A SURVEY OF THE THEORY OF GRAPHONS AND PERMUTONS

PIERRE-LOÏC MÉLIOT

Abstract. The purpose of this note is to present the theory of graphons and permutons.

Contents

1. Graphons and their topology

1.1. Graphs and morphisms. In this paper, a *graph* will be a finite undirected simple graph, that is to say a pair (V, E) with V finite set of *vertices*, and E subset of the set $\mathfrak{P}_2(V)$ of pairs of vertices. Thus, E is a finite set of pairs $\{v_1, v_2\}$ with $v_1, v_2 \in V$ and $v_1 \neq v_2$. These pairs are the *edges* of the graph.

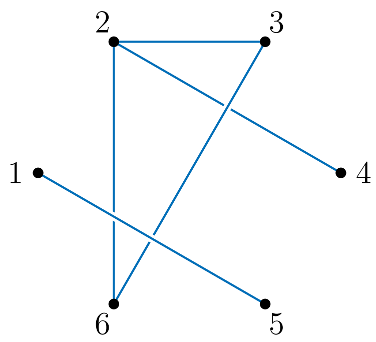

Figure 1. A graph G with vertex set $V = [\![1, 6]\!]$ and edge set $E = \{\{1,5\},\{2,3\},\{2,4\},\{2,6\},\{3,6\}\}$.

A *morphism* (*cf.* [LS06]) from a graph $F = (V_F, E_F)$ to a graph $G = (V_G, E_G)$ is a map $\phi : V_F \to V_G$ such that, if $(v_1, v_2) \in E_F$, then $(\phi(v_1), \phi(v_2)) \in E_G$. We denote $\hom(F, G)$ the set of morphisms from F to G, and the *morphism density* from F to G is defined by

$$t(F, G) = \frac{|\hom(F, G)|}{|V_G|^{|V_F|}},$$

where $|A|$ denotes the cardinality of a set A. This is a real number between 0 and 1, which measures the number of copies of F inside G. One can also work with *embeddings* of F into G, that is morphisms that are injective maps $V_F \to V_G$. Set

$$t_0(F, G) = \frac{|\mathrm{emb}(F, G)|}{|V_G|^{\downarrow |V_F|}},$$

where $\mathrm{emb}(F, G)$ is the set of embeddings of F into G, and $n^{\downarrow k} = n(n-1)\cdots(n-k+1)$ denotes a falling factorial — thus, $|V_G|^{\downarrow |V_F|}$ is the number of injective maps from V_F to V_G. The two quantities $t(F, G)$ and $t_0(F, G)$ are close when G is sufficiently large:

Lemma 1. *For any finite graphs F and G,*

$$|t(F, G) - t_0(F, G)| \leq \frac{1}{|V_G|} \binom{|V_F|}{2}.$$

Proof. We have:

$$\begin{aligned} t(F, G) - t_0(F, G) &= \frac{|\hom(F, G)|}{|V_G|^{|V_F|}} - \frac{|\mathrm{emb}(F, G)|}{|V_G|^{\downarrow |V_F|}} \\ &\leq \frac{|\hom(F, G)|}{|V_G|^{|V_F|}} - \frac{|\mathrm{emb}(F, G)|}{|V_G|^{|V_F|}} \\ &\leq \frac{|\text{number of non-injective morphisms } F \to G|}{|V_G|^{|V_F|}}. \end{aligned}$$

Set $n = |V_G|$ and $k = |V_F|$. To construct a non-injective map from V_F to V_G, it suffices to choose a pair $\{a, b\}$ of vertices in V_F that will be sent to the same image in V_G ($\binom{k}{2}$ possibilities for the pair, and n possibilities for the image), and then to choose the $k - 2$ other images (n^{k-2} possibilities).

So, the number of non-injective maps, and therefore the number of non-injective morphisms from F to G is smaller than $\binom{k}{2}\, n^{k-1}$, and

$$t(F,G) - t_0(F,G) \leq \frac{1}{n^k}\left(\binom{k}{2}\, n^{k-1}\right) = \frac{1}{n}\binom{k}{2}.$$

Similarly,

$$\begin{aligned} t(F,G) - t_0(F,G) &= \frac{|\hom(F,G)|}{|V_G|^{|V_F|}} - \frac{|\mathrm{emb}(F,G)|}{|V_G|^{\downarrow|V_F|}} \\ &\geq |\mathrm{emb}(F,G)|\left(\frac{1}{|V_G|^{|V_F|}} - \frac{1}{|V_G|^{\downarrow|V_F|}}\right) = t_0(F,G)\left(\frac{|V_G|^{\downarrow|V_F|}}{|V_G|^{|V_F|}} - 1\right) \\ &\geq \frac{|V_G|^{\downarrow|V_F|}}{|V_G|^{|V_F|}} - 1 \geq -\frac{1}{n}\binom{k}{2}, \end{aligned}$$

the last inequality coming from the same argument as before. □

Definition 2. *Let $(G_n)_{n\in\mathbb{N}}$ be a sequence of graphs. One says that $(G_n)_{n\in\mathbb{N}}$ converges if, for any fixed graph F, the density of morphisms $t(F,G_n)$ admits a limit when n goes to infinity. If $|V_{G_n}| \to \infty$, then by the previous lemma this is equivalent to ask that $t_0(F,G_n)$ admits a limit for any fixed graph F.*

We call *graph parameter* a family of real numbers $(t(F))_{F \text{ graph}}$ indexed by the countable set of (isomorphism classes of) finite graphs, such that there exists a sequence of finite graphs G_n with

$$\lim_{n\to\infty} t(F,G_n) = t(F)$$

for any F. The theory of graphons will allow us to identify all the graph parameters.

1.2. **Graph parameters and graph functions.** A *graph function* is a function $f : [0,1]^2 \to [0,1]$ that is measurable and symmetric: $f(x,y) = f(y,x)$ Lebesgue almost surely on $[0,1]^2$. Thus, the graph functions form a subset $\mathcal{W}$ of the space $\mathrm{L}^\infty([0,1]^2)$ of essentially bounded measurable functions on the square $[0,1]$. If f is a graph function, then one can associate to it a family $(t(F,f))_{F \text{ graph}}$ indexed by finite graphs:

$$t(F,f) = \int_{[0,1]^k}\left(\prod_{e=(i,j)\in E_F} f(x_i,x_j)\right) dx_1\, dx_2 \cdots dx_k,$$

where V_F is identified with $[\![1,k]\!]$ if $k = |V_F|$. For instance, if F is the graph of Figure 1, then

$$t(F,f) = \int_{[0,1]^6} f(x_1,x_5) f(x_2,x_3) f(x_2,x_4) f(x_2,x_6) f(x_3,x_6)\, dx.$$

Notice that if $\sigma : [0,1] \to [0,1]$ is a map that preserves the Lebesgue measure, then $t(F, f(\sigma(\cdot),\sigma(\cdot))) = t(F, f(\cdot,\cdot))$. Therefore, the map $t(F,\cdot) : \mathcal{W} \to [0,1]$ is invariant by the action of the Lebesgue isomorphisms of $[0,1]$. In a moment, we shall define graphons as orbits in $\mathcal{W}$ under this action. We first describe the connection between graph functions and graph parameters:

Theorem 3 (Theorem 2.2 in [LS06]). *A family $(t(F))_F$ is a graph parameter if and only if there exists a graph function f such that $t(F,f) = t(F)$ for any finite graph F.*

Let us first see why graph functions give rise to graph parameters. If G is a finite graph with vertex set $V_G = [\![1,n]\!]$, then one can associate to it a graph function g as follows: g is the function on the square that takes its values in $\{0,1\}$, and is such that

$$g(x,y)=1 \text{ if } x\in\left[\frac{i-1}{n},\frac{i}{n}\right),\ y\in\left[\frac{j-1}{n},\frac{j}{n}\right) \text{ and } i\sim j \text{ in } G,$$

and 0 otherwise.

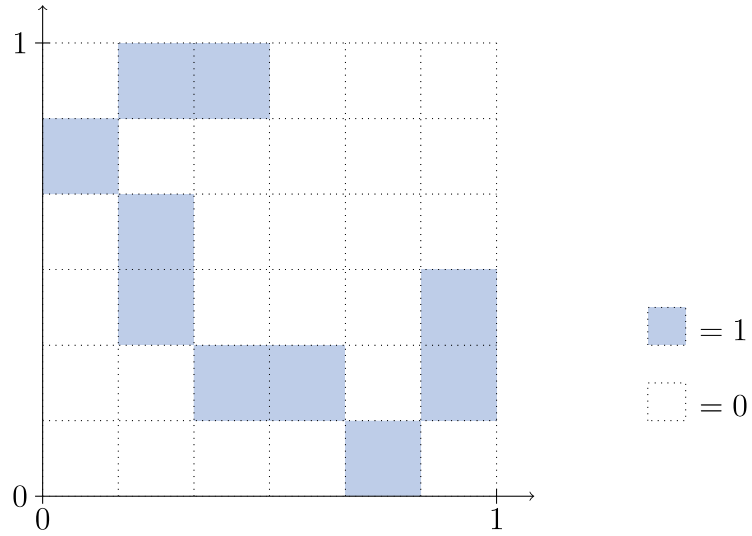

FIGURE 2. The graph function associated to the graph of Figure 1.

