)$ in $\widehat{\mathfrak{G}}$, i.e., if and only if $\xi(\mathsf{Lbl}_s(G_n)) \xrightarrow{\mathrm{d}} \xi(\mathsf{Lbl}_s(G))$ in $\mathcal{M}(\mathbb{R}_+^2)$.*

We will not repeat the proof of [16, Lemma 4.11], but we note that it can be interpreted as defining a map $\Phi_s : \mathcal{P}(\mathfrak{G}_{\mathsf{f}}) \to \mathcal{P}(\mathcal{M}(\mathbb{R}_+^2))$, by taking for

$\mu \in \mathcal{P}(\mathfrak{G}_{\mathsf{f}})$ a random graph $G \sim \mu$, and defining $\Phi_s(\mu) := \mathcal{L}(\xi(\mathsf{Lbl}_s(G))) \in \mathcal{P}(\mathcal{M}(\mathbb{R}_+^2))$. It is then shown that Φ_s is continuous, injective and proper. Finally, any continuous and proper map to a metric space is closed, and a continuous and closed injection is a homeomorphism onto a closed subset.

3.2. **Cut metric: $\delta_\square$.** The invariant cut metric $\delta_\square$ is defined only for integrable graphons. It is the standard metric for classical graphons, see e.g. [4], [14], [9]. The definition was extended to graphons defined on arbitrary σ-finite measure spaces by Borgs, Chayes, Cohn and Holden [3], to which we refer for details. Here we only consider graphons on $\mathbb{R}_+$, and then the results simplify as follows, see [3].

First, for an integrable function F on $\mathbb{R}_+^2$, we define its *cut norm* by

$$\|F\|_\square := \sup_{T,U} \left| \int_{T\times U} F(x,y)\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) \right|, \tag{3.4}$$

If W is a graphon and $\varphi : \mathbb{R} \to \mathbb{R}$ is measure preserving, let $W^\varphi(x,y) := W(\varphi(x),\varphi(y))$. Then, for two integrable graphons W_1 and W_2, define

$$\delta_\square(W_1,W_2) := \inf_{\varphi_1,\varphi_2} \|W_1^{\varphi_1} - W_2^{\varphi_2}\|_\square, \tag{3.5}$$

taking the infimum over all pairs of measure preserving maps $\varphi_1, \varphi_2 : \mathbb{R}_+ \to \mathbb{R}_+$. Moreover, it is shown in [3, Proposition 4.3(c)] that

$$\delta_\square(W_1,W_2) := \inf_{\varphi} \|W_1^{\varphi} - W_2\|_\square, \tag{3.6}$$

taking the infimum over all measure-preserving bijections $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$.

By [3, Theorem 2.22], $\delta_\square(W_1,W_2) = 0$ if and only if $W_1 \cong W_2$, i.e., W_1 and W_2 are equivalent in the sense defined above. (For characterizations of this, see [3] and [10].)

If W_n and W are integrable graphons, we write $W_n \to_{\delta_\square} W$ as $n \to \infty$ if $\delta_\square(W_n,W) \to 0$. We shall show that this is at least as strong as $\to_{\mathsf{GP}}$. We begin with the special case of classical graphons, where we have equivalence.

Lemma 3.3. *Suppose that W_n, $n \geqslant 1$, and W are graphons with support on $[0,1]^2$. Also suppose that they all vanish on the diagonal $\Delta := \{(x,x)\}_{x\in[0,1]}$. Then the following are equivalent, as $n \to \infty$:*

(i) $\delta_\square(W_n,W) \to 0$

(ii) *for every $s \geqslant 0$, $G_s(W_n) \xrightarrow{\mathrm{d}} G_s(W)$.*

(iii) $W_n \to_{\mathsf{GP}} W$.

Proof. (i) $\implies$ (ii). Using the notation in Section 2.3, (i) implies (and is equivalent to) $\overline{G}_N(W_n) \xrightarrow{\mathrm{d}} \overline{G}_N(W)$ as $n \to \infty$ for every $N \in \mathbb{N}$, see e.g. [4; 7]. Hence the same holds if we let $N = N_s \sim \mathrm{Po}(s)$ be random, and since the space of finite graphs is discrete, the result holds also if we remove all isolated vertices from the graphs, which yields (ii), see Section 2.3.

(ii) $\iff$ (iii). By Theorem 3.1.

(ii) $\implies$ (i). The implication (i) $\implies$ (ii) shows that the map $W \mapsto (\mathcal{L}(G_s(W)))_{s\geqslant 0} \in \mathcal{P}(\mathfrak{G}_{\mathsf{f}})^{\mathbb{R}_+}$ is continuous. Since the space of (equivalence classes) of classical graphons is compact, it suffices to show that the map is injective.

