Graph Theory and Additive Combinatorics

Lecturer: Prof. Yufei Zhao

5
Graph limits

5.1 Introduction and statements of main results

Graph limits seeks a generalization of analytic limits to graphs. Consider the following two examples that shows the potential parallel between the set of rational numbers and graphs:

Example 5.1. For $x \in [0,1]$, the minimum of $x^3 - x$ occurs at $x = 1/\sqrt{3}$. But if we restrict ourselves in $\mathbb{Q}$ (pretending that we don't know about real numbers), a way to express this minimum is to find a sequence $x_1, x_2, \ldots$ of rational numbers that converges to $1/\sqrt{3}$.

Example 5.2. Given $p \in (0,1)$, we want to minimize the density of C_4's among all graphs with edge density p. From Theorem 4.1 we see that the minimum is p^4, which is obtained via a sequence of quasirandom graphs. (There is no single finite graph that obtains this minimum.)

We can consider the set of all graphs as a set of discrete objects (analogous to $\mathbb{Q}$), and seek its "completion" (analogously $\mathbb{R}$).

Definition 5.3. A ***graphon*** ("graph function") is a symmetric measurable function $W : [0,1]^2 \to [0,1]$.

Remark 5.4. Definition 5.3 can be generalized to $\Omega \times \Omega \to [0,1]$ where Ω is any measurable probability space, but for simplicity we will usually work with $\Omega = [0,1]$. (In fact, most "nice" measurable probability space can be represented by $[0,1]$.)

The codomain of the function can also be generalized to $\mathbb{R}$, in which case we will refer to the function as a ***kernel***. Note that this naming convention is not always consistent in literature.

Graphons can be seen as a generalized type of graphs. In fact, we can convert any graph into a graphon, which allow us to start imagining what the limits of some sequences of graph should look like.

Example 5.5. Consider a ***half graph*** G_n, which is a bipartite graph where one part is labeled $1, 2, \ldots, n$ and the other part is labeled $n+1, \ldots, 2n$, and vertices i and $n+j$ is connected if and only if $i \leq j$. If we treat the adjacency matrix $\mathrm{Adj}(G_n)$ as a $0/1$ bit image, we can define graphon $W_{G_n} : [0,1]^2 \to [0,1]$ (which consists of $(2n)^2$ "pixels" of size $1/(2n) \times 1/(2n)$ each). When n goes to infinity, the graphon converges (pointwise) to a function that looks like Figure 5.2.

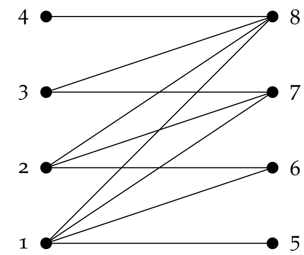

Figure 5.1: The half graph G_n for $n = 4$

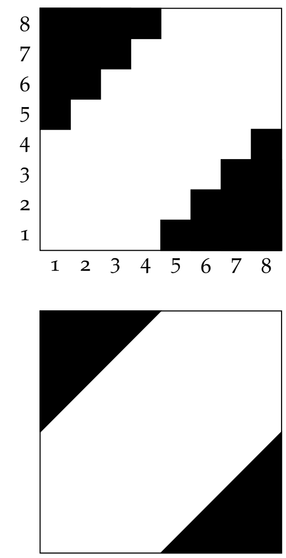

Figure 5.2: The graph of W_{G_n} (for $n = 4$) and the limit as n goes to infinity (black is 1, white is 0)

This process of converting graphs to graphons can be easily generalized.

Definition 5.6. Given a graph G with n vertices (labeled $1, \ldots, n$), we define its ***associated graphon*** as $W_G : [0,1]^2 \to [0,1]$ obtained by partitioning $[0,1] = I_1 \cup I_2 \cup \cdots I_n$ with $\lambda(I_i) = 1/n$ such that if $(x,y) \in I_i \times I_j$, then $W(x,y) = 1$ if i and j are connected in G and 0 otherwise. (Here $\lambda(I)$ is the Lebesgue measure of I.)

However, as we experiment with more examples, we see that using pointwise limit as in Example 5.5 does not suffice for our purpose in general.

Example 5.7. Consider any sequence of random (or quasirandom) graphs with edge density $1/2$ (with number of vertices approaching infinity), then the limit (should) approach the constant function $W = 1/2$, though it certainly does not do so pointwise.