It is then easily seen that $t(F,G)=t(F,g)$ for any finite graph F, so a finite graph G can be embedded in the space $\mathcal{W}$ of graph functions in a way that is compatible with graph parameters. There is a reciprocal to this construction, which associates to any graph function w a model of *random* graphs. Fix a graph function w, and for $n\geq 1$, consider a family $(X_1,\ldots,X_n)$ of independent uniform random variables with values in $[0,1]$. We denote $G_n(w)$ the random graph with vertex set $[\![1,n]\!]$, and with i connected to j with probability $w(X_i,X_j)$. Thus, the random variables $X_1,\ldots,X_n$ being drawn, we consider new independent Bernoulli random variables $B_{i\neq j}$ of parameters $w(X_i,X_j)$, and we connect i to j in $G_n(w)$ if and only if $B_{ij}=1$. Again, the laws of these random graphs $G_n(w)$ are invariant under the action of any Lebesgue isomorphism of $[0,1]$ on w.

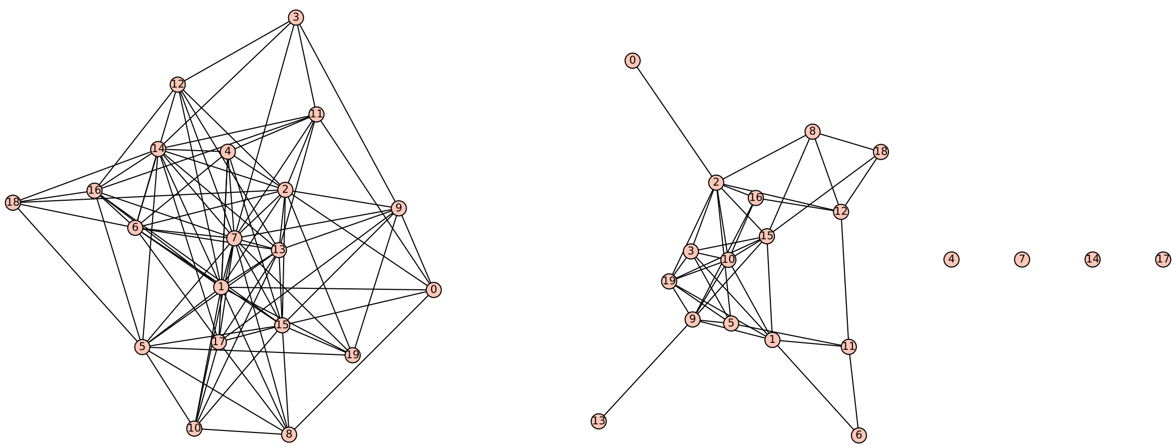

FIGURE 3. Two random graphs of size $n=20$ associated to the graph functions $w(x,y)=\frac{x+y}{2}$ and $w'(x,y)=xy$.

Proposition 4. *If $w \in \mathcal{W}$, then for any $n \geq 1$,*

$$\mathbb{E}[t_0(F, G_n(w))] = t(F, w);$$
$$\mathrm{var}(t(F, G_n(w))) \leq \frac{3\,|V_F|^2}{n}.$$

Proof. Set $k = |V_F|$, and let ϕ be an injective map from $[\![1, k]\!]$ to $[\![1, n]\!]$. Conditionally to the random variables $X_1, \ldots, X_n$, the probability that ϕ is an embedding of F into the random graph $G_n(w)$ is $\prod_{(i,j)\in E_F} w(X_{\phi(i)}, X_{\phi(j)})$. Therefore,

$$\begin{aligned}\mathbb{P}[\phi \text{ is an embedding}] &= \int_{[0,1]^n} \left(\prod_{(i,j)\in E_F} w(x_{\phi(i)}, x_{\phi(j)})\right) dx_1 \cdots dx_n \\ &= \int_{[0,1]^k} \left(\prod_{(i,j)\in E_F} w(x_i, x_j)\right) dx_1 \cdots dx_k = t(F, w).\end{aligned}$$

As a consequence,

$$\mathbb{E}[t_0(F, G_n(w))] = \frac{1}{n^{\downarrow k}} \sum_{\phi \text{ injective map}} t(F, w) = t(F, w).$$

To compute the variance, introduce $F_2 = F \sqcup F$, which is the disjoint union of two copies of F. Then, $\hom(F_2, G) = \hom(F, G) \times \hom(F, G)$, and as a consequence, $t(F_2, G) = (t(F, G))^2$ for any finite graph F. We also have $t(F_2, w) = (t(F, w))^2$ for any graph function w. So, by using Lemma 1,

$$\begin{aligned}\mathbb{E}[(t(F, G_n(w)))^2] = \mathbb{E}[t(F_2, G_n(w))] &\leq \mathbb{E}[t_0(F_2, G_n(w))] + \frac{1}{n}\binom{2k}{2} \\ &\leq t(F_2, w) + \frac{2k^2}{n} = (t(F, w))^2 + \frac{2k^2}{n}; \\ (\mathbb{E}[t(F, G_n(w))])^2 &\geq \left(t(F, w) - \frac{k^2}{2n}\right)^2 \geq (t(F, w))^2 - \frac{k^2}{n}\end{aligned}$$

and $\mathrm{var}(t(F, G_n(w))) \leq \frac{3\,k^2}{n} = \frac{3\,|V_F|^2}{n}$. □

Fix $\varepsilon > 0$, and let n be large enough so that $\frac{|V_F|^2}{2n} < \frac{\varepsilon}{2}$. We then have

$$|\mathbb{E}[t(F, G_n(w))] - t(F, w)| \leq \mathbb{E}[|t(F, G_n(w)) - t_0(F, G_n(w))|] \leq \frac{\varepsilon}{2},$$

and a direct consequence of the previous proposition is

$$\begin{aligned}\mathbb{P}[|t(F, G_n(w)) - t(F, w)| \geq \varepsilon] &\leq \mathbb{P}\left[|t(F, G_n(w)) - \mathbb{E}[t(F, G_n(w))]| \geq \frac{\varepsilon}{2}\right] \\ &\leq \frac{4\,\mathrm{var}(t(F, G_n(w)))}{\varepsilon^2} \leq 12\left(\frac{|V_F|}{\varepsilon}\right)^2 \frac{1}{n}.\end{aligned}$$

So:

Corollary 5. *For any graph function $w \in \mathcal{W}$, the model of random graphs $(G_n(w))_{n\in\mathbb{N}}$ has the property that $t(F, G_n(w))$ converges in probability to $t(F, w)$ for any finite graph F.*

A classical consequence of convergence in probability is the existence of a subsequence that converges almost surely (see [Bil95, Theorem 20.5]). Since the set of isomorphism classes of finite graphs is countable, by diagonal extraction, one can find a subsequence $(G_{n_k}(w))_{k\in\mathbb{N}}$ such that for *any* finite graph F,

$$\lim_{k\to\infty} t(F,G_{n_k}(w)) = t(F,w) \quad \text{almost surely.}$$

In particular, there exists a sequence of graphs $(G_{n_k})_{k\in\mathbb{N}}$ whose observables $t(F,G_{n_k})$ converge to the observables $t(F,w)$, so $(t(F,w))_F$ is indeed a graph parameter. This ends the proof of the first half of Theorem 3.

1.3. **The space of graphons.** We now want to prove the second part of Theorem 3: if a sequence of graphs $(G_n)_{n\in\mathbb{N}}$ has all its observables $t(F,G_n)$ that converge, then the limits of the observables correspond to a graph function $w\in\mathcal{W}$. This is clearly a completeness result, so it is natural to try to detail the topology on $\mathcal{W}$ that is associated to the observables $t(F,\cdot)$. Given $w\in \mathrm{L}^\infty([0,1]^2)$, we set:

$$\|w\|_\square = \sup_{S,T\subset[0,1]} \left|\int_{S\times T} w(x,y)\,dx\,dy\right|.$$

This is a norm on the space $\mathrm{L}^\infty([0,1]^2)$, and one can show that it is equivalent to the norm of operator $\|\cdot\|_{\mathrm{L}^\infty([0,1])\to\mathrm{L}^1([0,1])}$ (here, $\mathrm{L}^\infty([0,1]^2)$ acts on these spaces by convolution).