The only complication is caused by the removal of isolated vertices. Thus, let H be a finite graph without isolated vertices, and let $H + mK_1$ be H with m isolated vertices added. Then, with $v := v(H)$,

$$\begin{aligned}\mathbb{P}(G_s(W) = H) &= \sum_{m=0}^{\infty} \mathbb{P}\big(\overline{G}_{N_s}(W) = H + mK_1\big) \\ &= e^{-s} \sum_{m=0}^{\infty} \frac{s^{m+v}}{(m+v)!}\, \mathbb{P}\big(\overline{G}_{m+v}(W) = H + mK_1\big). \end{aligned} \tag{3.7}$$

Hence, if W and W' are two classical graphons and $\mathbb{P}(G_s(W) = H) = \mathbb{P}(G_s(W') = H)$, then multiplication by e^s and identification of the coefficients of the power series in (3.7) show that $\mathbb{P}(\overline{G}_{m+v}(W) = H + mK_1) = \mathbb{P}(\overline{G}_{m+v}(W') = H+mK_1)$ for every $m \geqslant 0$. Consequently, $G_s(W) \stackrel{\mathrm{d}}{=} G_s(W')$ for every $s > 0$ implies that $\overline{G}_N(W) \stackrel{\mathrm{d}}{=} \overline{G}_N(W')$ for every N, and thus $W \cong W'$ and $\delta_\square(W, W') = 0$. □

Theorem 3.4. *Suppose that W_n, $n \geqslant 1$, and W are integrable graphons. If $W_n \to_{\delta_\square} W$ as $n \to \infty$, then $W_n \to_{\mathsf{GP}} W$.*

Proof. Assume $\delta_\square(W_n, W) \to 0$. By (3.6), we may then replace each W_n by an equivalent $W_n^{\varphi_n}$ such that $\|W_n - W\|_\square \to 0$; note that the measurable rearrangement $W_n^{\varphi_n}$ defines the same random graphs as W_n in distribution, i.e., $\xi(W_n^{\varphi_n}) \stackrel{\mathrm{d}}{=} \xi(W_n)$, and thus $d_{\mathsf{GP}}(W_n, W_n^{\varphi_n}) = 0$, see (3.1).

In the sequel we thus assume $\varepsilon_n := \|W_n - W\|_\square \to 0$. Define the truncations $W_n^{(N)} := W_n \mathbf{1}_{[0,N]^2}$ and $W^{(N)} := W\mathbf{1}_{[0,N]^2}$. Let $\varepsilon > 0$ and fix a large N such that $\|W^{(N)} - W\|_{L^1} < \varepsilon$. Then,

$$\|W_n^{(N)} - W^{(N)}\|_\square \leqslant \|W_n - W\|_\square = \varepsilon_n \to 0 \tag{3.8}$$

as $n \to \infty$. Consequently, by Lemma 3.3 and a rescaling, for every $s \geqslant 0$,

$$G_s(W_n^{(N)}) \xrightarrow{\mathrm{d}} G_s(W^{(N)}) \qquad \text{as } n \to \infty. \tag{3.9}$$

Furthermore,

$$\begin{aligned}\int (W_n - W_n^{(N)})\,\mathrm{d}\lambda &= \int (W - W^{(N)})\,\mathrm{d}\lambda + \int_{\mathbb{R}_+^2 \setminus [0,N]^2} (W_n - W)\,\mathrm{d}\lambda \\ &< \varepsilon + 2\|W_n - W\|_\square = \varepsilon + 2\varepsilon_n. \end{aligned} \tag{3.10}$$

$G_s(W)$ and $G_s(W^{(N)})$ differ only if the labelled graph $\Gamma_s(W)$ contains some edge with at least one label $> N$. The expected number of such edges is

$$\mathbb{E}\big(e(G_s(W)) - e(G_s(W^{(N)}))\big) = \frac{s^2}{2} \int_{\mathbb{R}_+^2} (W - W^{(N)})\,\mathrm{d}\lambda, \tag{3.11}$$

and thus

$$\mathbb{P}\big(G_s(W) \neq G_s(W^{(N)})\big) \leqslant s^2 \int_{\mathbb{R}_+^2} (W - W^{(N)})\,\mathrm{d}\lambda < s^2\varepsilon, \tag{3.12}$$

and similarly, using (3.10),

$$\mathbb{P}\big(G_s(W_n) \neq G_s(W_n^{(N)})\big) \leqslant s^2 \int_{\mathbb{R}_+^2} (W_n - W_n^{(N)})\,\mathrm{d}\lambda < s^2(\varepsilon + 2\varepsilon_n). \tag{3.13}$$

Consequently, if $f:\mathfrak{G}_{\mathsf{f}}\to[0,1]$ is any function, then by (3.9) and (3.13),

$$\begin{aligned}\limsup_{n\to\infty}&|\,\mathbb{E}\,f(G_s(W_n))-\mathbb{E}\,f(G_s(W))|\\&\leqslant\limsup_{n\to\infty}\bigl(|\,\mathbb{E}\,f(G_s(W_n^{(N)}))-\mathbb{E}\,f(G_s(W^{(N)}))|+s^2(2\varepsilon+2\varepsilon_n)\bigr)\\&=2s^2\varepsilon.\end{aligned}$$

Since $\varepsilon>0$ is arbitrary, this shows $\mathbb{E}\,f(G_s(W_n))\to\mathbb{E}\,f(G_s(W))$, and thus $G_s(W_n)\xrightarrow{\mathrm{d}}G_s(W)$ as $n\to\infty$. The result follows by Theorem 3.1. □

Remark 3.5. In Lemma 3.3(ii), it is not necessary to assume the condition $G_s(W_n)\xrightarrow{\mathrm{d}}G_s(W)$ for every $s\geqslant 0$. In fact, it suffices to assume this for s in an arbitrary non-empty interval (a,b), or even more generally, for s in any infinite set having a cluster point in $[0,\infty)$; this follows by the same proof and the uniqueness theorem for analytic functions.

