4

Graph Limits

Chapter Highlights

- An analytic language for studying dense graphs
- Convergence and limit for a sequence of graphs
- Compactness of the graphon space with respect to the cut metric
- Applications of compactness
- Equivalence of cut metric convergence and left-convergence

The theory of graph limits was developed by Lovász and his collaborators in a series of works starting around 2003. The researchers were motivated by questions about very large graphs from several different angles, including from combinatorics, statistical physics, computer science, and applied math. Graph limits give an analytic framework for analyzing large graphs. The theory offers both a convenient mathematical language as well as powerful theorems.

Motivation

Suppose we live in a hypothetical world where we only had access to rational numbers and had no language for irrational numbers. We are given the following optimization problem:

$$\text{minimize } x^3 - x \text{ subject to } 0 \le x \le 1.$$

The minimum occurs at $x = 1/\sqrt{3}$, but this answer does not make sense over the rationals. With only access to rationals, we can state a progressively improving sequence of answers that converge to the optimum. This is rather cumbersome. It is much easier to write down a single real number expressing the answer.

Now consider an analogous question for graphs. Fix some real $p \in [0, 1]$. We want to

$$\begin{aligned} \text{minimize} \quad & (\#\text{ closed walks of length } 4)/n^4 \\ \text{among} \quad & n\text{-vertex graphs with} \ge pn^2/2 \text{ edges.} \end{aligned}$$

We know from Proposition 3.1.14 that every n-vertex graph with edge density $\ge p$ has at least $n^4 p^4$ closed walks of length 4. On the other hand, every sequence of quasirandom graphs with edge density $p + o(1)$ has $p^4 n^4 + o(n^4)$ closed walks of length 4. It follows that the minimum (or rather, infimum) is p^4 and is attained not by any single graph, but rather by a sequence of quasirandom graphs.

One of the purposes of graph limits is to provide an easy-to-use mathematical object that captures the limit of such graph sequences. The central object in the theory of dense graph limits is called a **graphon** (the word comes from combining *graph* and *function*), to be defined shortly. Graphons can be viewed as an analytic generalization of graphs.

Here are some questions that we will consider:

(1) What does it mean for a sequence of graphs (or graphons) to converge?
(2) Are different notions of convergence equivalent?
(3) Does every convergent sequence of graphs (or graphons) have a limit?

Note that it is possible to talk about convergence without a limit. In a first real analysis course, one learns about a ***Cauchy sequence*** in a metric space $(\mathcal{X}, d)$, which is some sequence $x_1, x_2, \dots \in \mathcal{X}$ such that for every $\varepsilon > 0$, there is some N so that $d(x_m, x_n) < \varepsilon$ for all $m, n \geq N$. For instance, one can have a Cauchy sequence without a limit in $\mathbb{Q}$. A metric space is ***complete*** if every Cauchy sequence has a limit. The ***completion*** of $\mathcal{X}$ is some complete metric space $\widetilde{\mathcal{X}}$ such that $\mathcal{X}$ is isometrically embedded in $\widetilde{\mathcal{X}}$ as a dense subset. The completion of $\mathcal{X}$ is in some sense the smallest complete space containing $\mathcal{X}$. For example, $\mathbb{R}$ is the completion of $\mathbb{Q}$. Intuitively, the completion of a space fills in all of its gaps. A basic result in analysis says that every space has a unique completion.

Here is a key result about graph limits that we will prove:

> The space of graphons is compact, and is the completion of the set of graphs.

To make this statement precise, we also need to define a notion of similarity (i.e., distance) between graphs, and also between graphons. We will see two different notions, one based on the *cut metric*, and another based on *subgraph densities*. Another important result in the theory of graph limits is that these two notions are equivalent. We will prove it at the end of the chapter once we have developed some tools.

4.1 Graphons

Here is the central object in the theory of dense graph limits.

Definition 4.1.1 (Graphon)
A ***graphon*** is a symmetric measurable function $W : [0,1]^2 \to [0,1]$. Here ***symmetric*** means $W(x,y) = W(y,x)$ for all x, y.

***Remark* 4.1.2.** More generally, we can consider an arbitrary probability space Ω and study symmetric measurable functions $\Omega \times \Omega \to [0,1]$. In practice, we do not lose much by restricting to $[0,1]$.

We will also sometimes consider symmetric measurable functions $[0,1]^2 \to \mathbb{R}$ (e.g., arising as the difference between two graphons). Such an object is sometimes called a ***kernel*** in the literature.

***Remark* 4.1.3** (Measure theoretic technicalities)**.** We try to sweep measure theoretic technicalities under the rug in order to focus on key ideas. If you have not seen measure theory before, do not worry. Just view "measure" as lengths of intervals or areas of boxes (or countable unions thereof) in the most natural sense. We always ignore measure zero differences. For example, we shall treat two graphons as the same if they only differ on a measure zero subset of the domain.

Turning a Graph into a Graphon

Here is a procedure to turn any graph G into a graphon W_G:

(1) Write down the adjacency matrix A_G of the graph;
(2) Replace the matrix by a black-and-white pixelated picture on $[0,1]^2$, by turning every 1-entry into a black square and every 0-entry into a white square.
(3) View the resulting picture as a graphon $W_G\colon [0,1]^2 \to [0,1]$ (with the axes labeled like a matrix with $x \in [0,1]$ running from top to bottom and $y \in [0,1]$ running from left to right), where we write $W_G(x,y) = 1$ if (x,y) is black and $W_G(x,y) = 0$ if (x,y) is white.

As with everything in this chapter, we ignore measure zero differences, and so it does not matter what we do with boundaries of the pixels.

Definition 4.1.4 (Associated graphon of a graph)

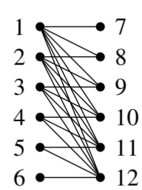
Given a graph G with n vertices labeled $1,\dots,n$, we define its ***associated graphon*** $\boldsymbol{W_G}\colon [0,1]^2 \to [0,1]$ by first partitioning $[0,1]$ into n equal-length intervals $I_1,\dots,I_n$ and setting W_G to be 1 on all $I_i \times I_j$ where ij is an edge of G, and 0 on all other $I_i \times I_j$s.

More generally, we can encode nonnegative vertex and edge weights in a graphon.

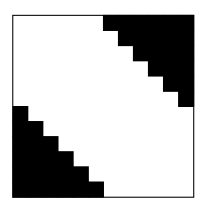
Definition 4.1.5 (Step graphon)

A ***step graphon*** W with k steps consists of first partitioning $[0,1]$ into k intervals $I_1,\dots,I_k$, and then setting W to be a constant on each $I_i \times I_j$.

***Example* 4.1.6** (Half-graph)**.** Consider the bipartite graph on $2n$ vertices, with one vertex part $\{v_1,\dots,v_n\}$ and the other vertex part $\{w_1,\dots,w_n\}$, and edges v_iw_j whenever $i \le j$. Its adjacency matrix and associated graphon are illustrated below.

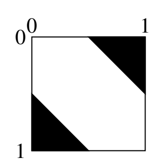


$$\begin{matrix}0&0&0&0&0&0&1&1&1&1&1&1\\0&0&0&0&0&0&0&1&1&1&1&1\\0&0&0&0&0&0&0&0&1&1&1&1\\0&0&0&0&0&0&0&0&0&1&1&1\\0&0&0&0&0&0&0&0&0&0&1&1\\0&0&0&0&0&0&0&0&0&0&0&1\\1&0&0&0&0&0&0&0&0&0&0&0\\1&1&0&0&0&0&0&0&0&0&0&0\\1&1&1&0&0&0&0&0&0&0&0&0\\1&1&1&1&0&0&0&0&0&0&0&0\\1&1&1&1&1&0&0&0&0&0&0&0\\1&1&1&1&1&1&0&0&0&0&0&0\end{matrix}$$



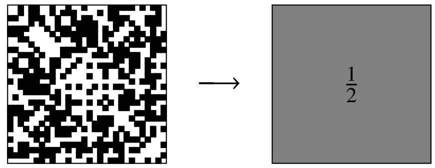
As $n \to \infty$, the associated graphons converge pointwise almost everywhere to the graphon

$$W(x,y) = \begin{cases} 1 & \text{if } x+y \le 1/2 \text{ or } x+y \ge 3/2, \\ 0 & \text{otherwise.} \end{cases}$$

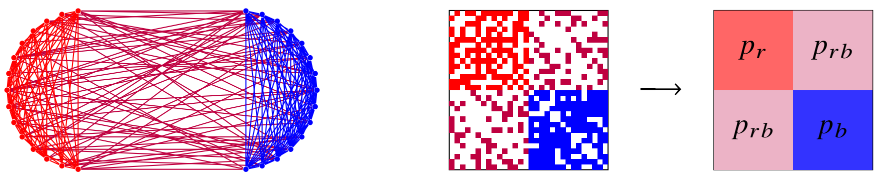


In general, pointwise convergence turns out to be too restrictive. We will need a more flexible notion of convergence, which we will discuss more in depth in the next section. Let us first give some more examples to motivate subsequent definitions.

***Example* 4.1.7** (Quasirandom graphs)**.** Let G_n be a sequence of quasirandom graphs with edge density approaching $1/2$, and $v(G_n) \to \infty$. The constant graphon $W \equiv 1/2$ seems like a reasonable candidate for its limit, and later we will see that this is indeed the case.

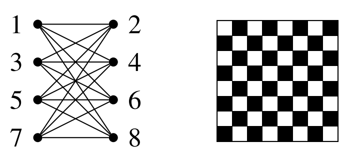


***Example* 4.1.8 (Stochastic block model).** Consider an n vertex graph with two types of vertices: red and blue. Half of the vertices are red and half of the vertices are blue. Two red vertices are adjacent with probability p_r, two blue vertices are adjacent with probability p_b, and finally, a red vertex and a blue vertex are adjacent with probability p_{rb}, all independently. Then as $n \to \infty$, the graphs converge to the step graphon shown below.

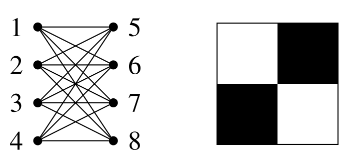


The above examples suggest that the limiting graphon looks like a blurry image of the adjacency matrix. However, there is an important caveat as illustrated in the next example.

***Example* 4.1.9 (Checkerboard).** Consider the $2n \times 2n$ "checkerboard" graphon shown below (for $n = 4$).



Since the 0s and 1s in the adjacency matrix are evenly spaced, one might suspect that this sequence converges to the constant $1/2$ graphon. However, this is not so. The checkerboard graphon is associated to the complete bipartite graph $K_{n,n}$, with the two vertex parts interleaved. By relabeling the vertices, we see that below is another representation of the associated graphon of the same graph.



So the graphon is the same for all n. So the graphon shown on the right, which is also W_{K_2}, must be the limit of the sequence, and not the constant $1/2$ graphon. This example tells us that we must be careful about the possibility of rearranging vertices when studying graph limits.

A graphon is an infinite-dimensional object. We would like some ways to measure the *similarity* between two graphons. We will explain two different approaches:

- cut distance, and

- homomorphism densities.

One of the main results in the theory of graph limits is that these two approaches are equivalent – we will show this later in the chapter.