Definition 6. *The* cut-metric *on graph functions $w\in\mathcal{W}$ is defined by*

$$d_\square(w,w') = \inf_\sigma \|w^\sigma - w'\|_\square,$$

where the infimum runs over Lebesgue isomorphisms σ of the interval $[0,1]$, and where

$$w^\sigma(x,y) = w(\sigma(x),\sigma(y)).$$

Notice that $d_\square(w,w')$ is also the infimum $\inf_{\sigma,\tau}\|w^\sigma-(w')^\tau\|_\square$ over pairs of Lebesgue isomorphisms; as a consequence, $d_\square$ satisfies the triangular inequality. We define an equivalence relation on $\mathcal{W}$ by

$$w\sim w' \iff d_\square(w,w')=0.$$

If ω and ω' are the equivalence classes of the graph functions w and w', then the quotient space $\mathcal{G}=\mathcal{W}/\sim$ is endowed with the distance $\delta_\square(\omega,\omega')=d_\square(w,w')$. We call *graphon* an equivalence class of graph functions in $\mathcal{G}$, and the space of graphons $(\mathcal{G},\delta_\square)$ is a metric space. Furthermore,

- the observables $t(F,\cdot)$,
- and the models of random graphs $(G_n(w))_{n\in\mathbb{N}}$

are invariant by Lebesgue isomorphisms, so they are well-defined on the space of graphons. Then, we have the following fundamental result:

Theorem 7 (Theorem 5.1 in [LS07] and Theorem 3.8 in [Bor+08])**.** *The space of graphons $(\mathcal{G},\delta_\square)$ is a compact metric space. A sequence of graphons $(\omega_n)_{n\in\mathbb{N}}$ converges in this space towards ω if and only if, for any finite finite graph F, $t(F,\omega_n)\to t(F,\omega)$.*

Before we prove Theorem 7, let us see why it implies the second half of Theorem 3. Let $(G_n)_{n\in\mathbb{N}}$ be a sequence of graphs whose observables converge: $\lim_{n\to\infty} t(F,G_n) = t(F)$ for some graph parameter $(t(F))_F$. One identifies the graphs G_n with their graph functions g_n, and then with the graphons γ_n that are the equivalence classes of the functions g_n. By compacity of $\mathcal{G}$, up to extraction, one can assume that $\gamma_n\to\gamma$ for some graphon $\gamma\in\mathcal{G}$. However, this convergence in the space of graphons is equivalent to the convergence of observables, so $t(F)=t(F,\gamma)$. This proves that the graph parameter $(t(F))_F$ comes from a graph function in $\mathcal{W}$ (any graph function in the equivalence class γ).

The proof of the compacity part of Theorem 7 relies on several approximation lemmas in the space of graph functions, which are variants of Szemerédi's regularity lemma (see [Sze78] for the original paper by Szemerédi; [Kom+02] for a survey of the applications of this result in graph theory; and [LS07] for the applications of the regularity lemma to the study of graphons). Let w be a graph function. If Π is a set partition of $[0,1]$ in $\ell = \ell(\Pi)$ measurable parts $P_1, P_2, \ldots, P_\ell$, we denote w_Π the graph function that is constant on each rectangle $P_i \times P_j$, and equal on this rectangle to the average

$$\frac{\int_{P_i \times P_j} w(x,y)\,dx\,dy}{\int_{P_i \times P_j} 1\,dx\,dy}.$$

Lemma 8. *For any graph function $w \in \mathcal{W}$ and any $\varepsilon > 0$, there exists a set partition Π of $[0,1]$ with at most $4^{1/\varepsilon^2}$ parts, such that*

$$\|w - w_\Pi\|_\square \leq \varepsilon.$$

Proof. Fix an integer ℓ and a set partition Π of $[0,1]$ into ℓ measurable parts. If S and T are fixed measurable subsets of $[0,1]$, let us consider the set partition Π' that is generated by Π and by the parts S and T. Thus, Π' is the coarsest set partition that is finer than Π and than the two set partitions $S \sqcup ([0,1] \setminus S)$ and $T \sqcup ([0,1] \setminus T)$. One sees at once that Π' has at most 4ℓ parts. Now, notice that among all *step functions* v on $[0,1]^2$ that are constant on the rectangles associated to the parts of Π', the function $w_{\Pi'}$ is the one that is the closest to w in L^2-norm (this can be seen by computing the derivative of v with respect to its value on a rectangle). Therefore, for any $t \in \mathbb{R}$,

$$\begin{aligned}\|w - w_{\Pi'}\|_{\mathrm{L}^2}^2 &\leq \|w - w_\Pi - t\,1_{S\times T}\|_{\mathrm{L}^2}^2 \\ &\leq \|w - w_\Pi\|_{\mathrm{L}^2}^2 - 2t \int_{S\times T} (w - w_\Pi)(x,y)\,dx\,dy + t^2.\end{aligned}$$

Choosing the optimal $t = \int_{S\times T}(w - w_\Pi)(x,y)\,dx\,dy$, we conclude that

$$\begin{aligned}\left|\int_{S\times T} (w - w_\Pi)(x,y)\,dx\,dy\right|^2 &\leq \|w - w_\Pi\|_{\mathrm{L}^2}^2 - \|w - w_{\Pi'}\|_{\mathrm{L}^2}^2 \\ &\leq \|w_{\Pi'}\|_{\mathrm{L}^2}^2 - \|w_\Pi\|_{\mathrm{L}^2}^2; \\ (\|w - w_\Pi\|_\square)^2 &\leq \sup_{\Pi'} \left(\|w_{\Pi'}\|_{\mathrm{L}^2}^2 - \|w_\Pi\|_{\mathrm{L}^2}^2\right)\end{aligned}$$

with the supremum on the last line that is taken over all set partitions Π' of $[0,1]$ that have at most 4ℓ measurable parts.

Starting from the trivial set partition $\Pi_0 = \{[0,1]\}$ of $[0,1]$, suppose that for any $k \leq \frac{1}{\varepsilon^2}$, one can find recursively a measurable set partition Π_{k+1} of $[0,1]$ with at most $4\ell(\Pi_k)$ measurable parts, and such that

$$\left(\|w_{\Pi_{k+1}}\|_{\mathrm{L}^2}^2 - \|w_{\Pi_k}\|_{\mathrm{L}^2}^2\right) > \varepsilon^2.$$

Then, for any $k \leq \frac{1}{\varepsilon^2}$,

$$\|w_{\Pi_{k+1}}\|_{\mathrm{L}^2}^2 \geq (k+1)\varepsilon^2.$$

However, we also have $\|w\|_{\mathrm{L}^2} \leq 1$ for any graph function, so we obtain a contradiction by choosing $k = \lfloor \frac{1}{\varepsilon^2} \rfloor$. Therefore, there exists $k \leq \frac{1}{\varepsilon^2}$ such that

$$\sup_{\Pi'} \left(\|w_{\Pi'}\|_{\mathrm{L}^2}^2 - \|w_{\Pi_k}\|_{\mathrm{L}^2}^2\right) \leq \varepsilon^2.$$

By the previous argument, $\|w - w_{\Pi_k}\|_\square \leq \varepsilon$, and by construction, $\ell(\Pi_k) \leq 4^k \leq 4^{1/\varepsilon^2}$. □

Lemma 9. *Fix again $w \in \mathcal{W}$ and $\varepsilon > 0$. If k is an integer larger than $2^{20/\varepsilon^2}$, then there exists a set partition Π of $[0,1]$ in k parts of same measure $\frac{1}{k}$, such that*

$$\|w - w_\Pi\|_\square \leq \varepsilon.$$

Proof. By the previous approximation lemma, there exists a set partition Π' into $k' \leq 2^{81/(8\varepsilon^2)}$ parts, such that

$$\|w - w_{\Pi'}\|_\square \leq \frac{4\varepsilon}{9}.$$

By cutting the parts of Π' in smaller blocks, one can then find a measurable set partition Π with exactly k parts, all of the same size, and with at most k' parts that intersect more than one part of Π'. Let R be the union of all these exceptional parts, and u be the step function equal to $w_{\Pi'}$ on $([0,1] \setminus R)^2$, and to 0 on the complement of this set. Notice that the Lebesgue measure of R is smaller than

$$\frac{k'}{k} \leq 2^{-79/(8\varepsilon^2)} \leq \varepsilon^2\, 2^{-79/8}.$$

Then, for any measurable sets S and T,

$$\begin{aligned}\left|\int_{S\times T} (w-u)(x,y)\,dx\,dy\right| &\leq \|w - w_{\Pi'}\|_\square + \left|\int_{(S\times T)\cap[0,1]^2\setminus([0,1]\setminus R)^2} w'_\Pi(x,y)\,dx\,dy\right| \\ &\leq \frac{4\varepsilon}{9} + \sqrt{\lambda([0,1]^2 \setminus ([0,1]\setminus R)^2)} = \frac{4\varepsilon}{9} + \sqrt{1-(1-\lambda(R))^2} \\ &\leq \frac{4\varepsilon}{9} + \sqrt{2\lambda(R)} \leq \left(\frac{4}{9} + 2^{-\frac{71}{16}}\right)\varepsilon \leq \frac{\varepsilon}{2},\end{aligned}$$

so $\|w-u\|_\square \leq \frac{\varepsilon}{2}$. By construction, u is a step function relatively to the set partition Π, hence $u_\Pi = u$. However, for any function in $\mathrm{L}^\infty([0,1]^2)$, $\|w_\Pi\|_\square \leq \|w\|_\square$, so

$$\|w - w_\Pi\|_\square \leq \|w-u\|_\square + \|u - w_\Pi\|_\square \leq \|w-u\|_\square + \|(u-w)_\Pi\|_\square \leq 2\,\|w-u\|_\square \leq \varepsilon. \qquad \square$$

Corollary 10. *There exists a universal sequence of integers $(\ell_j)_{j\geq 1}$, such that for any graph function w, one can find a sequence of measurable set partitions $(\Pi_j)_{j\geq 1}$ with the following properties:*

(1) For any j, Π_{j+1} is a refinement of Π_j, $\ell(\Pi_j) = \ell_j$, and Π_j has all its parts with the same size $\frac{1}{\ell_j}$.