Moreover, the argument suggests that it might suffice to assume (ii) for a single $s>0$. We state this as an open problem. By the proof above, this is equivalent to the problem whether $G_s(W)\overset{\mathrm{d}}{=}G_s(W')$ for the random graphs without isolated vertices is equivalent to the corresponding equality in distribution of the random graphs $\overline{G}_{\mathrm{Po}(s)}(W)$ and $\overline{G}_{\mathrm{Po}(s)}(W')$ with isolated vertices.

Problem 3.6. In Lemma 3.3, does $G_s(W_n)\xrightarrow{\mathrm{d}}G_s(W)$ for a single $s>0$ imply $W_n\to_{\mathsf{GP}}W$?

We can also ask whether this holds for general graphexes and not just for classical graphons. We leave this too as open problems.

Problem 3.7. In Theorem 3.1, does $G_s(\mathcal{W}_n)\xrightarrow{\mathrm{d}}G_s(\mathcal{W})$ for, say, $0<s<1$ imply $\mathcal{W}_n\to_{\mathsf{GP}}\mathcal{W}$?

Problem 3.8. In Theorem 3.1, does $G_s(\mathcal{W}_n)\xrightarrow{\mathrm{d}}G_s(\mathcal{W})$ for a single $s>0$ imply $\mathcal{W}_n\to_{\mathsf{GP}}\mathcal{W}$?

3.3. **Stretched convergence: $\delta_{\square}^{\mathsf{s}}$ and $\to_{\mathsf{GS}}$.** We define, as in [3] and [16], given a graphon W or more generally a graphex $\mathcal{W}=(I,S,W)$ and a real number $c>0$, the *stretched* graphon or graphex by $W^{(c)}(x,y):=W(x/c,y/c)$ or $\mathcal{W}^{(c)}:=(c^2I,S^{(c)},W^{(c)})$ where further $S^{(c)}(x):=cS(x/c)$. It follows easily from the construction of the random graphs above that a stretched graphex defines the same random graph process $G_r(\mathcal{W})$ up to a change of parameter:

$$(G_r(\mathcal{W}^{(c)}))_r\overset{\mathrm{d}}{=}(G_{cr}(\mathcal{W}))_r. \tag{3.14}$$

Borgs, Chayes, Cohn and Holden [3] define the *stretched cut metric* $\delta_{\square}^{\mathsf{s}}$ by, for two nonzero integrable graphons W_1 and W_2,

$$\delta_{\square}^{\mathsf{s}}(W_1,W_2):=\delta_{\square}\bigl(W_1^{(c_1)},W_2^{(c_2)}\bigr) \tag{3.15}$$

where $c_i:=\|W_i\|_{L_1}^{-1/2}$. We write $W_n\to_{\delta_{\square}^{\mathsf{s}}}W$ as $n\to\infty$ if $\delta_{\square}^{\mathsf{s}}(W_n,W)\to0$. Since $\delta_{\square}(W_n,W)\to0$ implies $\|W_n\|_{L^1}\to\|W\|_{L^1}$, it is easily seen that, for any integrable graphons W_n and W,

$$W_n\to_{\delta_{\square}^{\mathsf{s}}}W\iff W_n^{(c_n)}\to_{\delta_{\square}}W^{(c)}\text{ for some constants }c_n,c>0. \tag{3.16}$$

Moreover, note that the random graph process $r \mapsto \Gamma_r(\mathcal{W})$ of labelled graphs is increasing and right-continuous (and thus cadlag), and has a.s. only a finite number of jumps τ_k in each finite interval. Hence the unlabelled graph process $r \mapsto G_r(\mathcal{W}) = \mathcal{G}(\Gamma_r(\mathcal{W}))$ has the same properties. Veitch and Roy [16] consider for any non-zero graphex $\mathcal{W}$ the sequence $(G_{\tau_k}(\mathcal{W}))_k$ of different (finite and unlabelled) graphs that appear in $\{G_r(\mathcal{W}) : 0 \leqslant r < \infty\}$; they define for $\mathcal{W}_n, \mathcal{W} \in \mathfrak{W}' := \mathfrak{W} \setminus \{0\}$,

$$\mathcal{W}_n \to_{\mathsf{GS}} \mathcal{W} \qquad \text{if } (G_{\tau_k}(\mathcal{W}_n))_{k=1}^{\infty} \xrightarrow{\mathrm{d}} (G_{\tau_k}(\mathcal{W}))_{k=1}^{\infty}. \tag{3.17}$$