4.2 Cut Distance

There are many ways to measure the distance between two graphs. Different methods may be useful for different applications. For example, we can consider the ***edit distance*** between two graphs (say on the same set of vertices), defined to be the number of edges needed to be added/deleted to obtain one graph from the other. The notion of edit distance arose when discussing the induced graph removal lemmas in Section 2.8. However, edit distance is not suitable for graph limits since it is incompatible with (quasi)random graphs. For example, given two n-vertex random graphs, independently generated with edge-probability $1/2$, we would like to say that they are similar as these graphs will end up converging to the constant $1/2$ graphon as $n \to \infty$ (e.g., Example 4.1.7). However, two independent random graphs typically only agree on around half of their edges (even if we allow permuting vertices), and so it takes $(1/4 + o(1))n^2$ edge additions/deletions to obtain one from the other.

A more suitable notion of distance is motivated by the discrepancy condition from Theorem 3.1.1 on quasirandom graphs. Inspired by the condition **DISC**, we would like to say that a graph G is ε-close to the constant p graphon if

$$|e_G(X,Y) - p\,|X|\,|Y|| \le \varepsilon\,|V(G)|^2 \quad \text{for all } X, Y \subseteq V(G).$$

Inspired by this notion, we now compare a pair of graphs G and G' on a common vertex set $V = V(G) = V(G')$. We say that ***G and G' are ε-close in cut norm*** if

$$|e_G(X,Y) - e_{G'}(X,Y)| \le \varepsilon\,|V|^2 \quad \text{for all } X, Y \subseteq V. \tag{4.1}$$

(This term "cut" is often used to refer to the set of edges in a graph G between some $X \subseteq V(G)$ and its complement. The cut norm builds on this concept.) With this notion, two independent n-vertex random graphs with the same edge-probability are $o(1)$-close in cut norm as $n \to \infty$.

As illustrated in Example 4.1.9, we also need to consider possible relabelings of vertices. Intuitively, the cut distance between two graphs will come from the relabeling of vertices that gives the greatest alignment. The actual definition will be a bit more subtle, allowing vertex fractionalization. The general definition of cut distance will allow us to compare graphs with different numbers of vertices. It is conceptually easier to define cut distance using graphons.

The edit distance of graphs corresponds to the L^1 distance for graphons. For every $p \ge 1$, we define the ***L^p norm*** of a function $W\colon [0,1]^2 \to \mathbb{R}$ by

$$\|W\|_p := \left(\int_{[0,1]^2} |W(x,y)|^p \, dxdy \right)^{1/p},$$

and the ***L^∞ norm*** by

$$\|W\|_\infty := \sup \left\{ t : W^{-1}([t,\infty)) \text{ has positive measure} \right\}.$$

(This is not simply the supremum of W; the definition should be invariant under measure zero changes of W.)

Definition 4.2.1 (Cut norm)
The ***cut norm*** of a measurable $W\colon [0,1]^2 \to \mathbb{R}$ is defined as

$$\|W\|_\square := \sup_{S,T\subseteq[0,1]} \left| \int_{S\times T} W \right|,$$

where S and T are measurable sets.

Let G and G' be two graphs sharing a common vertex set. Let W_G and $W_{G'}$ be their associated graphons (using the same ordering of vertices when constructing the graphons). Then G and G' are ε-close in cut norm (see (4.1)) if and only if

$$\|W_G - W_{G'}\|_\square \le \varepsilon.$$

(There is a subtlety in this claim that is worth thinking about: should we be worried about sets $S, T \subseteq [0,1]$ in Definition 4.2.1 of cut norm that contain fractions of some intervals that represent vertices? See Lemma 4.5.3 for a reformulation of the cut norm that may shed some light.)

We need a concept for an analog of a vertex set permutation for graphons. We write

$$\lambda(A) := \text{the Lebesgue measure of } A.$$

Intuitively, this is the "length" or "area" of A. We will always be referring to Lebesgue measurable sets. (Measure theoretic technicalities are not central to the discussions here, so feel free to ignore them.)

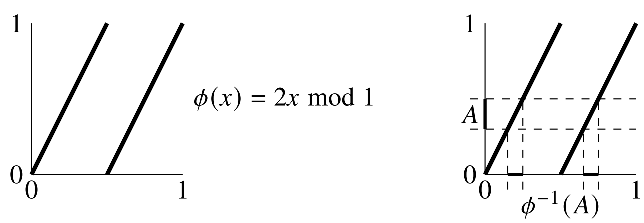
Definition 4.2.2 (Measure preserving map)
We say that $\phi : [0,1] \to [0,1]$ is a ***measure preserving map*** if

$$\lambda(A) = \lambda(\phi^{-1}(A)) \qquad \text{for all measurable } A \subseteq [0,1].$$

We say that ϕ is an ***invertible*** measure preserving map if there is another measure preserving map $\psi\colon [0,1] \to [0,1]$ such that $\phi\circ\psi$ and $\psi\circ\phi$ are both identity maps outside sets of measure zero.

***Example* 4.2.3.** For any constant $\alpha \in \mathbb{R}$, the function $\phi(x) = x + \alpha \bmod 1$ is measure preserving (this map rotates the circle $\mathbb{R}/\mathbb{Z}$ by α).

A more interesting example is, $\phi(x) = 2x \bmod 1$, illustrated below.



This map is also measure preserving. This might not seem to be the case at first, since ϕ seems to shrink some intervals by half. However, the definition of measure preserving actually says $\lambda(\phi^{-1}(A)) = \lambda(A)$ and not $\lambda(\phi(A)) = \lambda(A)$. For any interval $[a,b] \subseteq [0,1]$, we have

$\phi^{-1}([a,b]) = [a/2, b/2] \cup [1/2 + a/2, 1/2 + b/2]$, which does have the same measure as $[a,b]$. This map is 2-to-1, and it is not invertible.

Given $W : [0,1]^2 \to \mathbb{R}$ and an invertible measure preserving map $\phi : [0,1] \to [0,1]$, we write

$$\boldsymbol{W^{\phi}(x,y)} := W(\phi(x), \phi(y)).$$

Intuitively, this operation relabels the vertex set.

Definition 4.2.4 (Cut metric)

Given two symmetric measurable functions $U, W : [0,1]^2 \to \mathbb{R}$, we define their ***cut distance*** (or ***cut metric***) to be

$$\begin{aligned}\boldsymbol{\delta_\square(U,W)} &:= \inf_{\phi} \left\| U - W^{\phi} \right\|_\square \\ &= \inf_{\phi} \sup_{S,T \subseteq [0,1]} \left| \int_{S \times T} (U(x,y) - W(\phi(x), \phi(y)))\, dx dy \right|,\end{aligned}$$

where the infimum is taken over all invertible measure preserving maps $\phi : [0,1] \to [0,1]$. Define the cut distance between two graphs G and G' by the cut distance of their associated graphons:

$$\boldsymbol{\delta_\square(G,G')} := \delta_\square(W_G, W_{G'}).$$

Likewise, we can also define the cut distance between a graph and a graphon U:

$$\boldsymbol{\delta_\square(G,U)} := \delta_\square(W_G, U).$$

Definition 4.2.5 (Convergence in cut metric)

We say that a sequence of graphs or graphons ***converges in cut metric*** if they form a Cauchy sequence with respect to $\delta_\square$. Furthermore, we say that W_n ***converges to W in cut metric*** if $\delta_\square(W_n, W) \to 0$ as $n \to \infty$.

Note that in $\delta_\square(G,G')$, we are doing more than just permuting vertices. A measure preserving map on $[0,1]$ is also allowed to split a single node into fractions.

It is possible for two different graphons to have cut distance zero. For example, they could differ on a measure-zero set, or they could be related via measure preserving maps.

Space of Graphons

We can form a metric space by identifying graphons with measure zero (i.e., treating such two graphs with cut distance zero as the same point).

Definition 4.2.6 (Graphon space)

Let $\widetilde{\mathcal{W}}_0$ be the set of graphons (i.e., symmetric measurable functions $[0,1]^2 \to [0,1]$) where any pair of graphons with cut distance zero are considered the same point in the space. This is a metric space under cut distance $\delta_\square$.

We view every graph G as a point in $\widetilde{\mathcal{W}}_0$ via its associated graphon (note that several graphs can be identified as the same point in $\widetilde{\mathcal{W}}_0$).

(The subscript 0 in $\widetilde{\mathcal{W}}_0$ is conventional. Sometimes, without the subscript, $\widetilde{\mathcal{W}}$ is used to denote the space of symmetric measurable functions $[0,1]^2 \to \mathbb{R}$.)

Here is a central theorem in the theory of graph limits, proved by Lovász and Szegedy (2007).

Theorem 4.2.7 (Compactness of graphon space)
The metric space $(\widetilde{\mathcal{W}}_0, \delta_\square)$ is compact.

One of the main goals of this chapter is to prove this theorem and show its applications.

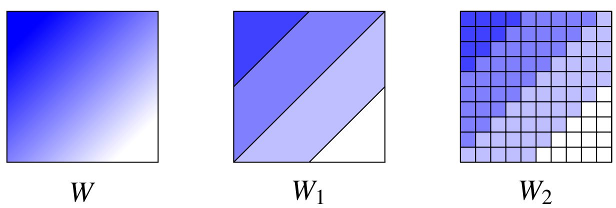
The compactness of graphon space is related to the graph regularity lemma. In fact, we will use the regularity method to prove compactness. Both compactness and the graph regularity lemma tell us that despite the infinite variability of graphs, every graph can be ε-approximated by a graph from a finite set of templates.

We close this section with the following observation.

Theorem 4.2.8 (Graphs are dense in the space of graphons)
The set of graphs is dense in $(\widetilde{\mathcal{W}}_0, \delta_\square)$.

Proof. Let $\varepsilon > 0$. It suffices to show that for every graphon W there exists a graph G such that $\delta_\square(G, W) < \varepsilon$.

We approximate W in several steps, illustrated below.



First, by rounding down the values of $W(x, y)$, we construct a graphon W_1 whose values are all integer multiples of $\varepsilon/3$, such that

$$\|W - W_1\|_\infty \le \varepsilon/3.$$

Next, since every Lebesgue measurable subset of $[0,1]^2$ can be arbitrarily well approximated using a union of boxes, we can find a step graphon W_2 approximating W_1 in L^1 norm:

$$\|W_1 - W_2\|_1 \le \varepsilon/3.$$

Finally, by replacing each block of W_2 by a sufficiently large quasirandom (bipartite) graph of edge density equal to the value of W_2, we find a graph G so that

$$\|W_2 - W_G\|_\square \le \varepsilon/3.$$

Then $\delta_\square(W, G) < \varepsilon$. □

***Remark* 4.2.9.** In the above proof, to obtain $\|W_1 - W_2\|_1 \le \varepsilon/3$, the number of steps of W_2 cannot be uniformly bounded as a function of ε (i.e., it must depend on W as well – think about what happens for a random graph). Consequently the number of vertices of the final graph G produced by this proof is not bounded by a function of ε.

Later on, we will see a different proof showing that for every $\varepsilon > 0$, there is some $N(\varepsilon)$ so that every graphon lies within cut distance ε of some graph with $\leq N(\varepsilon)$ vertices (Proposition 4.8.1).

Since every compact metric space is complete, we have the following corollary.

Corollary 4.2.10 (Graphons complete graphs)
The graphon space $(\widetilde{\mathcal{W}}_0, \delta_\square)$ is the completion of the space of graphs with respect to the cut metric.

Exercise 4.2.11 (Zero-one valued graphons)**.** Let W be a $\{0,1\}$-valued graphon. Suppose graphons W_n satisfy $\|W_n - W\|_\square \to 0$ as $n \to \infty$. Show that $\|W_n - W\|_1 \to 0$ as $n \to \infty$.