(2) For any j, $\|w - w_{\Pi_j}\|_\square \leq \frac{1}{j}$.

Proof. We can take $\ell_1 = 1$ and $\Pi_1 = \{[0,1]\}$ for any graph function. Suppose that the sequence of integers $\ell_1, \ell_2, \ldots$ is determined up to rank j, and fix a graph function w and the corresponding set partitions $\Pi_1, \ldots, \Pi_j$, that are already constructed by induction hypothesis. In the proof of the previous lemma, we set $\varepsilon = \frac{1}{j+1}$, and choose Π' such that

$$\|w - w_{\Pi'}\|_\square \leq \frac{4\varepsilon}{9}.$$

One can then choose $\Pi = \Pi_{j+1}$ with $\ell_j \times k = \ell_{j+1}$ parts of the same size, that is finer than Π_j, and such that the number of parts of Π that intersect more than one part of $\Pi_j \wedge \Pi'$ is smaller than $\ell_j \times k'$, where $\Pi_j \wedge \Pi'$ is the coarsest common refinement of Π_j and Π'. The proof of the inequality $\|w - w_{\Pi_{j+1}}\|_\square \leq \varepsilon = \frac{1}{j+1}$ is then exactly the same as before, so we have indeed found an integer ℓ_{j+1} independent of w, and then a set partition Π_{j+1} with the properties required. $\square$

Proof of Theorem 7: compacity. Let $(\gamma^n)_{n\in\mathbb{N}}$ be a sequence of graphons. For any n, we fix a representative $g^n \in \mathcal{W}$ of the graphon γ^n, and then a sequence of set partitions $(\Pi_j^n)_{j\geq 1}$ with the properties listed in the previous corollary. Thus,

$$\left\|g^n - (g^n)_{\Pi_j^n}\right\|_\square \leq \frac{1}{j},$$

and moreover, the graph functions $(g^n)_{\Pi_j^n}$ have the following property of averaging: if P, Q are parts of $\Pi_{n,j}$, then the value of $(g_n)_{\Pi_j^n}$ on $P \times Q$ is the average of the values of $(g^n)_{\Pi_{j'}^n}$ on this rectangle, for any $j' \geq j$. This statement is an immediate consequence of the fact that the set partition $\Pi_{j'}^n$ is a refinement of the set partition Π_j^n. Now, as the set partitions Π_j^n have parts with the same size $(\ell_j)^{-1}$, we can also find for any n a Lebesgue isomorphism σ^n that conjugates the parts of Π_j^n to the intervals of size $(\ell_j)^{-1}$ (notice that we can choose a *common* Lebesgue isomorphism σ^n for all the values of j; this is not very hard to see). Then, $g_j^n = ((g^n)_{\Pi_j^n})^{\sigma^n}$ is a function that is constant on all the squares of the grid with mesh size $\frac{1}{\ell_j}$; and the corresponding graphon γ_j^n satisfies

$$\delta_\Box(\gamma^n, \gamma_j^n) \leq \left\| g^n - (g^n)_{\Pi_j^n} \right\|_\Box \leq \frac{1}{j}.$$

Moreover, for any n, the sequence of graph functions $(g_j^n)_{j \geq 1}$ has the same averaging property as stated before. Now, the space of graph functions that are constant on the squares of a fixed grid is isomorphic to a finite product of intervals $[0,1]$, so there is an extraction such that $(g_1^{n_k})_{k \in \mathbb{N}}$ converges on all the squares of the grid with mesh size $(\ell_1)^{-1}$. By diagonal extraction, we can in fact assume that $g_2^{n_k}, g_3^{n_k}, \ldots$ are also convergent. So, there exists an extraction $(n_k)_{k \in \mathbb{N}}$, as well as limits $g_1, g_2, \ldots$ that are constant functions on grids, such that $\lim_{k \to \infty} g_j^{n_k} = g_j$ for any j. Moreover, the limiting graph functions g_j have the same averaging property as before.

If (X, Y) is a uniform random variable in the square $[0,1]^2$, then $(g_j(X,Y))_{j \geq 1}$ is a martingale, because of the averaging property. It is bounded, so it admits a limit almost surely (see [Bil95, Theorem 35.5]). It means that $g_j(x,y) \to g(x,y)$ for almost any $(x,y) \in [0,1]^2$, and some graph function g. Let γ be the graphon corresponding to g, and $\varepsilon > 0$. For j large enough,

$$\delta_\Box(\gamma^{n_k}, \gamma_j^{n_k}) \leq \frac{1}{j} \leq \varepsilon,$$

and we also have $\|g_j - g\|_\Box \leq \|g_j - g\|_{\mathrm{L}^1([0,1]^2)} \leq \varepsilon$ by dominated convergence. Then, j being fixed, for k large enough,

$$\begin{aligned}\delta_\Box(\gamma_j^{n_k}, \gamma) &\leq \left\| g_j^{n_k} - g \right\|_\Box \leq \left\| g_j^{n_k} - g_j \right\|_\Box + \|g_j - g\|_\Box \\ &\leq \left\| g_j^{n_k} - g_j \right\|_\Box + \varepsilon \\ &\leq 2\varepsilon,\end{aligned}$$

so $\delta_\Box(\gamma^{n_k}, \gamma) \leq 3\varepsilon$ for k large enough. This ends the proof of the compacity of the metric space $(\mathcal{G}, \delta_\Box)$. ☐

1.4. **Concentration of the graphon models.** In order to prove the second part of Theorem 7, note first that the observables $t(F, \cdot)$ are continuous with respect to the distance $\delta_\Box$, and even Lipschitz:

Lemma 11. *For any finite graph F and any graph functions w, w',*

$$|t(F,w) - t(F,w')| \leq |E_F| \, \|w - w'\|_\Box.$$

Proof. We enumerate the edges of F as follows: $E_F = \{e_1, e_2, \ldots, e_m\}$ with $e_s = (i_s, j_s)$. Then,

$$\begin{aligned}|t(F,w) - t(F,w')| &= \left| \int_{[0,1]^k} \left(\prod_{s=1}^m w(x_{i_s}, x_{j_s}) - \prod_{s=1}^m w'(x_{i_s}, x_{j_s}) \right) dx_1 \cdots dx_k \right| \\ &\leq \sum_{t=1}^m \left| \int_{[0,1]^k} \left(\prod_{s=1}^{t-1} w'(x_{i_s}, x_{j_s}) \right) (w(x_{i_t}, y_{i_t}) - w'(x_{i_t}, y_{i_t})) \left(\prod_{s=t+1}^m w(x_{i_s}, x_{j_s}) \right) dx_1 \cdots dx_k \right| \\ &\leq m \sup_{0 \leq f, g \leq 1} \left| \int_{[0,1]^2} f(x) g(y) \, (w(x,y) - w'(x,y)) \, dx \, dy \right|,\end{aligned}$$

by integrating on the last line the variables different from x_{i_t} and x_{j_t}. The supremum over pairs of functions (f,g) is then easily seen to be equal to $\|w-w'\|_\Box$. □

As a consequence, for any graphons γ and γ', $|t(F,\gamma)-t(F,\gamma')| \leq |E_F|\,\delta_\Box(\gamma,\gamma')$. A converse of this inequality is:

Proposition 12 (Theorem 3.7 in [Bor+08]). *Let γ and γ' be two graphons in $\mathcal{G}$, such that $|t(F,\gamma)-t(F,\gamma')| \leq 3^{-k^2}$ for any simple graph F on k vertices. Then,*

$$\delta_\Box(\gamma,\gamma') \leq \frac{22}{\sqrt{\log_2 k}}.$$

This proposition and the previous lemma ensure that convergence with respect to the metric $\delta_\Box$ is equivalent to the convergence of all the observables $t(F,\cdot)$, hence the second part of Theorem 7. In turn, Proposition 12 relies on a concentration result for the model of random graphs $(G_n(\gamma))_{n\in\mathbb{N}}$ associated to the graphon γ, which we shall just call *graphon model*. Thus:

Theorem 13 (Theorem 4.7 in [Bor+08]). *Let γ be any graphon in $\mathcal{G}$. One has*

$$\mathbb{E}[\delta_\Box(\gamma,G_k(\gamma))] \leq \frac{5}{\sqrt{\log_2 k}},$$

where a (random) graph $G_k(\gamma)$ is identified with the corresponding graph function and graphon.

Remark. One can show that with probability larger than $1-\mathrm{e}^{-\frac{k^2}{2\log_2 k}}$, the distance $\delta_\Box(\gamma,G_k(\gamma))$ is smaller than $10/\sqrt{\log_2 k}$. For our purpose, it will be sufficient to have a bound on the expectation of the distance.

For the proof of Theorem 13, we refer again to [Bor+08]; the proof uses once more the approximation Lemma 8. Let us then see why Theorem 13 implies Proposition 12.