This too is a metric convergence. In analogy with the definition of d_{GP} above, we can define a map $\Psi' : \mathfrak{W} \setminus \{0\} \to \mathcal{P}(\mathfrak{G}_{\mathsf{f}}^{\infty})$ by $\Psi(\mathcal{W}) = \mathcal{L}\big((G_{\tau_k}(\mathcal{W}))_{k\geqslant 1}\big)$, and by fixing a metric on $\mathcal{P}(\mathfrak{G}_{\mathsf{f}}^{\infty})$, we define a pseudometric d_{GS} on $\mathfrak{W}'$. Then obviously

$$\mathcal{W}_n \to_{\mathsf{GS}} \mathcal{W} \iff d_{\mathsf{GS}}(\mathcal{W}_n, \mathcal{W}) \to 0. \tag{3.18}$$

It follows by (3.14) that $(G_{\tau_k}(\mathcal{W}))_k \stackrel{\mathrm{d}}{=} (G_{\tau_k}(\mathcal{W}^{(c)}))_k$ for any stretching of a graphex $\mathcal{W}$, and thus $d_{\mathsf{GS}}(\mathcal{W}, \mathcal{W}^{(c)}) = 0$, cf. [16, Corollary 5.5]. Furthermore, it follows from Theorem 3.1 that, assuming $\mathcal{W}, \mathcal{W} \in \mathfrak{W}'$, see [16, Lemma 5.6] for a detailed proof,

$$\mathcal{W}_n \to_{\mathsf{GP}} \mathcal{W} \implies \mathcal{W}_n \to_{\mathsf{GS}} \mathcal{W}. \tag{3.19}$$

Consequently, we have the following partial analogue of (3.16).

Lemma 3.9 ([16])**.** *If $\mathcal{W}_n, \mathcal{W} \in \mathfrak{W}$ and there exist $c_n, c > 0$ such that $\mathcal{W}_n^{(c_n)} \to_{\mathsf{GP}} \mathcal{W}^{(c)}$, then $\mathcal{W}_n \to_{\mathsf{GS}} \mathcal{W}$.* □

Problem 3.10. Does the converse to Lemma 3.9 hold? (We conjecture so.)

We also have a result corresponding to Theorem 3.4 for the stretched metrics.

Theorem 3.11. *Suppose that W_n, $n \geqslant 1$, and W are integrable non-zero graphons. If $W_n \to_{\delta_{\square}^{\mathrm{s}}} W$ as $n \to \infty$, then $W_n \to_{\mathsf{GS}} W$.*

Proof. By (3.16), we have $W_n^{(c_n)} \to_{\delta_{\square}} W^{(c)}$ for some constants c_n and c, and the result follows by Theorem 3.4 and Lemma 3.9. □

Problem 3.12. Does the converses of Theorems 3.4 and 3.11 hold. (We conjecture not.)

4. Random relabellings

We show in this section that [16, Theorem 4.3] extends to convergence as $s \to \infty$ through the set of all positive real numbers.

Theorem 4.1 (Extension of [16, Theorem 4.3])**.** *Let $\mathcal{W}$ be a graphex and let $G_s := \mathcal{G}(\Gamma_s(\mathcal{W}))$, $s \geqslant 0$, be the corresponding process of unlabelled graphs. Then, a.s., as $s \to \infty$, $\big(\mathsf{Lbl}_s(G_s) \mid G_s\big) \xrightarrow{\mathrm{d}} \Gamma(\mathcal{W})$ in $\widehat{\mathfrak{G}}$, i.e., $\mathcal{L}(\mathsf{Lbl}_s(G_s) \mid G_s) \to \mathcal{L}(\Gamma(\mathcal{W}))$ in $\mathcal{P}(\widehat{\mathfrak{G}})$. Equivalently, $\big(\xi(\mathsf{Lbl}_s(G_s)) \mid G_s\big) \xrightarrow{\mathrm{d}} \xi(\mathcal{W})$ in $\mathcal{M}(\mathbb{R}_+^2)$.*

Note that Theorem 4.1 not only generalizes [16, Theorem 4.3]; also [16, Theorem 4.4] is an immediate corollary.

In the proof we use the following lemma, which extends a standard lemma used by [16] from a parameter $s \in \mathbb{N}$ to a parameter $s \in \mathbb{Q}_+$. We guess that also the version here is known, but since we do not know a reference, we give a simple proof for completeness.

Lemma 4.2. *Assume that X_s is a random variable for each $s \in \mathbb{Q}_+$ such that $X_s \to X$ a.s. as $s \to \infty$, and further a.s. $|X_s| \leqslant Y$ for some random variable Y with $\mathbb{E}\,Y < \infty$. Let $(\mathcal{F}_s)_{s\in\mathbb{Q}_+}$ be a decreasing family of σ-fields and let $\mathcal{F}_\infty := \bigcap_{s\in\mathbb{Q}_+} \mathcal{F}_s$. Then $\mathbb{E}\bigl(X_s \mid \mathcal{F}_s\bigr) \to \mathbb{E}\bigl(X \mid \mathcal{F}_\infty\bigr)$ a.s. as $s \to \infty$ with $s \in \mathbb{Q}_+$.*

Proof. Let $Z_t := \sup_{s\geqslant t} |X_s - X|$, where as in the rest of the proof we consider only $s \in \mathbb{Q}_+$. The assumption implies a.s. $|X| \leqslant Y$ and thus $0 \leqslant Z_t \leqslant 2Y$. Moreover, $Z_t \to 0$ a.s. as $t \to \infty$ by assumption. Hence, $\mathbb{E}\,Z_t < \infty$ and $\mathbb{E}\,Z_t \to 0$ as $t \to \infty$ by dominated convergence.