Example 5.8. Consider a complete bipartite graph $K_{n,n}$ with the two parts being odd-indexed and even-indexed vertices. Since the adjacency matrix looks like a checkerboard, we may expect limit to look like the $1/2$ constant function as well, but this is not the case: if we instead label the two parts $1, \ldots, n$ and $n+1, \ldots 2n$, then we see that the graphons should in fact converge to a 2×2 checkerboard instead.

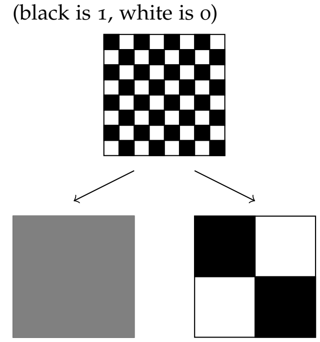
Figure 5.3: A graph of $W_{K_{n,n}}$ and two possible limits of $W_{K_{n,n}}$ as n goes to infinity

The examples above show that we need to (at the very least) take care of relabeling of the vertices in our definition of graph limits.

Definition 5.9. A ***graph homomorphism*** from H to G is a map $\phi : V(H) \to V(G)$ such that if $uv \in E(H)$ then $\phi(u)\phi(v) \in E(G)$. (Maps edges to edges.) Let $\mathrm{Hom}(H,G)$ be the set of all such homomorphisms. and let $\hom(H,G) = |\mathrm{Hom}(H,G)|$. Define ***homomorphism density*** as

$$t(H,G) = \frac{\hom(H,G)}{|V(G)|^{|V(H)|}}.$$

This is also the probability that a uniformly random map is a homomorphism.

Example 5.10. • $\hom(K_1, G) = |V(G)|$,

- $\hom(K_2, G) = 2|E(G)|$,
- $\hom(K_3, G)$ is 6 times the number of triangles in G,
- $\hom(G, K_3)$ is the number of proper 3-colorings of G (where the colors are labeled, say red/green/blue).

Remark 5.11. Note that the homomorphisms from H to G do not quite correspond to copies of subgraphs H inside G, because the homomorphisms can be non-injective. Since the number of non-injective homomorphisms contribute at most $O_H(n^{|V(H)-1|})$ (where $n = |V(G)|$), they form a lower order contribution as $n \to \infty$ when H is fixed.

Definition 5.12. Given a symmetric measurable function $W : [0,1]^2 \to \mathbb{R}$, define

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

Note that $t(H, G) = t(H, W_G)$ for every G and H.

Example 5.13. When $H = K_3$, we have

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

This can be viewed as the "triangle density" of W.

We may now define what it means for graphs to converge and what the limit is.

Definition 5.14. We say that a sequence of graphs G_n (or graphons W_n) is ***convergent*** if $t(H, G_n)$ (or $t(H, W_n)$) converges as n goes to infinity for every graph H. The sequence ***converges to*** W if $t(H, G_n)$ (or $t(H, W_n)$) converges to $t(H, W)$ for every graph H.

Remark 5.15. Though not necessary for the definition, we can think of $|V(G_n)|$ going to infinity as n goes to infinity.

A natural question is whether a convergent sequence of graphs has a "limit". (Spoiler: yes.) We should also consider whether the "limit" we defined this way is consistent with what we expect. To this end, we need a notion of "distance" between graphs.

One simple way to define the distance between G and G' to be $\sum_k 2^{-k}|t(H_k, G) - t(H_k, G')|$ for some sequence $H_1, H_2, \ldots$ of all the graphs. (Here 2^{-k} is added to make sure the sum converges to a number between 0 and 1.) This is topologically equivalent to the concept of convergence in Definition 5.14, but it is not useful.

Another possibility is to consider the ***edit distance*** between two graphs (number of edge changes needed), normalized by a factor of

$1/|V(G)|^2$. This is also not very useful, since the distance between any two $G(n, 1/2)$ is around $1/4$, but we should expect them to be similar (and hence have $o(1)$ distance).

This does, however, inspires us to look back to our discussion of quasirandom graphs and consider when a graph is close to constant p (i.e. similar to $G(n, p)$). Recall the DISC criterion in Theorem 4.1, where we expect $|e(X, Y) - p|X||Y||$ to be small if the graph is sufficiently random. We can generalize this idea to compare the distance between two graphs: intuitively, two graphs (on the same vertex set, say) are close if $|e_G(X, Y) - e_{G'}(X, Y)|/n^2$ is small for all subsets X and Y. We do, however, need some more definitions to handle (for example) graph isomorphisms (which should not change the distances) and graphs of different sizes.