Proof of Proposition 12. Let w and w' be graph functions in the equivalence classes γ and γ', and $u=\frac{1+w}{2}$, $u'=\frac{1+w'}{2}$. Clearly, $\delta_\Box(w,w')=2\,\delta_\Box(u,u')$. We are going to construct a coupling of the random graphs $G_k(u)$ and $G_k(u')$, such that $G_k(u)=G_k(u')$ with very high probability. To this purpose, we introduce the notion of *induced subgraph* of a graph: a morphism $\phi : F\to G$ gives rise to an induced subgraph if it is injective from V_F to V_G (embedding), and if $\phi(i)\sim\phi(j)$ in G if and only if $i\sim j$ in F. The difference with embeddings is that for an embedding, one can have $\phi(i)\sim\phi(j)$ although $i\not\sim j$ in F. Let $\mathrm{ind}(F,G)$ be the set of embeddings as induced subgraphs of F into G. Then,

$$|\mathrm{emb}(F,G)| = \sum_{F\subset F'} |\mathrm{ind}(F',G)|,$$

where the sum runs over graphs F' with the same vertex set as F, and with more edges. By inclusion-exclusion,

$$|\mathrm{ind}(F,G)| = \sum_{F\subset F'} (-1)^{|E_{F'}|-|E_F|}\,|\mathrm{emb}(F',G)|.$$

If $t_1(F,G)=\frac{|\mathrm{ind}(F,G)|}{|V_G|^{\downarrow |V_F|}}$ is the density of induced subgraphs, then we have similarly

$$t_0(F,G)=\sum_{F\subset F'} t_1(F',G) \qquad ; \qquad t_1(F,G)=\sum_{F\subset F'}(-1)^{|E_{F'}|-|E_F|}\,t_0(F',G).$$

On the other hand, notice that given two graphs G and H with the same number k of vertices, we have $|\mathrm{ind}(G,H)|=0$ unless G and H are isomorphic. Fix a graph F with k vertices. We have by

Proposition 4

$$\begin{aligned}
t(F,u) = \mathbb{E}[t_0(F,G_k(u))] &= \sum_{G \text{ graph on } k \text{ vertices}} \mathbb{P}[G_k(u)=G]\, t_0(F,G) \\
&= \sum_{\substack{F' \,|\, F\subset F' \\ G \text{ graph on } k \text{ vertices}}} \mathbb{P}[G_k(u)=G]\, t_1(F',G) \\
&= \sum_{\substack{F' \,|\, F\subset F' \\ G \text{ isomorphic to } F'}} \mathbb{P}[G_k(u)=G]\, \frac{|\mathrm{aut}(F')|}{k!} \\
&= \sum_{F' \,|\, F\subset F'} \mathbb{P}[G_k(u)=F']\, \frac{|\mathrm{aut}(F')|^2}{k!},
\end{aligned}$$

where $\mathrm{aut}(F')$ is the group of automorphism of the graph F'. Therefore, by inclusion-exclusion,

$$\mathbb{P}[G_k(u)=F] = \frac{k!}{|\mathrm{aut}(F)|^2} \sum_{F' \,|\, F\subset F'} (-1)^{|E_{F'}|-|E_F|}\, t(F',u),$$

and as a consequence,

$$|\mathbb{P}[G_k(u)=F]-\mathbb{P}[G_k(u')=F]| \leq \frac{k!}{|\mathrm{aut}(F)|^2} \sum_{F' \,|\, F\subset F'} |t(F',u)-t(F',u')|$$

$$\sum_F |\mathbb{P}[G_k(u)=F]-\mathbb{P}[G_k(u')=F]| \leq k! \sum_{F,F' \,|\, F\subset F'} |t(F',u)-t(F',u')|.$$

Notice that the left-hand side of the last inequality is twice the total variation distance between the two random graphs $G_k(u)$ and $G_k(u')$. The theory of coupling ensures that there is a way to realise the two random graphs $G_k(u)$ and $G_k(u')$, in other words a common probability space such that $\mathbb{P}[G_k(u) = G_k(u')] = 1 - d_{\mathrm{TV}}(G_k(u), G_k(u'))$ (see Section 4.12 in [GS01]). Thus, if we can compute a good upper bound of the quantity $k! \sum_{F,F' \,|\, F\subset F'} |t(F,u)-t(F,u')|$, then with high probability we shall have $G_k(u) = G_k(u')$, and therefore $\delta_\square(G_k(u), G_k(u')) = 0$.

Since $u = \frac{1+w}{2}$, we have $t(F',u) = 2^{-|E_{F'}|} \sum_{F'' \,|\, F''\subset F'} t(F'',w)$, and therefore

$$|t(F',u)-t(F',u')| \leq 2^{-|E_{F'}|} \sum_{F'' \,|\, F''\subset F'} 3^{-k^2} = 3^{-k^2}.$$

So,

$$2\, d_{\mathrm{TV}}(G_k(u), G_k(u')) \leq k! \sum_{F,F' \,|\, F\subset F'} 3^{-k^2} = k!\, 3^{\frac{k(k-1)}{2}-k^2} = k!\, 3^{-\frac{k(k+1)}{2}};$$

$$\mathbb{P}[G_k(u) \neq G_k(u)'] \leq 3^{-\frac{k}{2}}$$

by using on the last line the trivial inequality $k! \leq 3^{k^2/2}$. This implies

$$\begin{aligned}
\delta_\square(u,u') &\leq \mathbb{E}[\delta_\square(u,G_k(u))] + \mathbb{E}[\delta_\square(G_k(u),G_k(u'))] + \mathbb{E}[\delta_\square(G_k(u'),u')] \\
&\leq \frac{10}{\sqrt{\log_2 k}} + 3^{-\frac{k}{2}} \leq \frac{11}{\sqrt{\log_2 k}}.
\end{aligned}$$

□

An important corollary of the second part of Theorem 7 is:

Corollary 14. *Let $\gamma \in \mathcal{G}$ be any graphon, and $(G_n(\gamma))_{n\in\mathbb{N}}$ be the corresponding graphon model. In the space of graphons $(\mathcal{G}, \delta_\square)$, $G_n(\gamma)$ converges in probability towards γ.*

Proof. Indeed, we saw that there was convergence in probability of all the observables $t(F, G_n(\gamma)) \to t(F, \gamma)$, and the convergence of observables is equivalent to the convergence for the metric. □

To conclude our presentation of the theory of graphons, let us propose a characterisation of the graphon models. If $\gamma \in \mathcal{G}$, then the graphon model $(G_n(\gamma))_{n\in\mathbb{N}}$ has the following properties:

(1) For any permutation $\sigma \in \mathfrak{S}(n)$, the graph $(G_n(\gamma))^\sigma$ obtained by permutation of the n vertices of $G_n(\gamma)$ has the same distribution as $G_n(\gamma)$.

(2) If one removes from $G_n(\gamma)$ the vertex n and all the edges coming from n, then one obtains a random graph on $n-1$ vertices with the same distribution as $G_{n-1}(\gamma)$.

(3) For any subset $S \subset [\![1, n]\!]$, the graphs induced by $G_n(\gamma)$ on S and on its complement $[\![1, n]\!] \setminus S$ are independent.

Theorem 15 (Theorem 2.7 in [LS06]). *A model of random graphs $(G_n)_{n\in\mathbb{N}}$ has the three properties above if and only if it is a graphon model.*

2. Permutons and their topology

2.1. Permutations and patterns. In [Hop+13], Hoppen, Kohayakawa, Moreira, Ráth and Sampaio developed a theory analoguous to the theory of graphons, and that allowed them to study sequences of (random) permutations, and their densities of patterns. Recall that a *permutation* of size n is a bijection $\sigma : [\![1, n]\!] \to [\![1, n]\!]$. The set of all permutations of size n is the *symmetric group* of order n, denoted $\mathfrak{S}(n)$, and of cardinality $n!$. If $\tau \in \mathfrak{S}(k)$ and $\sigma \in \mathfrak{S}(n)$ with $k \leq n$, we say that τ is a *pattern* in σ if there exists a part $\{a_1 < a_2 < \cdots < a_k\} \subset [\![1, n]\!]$ such that $\sigma(a_i) < \sigma(a_j)$ if and only if $\tau(i) < \tau(j)$. This definition is better understood on a picture: if one draws the graph of σ, then one can isolate points $a_1 < a_2 < \cdots < a_k$ such that the restriction of the graph of σ to these points is the graph of the permutation τ; see Figure 4 hereafter.

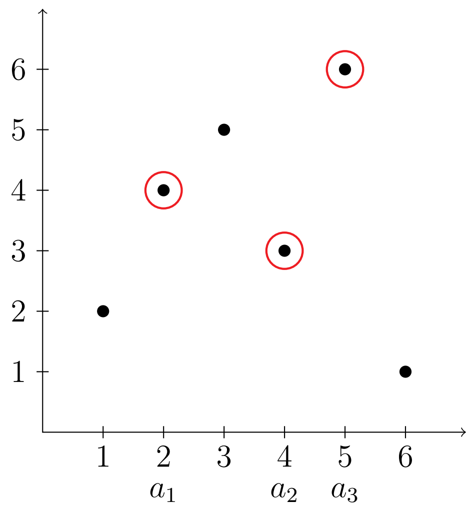

FIGURE 4. The permutation 213 is a pattern in $\sigma = 245361$.