Fix t. For $s \geqslant t$ we have $|X_s - X| \leqslant Z_t$, and thus $\mathbb{E}\bigl(|X_s - X| \mid \mathcal{F}_s\bigr) \leqslant \mathbb{E}\bigl(Z_t \mid \mathcal{F}_s\bigr)$ a.s. Consequently, using the convergence theorem for (reverse) martingales, [12, Theorem 7.23], a.s.,

$$\limsup_{s\to\infty} \mathbb{E}\bigl(|X_s - X| \mid \mathcal{F}_s\bigr) \leqslant \limsup_{s\to\infty} \mathbb{E}\bigl(Z_t \mid \mathcal{F}_s\bigr) = \mathbb{E}\bigl(Z_t \mid \mathcal{F}_\infty\bigr). \tag{4.1}$$

Hence, for every $t \geqslant 0$,

$$\mathbb{E}\limsup_{s\to\infty} \mathbb{E}\bigl(|X_s - X| \mid \mathcal{F}_s\bigr) \leqslant \mathbb{E}\bigl(\mathbb{E}\bigl(Z_t \mid \mathcal{F}_\infty\bigr)\bigr) = \mathbb{E}\,Z_t. \tag{4.2}$$

However, we have shown that $\mathbb{E}\,Z_t \to 0$ as $t \to \infty$, and thus (4.2) implies

$$\mathbb{E}\limsup_{s\to\infty} \mathbb{E}\bigl(|X_s - X| \mid \mathcal{F}_s\bigr) = 0. \tag{4.3}$$

Consequently, a.s., $\mathbb{E}\bigl(|X_s - X| \mid \mathcal{F}_s\bigr) \to 0$ as $s \to \infty$, and thus

$$\bigl|\mathbb{E}\bigl(X_s - X \mid \mathcal{F}_s\bigr)\bigr| \leqslant \mathbb{E}\bigl(|X_s - X| \mid \mathcal{F}_s\bigr) \to 0. \tag{4.4}$$

Furthermore, by the (reverse) martingale convergence theorem again and (4.4), a.s.,

$$\mathbb{E}\bigl(X_s \mid \mathcal{F}_s\bigr) = \mathbb{E}\bigl(X \mid \mathcal{F}_s\bigr) + \mathbb{E}\bigl(X_s - X \mid \mathcal{F}_s\bigr) \to \mathbb{E}\bigl(X \mid \mathcal{F}_\infty\bigr) + 0. \tag{4.5}$$

□

Proof of Theorem 4.1. Let $(\tau_k)_k$ be the jump times, where $G_s \neq G_{s-}$. Then G_s is constant for $s \in [\tau_k, \tau_{k+1})$, and it is easily seen that the distribution $\mathcal{L}(\mathsf{Lbl}_s(G_s) \mid G_s)$ is a continuous function of $s \in [\tau_k, \tau_{k+1})$. It follows that it suffices to prove the result as $s \to \infty$ through the countable set of rational numbers. Thus, in the sequel of the proof, we assume that $s \in \mathbb{Q}_+$, and we consider limits as $s \to \infty$ through $\mathbb{Q}_+$.

Except for this, we follow the proof of [16, Theorem 4.3]; for completeness we repeat most of the arguments.

First, by [16, Lemma 4.2], a consequence of [12, Theorems 16.28–29], it suffices to prove that if U is a finite union of rectangles with rational coordinates, then, with $\xi_s := \xi(\mathsf{Lbl}_s(G_s))$ and $\xi := \xi(\mathcal{W})$,

$$\mathcal{L}(\xi_s(U) \mid G_s) \to \mathcal{L}\bigl(\xi(U)\bigr) \qquad \text{a.s.} \tag{4.6}$$

as $s \to \infty$. Note that [16, Lemma 4.2], although stated for sequences, immediately extends to a parameter $s \in \mathbb{Q}_+$ (or even $\mathbb{R}_+$) since convergence in $\mathcal{P}(\mathcal{M}(\mathbb{R}_+^2))$ can be defined by a metric.

Next, (4.6) means that for every function $f : \mathbb{Z}_+ \to \mathbb{R}_+$ of the form $f(x) = \mathbf{1}_{\{k\}}(x) = \mathbf{1}\{x = k\}$,

$$\mathbb{E}\big(f(\xi_s(U)) \mid G_s\big) \to \mathbb{E}\, f(\xi(U)) \qquad \text{a.s.} \tag{4.7}$$