4.3 Homomorphism Density

Subgraph densities give another way of measuring graphs. It will be technically more convenient to work with graph homomorphisms instead of subgraphs.

Definition 4.3.1 (Homomorphism density)
A ***graph homomorphism*** from F to G is a map $\phi\colon V(F) \to V(G)$ such that if $uv \in E(F)$ then $\phi(u)\phi(v) \in E(G)$ (i.e., ϕ maps edges to edges). Define

$$\mathbf{Hom}(F,G) := \{\text{homomorphisms from } F \text{ to } G\}$$

and

$$\mathbf{hom}(F,G) := |\mathrm{Hom}(F,G)|\,.$$

Define the ***F-homomorphism density in G*** (or ***F-density in G*** for short) as

$$t(F,G) := \frac{\hom(F,G)}{v(G)^{v(F)}}.$$

This is also the probability that a uniformly random map $V(F) \to V(G)$ induces a graph homomorphism from F to G.

***Example* 4.3.2** (Homomorphism counts)**.**

- $\hom(K_1, G) = v(G)$.
- $\hom(K_2, G) = 2e(G)$.
- $\hom(K_3, G) = 6 \cdot \#\text{triangles in } G$
- $\hom(G, K_3)$ is the number of proper colorings of G using three labeled colors such as $\{\text{red, green, blue}\}$ (corresponding to the vertices of K_3).

Remark* 4.3.3** (Subgraphs vs. homomorphisms)**.** Note that homomorphisms from F to G do not quite correspond to copies of subgraphs F inside G, because these homomorphisms can be noninjective. Define the ***injective homomorphism density

$$t_{\text{inj}}(F,G) := \frac{\#\text{injective homomorphisms from } F \text{ to } G}{v(G)(v(G)-1)\cdots(v(G)-v(F)+1)}.$$

Equivalently, this is the fraction of injective maps $V(F) \to V(G)$ that are graph homomorphisms (i.e., send edges to edges). The fraction of maps $V(F) \to V(G)$ that are noninjective is $\le \binom{v(F)}{2}/v(G)$ (for every fixed pair of vertices of F, the probability that they collide is exactly $1/v(G)$). So

$$\left|t(F,G) - t_{\text{inj}}(F,G)\right| \le \frac{1}{v(G)}\binom{v(F)}{2}.$$

If F is fixed, the right-hand side tends to zero as $v(G) \to \infty$. So all but a negligible fraction of such homomorphisms correspond to subgraphs. This is why we often treat subgraph densities interchangeably with homomorphism densities as they agree in the limit.

Now we define the corresponding notion of homomorphism density in graphons. We first give an example and then the general formula.

***Example* 4.3.4** (Triangle density in graphons)**.** The following quantity is the triangle density in a graphon W.

$$t(K_3, W) = \int_{[0,1]^3} W(x,y)W(y,z)W(z,x)\,dxdydz.$$

This definition agrees with Definition 4.3.1 for the triangle density in graphs. Indeed, for every graph G, the triangle density in G equals the triangle density in the associated graphon W_G; that is, $t(K_3, W_G) = t(K_3, G)$.

Definition 4.3.5 (Homomorphism density in graphon)

Let F be a graph and W a graphon. The ***F-density in W*** is defined to be

$$t(F,W) = \int_{[0,1]^{V(F)}} \prod_{ij \in E(F)} W(x_i, x_j) \prod_{i \in V(F)} dx_i.$$

We also use the same formula when W is a symmetric measurable function.

Note that for all graphs F and G, letting W_G be the graphon associated to G,

$$t(F,G) = t(F, W_G). \tag{4.2}$$

So the two definitions of F-density agree.

Definition 4.3.6 (Left convergence)

We say that a sequence of graphons W_n is ***left-convergent*** if for every graph F, $t(F, W_n)$ converges as $n \to \infty$. We say that this sequence ***left-converges*** to a graphon W if $\lim_{n\to\infty} t(F, W_n) = t(F, W)$ for every graph F.

For a sequence of graphs, we say that it is ***left-convergent*** if the sequence of associated graphons $W_n = W_{G_n}$ is left-convergent, and that it ***left-converges*** to W if W_n does.

One usually has $v(G_n) \to \infty$, but it is not strictly necessary for this definition. Note that when $v(G_n) \to \infty$, homomorphism densities and subgraph densities coincide (see Remark 4.3.3).

It turns out that left-convergence is equivalent to convergence in cut metric. This foundational result in graph limits is due to Borgs, Chayes, Lovász, Sós, and Vesztergombi (2008).

Theorem 4.3.7 (Equivalence of convergence)
A sequence of graphons is left-convergent if and only if it is a Cauchy sequence with respect to the cut metric $\delta_\square$.

The sequence left-converges to some graphon W if and only if it converges to W in cut metric.

The implication that convergence in cut metric implies left-convergence is easier; it follows from the counting lemma (Section 4.5). The converse is more difficult, and we will establish it at the end of the chapter.

This allows us to talk about ***convergent sequences*** of graphs or graphons without specifying whether we are referring to left-convergence or convergence in cut metric. However, since a major goal of this chapter is to prove the equivalence between these two notions, we will be more specific about the notion of convergence.

From the compactness of the space of graphons and the equivalence of convergence (actually only needing the easier implication), we will be able to quickly deduce the existence of limit for a left-convergent sequence, which was first proved by Lovász and Szegedy (2006). Note that the following statement does not require knowledge of the cut metric.

Theorem 4.3.8 (Existence of limit for left-convergence)
Every left-convergent sequence of graphs or graphons left-converges to some graphon.

***Remark* 4.3.9.** One can artificially define a metric that coincides with left-convergence. Let $(F_n)_{n\geq 1}$ enumerate over all graphs. One can define a distance between graphons U and W by

$$\sum_{k\geq 1} 2^{-k} \left| t(F_k, W) - t(F_k, U) \right|.$$

We see that a sequence of graphons converges under this notion of distance if and only if it is left-convergent. This shows that left-convergence defines a metric topology on the space of graphons, but in practice the above distance is pretty useless.

Exercise 4.3.10 (Counting Eulerian orientations). Define $W\colon [0,1]^2 \to \mathbb{R}$ by $W(x,y) = 2\cos(2\pi(x-y))$. Let F be a graph. Show that $t(F,W)$ is the number of ways to orient all edges of F so that every vertex has the same number of incoming edges as outgoing edges.

4.4 W-Random Graphs

In this section, we explain how to use a graphon to create a random graph model. This hopefully gives more intuition about graphons.

The most common random graph model is the Erdős–Rényi random graph $\mathbf{G}(n,p)$, which is an n-vertex graph with every edge chosen with probability p.

Stochastic Block Model

The ***stochastic block model*** is a random graph model that generalizes the Erdős–Rényi random graph. We already saw an example in Example 4.1.8. Let us first illustrate the ***two-block model***, which has several parameters:

	q_r	q_b
q_r	p_{rr}	p_{rb}
q_b	p_{rb}	p_{bb}

with all the numbers lying in $[0, 1]$, and subject to $q_r + q_b = 1$. We form a n-vertex random graph as follows:

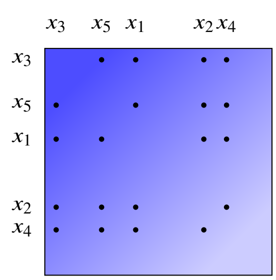
(1) Color each vertex red with probability q_r and blue with probability q_b, independently at random. These vertex colors are "hidden states" and are not part of the data of the output random graph (this step is slightly different from Example 4.1.8 in an unimportant way);

(2) For every pair of vertices, independently place an edge between them with probability
- p_{rr} if both vertices are red,
- p_{bb} if both vertices are blue, and
- p_{rb} if one vertex is red and the other is blue.

One can easily generalize the above to a ***k-block model***, where vertices have k hidden states, with $q_1, \dots, q_k$ (adding up to 1) being the vertex state probabilities, and a symmetric $k \times k$ matrix $(p_{ij})_{1 \le i,j \le k}$ of edge probabilities for pairs of vertices between various states.

W-Random Graph

The W-random graph is a further generalization. The stochastic block model corresponds to step graphons W.



Definition 4.4.1 (W-random graph)
Let W be a graphon. The n-vertex ***W-random graph*** $\mathbf{G}(n, W)$ denotes the n-vertex random graph (with vertices labeled $1, \dots, n$) obtained by first picking $x_1, \dots, x_n$ uniformly at random from $[0, 1]$, and then putting an edge between vertices i and j with probability $W(x_i, x_j)$, independently for all $1 \le i < j \le n$.

Let us show that these W-random graphs left-converge to W with probability 1.

Theorem 4.4.2 (W-random graphs left-converge to W)
Let W be a graphon. For each n, let G_n be a random graph distributed as $\mathbf{G}(n, W)$. Then G_n left-converges to W with probability 1.

***Remark* 4.4.3.** The theorem does not require each G_n to be sampled independently. For example, we can construct the sequence of random graphs, with G_n distributed as $\mathbf{G}(n, W)$, by revealing one vertex at a time without resampling the previous vertices and edges. In this case, each G_n is a subgraph of the next graph G_{n+1}.

We will need the following standard result about concentration of Lipschitz functions. This can be proved using Azuma's inequality (e.g., see Chapter 7 of *The Probabilistic Method* by Alon and Spencer).

Theorem 4.4.4 (Bounded differences inequality)
Let $X_1 \in \Omega_1, \ldots, X_n \in \Omega_n$ be independent random variables. Suppose $f \colon \Omega_1 \times \cdots \times \Omega_n \to \mathbb{R}$ is L-Lipschitz for some constant L in the sense of satisfying

$$\left| f(x_1, \ldots, x_n) - f(x_1', \ldots, x_n') \right| \le L \tag{4.3}$$

whenever $(x_1, \ldots, x_n)$ and $(x_1', \ldots, x_n')$ differ on exactly one coordinate. Then the random variable $Z = f(X_1, \ldots, X_n)$ satisfies, for every $\lambda \ge 0$,

$$\mathbb{P}(Z - \mathbb{E}Z \ge \lambda L) \le e^{-2\lambda^2/n} \quad \text{and} \quad \mathbb{P}(Z - \mathbb{E}Z \le -\lambda L) \le e^{-2\lambda^2/n}.$$

Let us show that the F-density in a W-random graph rarely differs significantly from $t(F, W)$.

Theorem 4.4.5 (Sample concentration for graphons)
For every $\varepsilon > 0$, positive integer n, graph F, and graphon W, we have

$$\mathbb{P}\left(|t(F, \mathbf{G}(n, W)) - t(F, W)| > \varepsilon\right) \le 2 \exp\left(\frac{-\varepsilon^2 n}{8 v(F)^2}\right). \tag{4.4}$$

Proof. Recall from Remark 4.3.3 that the injective homomorphism density $t_{\text{inj}}(F, G)$ is defined to be the fraction of injective maps $V(F) \to V(G)$ that carry every edge of F to an edge of G. We will first prove that

$$\mathbb{P}\left(\left|t_{\text{inj}}(F, \mathbf{G}(n, W)) - t(F, W)\right| > \varepsilon\right) \le 2 \exp\left(\frac{-\varepsilon^2 n}{2 v(F)^2}\right). \tag{4.5}$$

Let $y_1, \ldots, y_n$, and z_{ij} for each $1 \le i < j \le n$, be independent uniform random variables in $[0, 1]$. Let G be the graph on vertices $\{1, \ldots, n\}$ with an edge between i and j if and only if $z_{ij} \le W(y_i, y_j)$, for every $i < j$. Then G has the same distribution as $\mathbf{G}(n, W)$. Let us group variables y_i, z_{ij} into $x_1, x_2, \ldots, x_n$ where

$$x_1 = (y_1), \quad x_2 = (y_2, z_{12}), \quad x_3 = (y_3, z_{13}, z_{23}), \quad x_4 = (y_4, z_{14}, z_{24}, z_{34}), \quad \ldots.$$

This amounts to exposing the graph G one vertex at a time. Define the function $f(x_1, \ldots, x_n) = t_{\text{inj}}(F, G)$. Note that $\mathbb{E} f = \mathbb{E}\, t_{\text{inj}}(F, \mathbf{G}(n, W)) = t(F, W)$ by linearity of expectations (in this step, it is important that we are using the injective variant of homomorphism densities). Note changing a single coordinate of f changes the value of the function by at most $v(F)/n$, since exactly a $v(F)/n$ fraction of injective maps $V(F) \to V(G)$ includes a fixed $v \in V(G)$ in the image. Then (4.5) follows from the bounded differences inequality, Theorem 4.4.4.