As for graphs, we can define the *pattern density* of τ in σ by

$$t(\tau,\sigma) = \frac{|\mathrm{patt}(\tau,\sigma)|}{\binom{n}{k}},$$

where the numerator of this fraction is the number of parts $\{a_1 < \cdots < a_k\}$ of $[\![1,n]\!]$ that make appear τ as a pattern of σ. We then have the analogue of Definition 2:

Definition 16. *Let $(\sigma_n)_{n\in\mathbb{N}}$ be a sequence of permutations of arbitrary order. One says that $(\sigma_n)_{n\in\mathbb{N}}$ converges if $|\sigma_n|$ goes to infinity, and if for any fixed permutation τ, the density of patterns $t(\tau,\sigma_n)$ admits a limit when n goes to infinity.*

We also call *permutation parameter* a family of real numbers $(t(\tau))_{\tau \text{ permutation}}$ indexed by the permutations $\tau \in \bigsqcup_{n\in\mathbb{N}} \mathfrak{S}(n)$, such that there exists a sequence of permutations $(\sigma_n)_{n\in\mathbb{N}}$ with $|\sigma_n| \to +\infty$ and

$$\lim_{n\to\infty} t(\tau,\sigma_n) = t(\tau)$$

for any τ. Again, we shall present a theory that allows one to identify all the permutation parameters.

2.2. Probability measures on the square and permutons.

Denote $\mathcal{M}([0,1]^2)$ the set of borelian probability measures on the square $[0,1]^2$. It is a topological space for the topology of weak convergence of measures; and this topology is metrizable and yields a compact space, see [Bil69]. Let p_1 and p_2 be the two projections $[0,1]^2 \to [0,1]$ associated to the first and second coordinates. These are continuous maps, which yield continuous maps $p_{1,*}$ and $p_{2,*}$ from $\mathcal{M}([0,1]^2)$ to $\mathcal{M}([0,1])$.

Definition 17. *A* permuton *is a probability measure $\pi \in \mathcal{M}([0,1]^2)$, such that $p_{1,*}(\pi) = p_{2,*}(\pi) = \lambda$ is the Lebesgue measure on $[0,1]$.*

Since $p_{1,*}$ and $p_{2,*}$ are continuous, the space of permutons $\mathcal{P}$ is the reciprocal image of a point by a continuous map, hence is closed, and a compact subspace of $\mathcal{M}([0,1]^2)$ for the topology of weak convergence.

Let $(x_1,y_1),\ldots,(x_k,y_k)$ be a family of points in the square $[0,1]^2$. We say that these points are in a general configuration if all the x_i's are distinct, and if all the y_i's are also distinct. To a general family of k points, we can associate a unique permutation $\tau \in \mathfrak{S}(k)$ with the following property: if $\psi_1 : \{x_1,\ldots,x_k\} \to [\![1,k]\!]$ and $\psi_2 : \{y_1,\ldots,y_k\} \to [\![1,k]\!]$ are increasing bijections, then

$$\tau(\psi_1(x_i)) = \psi_2(y_i)$$

for any $i \in [\![1,k]\!]$. We then say that τ is the *configuration* of the set of points; and we denote $\tau = \mathrm{conf}((x_1,y_1),\ldots,(x_k,y_k))$. This notion allows one to define the pattern density of a permuton π. If τ is a permutation of size k, we set

$$t(\tau,\pi) = \int_{([0,1]^2)^k} 1_{\mathrm{conf}((x_1,y_1),\ldots,(x_k,y_k))=\tau}\, \pi^{\otimes k}(dx_1,dy_1,\ldots,dx_k,dy_k).$$

One can give a probabilistic interpretation to this definition. Let $(X_1,Y_1),\ldots,(X_k,Y_k)$ be independent random points in $[0,1]$, all following the same law π. Since the marginal laws of π on $[0,1]$ are the uniform laws, with probability 1, the random family of points $(X_1,Y_1),\ldots,(X_k,Y_k)$ is in a general configuration. Then,

$$t(\tau,\pi) = \mathbb{P}[\mathrm{conf}((X_1,Y_1),\ldots,(X_k,Y_k)) = \tau].$$

Now, the analogue of Theorem 3 in the setting of permutations is:

Theorem 18 (Theorem 1.6 in [Hop+13]). *A family $(t(\tau))_\tau$ is a permutation parameter if and only if there exists a permuton π such that $t(\tau,\pi) = t(\tau)$ for any permutation τ.*

Again, the easy part of Theorem 18 is the construction of permutations that converge to π for any $\pi \in \mathcal{P}$. Given an integer n and a permuton π, we denote $\sigma_n(\pi)$ the random permutation of size n that is the configuration of independent random points $(X_1,Y_1),\ldots,(X_n,Y_n)$ in the square, all chosen according to the probability measure π.

Proposition 19. *If $\pi \in \mathcal{P}$ and $\tau \in \mathfrak{S}(k)$, then for any $n \geq 2k$,*

$$\mathbb{E}[t(\tau,\sigma_n(\pi))] = t(\tau,\pi);$$
$$\mathrm{var}(t(\tau,\sigma_n(\pi))) \leq \frac{k^2}{n}.$$

Proof. Notice that if $((X_1,Y_1),\ldots,(X_n,Y_n))$ follows the law $\pi^{\otimes n}$, then for any part $\{a_1 < a_2 < \cdots < a_k\}$, the family of points $((X_{a_1},Y_{a_1}),\ldots,(X_{a_k},Y_{a_k}))$ follows the law $\pi^{\otimes k}$. Therefore,

$$\begin{aligned}\mathbb{E}[t(\tau,\sigma_n(\pi))] &= \frac{1}{\binom{n}{k}} \sum_{\{a_1<\cdots<a_k\}\subset [\![1,n]\!]} \mathbb{P}[\mathrm{conf}((X_{a_1},Y_{a_1}),\ldots,(X_{a_k},Y_{a_k})) = \tau] \\ &= \frac{1}{\binom{n}{k}} \sum_{\{a_1<\cdots<a_k\}\subset [\![1,n]\!]} t(\tau,\pi) \\ &= t(\tau,\pi).\end{aligned}$$

To compute the variance, we introduce the random variables $C_{A,\tau}$, defined as follows: if $A = \{a_1 < a_2 < \cdots < a_k\}$, then

$$C_{A,\tau} = \begin{cases} 1 & \text{if } \mathrm{conf}((X_{a_1},Y_{a_1}),\ldots,(X_{a_k},Y_{a_k})) = \tau, \\ 0 & \text{otherwise.}\end{cases}$$

We then have to compute

$$\mathbb{E}\big[(t(\tau,\sigma_n(\pi)))^2\big] = \frac{1}{\binom{n}{k}^2} \sum_{A,B} \mathbb{E}[C_{A,\tau}C_{B,\tau}],$$

where the sum runs over pairs of subsets (A,B) of size k in $[\![1,n]\!]$. Suppose first that A and B are disjoint. Then, $C_{A,\tau}$ and $C_{B,\tau}$ are independent, since they involve independent families of points. So, the part of the sum that corresponds to disjoint subsets is

$$\frac{1}{\binom{n}{k}^2} \sum_{A,B\,|\,A\cap B=\emptyset} \mathbb{E}[C_{A,\tau}]\,\mathbb{E}[C_{B,\tau}] = \frac{1}{\binom{n}{k}^2} \sum_{A,B\,|\,A\cap B=\emptyset} (t(\tau,\pi))^2 = \frac{\binom{n-k}{k}}{\binom{n}{k}}\,(t(\tau,\pi))^2.$$

On the other hand, if A and B are not disjoint, then we can still bound $\mathbb{E}[C_{A,\tau}\,C_{B,\tau}]$ by 1. Therefore,

$$\begin{aligned}\mathbb{E}[(t(\tau,\sigma_n(\pi)))^2] &\leq \frac{\binom{n-k}{k}}{\binom{n}{k}}\,(t(\tau,\pi))^2 + \frac{\binom{n}{k}-\binom{n-k}{k}}{\binom{n}{k}} \\ \mathrm{var}(t(\tau,\sigma_n(\pi))) &\leq \frac{\binom{n}{k}-\binom{n-k}{k}}{\binom{n}{k}}(1-(t(\tau,\pi))^2) \leq \frac{\binom{n}{k}-\binom{n-k}{k}}{\binom{n}{k}} = 1 - \frac{(n-k)^{\downarrow k}}{n^{\downarrow k}}.\end{aligned}$$

The right-hand side of the last inequality is the probability that a random arrangement $(a_1,\ldots,a_k)$ in $[\![1,n]\!]$ meets $[\![1,k]\!]$. This probability is smaller than the sum of probabilities $\mathbb{P}[a_i \in [\![1,k]\!]] = \frac{k}{n}$, hence it is smaller than $\frac{k^2}{n}$. ☐

Corollary 20. *For any permuton π, and any permutation τ, $(t(\tau,\sigma_n(\pi)))_{n\in\mathbb{N}}$ converges in probability to $t(\tau,\pi)$.*

Then, the same argument as for graphons allows one to construct a sequence of random permutations whose observables $t(\tau,\cdot)$ converge almost surely to $t(\tau,\pi)$. In particular, for any $\pi \in \mathcal{P}$, $(t(\tau,\pi))_\tau$ is a permutation parameter.