Thus, fix such a U and f (or any bounded $f : \mathbb{Z}_+ \to \mathbb{R}_+$), and fix some large enough r such that $U \subset [0,r]^2$. For $s \in \mathbb{Q}_+$, let Γ^s be the partially labelled graph obtained from $\Gamma(\mathcal{W})$ by forgetting all labels in $[0,s]$ (but keeping larger labels). Let $\mathcal{F}_s$ be the σ-field generated by Γ^s. Conditioned on Γ^s, $\Gamma(\mathcal{W})$ is obtained by randomly relabelling the unlabelled vertices by labels in $[0,s]$; note that the unlabelled part of Γ^s is G_s. Hence, if $s > r$,

$$\mathbb{E}\big(f(\xi(U)) \mid \mathcal{F}_s\big) = \mathbb{E}\big(f(\xi(\mathsf{Lbl}_s(G_s)) \mid G_s\big) = \mathbb{E}\big(f(\xi_s(U)) \mid G_s\big). \tag{4.8}$$

Define as in [16] $U_t := U + (t,t)$ and, for $s > r$,

$$X_s^{(r)} := \frac{1}{s-r} \int_0^{s-r} f(\xi(U_t))\, \mathrm{d}t. \tag{4.9}$$

Since $t \mapsto Y_t := f(\xi(U_t))$ is a stationary stochastic process, which is r-dependent in the sense that $\{Y_t\}_{t \leqslant t_0}$ is independent of $\{Y_t\}_{t > t_0 + r}$ for every t_0, it follows from the ergodic theorem [12, Corollary 10.9] that $X_s^{(r)} \xrightarrow{\text{a.s.}} \mathbb{E}\, f(\xi(U))$ as $s \to \infty$. Lemma 4.2 now shows that

$$\mathbb{E}\big(X_s^{(r)} \mid \mathcal{F}_s\big) \xrightarrow{\text{a.s.}} \mathbb{E}\, f(\xi(U)). \tag{4.10}$$

Furthermore, by symmetry, for every s and every $t < s - r$, $\mathbb{E}\big(f(\xi(U_t)) \mid \mathcal{F}_s\big) = \mathbb{E}\big(f(\xi(U)) \mid \mathcal{F}_s\big)$ a.s., and thus (4.9) implies

$$\mathbb{E}\big(X_s^{(r)} \mid \mathcal{F}_s\big) = \mathbb{E}\big(f(\xi(U)) \mid \mathcal{F}_s\big) \qquad \text{a.s.} \tag{4.11}$$

We obtain (4.7) by combining (4.8), (4.11) and (4.10), which completes the proof. □

5. Convergence of empirical graphons

If G is any finite graph, and $s > 0$, we define as in [16] the *stretched empirical graphon* $\hat{W}_{G,s}$ as the graphon obtained from the adjacency matrix of G by replacing each vertex by an interval of length $1/s$. (This assumes some ordering of the vertices, but different orderings give equivalent graphons.) In other words, using the notation of Section 3.3, $\hat{W}_{G,s} = \hat{W}_G^{(1/s)}$ where $\hat{W}_G := \hat{W}_{G,1}$. Note that $\hat{W}_{G,s}$ only takes the values $\{0,1\}$.

We identify as usual the graphon $\hat{W}_{G,s}$ with the graphex $(0,0,\hat{W}_{G,s})$.

Borgs, Chayes, Cohn and Holden [3, Theorem 2.23] show that if W is a non-zero integrable graphon, then a.s. $\delta_\square^{\mathsf{s}}\big(\hat{W}_{G_s(W)}, W\big) \to 0$ and (implicitly, or as a consequence) $\delta_\square\big(\hat{W}_{G_s(W),s}, W\big) \to 0$ as $s \to \infty$.

Similarly, Veitch and Roy [16, Theorems 4.8 and 4.12] show convergence in $\to_{\mathsf{GP}}$ for a general graphex; however, only for sequences $s_k \to \infty$. We show in this section an extension of their result to convergence as $s \to \infty$ through the set of all positive real numbers.

Theorem 5.1 (Extension of [16, Theorems 4.8 and 4.12])**.** *Let $\mathcal{W}$ be any graphex and let $G_s := \mathcal{G}(\Gamma_s(\mathcal{W}))$, $s \geqslant 0$, be the corresponding process of unlabelled graphs. Then, $\hat{W}_{G_s,s} \to_{\mathsf{GP}} \mathcal{W}$ a.s. as $s \to \infty$.*

Proof. Again we follow the proof in [16] with some modifications. Let $0 < r \leqslant s$.

Given G_s we define two random induced subgraphs $X_r^{(s)}$ and $M_r^{(s)}$ of G_s; in $X_r^{(s)}$ we select each vertex of G_s with probability r/s, and in $M_r^{(s)}$ we take $\mathrm{Po}(r/s)$ copies of each vertex in G_s (thus allowing repetitions); in both cases this is done independently for all vertices. In both cases, we then add the edges induced by G_s and remove any isolated vertex.