Definition 5.16. The ***cut norm*** of $W : [0, 1]^2 \to \mathbb{R}$ 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.

For future reference, we also define some related norms.

Definition 5.17. For $W : [0, 1]^2 \to \mathbb{R}$, define the ***$L^p$ norm*** as $\|W\|_p = \left(\int |W|^p\right)^{1/p}$, and the ***$L^\infty$ norm*** as the infimum of all the ream numbers m such that the set of all the points (x, y) for which $W(x, y) > m$ has measure zero. (This is also called the ***essential supremum*** of W.)

Definition 5.18. We say that $\phi : [0, 1] \to [0, 1]$ is ***measure-preserving*** if $\lambda(A) = \lambda(\phi^{-1}(A))$ for all measurable $A \subseteq [0, 1]$.

Example 5.19. The function $\phi(x) = x + 1/2 \bmod 1$ is clearly measure-preserving. Perhaps less obviously, $\phi(x) = 2x \bmod 1$ is also measure-preserving, since while each interval is dilated by a factor of 2 under ϕ, every point has two pre-images, so the two effects cancel out. This only works because we compare A with $\phi^{-1}(A)$ instead of $\phi(A)$.

Definition 5.20. Write $W^\phi(x, y) = W(\phi(x), \phi(y))$ (intuitively, "relabelling the vertices"). We define the ***cut distance***

$$\delta_\square(U, W) = \inf_\phi \|U - W^\phi\|_\square$$

where ϕ is a measure-preserving bijection.

For graphs G, G', define the ***cut distance*** $\delta_\square(G, G') = \delta_\square(W_G, W_{G'})$. We also define the cut distance between a graph and a graphon as $\delta_\square(G, U) = \delta_\square(W_G, U)$.

Note that ϕ is not quite the same as permuting vertices: it is allowed to also split vertices or overlay different vertices. This allows us to optimize the minimum discrepancy/cut norm better than simply considering graph isomorphisms.

Remark 5.21. The inf in the definition is indeed necessary. Suppose $U(x, y) = xy$ and $W = U^{\phi}$, where $\phi(x) = 2x \bmod 1$, we cannot attain $\|U - W^{\phi'}\|_{\square} = 0$ for any ϕ' (although the cut distance is 0) since ϕ is not bijective.

Now we present the main theorems in graph limit theory that we will prove later. First of all, one might suspect that there is an alternative definition of convergence using the cut distance metric, but it turns out that this definition is equivalent to Definition 5.14.

Theorem 5.22 (Equivalence of convergence)**.** *A sequence of graphs or graphons is convergent if and only if it is a Cauchy sequence with respect to the cut (distance) metric.*

Borgs, Chayes, Lovász, Sós, and Vesztergombi (2008)

(A Cauchy sequence with respect to metric d is a sequence $\{x_i\}$ that satisfies $\sup_{m \geq 0} d(x_n, x_{n+m}) \to 0$ as $n \to \infty$.)

Theorem 5.23 (Existence of limit)**.** *Every convergent sequence of graphs or graphons has a limit graphon.*

Lovász and Szegedy (2006)

Denote $\tilde{\mathcal{W}}_0$ as the space of graphons, where graphons with cut distance 0 are identified.

Theorem 5.24 (Compactness of the space of graphons)**.** *The set $\tilde{\mathcal{W}}_0$ is a compact metric space under the cut metric.*

Lovász and Szegedy (2007)

Remark 5.25. Intuitively, this means that the spaces of "essentially different" graphs is not very large. This is similar to the regularity lemma, where every graph has a constant-size description that approximates the graph well. In fact, we can consider this compactness theorem as a qualitative analytic version of the regularity lemma.

5.2 *W-random graphs*

Recall the Erdős-Rényi random graphs $G(n, p)$ we've seen before. We now introduce its graphon generalization. Let's start with a special case, **the stochastic block model**. It is a graph with vertices colored randomly (blue or red), and two red vertices are connected with probability p_{rr}, a red vertex and a blue vertex are connected with probability $p_{rb} = p_{br}$, and two blue vertices are connected with probability p_{bb}.