2.3. **Convergence in the space of permutons.** To prove the second part of Theorem 18, we shall use the following topological result:

Theorem 21. *Let $(\pi_n)_{n\in\mathbb{N}}$ be a sequence of permutons. The following are equivalent:*

(1) The sequence $(\pi_n)_{n\in\mathbb{N}}$ converges weakly to π.

(2) The rectangular distance

$$d_\square(\pi_n,\pi) = \sup_{\substack{0\leq a<b\leq 1\\ 0\leq c<d\leq 1}} |\pi_n([a,b]\times[c,d]) - \pi([a,b]\times[c,d])|$$

goes to 0.

(3) For any permutation τ, $t(\tau,\pi_n)$ converges towards $t(\tau,\pi)$.

Let us first explain why this implies the second part of Theorem 18. If σ is a permutation of size n, then one can associate to it a canonical permuton, namely, the measure π_σ on $[0,1]^2$ with density

$$f_\sigma(x,y) = n\, 1_{\sigma(\lceil nx\rceil)=\lceil ny\rceil}.$$

For any x, the set of y's such that $f_\sigma(x,y)=n$ has measure $\frac{1}{n}$, so

$$\frac{d(p_{1,*}(\pi_\sigma))(x)}{dx} = \int_{y=0}^1 f_\sigma(x,y)\,dy = 1$$

hence $p_{1,*}(\pi_\sigma)=\lambda$. Similarly, $p_{2,*}(\pi_\sigma)=\lambda$, and π_σ is indeed a measure whose marginal laws are uniform. We refer to Figure 5 for an example.

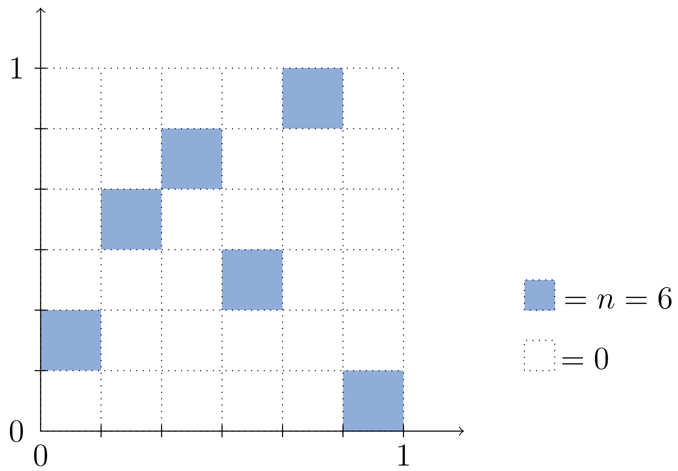

FIGURE 5. The density of the permuton π_σ associated to the permutation $\sigma = 245361$.

Consider now a permutation τ of size $k \leq n$.

Lemma 22. *We have*

$$|t(\tau,\sigma) - t(\tau,\pi_\sigma)| \leq \frac{1}{n}\binom{k}{2}.$$

Proof. Let $(X_1,Y_1),\ldots,(X_k,Y_k)$ be independent random variables with law π_σ; their configuration is τ with probability $t(\tau,\pi_\sigma)$. If $n_i=\lceil nX_i\rceil$, then $\sigma(n_i)=\lceil nY_i\rceil$ by definition of the probability distribution π_σ. We introduce the two following events:

$$A=\{\mathrm{conf}((X_1,Y_1),\ldots,(X_k,Y_k))=\tau\};$$
$$B=\{\forall 1\leq i<j\leq k,\ n_i\neq n_j\}.$$

We then have $\mathbb{P}[A|B]-\mathbb{P}[A]=\mathbb{P}[A|B](1-\mathbb{P}[B])$, hence

$$|\mathbb{P}[A|B]-\mathbb{P}[A]|\leq 1-\mathbb{P}[B]=\mathbb{P}[B^{\mathrm{c}}]\leq\sum_{1\leq i<j\leq k}\mathbb{P}[n_i=n_j]=\frac{1}{n}\binom{k}{2}$$

since the X_i's are uniformly distributed on $[0,1]$ and independent. By the previous discussion, $\mathbb{P}[A]=t(\tau,\pi_\sigma)$. On the other hand, conditionnally to B, the random vector $(n_1,\ldots,n_k)$ is uniformly distributed on the set of arrangements of size k in $[\![1,n]\!]$, and then A is equivalent to the fact that this arrangement allows one to read τ as a pattern of σ. So, $\mathbb{P}[A|B]=t(\tau,\sigma)$, which ends the proof. □

Consider now a sequence of permutations $(\sigma_n)_{n\in\mathbb{N}}$ such that $|\sigma_n|\to\infty$. Since $\mathcal{P}$ is a compact set for the topology of weak convergence of probability measures, up to extraction, we can assume that $\pi_{\sigma_n}\to\pi$ in the sense of weak convergence, where π is some permuton. By Theorem 21, this is equivalent to the fact that $t(\tau,\pi_{\sigma_n})\to t(\tau,\pi)$ for any τ, and by the previous lemma, we have in fact $t(\tau,\sigma_n)\to t(\tau,\pi)$. Hence, any permutation parameter corresponds indeed to a permuton $\pi\in\mathcal{P}$, which ends the proof of Theorem 18. Let us now attack the proof of Theorem 21. We start with:

Proof of Theorem 21: (1) $\Leftrightarrow$ (2). Suppose that $(\pi_n)_{n\in\mathbb{N}}$ is a sequence of permutons that converges to π with respect to the rectangular distance. We fix a continuous function f on $[0,1]^2$, and we want to show that $\pi_n(f)$ converges to $\pi(f)$. If $\varepsilon>0$, then by compacity of $[0,1]^2$, f is uniformly continuous and there exists a partition of $[0,1]^2$ in N^2 small squares S_i of size $\frac{1}{N}$, such that

$$\forall i,\ \sup_{p,q\in S_i}|f(p)-f(q)|\leq\varepsilon.$$

Consequently, there exists an approximation f_ε of f that is constant on each of the squares S_i, and such that $\|f_\varepsilon-f\|_\infty\leq\varepsilon$ and $\|f_\varepsilon\|_\infty\leq\|f\|_\infty$. Then,

$$\begin{aligned}|\pi_n(f)-\pi(f)|&\leq 2\varepsilon+|\pi_n(f_\varepsilon)-\pi(f_\varepsilon)|\\&\leq 2\varepsilon+\sum_{i=1}^{N^2}|f_\varepsilon(S_i)|\,|\pi_n(S_i)-\pi(S_i)|\\&\leq 2\varepsilon+N^2\,\|f\|_\infty\,d_\square(\pi_n,\pi),\end{aligned}$$

so $\lim_{n\to\infty}\pi_n(f)=\pi(f)$. So, the convergence with respect to $d_\square$ is stronger than the weak convergence of probability measures.

Conversely, suppose that $(\pi_n)_{n\in\mathbb{N}}$ converges weakly towards π. Since π_n and π are permutons, their marginal laws are uniform, and in particular they do not have atoms; therefore, for any rectangle $R=[a,b]\times[c,d]$, $\pi_n(\partial R)=\pi(\partial R)=0$. Then, by Portmanteau's theorem (*cf.* [Bil69, Section 2]), $\lim_{n\to\infty}\pi_n(R)=\pi(R)$. Introduce the bivariate cumulative generating functions $F_n(x,y)=\pi_n([0,x]\times[0,y])$ and $F(x,y)=\pi([0,x]\times[0,y])$. The sequence of functions $(F_n)_{n\in\mathbb{N}}$ converges pointwise to F, and on the other hand, these functions are increasing in both variables. Fix an integer N, and n_0 such that for any point $(\frac{i}{N},\frac{j}{N})$ of the grid with mesh size $\frac{1}{N}$, and any $n\geq n_0$,

$$\left|F_n\left(\frac{i}{N},\frac{j}{N}\right)-F\left(\frac{i}{N},\frac{j}{N}\right)\right|\leq\frac{1}{N}.$$

Then, for any (x,y) in $[0,1]$, if $\frac{i}{N} \leq x \leq \frac{i+1}{N}$ and $\frac{j}{N} \leq y \leq \frac{j+1}{N}$, then

$$\begin{aligned}
F_n(x,y) - F(x,y) &\leq F_n\left(\frac{i+1}{N}, \frac{j+1}{N}\right) - F\left(\frac{i}{N}, \frac{j}{N}\right) \\
&\leq \frac{1}{N} + \left(F\left(\frac{i+1}{N}, \frac{j+1}{N}\right) - F\left(\frac{i+1}{N}, \frac{j}{N}\right)\right) + \left(F\left(\frac{i+1}{N}, \frac{j}{N}\right) - F\left(\frac{i}{N}, \frac{j}{N}\right)\right) \\
&\leq \frac{1}{N} + \pi\left(\left[0, \frac{i+1}{N}\right] \times \left[\frac{j}{N}, \frac{j+1}{N}\right]\right) + \pi\left(\left[\frac{i}{N}, \frac{i+1}{N}\right] \times \left[0, \frac{j}{N}\right]\right) \\
&\leq \frac{1}{N} + \pi\left([0,1] \times \left[\frac{j}{N}, \frac{j+1}{N}\right]\right) + \pi\left(\left[\frac{i}{N}, \frac{i+1}{N}\right] \times [0,1]\right) = \frac{3}{N},
\end{aligned}$$

by using on the last line the fact that π has uniform marginal laws. Similarly, one can show that $F_n(x,y) - F(x,y) \geq -\frac{3}{N}$, so for any N, one can find n_0 such that

$$\sup_{n \geq n_0} \sup_{x,y \in [0,1]} |F_n(x,y) - F(x,y)| \leq \frac{3}{N}.$$

However, the rectangular distance is directly related to this quantity, because

$$\pi_n([a,b] \times [c,d]) = F_n(c,d) - F_n(c,b) - F_n(a,d) + F_n(a,b),$$

and similarly for π and F. Therefore, $d_\Box(\pi_n, \pi) \to 0$, and the proof of the equivalence (1) $\Leftrightarrow$ (2) is completed. □