It follows that, conditionally given G_s, $\mathsf{Lbl}_r(X_r^{(s)}) \overset{\mathrm{d}}{=} \mathsf{Lbl}_s(G_s)|_r$. Hence, by Theorems 4.1 and 3.1 together with (2.3), a.s. as $s \to \infty$,

$$\bigl(\mathsf{Lbl}_r(X_r^{(s)}) \mid G_s\bigr) \overset{\mathrm{d}}{=} \bigl(\mathsf{Lbl}_s(G_s)|_r \mid G_s\bigr) \xrightarrow{\mathrm{d}} \Gamma(\mathcal{W})|_r = \Gamma_r(\mathcal{W}) = \mathsf{Lbl}_r(G_r(\mathcal{W})).$$

Consequently, using Lemma 3.2,

$$\bigl(X_r^{(s)} \mid G_s\bigr) \xrightarrow{\mathrm{d}} G_r(\mathcal{W}). \tag{5.1}$$

On the other hand, still conditionally given G_s, the random graph $G_r(\hat{W}_{G_s,s})$ constructed as in Section 2.2 from the empirical graphon $\hat{W}_{G_s,s}$ is precisely $M_r^{(s)}$. Hence,

$$\bigl(M_r^{(s)} \mid G_s\bigr) \overset{\mathrm{d}}{=} \bigl(G_r(\hat{W}_{G_s,s}) \mid G_s\bigr). \tag{5.2}$$

Lemma 5.2 below shows that, still conditionally given G_s, there exists a coupling of $X_r^{(s)}$ and $M_r^{(s)}$ such that $\mathbb{P}\bigl(M_r^{(s)} \neq X_r^{(s)} \mid X_r^{(s)}\bigr) \leqslant 2(r/s)v(X_r^{(s)})$. Since trivially also the probability is at most 1, this yields, for any constant $A > 0$, $\mathbb{P}\bigl(M_r^{(s)} \neq X_r^{(s)} \mid X_r^{(s)}\bigr) \leqslant 2Ar/s + \mathbf{1}\{v(X_r^{(s)}) > A\}$ and thus

$$\mathbb{P}\bigl(M_r^{(s)} \neq X_r^{(s)} \mid G_s\bigr) \leqslant 2Ar/s + \mathbb{P}\bigl(v(X_r^{(s)}) > A \mid G_s\bigr). \tag{5.3}$$

Moreover, (5.1) implies that

$$\bigl(v(X_r^{(s)}) \mid G_s\bigr) \xrightarrow{\mathrm{d}} v(G_r(\mathcal{W})) = v(\Gamma_r(\mathcal{W})) \tag{5.4}$$

a.s. as $s \to \infty$, and thus

$$\mathbb{P}\bigl(v(X_r^{(s)}) > A \mid G_s\bigr) \xrightarrow{\text{a.s.}} \mathbb{P}\bigl(v(\Gamma_r(\mathcal{W})) > A\bigr). \tag{5.5}$$

It follows from (5.3) and (5.5), that for every fixed r and A, a.s.,

$$\limsup_{s\to\infty} \mathbb{P}\bigl(M_r^{(s)} \neq X_r^{(s)} \mid G_s\bigr) \leqslant 0 + \mathbb{P}\bigl(v(\Gamma_r(\mathcal{W})) > A\bigr). \tag{5.6}$$

The left-hand side does not depend on A, and as $A \to \infty$, the right-hand side tends to 0. Hence,

$$\mathbb{P}\bigl(M_r^{(s)} \neq X_r^{(s)} \mid G_s\bigr) \to 0 \tag{5.7}$$

a.s. as $s \to \infty$, for every fixed r. Combining (5.7) with (5.1) and (5.2), we obtain

$$\bigl(G_r(\hat{W}_{G_s,s}) \mid G_s\bigr) \xrightarrow{\mathrm{d}} G_r(\mathcal{W}) \tag{5.8}$$

a.s. as $s \to \infty$, for every fixed r, and thus a.s. for every $r \in \mathbb{N}$. Consequently, the result follows by Theorem 3.1. □

Lemma 5.2. *Given a finite graph G and $p \in [0,1]$, let X_p and M_p be the random induced subgraphs of G obtained by independently taking* $\mathrm{Be}(p)$ *and* $\mathrm{Po}(p)$ *copies of each vertex of G, respectively, and then eliminating all resulting isolated vertices. Then X_p and M_p may be coupled such that $\mathbb{P}(M_p \neq X_p \mid X_p) \leqslant 2pv(X_p)$.*

Proof. Let, for $i \in V(G)$, $I_i \sim \mathrm{Be}(p)$ and $Y_i \sim \mathrm{Po}(p)$ be random variables with the pairs $(I_i, Y_i)_{i\in V(G)}$ independent, let $\bar{X}_p$ and $\bar{M}_p$ be the induced subgraphs of G obtained by taking I_i or Y_i copies of each vertex i, respectively, and let X_p and M_p be the subgraphs of $\bar{X}_p$ and $\bar{M}_p$ obtained by deleting all isolated vertices.