To deduce the theorem from (4.5), recall from Remark 4.3.3 that

$$\left|t(F,G) - t_{\mathrm{inj}}(F,G)\right| \le v(F)^2/(2v(G)).$$

If $\varepsilon < v(F)^2/n$, then the right-hand side of (4.4) is at least $2e^{-\varepsilon/8} \ge 1$, and so the inequality trivially holds. Otherwise, $|t(F,\mathbf{G}(n,W)) - t(F,W)| > \varepsilon$ implies $\left|t_{\mathrm{inj}}(F,\mathbf{G}(n,W)) - t(F,W)\right| > \varepsilon - v(F)^2/(2n) \ge \varepsilon/2$, and then we can apply (4.5) to conclude. □

Theorem 4.4.2 then follows from the Borel–Cantelli lemma, stated below, applied to Theorem 4.4.5 for all rational $\varepsilon > 0$.

Theorem 4.4.6 (Borel–Cantelli lemma)
Given a sequence of events $E_1, E_2, \dots,$ if $\sum_n \mathbb{P}(E_n) < \infty$, then with probability 1, only finitely many of them occur.

4.5 Counting Lemma

In Chapter 2 on the graph regularity lemma, we proved a counting lemma that gave a lower bound on the number of copies of some fixed graph H in a regularity partition. The same techniques can be modified to give a similar upper bound. Here we prove another graph counting lemma. The proof is more analytic, whereas the previous proofs in Chapter 2 were more combinatorial (embedding one vertex at a time).

Theorem 4.5.1 (Counting lemma)
Let F be a graph. Let W and U be graphons. Then

$$|t(F,W) - t(F,U)| \le |E(F)|\, \delta_\square(W,U).$$

Qualitatively, the counting lemma tells us that for every graph F, the function $t(F,\cdot)$ is continuous in $(\widetilde{\mathcal{W}}_0, \delta_\square)$, the graphon space with respect to the cut metric. It implies the easier direction of the equivalence in Theorem 4.3.7, namely that convergence in cut metric implies left-convergence.

Corollary 4.5.2 (Cut metric convergence implies left-convergence)
Every Cauchy sequence of graphons with respect to the cut metric is left-convergent.

In the rest of this section, we prove Theorem 4.5.1. It suffices to prove that

$$|t(F,W) - t(F,U)| \le |E(F)|\, \|W - U\|_\square\,. \tag{4.6}$$

Indeed, for every invertible measure preserving map $\phi\colon [0,1] \to [0,1]$, we have $t(F,U) = t(F,U^\phi)$. By considering the above inequality with U replaced by U^ϕ, and taking the infimum over all U^ϕ, we obtain Theorem 4.5.1.

The following reformulation of the cut norm is often quite useful.

Lemma 4.5.3 (Reformulation of cut norm)
For every measurable $W\colon [0,1]^2 \to \mathbb{R}$,

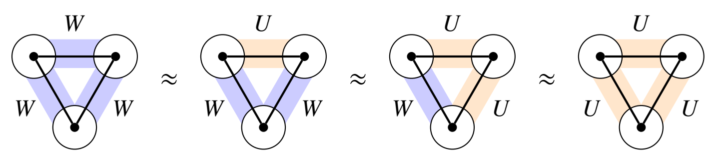
$$\|W\|_\square = \sup_{\substack{u,v:[0,1]\to[0,1] \\ \text{measurable}}} \left| \int_{[0,1]^2} W(x,y)u(x)v(y)\,dxdy \right|.$$

Proof. We want to show (left-hand side below is how we defined the cut norm in Definition 4.2.1)

$$\sup_{\substack{S,T\subseteq[0,1] \\ \text{measurable}}} \left| \int_{[0,1]^2} W(x,y)1_S(x)1_T(y)\,dxdy \right| = \sup_{\substack{u,v:[0,1]\to[0,1] \\ \text{measurable}}} \left| \int_{[0,1]^2} W(x,y)u(x)v(y)\,dxdy \right|.$$

The right-hand side is at least as large as the left-hand side since we can take $u = 1_S$ and $v = 1_T$. On the other hand, the integral on the right-hand side is bilinear in u and v, and so it is always possible to change u and v to $\{0,1\}$-valued functions without decreasing the value of the integral (e.g., think about what is the best choice for v with u held fixed, and vice versa). If u and v are restricted to $\{0,1\}$-valued functions, then the two sides are identical. □

As a warm up, let us illustrate the proof of the triangle counting lemma, which has all the ideas of the general proof but with simpler notation. As illustrated below, the main idea to "replace" W by U on the triangle one at a time using the cut norm.



Proposition 4.5.4 (Triangle counting lemma)
Let W and U be graphons. Then

$$|t(K_3,W) - t(K_3,U)| \le 3\,\|W-U\|_\square.$$

Proof. Given three graphons W_{12}, W_{13}, W_{23}, define

$$t(W_{12},W_{13},W_{23}) = \int_{[0,1]^3} W_{12}(x,y)W_{13}(x,z)W_{23}(y,z)\,dxdydz.$$

So

$$t(K_3,W) = t(W,W,W) \quad \text{and} \quad t(K_3,U) = t(U,U,U).$$

Observe that $t(W_{12},W_{13},W_{23})$ is trilinear in W_{12}, W_{13}, W_{23}. We have

$$t(W,W,W) - t(U,W,W) = \int_{[0,1]^3} (W-U)(x,y)W(x,z)W(y,z)\,dxdydz.$$

For any fixed z, note that $x \mapsto W(x,z)$ and $y \mapsto W(y,z)$ are both measurable functions $[0,1] \to [0,1]$. So applying Lemma 4.5.3 gives

$$\left| \int_{[0,1]^2} (W-U)(x,y)W(x,z)W(y,z)\,dxdy \right| \le \|W-U\|_\square$$

for every z. Now integrate over all z and applying the triangle inequality, we obtain

$$|t(W, W, W) - t(U, W, W)| \le \|W - U\|_\square .$$

We have similar inequalities in the other two coordinates. We can write

$$t(W, W, W) - t(U, U, U) = t(W, W, W - U) + t(W, W - U, U) + t(W - U, U, U).$$

Each term on the right-hand side is at most $\|W - U\|_\square$ in absolute value. So the result follows. □

The above proof generalizes in a straightforward way to a general graph counting lemma..

Proof of the counting lemma (Theorem 4.5.1). Given a collection of graphons W_e indexed by the edges e of F, define

$$t_F(W_e : e \in E(F)) = \int_{[0,1]^{V(F)}} \prod_{ij \in E(F)} W_{ij}(x_i, x_j) \prod_{i \in V(H)} dx_i .$$

In particular, this quantity equals $t(F, W)$ if $W_e = W$ for all $e \in E(F)$. A straightforward generalization of the triangle case shows that if we change exactly one argument in the above function from W to U, then its value changes by at most $\|W - U\|_\square$ in absolute value. Thus, starting with $t_F(W_e : e \in E(F))$ with every $W_e = W$, we can change each argument from W to U, one by one, resulting in a total change of at most $e(F)\,\|W - U\|_\square$. This proves (4.6), and hence the theorem. □

4.6 Weak Regularity Lemma

In Chapter 2, we defined an ε-regular vertex partition of a graph to be a partition such that all but ε-fraction of pairs of vertices lie between ε-regular pairs of vertex parts. The number of parts is at most an exponential tower of height $O(\varepsilon^{-5})$.

The goal of this section is to introduce a weaker version of the regularity lemma, requiring substantially fewer parts for the partition. The guarantee provided by the partition can be captured by the cut norm.

Let us first state this notion for a graph and then for a graphon.

Definition 4.6.1 (Weak regular partition for graphs)
Given graph G, a partition $\mathcal{P} = \{V_1, \dots, V_k\}$ of $V(G)$ is called ***weak ε-regular*** if for all $A, B \subseteq V(G)$,

$$\left| e(A, B) - \sum_{i,j=1}^{k} d(V_i, V_j)|A \cap V_i||B \cap V_j| \right| \le \varepsilon v(G)^2 .$$

***Remark* 4.6.2** (Interpreting weak regularity)**.** Given $A, B \subseteq V(G)$, suppose we only knew how many vertices from A and B lie in each part of the partition (and not specifically which vertices), and we are asked to predict the number of edges between A and B. Then the sum above is the number of edges between A and B that one would naturally expect based on the edge densities between vertex parts. Being weak regular says that this prediction is roughly correct.

Weak regularity is more "global" compared to the notion of an ε-regular partition from Chapter 2. Here A and B have size a constant order fraction of the entire vertex set, rather than subsets of individual parts of the partition. The edge densities between certain pairs $A \cap V_i$ and $B \cap V_j$ could differ significantly from that of V_i and V_j. All we ask is that on average these discrepancies mostly cancel out.

The following weak regularity lemma was proved by Frieze and Kannan (1999), initially motivated by algorithmic applications that we will mention in Remark 4.6.11.

Theorem 4.6.3 (Weak regularity lemma for graphs)
Let $0 < \varepsilon < 1$. Every graph has a weak ε-regular partition into at most $4^{1/\varepsilon^2}$ vertex parts.

Now let us state the corresponding notions for graphons.

Definition 4.6.4 (Stepping operator)
Given a symmetric measurable function $W \colon [0,1]^2 \to \mathbb{R}$, and a measurable partition $\mathcal{P} = \{S_1, \dots, S_k\}$ of $[0,1]$, define a symmetric measurable function $W_{\mathcal{P}} \colon [0,1]^2 \to \mathbb{R}$ by setting its value on each $S_i \times S_j$ to be the average value of W over $S_i \times S_j$ (since we only care about functions up to measure zero sets, we can ignore all parts S_i with measure zero).

In other words, $W_{\mathcal{P}}$ is a step graphon with steps given by $\mathcal{P}$ and values given by averaging W over the steps.

***Remark* 4.6.5.** The stepping operator is the orthogonal projection in the Hilbert space $L^2([0,1]^2)$ onto the subspace of functions constant on each step $S_i \times S_j$. It can also be viewed as the conditional expectation with respect to the σ-algebra generated by $S_i \times S_j$.

Definition 4.6.6 (Weak regular partition for graphons)
Given graphon W, we say that a measurable partition $\mathcal{P}$ of $[0,1]$ into finitely many parts is ***weak ε-regular*** if

$$\|W - W_{\mathcal{P}}\|_\square \le \varepsilon.$$

Theorem 4.6.7 (Weak regularity lemma for graphons)
Let $0 < \varepsilon < 1$. Then every graphon has a weak ε-regular partition into at most $4^{1/\varepsilon^2}$ parts.