Definition 5.26. Uniformly pick $x_1, \ldots, x_n$ from the interval $[0, 1]$. A ***W-random graph***, denoted $\mathbb{G}(n, W)$, has vertex set $[n]$ and vertices i and j are connected with probability $W(x_i, x_j)$.

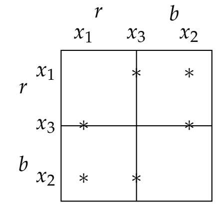

Figure 5.4: 2-block model

An important statistical question is that given a graph, whether there is a good model for where this graph comes from. This gives some motivation to study W-random graphs. We also learnt that the sequence of Erdős-Rényi random graphs converges to the constant graphon, where below is an analogous result.

Theorem 5.27. *Let W be a graphon. Suppose that for all n, G_n are chosen from W-random graphs independently, then $G_n \to W$ almost surely.*

Remark 5.28. In particular, every graphon W is the limit of some sequence of graphs. This gives us some form of graph approximations.

The proof for the above theorem uses Azuma's inequality in order to show that $t(F, G_n) \approx t(F, W)$ with high probability.

5.3 Regularity and counting lemmas

We now develop a series of tools to prove Theorem 5.24.

Theorem 5.29 (Counting Lemma)**.** *For graphons W, U and graph F, we have*

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

Proof. It suffices to prove $|t(F,W) - t(F,U)| \leq |E(F)| \|W-U\|_\square$. Indeed, by considering the above over U replaced by U^ϕ, and taking the infimum over all measure-preserving bijections ϕ, we obtain the desired result.

Recall that the cut norm $\|W\|_\square = \sup_{S,T\subseteq[0,1]} |\int_{S\times T} W|$. Now we prove its useful reformulation: for measurable functions u and v,

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

Here's the reason for the equality to hold: we take $u = 1_S$ and $v = 1_T$ so the left hand side is no more than the right hand side, and then the bilinearity of the integrand in u, v yields the other direction (the extrema are attained for u, v taking values at 0 or 1).

We now illustrate the case when $F = K_3$. Observe that

$$\begin{aligned} t(K_3,W) - t(K_3,U) &= \int ((W(x,y)W(x,z)W(y,z) - U(x,y)U(x,z)U(y,z))dxdydz \\ &= \int (W-U)(x,y)W(x,z)W(y,z)dxdydz \\ &\quad + \int U(x,y)(W-U)(x,z)W(y,z)dxdydz \\ &\quad + \int U(x,y)U(x,z)(W-U)(y,z)dxdydz. \end{aligned}$$

Take the first term as an example: for a fixed z,

$$\left|\int (W-U)(x,y)W(x,z)W(y,z)dxdydz\right| \leq \|W-U\|_\square$$

by the above reformulation. Therefore, the whole sum is bounded by $3\|W-U\|_\square$ as we desire.

For a general graph F, by the triangle inequality we have

$$|t(F,W)-t(F,U)| = \left| \int \Big(\prod_{u_iv_i\in E} W(u_i,v_i) - \prod_{u_iv_i\in E} U(u_i,v_i) \Big) \prod_{v\in V} dv \right|$$

$$\leq \sum_{i=1}^{|E|} \left| \int \left(\prod_{j=1}^{i-1} U(u_j,v_j)(W(u_i,v_i)-U(u_i,v_i)) \prod_{k=i+1}^{|E|} W(u_k,v_k) \right) \prod_{v\in V} dv \right|.$$

Here, each absolute value term in the sum is bounded by $\|W-U\|_\square$ the cut norm if we fix all other irrelavant variables (everything except u_i and v_i for the i-th term), altogether implying that $|t(F,W)-t(F,U)| \leq |E(F)|\ \delta_\square(W,U)$.

□

We now introduce an "averaging function" for graphon W.

Definition 5.30. For a partition $\mathcal{P}=\{S_1,\ldots,S_k\}$ of $[0,1]$ into measurable subsets, and $W:[0,1]^2\to\mathbb{R}$ a symmetrical measurable function, define the ***stepping operator*** $W_\mathcal{P}:[0,1]^2\to\mathbb{R}$ constant on each $S_i\times S_j$ such that $W_\mathcal{P}(x,y)=\frac{1}{\lambda(S_i)\lambda(S_j)}\int_{S_i\times S_j} W$ if $(x,y)\in S_i\times S_j$.

(We ignore the defined term when the denominator equals to 0, because the sets are measure-zero anyway).