We couple I_i and Y_i such that for each i, $I_i = 0 \implies Y_i = 0$. This is possible since $\mathbb{P}(Y_i = 0) = e^{-p} \geqslant 1-p = \mathbb{P}(I_i) = 0$, and we have by a simple calculation

$$\mathbb{P}\big(Y_i \neq I_i \mid I_i = 1\big) = \frac{\mathbb{P}(Y_i \neq I_i)}{\mathbb{P}(I_i = 1)} = \frac{2(p - pe^{-p})}{p} = 2(1 - e^{-p}) \leqslant 2p. \quad (5.9)$$

Using this coupling, the vertices of $\bar{M}_p$ form a subset of the vertices of $\bar{X}_p$, possibly with some repetitions, and it follows that if vertex i is selected for $\bar{X}_p$ but is isolated, and thus does not appear in X_p, then it also will not appear in M_p, since even if it appears in $\bar{M}_p$ it will be isolated there. Hence, if $X_p \neq M_p$, there must be some vertex $i \in X_p$ such that $Y_i \neq I_i$. Consequently, by (5.9),

$$\mathbb{P}\big(X_p \neq M_p \mid X_p\big) \leqslant \sum_{i\in V(X_p)} \mathbb{P}\big(Y_i \neq I_i \mid I_i = 1\big) \leqslant 2pv(X_p). \quad (5.10)$$

□

As in [16], we obtain a corollary on stretched convergence of unstretched empirical graphons, cf. Section 3.3.

Theorem 5.3 (Extension of [16, Theorem 5.7(2)])**.** *If $\mathcal{W}$ is any non-zero graphex, then $\hat{W}_{G_{\tau_k}(\mathcal{W}),1} \to_{\mathsf{GS}} \mathcal{W}$.*

Proof. Immediate by Theorem 5.1 and Lemma 3.9. □

Remark 5.4. Note that even if $\mathcal{W} = (I, S, W)$ is a general graphex with non-zero dust and star components I and S, Theorem 5.1 and [16, Theorem 4.12] yield convergence to $\mathcal{W}$ of a sequence of graphons, where the dust and star components are taken to be 0. This shows that for finite r, it is not possible to distinguish between isolated edges or stars coming from I and S and isolated edges or stars coming from the graphon part W. Note that in contrast, for the infinite random graph $\Gamma(\mathcal{W})$, there is an obvious difference between edges produced by I, S and W: they have a.s. 2, 1 and 0 endpoints of degree 1, respectively.

Acknowledgement

I thank Daniel Roy for an interesting discussion at the Isaac Newton Institute for Mathematical Sciences during the programme Theoretical Foundations for Statistical Network Analysis in 2016 (EPSCR Grant Number EP/K032208/1). This work was also supported by the Knut and Alice Wallenberg Foundation.

References

[1] Tim Austin. On exchangeable random variables and the statistics of large graphs and hypergraphs. *Probab. Surv.* **5** (2008), 80–145.

[2] Patrick Billingsley. *Convergence of Probability Measures*. Wiley, New York, 1968.

[3] Christian Borgs, Jennifer T. Chayes, Henry Cohn & Nina Holden. Sparse exchangeable graphs and their limits via graphon processes. *J. Mach. Learn. Res.* **18** (2017), Paper No. 210, 71 pp.

[4] Christian Borgs, Jennifer T. Chayes, László Lovász, Vera T. Sós & Katalin Vesztergombi. Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing, *Advances in Math.* **219** (2008), 1801–1851.

[5] Christian Borgs, Jennifer T. Chayes, László Lovász, Vera T. Sós & Katalin Vesztergombi. Convergent sequences of dense graphs II. Multiway cuts and statistical physics. *Ann. of Math. (2)* **176** (2012), no. 1, 151–219.

[6] François Caron & Emily B. Fox. Sparse graphs using exchangeable random measures. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** (2017), no. 5, 1295–1366.

[7] Persi Diaconis & Svante Janson. Graph limits and exchangeable random graphs. *Rend. Mat. Appl.* (7) **28** (2008), no. 1, 33–61.

[8] Tue Herlau, Mikkel N. Schmidt & Morten Mørup, Completely random measures for modelling block-structured sparse networks. In *Advances in Neural Information Processing Systems*, vol. 29 (editors D. D. Lee, M. Sugiyama, U. Von Luxburg, I. Guyon and R. Garnett), pp. 4260–4268. Curran Associates, 2016.

[9] Svante Janson. Graphons, cut norm and distance, rearrangements and coupling. *New York J. Math. Monographs*, **4**, 2013.

[10] Svante Janson. Graphons and cut metric on σ-finite measure spaces. Preprint, 2016. `arXiv:1608.01833`

[11] Olav Kallenberg. Exchangeable random measures in the plane. *J. Theoret. Probab.* **3** (1990), no. 1, 81–136.

[12] Olav Kallenberg. *Foundations of Modern Probability.* 2nd ed., Springer, New York, 2002.

[13] Olav Kallenberg. *Probabilistic Symmetries and Invariance Principles.* Springer, New York, 2005.

[14] László Lovász. *Large Networks and Graph Limits.* American Mathematical Society, Providence, RI, 2012.

[15] Victor Veitch & Daniel M. Roy. The class of random graphs arising from exchangeable random measures. Preprint, 2015. `arXiv:1512.03099v1`

[16] Victor Veitch & Daniel M. Roy. Sampling and estimation for (sparse) exchangeable graphs. *Ann. Statist.* **47** (2019), no. 6, 3274–3299.

Department of Mathematics, Uppsala University, PO Box 480, SE-751 06 Uppsala, Sweden

Email address: `svante.janson@math.uu.se`

URL: `http://www.math.uu.se/svante-janson`