***Remark* 4.6.8.** Technically speaking, Theorem 4.6.3 does not follow from Theorem 4.6.7 since the partition of $[0,1]$ for W_G could split intervals corresponding to individual vertices of G. However, the proofs of the two claims are exactly the same. Alternatively, one can allow a more flexible definition of a graphon as a symmetric measurable function $W \colon \Omega \times \Omega \to [0,1]$, and then take Ω to be the discrete probability space $V(G)$ endowed with the uniform measure.

Like the proof of the regularity lemma in Section 2.1, we use an energy increment strategy. Recall from Definition 2.1.10 that the energy of a vertex partition is the mean-squared edge-density between parts. Given a graphon W, we define the ***energy*** of a measurable partition

$\mathcal{P} = \{S_1, \ldots, S_k\}$ of $[0,1]$ by

$$\|W_{\mathcal{P}}\|_2^2 = \int_{[0,1]^2} W_{\mathcal{P}}(x,y)^2\, dxdy = \sum_{i,j=1}^{k} \lambda(S_i)\lambda(S_j)(\text{avg of } W \text{ on } S_i \times S_j)^2.$$

Given $W, U \colon [0,1]^2 \to \mathbb{R}$, we write

$$\langle \mathbf{W}, \mathbf{U} \rangle := \int WU = \int_{[0,1]^2} W(x,y)U(x,y)\, dxdy.$$

Lemma 4.6.9 (L^2 energy increment)
Let W be a graphon. Let $\mathcal{P}$ be a finite measurable partition of $[0,1]$ that is not weak ε-regular for W. Then there is a measurable refinement $\mathcal{P}'$ of $\mathcal{P}$, dividing each part of $\mathcal{P}$ into at most 4 parts, such that

$$\|W_{\mathcal{P}'}\|_2^2 > \|W_{\mathcal{P}}\|_2^2 + \varepsilon^2.$$

Proof. Because $\|W - W_{\mathcal{P}}\|_\square > \varepsilon$, there exist measurable subsets $S, T \subseteq [0,1]$ such that

$$|\langle W - W_{\mathcal{P}}, 1_{S\times T}\rangle| > \varepsilon.$$

Let $\mathcal{P}'$ be the refinement of $\mathcal{P}$ by introducing S and T, dividing each part of $\mathcal{P}$ into ≤ 4 sub-parts. We know that

$$\langle W_{\mathcal{P}}, W_{\mathcal{P}}\rangle = \langle W_{\mathcal{P}'}, W_{\mathcal{P}}\rangle$$

because $W_{\mathcal{P}}$ is constant on each step of $\mathcal{P}$, and $\mathcal{P}'$ is a refinement of $\mathcal{P}$. Thus,

$$\langle W_{\mathcal{P}'} - W_{\mathcal{P}}, W_{\mathcal{P}}\rangle = 0.$$

By the Pythagorean Theorem (in the Hilbert space $L^2([0,1]^2)$),

$$\|W_{\mathcal{P}'}\|_2^2 = \|W_{\mathcal{P}}\|_2^2 + \|W_{\mathcal{P}'} - W_{\mathcal{P}}\|_2^2. \tag{4.7}$$

Note that $\langle W_{\mathcal{P}'}, 1_{S\times T}\rangle = \langle W, 1_{S\times T}\rangle$ since S and T are both unions of parts of the partition $\mathcal{P}'$. So, by the Cauchy–Schwarz inequality,

$$\|W_{\mathcal{P}'} - W_{\mathcal{P}}\|_2 \ge |\langle W_{\mathcal{P}'} - W_{\mathcal{P}}, 1_{S\times T}\rangle| = |\langle W - W_{\mathcal{P}}, 1_{S\times T}\rangle| > \varepsilon.$$

So by (4.7), we have $\|W_{\mathcal{P}'}\|_2^2 > \|W_{\mathcal{P}}\|_2^2 + \varepsilon^2$, as claimed. □

We will prove the following slight generalization of Theorem 4.6.7, allowing an arbitrary starting partition (this will be useful later).

Theorem 4.6.10 (Weak regularity lemma for graphons)
Let $0 < \varepsilon < 1$. Let W be a graphon. Let $\mathcal{P}_0$ be a finite measurable partition of $[0,1]$. Then every graphon has a weak ε-regular partition $\mathcal{P}$, such that $\mathcal{P}$ refines $\mathcal{P}_0$, and each part of $\mathcal{P}_0$ is partitioned into at most $4^{1/\varepsilon^2}$ parts under $\mathcal{P}$.

This proposition specifically tells us that starting with any given partition, the regularity argument still works.

Proof. Starting with $i = 0$:

(1) If $\mathcal{P}_i$ is weak ε-regular, then STOP.
(2) Else, by Lemma 4.6.9, there exists a measurable partition $\mathcal{P}_{i+1}$ refining each part of $\mathcal{P}_i$ into at most 4 parts, such that $\|W_{\mathcal{P}_{i+1}}\|_2^2 > \|W_{\mathcal{P}_i}\|_2^2 + \varepsilon^2$.
(3) Increase i by 1 and go back to Step (1).

Since $0 \le \|W_{\mathcal{P}}\|_2 \le 1$ for every $\mathcal{P}$, the process terminates with $i < 1/\varepsilon^2$, resulting in a terminal $\mathcal{P}_i$ with the desired properties. □

Remark* 4.6.11** (Additive approximation of maximum cut). One of the initial motivations for developing the weak regularity lemma was to develop a general efficient algorithm for estimating the maximum cut in a dense graph. The ***maximum cut problem is a central problem in algorithms and combinatorial optimization:

MAX CUT: Given a graph S, find a $S \subseteq V(G)$ that maximizes $e(S, V(G) \setminus S)$.

Goemans and Williamson (1995) found an efficient 0.878-approximation algorithm (this means that the algorithm outputs some S with $e(S, V(G) \setminus S)$ at least a factor 0.878 of the optimum). Their seminal algorithm uses a semidefinite relaxation. The Unique Games Conjecture (currently still open) would imply that it would be NP-hard to obtain a better approximation than the Goemans–Williamson algorithm (Khot, Kindler, Mossel, and O'Donnell 2007). It is also known that approximating beyond $16/17 \approx 0.941$ is NP-hard (Håstad 2001).

On the other hand, an algorithmic version of the weak regularity lemma gives us an efficient algorithm to approximate the maximum cut for dense graphs with an additive error. This means, given $\varepsilon > 0$, we wish to find a cut whose number of edges is within εn^2 of the optimum. The basic idea is to find a weak regular partition $V(G) = V_1 \cup \cdots \cup V_k$, and then do a brute-force search through all possibles size $|S \cap V_i|$. See Frieze and Kannan (1999) for more details. These ideas have been further developed into efficient sampling algorithms, sampling only $\mathrm{poly}(1/\varepsilon)$ random vertices, for estimating the maximum cut in a dense graph, (e.g., Alon, Fernandez de la Vega, Kannan, and Karpinski (2003b)).

The following exercise offers another approach to the weak regularity lemma. It gives an approximation of a graphon as a linear combination of $\le \varepsilon^{-2}$ indicator functions of boxes. The polynomial dependence of ε^{-2} is important for designing efficient approximation algorithms.

Exercise 4.6.12 (Weak regularity decomposition).

(a) Let $\varepsilon > 0$. Show that for every graphon W, there exist measurable $S_1, \dots, S_k, T_1, \dots, T_k \subseteq [0,1]$ and reals $a_1, \dots, a_k \in \mathbb{R}$, with $k < \varepsilon^{-2}$, such that

$$\left\| W - \sum_{i=1}^{k} a_i \mathbf{1}_{S_i \times T_i} \right\|_\square \le \varepsilon.$$

The rest of the exercise shows how to recover a regularity partition from the above approximation.

(b) Show that the stepping operator is contractive with respect to the cut norm, in the sense that if $W\colon [0,1]^2 \to \mathbb{R}$ is a measurable symmetric function, then $\|W_{\mathcal{P}}\|_\square \le \|W\|_\square$.

(c) Let $\mathcal{P}$ be a partition of $[0,1]$ into measurable sets. Let U be a graphon that is constant on $S \times T$ for each $S, T \in \mathcal{P}$. Show that for every graphon W, one has

$$\|W - W_{\mathcal{P}}\|_\square \le 2\|W - U\|_\square .$$

(d) Use (a) and (c) to give a different proof of the weak regularity lemma (with slightly worse bounds than the one given in class): show that for every $\varepsilon > 0$ and every graphon W, there exists a partition $\mathcal{P}$ of $[0,1]$ into $2^{O(1/\varepsilon^2)}$ measurable sets such that $\|W - W_{\mathcal{P}}\|_\square \le \varepsilon$.

Exercise 4.6.13* (Second neighborhood distance)**.** Let $0 < \varepsilon < 1/2$. Let W be a graphon. Define $\tau_{W,x} \colon [0,1] \to [0,1]$ by

$$\tau_{W,x}(z) = \int_{[0,1]} W(x,y)W(y,z)\,dy.$$

(This models the second neighborhood of x.) Prove that if a finite set $S \subseteq [0,1]$ satisfies

$$\|\tau_{W,s} - \tau_{W,t}\|_1 > \varepsilon \qquad \text{for all distinct } s, t \in S,$$

then $|S| \le (1/\varepsilon)^{C/\varepsilon^2}$, where C is some absolute constant.

Exercise 4.6.14 (Strong regularity lemma)**.** Let $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots)$ be a sequence of positive reals. By repeatedly applying the weak regularity lemma, show that there is some $M = M(\varepsilon)$ such that for every graphon W, there is a pair of partitions $\mathcal{P}$ and $\mathcal{Q}$ of $[0,1]$ into measurable sets, such that $\mathcal{Q}$ refines $\mathcal{P}$, $|\mathcal{Q}| \le M$ (here $|\mathcal{Q}|$ denotes the number of parts of $\mathcal{Q}$),

$$\|W - W_{\mathcal{Q}}\|_\square \le \varepsilon_{|\mathcal{P}|} \qquad \text{and} \qquad \|W_{\mathcal{Q}}\|_2^2 \le \|W_{\mathcal{P}}\|_2^2 + \varepsilon_1^2.$$

Furthermore, deduce the strong regularity lemma in the following form:

$$W = W_{\mathrm{str}} + W_{\mathrm{psr}} + W_{\mathrm{sml}},$$

where W_{str} is a k-step graphon with $k \le M$, $\|W_{\mathrm{psr}}\|_\square \le \varepsilon_k$, and $\|W_{\mathrm{sml}}\|_1 \le \varepsilon_1$. State your bounds on M explicitly in terms of ε. (Note: the parameter choice $\varepsilon_k = \varepsilon/k^2$ roughly corresponds to Szemerédi's regularity lemma, in which case your bound on M should be an exponential tower of 2's of height $\varepsilon^{-O(1)}$; if not then you are doing something wrong.)

4.7 Martingale Convergence Theorem

In this section we prove a result about martingales that will be used in the proof of the compactness of the graphon space.

Martingales are a standard notion in probability theory. It is a stochastic sequence where the expected change at each step is zero, even conditioned on all prior values of the sequence.