For the other equivalences of Theorem 21, we shall use the following lemma:

Lemma 23 (Lemma 5.1 in [Hop+13]). *Let π and π' be two permutons. If $t(\tau,\pi) = t(\tau,\pi')$ for any permutation τ, then $\pi = \pi'$ in $\mathcal{P}$.*

Sketch of proof. Let $F(x,y)$ be the bivariate cumulative distribution function of π. This function determines the probabilities under π of any rectangle $[a,b] \times [c,d] \subset [0,1]^2$, and therefore it determines π in $\mathcal{P} \subset \mathcal{M}([0,1]^2)$. So, it suffices to show that one can reconstruct F from the family $(t(\tau,\pi))_\tau$. However, if one knows $t(\tau,\pi)$ for any τ, then one knows the distribution of the random permutation $\sigma_n(\pi)$ for any $n \in \mathbb{N}$. As before, F is increasing in both variables, and it has the following regularity property:

$$\begin{aligned}
F(x+\varepsilon, y+\varepsilon) &= \pi([0, x+\varepsilon] \times [0, y+\varepsilon]) \\
&\leq \pi([0,x] \times [0,y]) + \pi([x, x+\varepsilon] \times [0, y+\varepsilon]) + \pi([0, x+\varepsilon] \times [y, y+\varepsilon]) \\
&\leq F(x,y) + \pi([x, x+\varepsilon] \times [0,1]) + \pi([0,1] \times [y, y+\varepsilon]) = F(x,y) + 2\varepsilon.
\end{aligned}$$

Set

$$F_n(x,y) = \frac{1}{n} \sum_{i=1}^{\lceil nx \rceil} 1_{(\sigma_n(\pi))(i) \leq \lceil ny \rceil},$$

which is a random permutation whose distribution is entirely determined by the observables $t(\tau,\pi)$. If $(X_n, Y_n)_{n \in \mathbb{N}}$ is a sequence of independent points of $[0,1]^2$ under π, denote $X_1^* < X_2^* < \cdots < X_n^*$ the increasing reordering of the X_i's, and $Y_1^* < Y_2^* < \cdots < Y_n^*$ the increasing reordering of the Y_i's. Then, with $k = \lceil nx \rceil$ and $l = \lceil ny \rceil$,

$$F_n(x,y) = \frac{1}{n} \sum_{i=1}^{n} 1_{(X_i < X_k^* \text{ and } Y_i < Y_l^*)}.$$

By using the Hoeffding inequalities, one can show that

$$\mathbb{P}\left[F_n(x,y) > F\left(\frac{k}{n}, \frac{l}{n}\right) + 3n^{-1/4}\right] \leq 3\,\mathrm{e}^{-2\sqrt{n}}.$$

For the same reasons,

$$\mathbb{P}\left[F_n(x,y) < F\left(\frac{k}{n},\frac{l}{n}\right) - 3n^{-1/4}\right] \leq 3\,\mathrm{e}^{-2\sqrt{n}}.$$

and by using the regularity properties of F_n and F, this implies that $F_n(x,y)$ converges in probability to $F(x,y)$, hence that F can be reconstructed from the observables $t(\tau,\pi)$. We refer to [Hop+13, Lemma 4.2] for the proof of the concentration inequality. □

Proof of Theorem 21: (1) ⇔ (3). Suppose that $(\pi_n)_{n\in\mathbb{N}}$ is a sequence of permutons that converges weakly to π, and fix a permutation τ of size k. If $((X_1^n,Y_1^n),\ldots,(X_k^n,Y_k^n))$ is a family of k independent points of $[0,1]$ chosen according to $(\pi_n)^{\otimes k}$, then we have the convergence in distribution of this family towards the law $\pi^{\otimes k}$. Now, the set of families $((x_1,y_1),\ldots,(x_k,y_k))$ in $([0,1]^2)^k$ with configuration τ has its boundary which has a measure 0 under $\pi^{\otimes k}$. Indeed, on the boundary of this set, $x_i = x_j$ or $y_i = y_j$ for some pair of indices (i,j), and this event has probability 0, because under $\pi^{\otimes k}$, the vectors $(x_1,\ldots,x_k)$ and $(y_1,\ldots,y_k)$ follow the uniform law λ^k on $[0,1]^k$, hence have distinct coordinates with probability 1. So, by Portmanteau's theorem,

$$\lim_{n\to\infty} \mathbb{P}[\mathrm{conf}((X_1^n,Y_1^n),\ldots,(X_k^n,Y_k^n)) = \tau] = \mathbb{P}[\mathrm{conf}((X_1,Y_1),\ldots,(X_k,Y_k)) = \tau],$$

where $((X_1,Y_1),\ldots,(X_k,Y_k))$ follows the law $\pi^{\otimes k}$. These probabilities can be rewritten as $t(\tau,\pi_n)$ and $t(\tau,\pi)$, so (1) ⇒ (3).

Conversely, suppose that we have the convergence of observables $t(\tau,\pi_n) \to t(\tau,\pi)$ for any permutation τ. If $(\pi_{n_k})_{k\in\mathbb{N}}$ is a subsequence of $(\pi_n)_{n\in\mathbb{N}}$ that converges weakly, then its limit π' satisfies $t(\tau,\pi') = t(\tau,\pi)$ for any permutation τ, so by Lemma 23, $\pi' = \pi$. The unicity of the limit of any convergent subsequence, and the compacity of $\mathcal{P}$ imply now that $\pi_n \to \pi$ in the sense of weak convergence. □

Again, an important corollary of the previous discussion is:

Corollary 24. *Let $\pi \in \mathcal{P}$ be any permuton, and $(\sigma_n(\pi))_{n\in\mathbb{N}}$ be the corresponding permuton model. In the space of permutons $\mathcal{P}$, we have the convergence in probability $\sigma_n(\pi) \to \pi$, where $\sigma_n(\pi)$ is identified with its canonical permuton as in Figure 5.*

Proof. We know that in the sense of convergence of observables, the permutations $\sigma_n(\pi)$ converge in probability towards π. By Lemma 22, the permutons associated to the permutations $\sigma_n(\pi)$ also converge in the sense of observables towards π. Finally, the convergence of observables is equivalent to the weak convergence by Theorem 21. □

Remark. The theory of permutons is sensibly easier than the theory of graphons, for two reasons: one does not have the problem of identifiability of graphons (one does not need to take a quotient space $\mathcal{G} = \mathcal{W}/\sim$), and the compacity of the space is immediately granted by standard results. On the other hand, a small difficulty that is specific to the theory of permutons is the following: if σ is a permutation and π_σ is the associated permuton, then the observables of σ are not exactly the same as the observables of π_σ (see Lemma 22).

References

[Bil69] P. Billingsley. *Convergence of Probability Measures*. John Wiley and Sons, 1969.

[Bil95] P. Billingsley. *Probability and Measure*. 3rd. John Wiley and Sons, 1995.

[Bor+08] C. Borgs, J. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi. “Convergent sequences of dense graphs I. Subgraph frequencies, metric properties and testing”. In: *Adv. Math.* 219.6 (2008), pp. 1801–1851.

[GS01] G. Grimmett and D. Stirzaker. *Probability and Random Processes*. 3rd. Oxford University Press, 2001.

[Hop+13] C. Hoppen, Y. Kohayakawa, C. G. Moreira, B. Ráth, and R. M. Sampaio. “Limits of permutation sequences”. In: *Journal of Combinatorial Theory, Series B* 103.1 (2013), pp. 93–113.

[Kom+02] J. Komlós, A. Shokoufandeh, M. Simonovits, and E. Szemerédi. “The regularity Lemma and its applications in graph theory”. In: *Theoretical aspects of computer science (Tehran, 2000)*. Vol. 2292. Lecture Notes in Computer Science. Springer-Verlag, 2002, pp. 84–112.

[LS06] L. Lovász and B. Szegedy. “Limits of dense graph sequences”. In: *Journal of Combinatorial Theory, Series B* 96 (2006), pp. 933–957.

[LS07] L. Lovász and B. Szegedy. “Szemerédi’s lemma for the analyst”. In: *Geom. Func. Anal.* 17 (2007), pp. 252–270.

[Sze78] E. Szemerédi. “Regular partitions of graphs”. In: *Problèmes combinatoires et théorie des graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976)*. 1978, pp. 399–401.