This is actually a projection in 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 $S_i\times S_j$.

Theorem 5.31 (Weak regularity lemma)**.** *For any $\epsilon>0$ and any graphon $W:[0,1]^2\to\mathbb{R}$, there exists a partition $\mathcal{P}$ of $[0,1]$ into no more than $4^{1/\epsilon^2}$ measurable sets such that $\|W-W_\mathcal{P}\|_\square\leq\epsilon$.*

Definition 5.32. Given graph G, a partition $\mathcal{P}=\{V_1,\ldots,V_k\}$ of $V(G)$ is called ***weakly ϵ-regular*** if for all $A,B\subset V(G)$,

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

These are similar but different notions we have seen when introducing Theorem 3.5.

Theorem 5.33 (Weak Regularity Lemma for Graphs)**.** *For all $\epsilon>0$ and graph G, there exists a weakly ϵ-regular partition of $V(G)$ into up to $4^{1/\epsilon^2}$ parts.*

Frieze-Kannan (1999)

Lemma 5.34 (L^2 energy increment)**.** *Let W be a graphon and $\mathcal{P}$ a partition of $[0,1]$, satisfying $\|W-W_\mathcal{P}\|_\square>\epsilon$. There exists a refinement $\mathcal{P}'$ of $\mathcal{P}$ dividing each part of $\mathcal{P}$ into no more than 4 parts, such that $\|W_{\mathcal{P}'}\|_2^2 > \|W_\mathcal{P}\|_2^2+\epsilon^2$.*

Proof. Because $\|W - W_{\mathcal{P}}\|_\square > \epsilon$, there exist subsets $S, T \subset [0,1]$ such that $|\int_{S\times T}(W - W_{\mathcal{P}})| > \epsilon$. Let $\mathcal{P}'$ be the refinement of $\mathcal{P}$ by introducing S and T (divide $\mathcal{P}$ based on whether it's in $S \setminus T, T \setminus S$, $S \cap T$ or $\overline{S} \cap \overline{T}$), and that gives at most 4 sub-parts each.

Define $\langle W, U\rangle$ to be $\int WU$. 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 Pythagorean Theorem,

$$\|W_{\mathcal{P}'}\|_2^2 = \|W_{\mathcal{P}'} - W_{\mathcal{P}}\|_2^2 + \|W_{\mathcal{P}}\|_2^2 > \|W_{\mathcal{P}}\|_2^2 + \epsilon^2,$$

where the latter inequality comes by the Cauchy–Schwarz inequality,

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

$\square$

Proposition 5.35. *For any $\epsilon > 0$, graphon W, and $\mathcal{P}_0$ partition of $[0,1]$, there exists partition $\mathcal{P}$ refining part of $\mathcal{P}_0$ into no more than $4^{1/\epsilon^2}$ parts, such that $\|W - W_{\mathcal{P}}\|_\square \leq \epsilon$.*

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

Proof. We repeatedly apply Lemma 5.34 to obtain $\mathcal{P}_0, \mathcal{P}_1, \ldots$ partitions of $[0,1]$. For each step, we either have $\|W - W_{\mathcal{P}}\|_\square \leq \epsilon$ and thus stop, or we know $\|W_{\mathcal{P}'}\|_2^2 > \|W_{\mathcal{P}}\|_2^2 + \epsilon^2$.

Because $\|W_{P_i}\|_2^2 \leq 1$, we are guaranteed to stop after fewer than than ϵ^{-2} steps. We also know that each part is subdivided into no more than 4 parts at each step, obtaining $4^{\epsilon^{-2}}$ as we desire. $\square$

We hereby introduce a related result in computer science, the MAXCUT problem: given a graph G, we want to find $\max e(S, \bar{S})$ among all vertex subsets $S \subset V(G)$. Polynomial-time approximation algorithms developed by Goemans and Williamson that finds a cut within around 0.878 fraction of the optimum. conjecture known as the Unique Games Conjecture would imply that the it would not be possible to obtain a better approximation than the Goemans–Williamson algorithm.2306295 states the impossibility of beating this. It is shown that approximating beyond $\frac{16}{17} \approx 0.941$ is NP-hard.