Definition 4.7.1 (Discrete time martingale)

A ***martingale*** is a random real sequence $X_0, X_1, X_2, \dots$ such that for all $n \ge 0$, $\mathbb{E}|X_n| < \infty$, and

$$\mathbb{E}[X_{n+1} | X_0, \dots, X_n] = X_n.$$

***Remark* 4.7.2.** The above definition is sufficient for our purposes. In order to give a more formal definition of a martingale, we need to introduce the notion of a *filtration*. See any standard measure theory based introduction to probability (Williams (1991, Chapters 10–11) has a particularly lucid discussion of martingales and their convergence theorem discussed below). This martingale is indexed by integers, and hence called "discrete-time." There are also continuous-time martingales (e.g., Brownian motion), which we will not discuss here.

***Example* 4.7.3 (Partial sum of independent mean zero random variables).** Let $Z_1, Z_2, \dots$ be a sequence of independent mean zero random variables (e.g., ± 1 with equal probability). Then $X_n = Z_1 + \cdots + Z_n$, $n \geq 0$, is a martingale.

***Example* 4.7.4 (Betting strategy).** Consider any betting strategy in a "fair" casino, where the expected value of each bet is zero. Let X_n be the balance after n rounds of betting. Then X_n is a martingale regardless of the betting strategy. So every betting strategy has zero expected gain after n rounds. Also see the **optional stopping theorem** for a more general statement (e.g., Williams (1991, Chapter 10)).

The original meaning of the word "martingale" refers to the following betting strategy on a sequence of fair coin tosses. Each round the better is allowed to bet an arbitrary amount Z: if heads, the better gains Z dollars, and if tails the better loses Z dollars.

Start betting 1 dollar. If one wins, stop. If one loses, then double one's bet for the next coin. And then repeat (i.e., keep doubling one's bet until the first win, at which point one stops).

A "fallacy" is that this strategy always results in a final net gain of $1, the supposed reason being that with probability 1 one eventually sees a head. This initially appears to contradict the earlier claim that all betting strategies have zero expected gain. Thankfully there is no contradiction. In real life, one starts with a finite budget and could possibly go bankrupt with this betting strategy, thereby leading to a forced stop. In the optional stopping theorem, there are some boundedness hypotheses that are violated by the above strategy.

The following construction of martingales is most relevant for our purposes.

***Example* 4.7.5 (Doob martingale).** Let X be some "hidden" random variable. Partial information is revealed about X gradually over time. For example, X is some fixed function of some random inputs. So the exact value of X is unknown but its distribution can be derived from the distribution of the inputs. Initially one does not know any of the inputs. Over time, some of the inputs are revealed. Let

$$X_n = \mathbb{E}[X \mid \text{all information revealed up to time } n].$$

Then $X_0, X_1, \dots$ is a martingale (why?). Informally, X_n is the best guess (in expectation) of X based on all the information available up to time n. We have $X_0 = \mathbb{E}X$ (when no information is revealed). All information is revealed as $n \to \infty$, and the martingale X_n converges to the random variable X with probability 1.

Here is a real-life example. Let $X \in \{0, 1\}$ be whether a candidate wins in a presidential election. Let X_n be the inferred probability that the candidate wins, given all the information known at time t_n. Then X_n converges to the "truth," a $\{0, 1\}$-value, eventually becoming deterministic when the election result is finalized.

Then X_n is a martingale. At time t_n, knowing X_n, if the expectation for X_{n+1} (conditioned on everything known at time t_n) were different from X_n, then one should have adjusted X_n accordingly in the first place.

The precise notion of "information" in the above formula can be formalized using the notion of *filtration* in probability theory.

Here is the main result of this section.

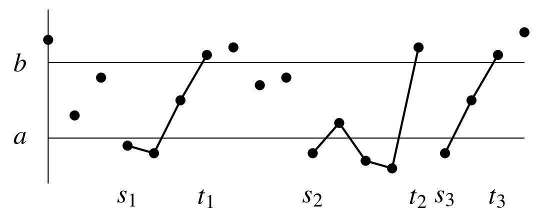
Theorem 4.7.6 (Martingale convergence theorem)
Every bounded martingale converges with probability 1.

In other words, if $X_0, X_1, \ldots$ is a martingale with $X_n \in [0,1]$ for every n, then the sequence is convergent with probability 1.

***Remark* 4.7.7.** The proof actually shows that the boundedness condition can be replaced by the weaker L^1-boundedness condition $\sup_n \mathbb{E}|X_n| < \infty$. Even more generally, a hypothesis called "uniform integrability" is enough.

Some boundedness condition is necessary. For example, in Example 4.7.3, a running sum of independent uniform ± 1 is a nonbounded martingale, and never converges.

Proof. If a sequence $X_0, X_1, \cdots \in [0,1]$ does not converge, then there exist a pair of rational numbers $0 < a < b < 1$ such that X_n "up-crosses" $[a, b]$ infinitely many times, meaning that there is an infinite sequence $s_1 < t_1 < s_2 < t_2 < \cdots$ such that $X_{s_i} < a < b < X_{t_i}$ for all i.



We will show that for each $a < b$, the probability that a bounded martingale $X_0, X_1, \cdots \in [0,1]$ up-crosses $[a, b]$ infinitely many times is zero. Then, by taking a union of all countably many such pairs (a, b) of rationals, we deduce that the martingale converges with probability 1.

Consider the following betting strategy. Imagine that X_n is a stock price. At any time, if X_n dips below a, we buy and hold one share until X_n reaches above b, at which point we sell this share. (Note that we always hold either zero or one share. We do not buy more until we have sold the currently held share). Start with a budget of $Y_0 = 1$ (so we will never go bankrupt). Let Y_n be the value of our portfolio (cash on hand plus the value of the share if held) at time n. Then Y_n is a martingale (why?). So $\mathbb{E}Y_n = Y_0 = 1$. Also $Y_n \geq 0$ for all n. If one buys and sells at least k times up to time n, then $Y_n \geq k(b-a)$ (this is only the net profit from buying and selling; the actual Y_n may be higher due to the initial cash balance and the value of the current share held). So, by Markov's inequality, for every n,

$$\mathbb{P}(\geq k \text{ up-crossings up to time } n) \leq \mathbb{P}(Y_n \geq k(b-a)) \leq \frac{\mathbb{E}Y_n}{k(b-a)} = \frac{1}{k(b-a)}.$$

By the monotone convergence theorem,

$$\mathbb{P}(\geq k \text{ up-crossings}) = \lim_{n\to\infty} \mathbb{P}(\geq k \text{ up-crossings up to time } n) \leq \frac{1}{k(b-a)}.$$

Letting $k \to \infty$, the probability of having infinitely many up-crossings is zero. □

4.8 Compactness of the Graphon Space

Using the weak regularity lemma and the martingale convergence theorem, let us prove that the space of graphons is compact with respect to the cut metric.

Proof of compactness of the graphon space (Theorem 4.2.7). As $\widetilde{\mathcal{W}}_0$ is a metric space, it suffices to prove sequential compactness. Fix a sequence $W_1, W_2, \ldots$ of graphons. We want to show that there is a subsequence which converges (with respect to $\delta_\square$) to some limit graphon.

Step 1. Regularize.

For each n, apply the weak regularity lemma (Theorem 4.6.7) repeatedly, to obtain a sequence of partitions $\mathcal{P}_{n,1}, \mathcal{P}_{n,2}, \mathcal{P}_{n,3}, \ldots$ (everything in this proof is measurable, and we will stop repeatedly mentioning it) such that

(a) $\mathcal{P}_{n,k+1}$ refines $\mathcal{P}_{n,k}$ for all n, k,
(b) $|\mathcal{P}_{n,k}| = m_k$ where m_k is a function of only k, and
(c) $\|W_n - W_{n,k}\|_\square \leq 1/k$ where $W_{n,k} = (W_n)_{\mathcal{P}_{n,k}}$.

The weak regularity lemma only guarantees that $|\mathcal{P}_{n,k}| \leq m_k$, but if we allow empty parts then we can achieve equality in (b).

Step 2. Passing to a subsequence.

Initially, each $\mathcal{P}_{n,k}$ partitions $[0,1]$ into arbitrary measurable sets. By restricting to a subsequence, we may assume that

- For each k and $i \in [m_k]$, the measure of the ith part of $\mathcal{P}_{n,k}$ converges to some value $\alpha_{k,i}$ as $n \to \infty$.
- For each k and $i, j \in [m_k]$, the value of $W_{n,k}$ on the product of the ith and jth parts of $\mathcal{P}_{n,k}$ converges to some value $\beta_{k,i,j}$ as $n \to \infty$.

Now construct, for each k, the following limiting objects as $n \to \infty$ along the above subsequence:

- Let $\mathcal{P}_k = \{I_{k,1}, \ldots, I_{k,m_k}\}$ denote a partition of $[0,1]$ into intervals with lengths $\lambda(I_{k,i}) = \alpha_{k,i}$ for each $i \in [m_k]$.
- Let U_k denote a step graphon with steps $\mathcal{P}_k$, and whose value on $I_{k,i} \times I_{k,j}$ is $\beta_{k,i,j}$ for each $i, j \in [m_k]$.

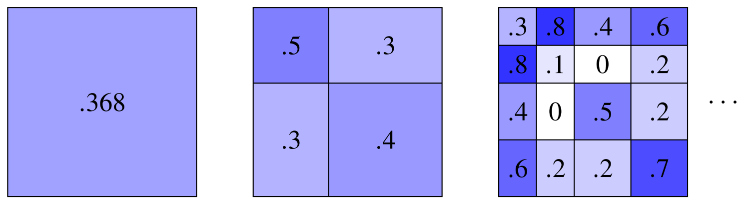
Then, for each k,

$$\delta_\square(W_{n,k}, U_k) \to 0, \quad \text{as } n \to \infty. \tag{4.8}$$

(In fact, some rearrangement of the step graphon $W_{n,k}$ converges pointwise almost everywhere to the step graphon U_k.)

For each k, since $W_{n,k} = (W_{n,k+1})_{\mathcal{P}_{n,k}}$ for every n, we have

$$U_k = (U_{k+1})_{\mathcal{P}_k}.$$



U_1 U_2 U_3

Step 3. Finding the limit.

Now each U_k can be thought of as a random variable on probability space $[0,1]^2$ (i.e., $U_k(X,Y)$ with $(X,Y) \sim \text{Uniform}([0,1]^2)$). The condition $U_k = (U_{k+1})_{\mathcal{P}_k}$ implies that the sequence $U_1, U_2, \ldots$ is a martingale. Since each U_k is bounded between 0 and 1, by the martingale convergence theorem (Theorem 4.7.6), there exists a graphon U such that $U_k \to U$ pointwise almost everywhere as $k \to \infty$.

We claim that $W_1, W_2, \ldots$ (which is a relabeled subsequence of the original sequence) converges to U in cut metric.

Let $\varepsilon > 0$. Then there exists some $k > 3/\varepsilon$ such that $\|U - U_k\|_1 < \varepsilon/3$, by pointwise convergence and the dominated convergence theorem. Then $\delta_\square(U, U_k) < \varepsilon/3$. By (4.8), there exists some $n_0 \in \mathbb{N}$ such that $\delta_\square(W_{n,k}, U_k) < \varepsilon/3$ for all $n > n_0$. Finally, since we chose $k > 3/\varepsilon$, we already know that $\delta_\square(W_n, W_{n,k}) < \varepsilon/3$ for all n. We conclude that

$$\delta_\square(U, W_n) \le \delta_\square(U, U_k) + \delta_\square(U_k, W_{n,k}) + \delta_\square(W_{n,k}, W_n) \le \varepsilon/3 + \varepsilon/3 + \varepsilon/3 = \varepsilon.$$

Since $\varepsilon > 0$ can be chosen to be arbitrarily small, we find that the subsequence W_n converges to U in cut metric. □

Quick Applications

The compactness of $(\widetilde{\mathcal{W}}_0, \delta_\square)$ is a powerful statement. We will use it to prove the equivalence of cut metric convergence and left-convergence in the next section. Right now, let us show how to use compactness to deduce the existence of limits for a left-convergent sequence of graphons.

Proof of Theorem 4.3.8 (existence of limit for a left-convergent sequence of graphons). Let $W_1, W_2, \ldots$ be a sequence of graphons such that the sequence of F-densities $\{t(F, W_n)\}_n$ converges for every graph F. Since $(\widetilde{\mathcal{W}}_0, \delta_\square)$ is a compact metric space by Theorem 4.2.7, it is also sequentially compact, and so there is a subsequence $(n_i)_{i=1}^\infty$ and a graphon W such that $\delta_\square(W_{n_i}, W) \to 0$ as $i \to \infty$. Fix any graph F. By the counting lemma, Theorem 4.5.1, it follows that $t(F, W_{n_i}) \to t(F, W)$. But by assumption, the sequence $\{t(F, W_n)\}_n$ converges. Therefore $t(F, W_n) \to t(F, W)$ as $n \to \infty$. Thus W_n left-converges to W. □

Let us now examine a different aspect of compactness. Recall that by definition, a set is compact if every open cover has a finite subcover.

Recall from Theorem 4.2.8 that the set of graphs is dense in the space of graphons with respect to the cut metric. This was proved by showing that for every $\varepsilon > 0$ and graphon W, one can find a graph G such that $\delta_\square(G, W) < \varepsilon$. However, the size of G produced by this proof depends on both ε and W, since the proof proceeds by first taking a discrete L^1 approximation of W, which could involve an unbounded number of steps to approximate. In contrast, we show below that the number of vertices of G needs to depend only on ε and not on W.

Proposition 4.8.1 (Uniform approximation of graphons by graphs)
For every $\varepsilon > 0$ there is some positive integer $N = N(\varepsilon)$ such that every graphon lies within cut distance ε of a graph on at most N vertices.

Proof. Let $\varepsilon > 0$. For a graph G, define the open ε-ball (with respect to the cut metric) around G:

$$B_\varepsilon(G) = \{W \in \widetilde{\mathcal{W}}_0 : \delta_\square(G, W) < \varepsilon\}.$$

Since every graphon lies within cut distance ε from some graph (Theorem 4.2.8), the balls $B_\varepsilon(G)$ cover $\widetilde{\mathcal{W}}_0$ as G ranges over all graphs. By compactness, this open cover has a finite subcover, and let N be the maximum number of vertices in graphs G of this subcover. Then every graphon lies within cut distance ε of a graph on at most N vertices. □

The following exercise asks to make the above proof quantitative.

Exercise 4.8.2. Show that for every $\varepsilon > 0$, every graphon lies within cut distance at most ε from some graph on at most C^{1/ε^2} vertices, where C is some absolute constant.

Hint: Use the weak regularity lemma.

Remark* 4.8.3** (Ineffective bounds from compactness)**.** Arguments using compactness usually do not generate quantitative bounds, meaning, for example, the proof of Proposition 4.8.1 does not give any specific function $n(\varepsilon)$, only that such a function always exists. In case where one does not have an explicit bound, we call the bound ***ineffective. Ineffective bounds also often arise from arguments involving ergodic theory and nonstandard analysis. Sometimes a different argument can be found that generates a quantitative bound (e.g., Exercise 4.8.2), but it is not always known how to do this. Here we illustrate a simple example of a compactness application (unrelated to dense graph limits) that gives an ineffective bound, but it remains an open problem to make the bound effective.

This example concerns bounded degree graphs. It is sometimes called a "regularity lemma" for bounded degree graphs, but it is very different from the regularity lemmas we have encountered so far.

A ***rooted graph*** (G, v) consists of a graph G with a vertex $v \in v(G)$ designated as the ***root***. Given a graph G and positive integer r, we can obtain a random rooted graph by first picking a vertex v of G as the root uniformly at random, and then removing all vertices more than distance r from v. We define the ***r-neighborhood-profile*** of G to be the probability distribution on rooted graphs generated by this process.

Recall that the ***total variation distance*** between two probability distributions μ and λ is defined by

$$d_{TV}(\mu, \lambda) = \sup_E |\mu(E) - \lambda(E)|,$$

where E ranges over all events. In the case of two discrete discrete random distributions μ

and λ, the above definition can be written as half the ℓ^1 distance between the two probability distributions:

$$d_{TV}(\mu, \lambda) = \frac{1}{2} \sum_x |\mu(x) - \lambda(x)| .$$

The following is an unpublished observation of Alon.

Theorem 4.8.4 ("Regularity lemma" for bounded degree graphs)
For every $\varepsilon > 0$ and positive integers Δ and r there exists a positive integer $N = N(\varepsilon, \Delta, r)$ such that for every graph G with maximum degree at most Δ, there exists a graph G' with at most N vertices, so that the total variation distance between the r-neighborhood-profiles of G and G' is at most ε.

Proof. Let $\mathcal{G} = \mathcal{G}_{\Delta,r}$ be the set of all possible rooted graphs with maximum degree Δ and radius at most r around the root. Then $|\mathcal{G}| < \infty$. The r-neighborhood-profile p_G of any rooted graph G can be represented as a point $p_G \in [0,1]^{\mathcal{G}}$ with coordinate sum 1, and let $A = \{p_G : \text{graph } G\} \subseteq [0,1]^{\mathcal{G}}$ be the set of all points that can arise this way. Since $[0,1]^{\mathcal{G}}$ is compact, the closure of A is compact. Since the union of the open ε-neighborhoods (with respect to d_{TV}) of p_G, ranging over all graphs G, covers the closure of A, by compactness there is some finite subcover. This subcover is a finite collection $\mathcal{X}$ of graphs so that for every graph G, p_G lies within ε total variance distance to some $p_{G'}$ with $G' \in \mathcal{X}$. We conclude by letting N be the maximum number of vertices of a graph from $\mathcal{X}$. □

Despite the short proof using compactness, it remains an open problem to make the above result quantitative.

Open Problem 4.8.5 (Effective "regularity lemma" for bounded degree graphs)
Find some specific $N(\varepsilon, \Delta, r)$ so that Theorem 4.8.4 holds.

4.9 Equivalence of Convergence

In this section, we prove Theorem 4.3.7, that left-convergence is equivalent to convergence in cut metric. The counting lemma (Theorem 4.5.1) already showed that cut metric convergence implies left-convergence. It remains to show the converse. In other words, we need to show that if $W_1, W_2, \ldots$ is a sequence of graphons such that $t(F, W_n)$ converges as $n \to \infty$ for every graph F, then W_n is a Cauchy sequence in $(\widetilde{\mathcal{W}}_0, \delta_\square)$.

By the compactness of the graphon space, there is always some (subsequential) limit point W of the sequence W_n under the cut metric. We want to show that this limit point is unique. Suppose U is another limit point. It remains to show that W and U are in fact the same point in $\widetilde{\mathcal{W}}_0$.

Let $(n_i)_{i=1}^\infty$ be a subsequence such that $W_{n_i} \to W$. By the counting lemma, $t(F, W_{n_i}) \to t(F, W)$ for all graphs F, and by convergence of F-densities, $t(F, W_n) \to t(F, W)$ for all graphs F. Similarly, $t(F, W_n) \to t(F, U)$ for all F. Hence, $t(F, U) = t(F, W)$ for all F. All it remains is to prove is the following claim.

Theorem 4.9.1 (Uniqueness of moments)
Let U and W be graphons such that $t(F, W) = t(F, U)$ for all graphs F. Then $\delta_\square(U, W) = 0$.

***Remark* 4.9.2.** The result is reminiscent of results from probability theory on the uniqueness of moments, which roughly says that if two "sufficiently well-behaved" real random variables X and Y share the same moments, (i.e., $\mathbb{E}[X^k] = \mathbb{E}[Y^k]$ for all nonnegative integers k), then X and Y must be identically distributed. One needs some technical conditions for the conclusion to hold. For example, Carleman's condition says that if the moments of X satisfy $\sum_{k=1}^{\infty} \mathbb{E}[X^{2k}]^{-1/(2k)} = \infty$, then the distribution of X is uniquely determined by its moments. This sufficient condition holds as long as the kth moment of X does not grow too quickly with k. It holds for many distributions in practice.

We need some preparation before proving the uniqueness of moments theorem.

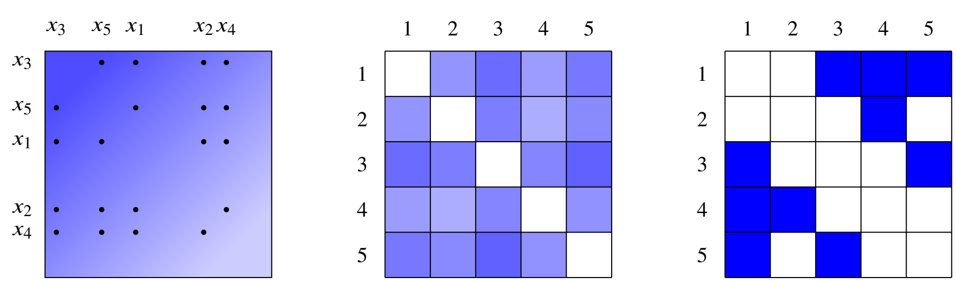
Lemma 4.9.3 (Tail bounds for U-statistics)
Let $U\colon [0,1]^2 \to [-1,1]$ be a symmetric measurable function. Let $x_1, \dots, x_k \in [0,1]$ be chosen independently and uniformly at random. Let $\varepsilon > 0$. Then

$$\mathbb{P}\left(\left| \frac{1}{\binom{k}{2}} \sum_{i<j} U(x_i, x_j) - \int_{[0,1]^2} U \right| \ge \varepsilon \right) \le 2e^{-k\varepsilon^2/8}.$$

Proof. Let $f(x_1, \dots, x_n)$ denote the expression inside the absolute value. So $\mathbb{E}f = 0$. Also f changes by at most $2(k-1)/\binom{k}{2} = 4/k$ whenever we change exactly one coordinate of f. By the bounded differences inequality, Theorem 4.4.4, we obtain

$$\mathbb{P}(|f| \ge \varepsilon) \le 2\exp\left(\frac{-2\varepsilon^2}{(4/k)^2 k} \right) = 2e^{-k\varepsilon^2/8}. \qquad \square$$

Let us now consider a variation of the W-random graph model from Section 4.4. Let $x_1, \dots, x_k \in [0,1]$ be chosen independently and uniformly at random. Let $\mathbf{H}(k, W)$ be an edge-weighted random graph on vertex set $[k]$ with edge ij having weight $W(x_i, x_j)$, for each $1 \le i < j \le n$. Note that this definition makes sense for any symmetric measurable $W\colon [0,1]^2 \to \mathbb{R}$. Furthermore, when W is a graphon, the W-random graph $\mathbf{G}(k, W)$ can be obtained by independently sampling each edge of $\mathbf{H}(k, W)$ with probability equal to its edge weight. We shall study the joint distributions of $\mathbf{G}(k, W)$ and $\mathbf{H}(k, W)$ coupled through the above two-step process.



W $\qquad$ $\mathbf{H}(k, W)$ $\qquad$ $\mathbf{G}(k, W)$

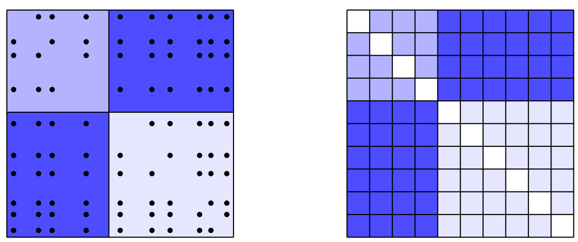
Similar to Definition 4.2.4 of the cut distance $\delta_\square$, define the distance based on the L^1 norm:

$$\boldsymbol{\delta_1(W,U)} := \inf_\phi \|W - U^\phi\|_1$$

where the infimum is taken over all invertible measure preserving maps $\phi\colon [0,1] \to [0,1]$. Since $\|\cdot\|_\square \le \|\cdot\|_1$, we have $\delta_\square \le \delta_1$.

Lemma 4.9.4 (1-norm convergence for $\mathbf{H}(k,W)$)
Let W be a graphon. Then $\delta_1(\mathbf{H}(k,W),W) \to 0$ as $k \to \infty$ with probability 1.

Proof. First we prove the result for step graphons W. In this case, with probability 1 the fraction of vertices of $\mathbf{H}(k,W)$ that fall in each step of W converges to the length of each step by the law of large numbers. If so, then after sorting the vertices of $\mathbf{H}(k,W)$, the associated graphon $\mathbf{H}(k,W)$ is obtained from W by changing the step sizes by $o(1)$ as $k \to \infty$, and then zeroing out the diagonal blocks, as illustrated below. Then $\mathbf{H}(k,W)$ converges to W pointwise almost everywhere as $k \to \infty$. In particular, $\delta_1(\mathbf{H}(k,W),W) \to 0$.



W $\mathbf{H}(k,W)$

Now let W be any graphon. For any other graphon W', by using the same random vertices for $\mathbf{H}(k,W)$ and $\mathbf{H}(k,W')$, the two random graphs are coupled so that with probability 1,

$$\|\mathbf{H}(k,W) - \mathbf{H}(k,W')\|_1 = \|\mathbf{H}(k,W-W')\|_1 = \|W - W'\|_1 + o(1) \quad \text{as } k \to \infty$$

by Lemma 4.9.3 applied to $U(x,y) = |W(x,y) - W'(x,y)|$.

For every $\varepsilon > 0$, we can find some step graphon W' so that $\|W - W'\|_1 \le \varepsilon$ (by approximating the Lebesgue measure using boxes). We saw earlier that $\delta_1(\mathbf{H}(k,W'),W') \to 0$. It follows that with probability 1,

$$\begin{aligned}\delta_1(\mathbf{H}(k,W),W) &\le \|\mathbf{H}(k,W) - \mathbf{H}(k,W')\|_1 + \delta_1(\mathbf{H}(k,W'),W') + \|W' - W\|_1 \\ &= 2\,\|W' - W\|_1 + o(1) \le 2\varepsilon + o(1)\end{aligned}$$

as $k \to \infty$. Since $\varepsilon > 0$ can be chosen to be arbitrarily small, we have $\delta_1(\mathbf{H}(k,W),W) \to 0$ with probability 1. □

Proof of Theorem 4.9.1 (uniqueness of moments). By inclusion-exclusion, for any k-vertex labeled graph F,

$$\Pr[\mathbf{G}(k,W) \cong F \text{ as labeled graphs}] = \sum_{F' \supseteq F} (-1)^{e(F')-e(F)} \Pr[\mathbf{G}(k,W) \supseteq F' \text{ as labeled graphs}],$$

where the sum ranges over all graphs F' with $V(F') = V(F)$ and $E(F') \supseteq E(F)$. Since

$$t(F', W) = \Pr[\mathbf{G}(k, W) \supseteq F' \text{ as labeled graphs}],$$

we see that the distribution of $\mathbf{G}(k, W)$ is determined by the values of $t(F, W)$ over all F. Since $t(F, W) = t(F, U)$ for all F, $\mathbf{G}(k, W)$ and $\mathbf{G}(k, U)$ are identically distributed.

Our strategy is to prove

$$W \overset{\delta_1}{\approx} \mathbf{H}(k, W) \overset{\delta_\square}{\approx} \mathbf{G}(k, W) \overset{D}{\equiv} \mathbf{G}(k, U) \overset{\delta_\square}{\approx} \mathbf{H}(k, U) \overset{\delta_1}{\approx} U.$$

By Lemma 4.9.4, $\delta_1(\mathbf{H}(k, W), W) \to 0$ with probability 1.

By coupling $\mathbf{H}(k, W)$ and $\mathbf{G}(k, W)$ using the same random vertices as noted earlier, so that $\mathbf{G}(k, W)$ is generated from $\mathbf{H}(k, W)$ by independently sampling each edge with probability equal to the edge weight, we have

$$\mathbb{P}(\delta_\square(\mathbf{G}(k, W), \mathbf{H}(k, W)) \geq \varepsilon) \to 0 \quad \text{as } k \to 0 \text{ for every fixed } \varepsilon > 0.$$

We leave the details of this claim as an exercise, below. It can be proved via the Chernoff bound and the union bound. We need to be a bit careful about the definition of the cut norm as one needs to consider fractional vertices.

Exercise 4.9.5 (Edge-sampling an edge-weighted graph and cut norm)**.** Let H be an edge-weighted graph on k vertices, with edge weights in $[0, 1]$, and let G be a random graph obtained from H by independently keeping each edge with probability equal to its edge-weight. Prove that for every $\varepsilon > 0$ and $\delta > 0$, there exists k_0 such that $\delta_\square(G, H) < \varepsilon$ with probability $> 1 - \delta$, provided that $k \geq k_0$.

So with probability 1,

$$\delta_\square(\mathbf{H}(k, W), \mathbf{G}(k, W)) \to 0 \qquad \text{as } k \to \infty.$$

Since $\delta_\square \leq \delta_1$, we have, with probability 1,

$$\delta_\square(W, \mathbf{G}(k, W)) \leq \delta_1(W, \mathbf{H}(k, W)) + \delta_\square(\mathbf{H}(k, W), \mathbf{G}(k, W)) = o(1).$$

Likewise $\delta_\square(U, \mathbf{G}(k, U)) = o(1)$ with probability 1. Since $\mathbf{G}(k, W)$ and $\mathbf{G}(k, U)$ are identically distributed as noted earlier, we deduce that $\delta_\square(W, U) = 0$. $\square$

This finishes the proof of the equivalence between left-convergence and cut metric convergence. This equivalence can be recast as counting and inverse counting lemmas. We state the inverse counting lemma below, and leave the proof as an instructive exercise in applying the compactness of the graphon space. (One need not invoke anything from the proof of the uniqueness of moments theorem. You may wish to review the discussions on applying compactness at the end of the previous section and the beginning of this section.)

Corollary 4.9.6 (Inverse counting lemma)

For every $\varepsilon > 0$ there is some $\eta > 0$ and integer $k > 0$ such that if U and W are graphons with

$$|t(F, U) - t(F, W)| \leq \eta \quad \text{whenever } v(F) \leq k,$$

then $\delta_\square(U, W) \leq \varepsilon$.

Exercise 4.9.7. Prove the inverse counting lemma Corollary 4.9.6 using the compactness of the graphon space (Theorem 4.2.7) and the uniqueness of moments (Theorem 4.9.1).

Hint: Consider a hypothetical sequence of counterexamples.

***Remark* 4.9.8.** The inverse counting lemma was first proved by Borgs, Chayes, Lovász, Sós, and Vesztergombi (2008) in the following quantitative form:

Theorem 4.9.9 (Inverse counting lemma)
Let k be a positive integer. Let U and W be graphons with

$$|t(F,U) - t(F,W)| \le 2^{-k^2} \quad \text{whenever } v(F) \le k,$$

then (here C is some absolute constant)

$$\delta_\square(U,W) \le \frac{C}{\sqrt{\log k}}.$$

Exercise 4.9.10. Prove that there exists a function $f\colon (0,1] \to (0,1]$ such that for all graphons U and W, there exists a graph F with

$$\frac{|t(F,U) - t(F,W)|}{e(F)} \ge f(\delta_\square(U,W)).$$

Exercise 4.9.11* (Generalized maximum cut)**.** For symmetric measurable functions $W, U\colon [0,1]^2 \to \mathbb{R}$, define

$$C(W,U) := \sup_\phi \langle W, U^\phi \rangle = \sup_\phi \int W(x,y)U(\phi(x),\phi(y))\,dxdy,$$

where ϕ ranges over all invertible measure preserving maps $[0,1] \to [0,1]$. Extend the definition of $C(\cdot,\cdot)$ to graphs via $C(G,\cdot) := C(W_G,\cdot)$ and so on.

(a) Is $C(U,W)$ continuous jointly in (U,W) with respect to the cut norm? Is it continuous in U if W is held fixed?

(b) Show that if W_1 and W_2 are graphons such that $C(W_1,U) = C(W_2,U)$ for all graphons U, then $\delta_\square(W_1,W_2) = 0$.

(c) Let $G_1, G_2, \ldots$ be a sequence of graphs such that $C(G_n,U)$ converges as $n \to \infty$ for every graphon U. Show that $G_1, G_2, \ldots$ is convergent.

(d) Can the hypothesis in (c) be replaced by "$C(G_n,H)$ converges as $n \to \infty$ for every graph H"?

Exercise 4.9.12 (Characterizing graphs in terms of homomorphism counts)**.**

(a) Let G_1 and G_2 be two graphs such that $\hom(F,G_1) = \hom(F,G_2)$ for every graph F. Show that G_1 and G_2 are isomorphic.

(b) Let G_1 and G_2 be two graphs such that $\hom(G_1,H) = \hom(G_2,H)$ for every graph H. Show that G_1 and G_2 are isomorphic.

Further Reading

The book *Large Networks and Graph Limits* by Lovász (2012) is the authoritative reference on the subject. His survey article titled *Very Large Graphs* (2009) also gives an excellent overview.

One particularly striking application of the theory of dense graph limits is to large deviations for random graphs by Chatterjee and Varadhan (2011). See the survey article *An Introduction to Large Deviations for Random Graphs* by Chatterjee (2016) as well as his book (Chatterjee 2017).

Chapter Summary

- A **graphon** is a symmetric measurable function $W\colon [0,1]^2 \to [0,1]$.
  - Every graph G can be turned into an associated graphon W_G.
  - A graphon can be turned into a random graph model known a W**-random graph**, generalizing the **stochastic block model**.
- The **cut metric** of two graphons U and W is defined by

$$\delta_\square(U,W) = \inf_\phi \|U - W^\phi\|_\square = \inf_\phi \sup_{S,T\subseteq[0,1]} \left| \int_{S\times T} (U(x,y) - W(\phi(x),\phi(y)))\,dxdy \right|,$$

  where the infimum is taken over all invertible measure preserving maps $\phi\colon [0,1] \to [0,1]$.
- Given a sequence of graphons (or graphs) $W_1, W_2, \dots$, we say that it
  - **converges in cut metric** if it is a Cauchy sequence with respect to the cut metric $\delta_\square$;
  - **left-converges** if the homomorphism density $t(F, W_n)$ converges for every fixed graph F as $n \to \infty$.
- The **graphon space is compact** under the cut metric.
  - Proof uses the weak regularity lemma and the martingale convergence theorem.
  - Compactness has powerful consequences.
- Convergence in cut metric and left-convergence are **equivalent** for a sequence of graphons.
  - ($\Rightarrow$) follows from a counting lemma.
  - ($\Leftarrow$) was proved here using compactness.