Goemans and Williamson (1995)

Khot, Kindler, Mossel, and O'Donnell (2007)

Håstad (2001)

On the other hand, the MAXCUT problem becomes easy to approximate for dense graphs, i.e., approximating the size of the maximum cut of an n-vertex graph with in to ϵn^2 additive error in time polynomial in n, where $\epsilon > 0$ is a fixed constant. One can apply an algorithmic version of the weak regularity lemma and brute-force search through all possible partition sizes of the parts. This application was one of the original motivations of the weak regularity lemma.

5.4 *Compactness of the space of graphons*

Definition 5.36. A ***martingale*** is a random sequence $X_0, X_1, X_2, \ldots$ such that for all n, $\mathbb{E}[X_n | X_{n-1}, X_{n-2}, \ldots, X_0] = X_{n-1}$.

Example 5.37. Let X_n denotes the time n balance at a fair casino, where the expected value of each round's gain is 0. Then $\{X_n\}_{n \geq 0}$ is a martingale.

Example 5.38. For a fixed random variable X, we define $X_n = \mathbb{E}(X |$ information up to time n$)$, so that this sequence also forms a martingale.

Theorem 5.39 (Martingale Convergence Theorem)**.** *Every bounded martingale converges almost surely.*

Remark 5.40. Actually, instead of bounded, it is enough for the martingales to be L^1-bounded or uniform integrable, both of which gives $\sup \mathbb{E}(X_n^+) < \infty$.

We sketch a idea inspired by a betting strategy. The proof below omits some small technical details that can be easily filled in for those who are familiar with the basic language of probability theory.

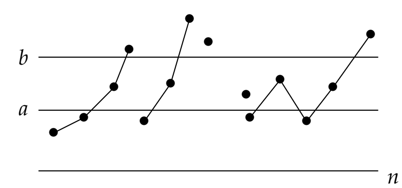

Figure 5.5: examples of "upcrossings"

Proof. An "upcrossing" of $[a, b]$ consists of an interval $[n, n+t]$ such that $X_n < a$, and X_{n+t} is the first instance after X_n such that $X_{n+t} > a$. We refer to the figure on the right instead of giving a more precise definition.

Suppose there is a sequence of bounded martingale $\{X_n\}$ that doesn't converge. Then there exists rational numbers $0 < a < b < 1$ such that $\{X_n\}$ upcrosses the interval $[a, b]$ infinitely many times. We will show that this event occurs with probability 0 (so that after we sum over $a, b \in \mathbb{Q}$, $\{X_n\}$ converges with probability 1).

Denote u_N to be the number of upcrossings (crossings from below to above the interval) up to time N. Consider the following betting strategy: at any time, we hold either 0 or 1 share. If $X_n < a$, then buy 1 share and hold it until the first time that the price (X_n) reads more than b (i.e. we sell at time m such that $X_m > b$ for the first time and $m > n$).

How much profit do we make from this betting strategy? We pocket $b - a$ for each upcrossing. Accounting for difference between our initial and final balance, our profit is at least $(b - a)u_N - 1$. On the other hand, the optional stopping theorem tells us that every "fair" betting strategy on a margingale has zero expected profit. So because the profits of a martingale is zero,

$$0 = \mathbb{E} \text{ profit} \geq (b - a)\mathbb{E} u_N - 1,$$

which implies $\mathbb{E} u_N \leq \frac{1}{b-a}$. Let $u_\infty = \lim_{N\to\infty} u_N$ denotes the total number of upcrossings. By the monotone convergence theorem, we have $\mathbb{E} u_\infty \leq \frac{1}{b-a}$ too, hence $\mathbb{P}(u_\infty = \infty) = 0$, implying our result. $\square$

We now prove the main theorems of graph limits using the tools developed in previous sections, namely the weak regularity lemma (Theorem 5.31) and the martingale convergence theorem (Theorem 5.39). We will start by proving that the space of graphons is compact (Theorem 5.24). In the next section we will apply this result to prove Theorem 5.23 and Theorem 5.22, in that order. We will also see how compactness can be used to prove a graphon-reformulation of the strong regularity lemma.

Recall that $\widetilde{\mathcal{W}}_0$ is the space of graphons modulo the equivalence relation $W \sim U$ if $\delta_\square(W,U) = 0$. We can see that $(\widetilde{\mathcal{W}}_0, \delta_\square)$ is a metric space.

Theorem 5.41 (Compactness of the space of graphons)**.** *The metric space $(\widetilde{\mathcal{W}}_0, \delta_\square)$ is compact.*

Lovász and Szegedy (2007)

Proof. 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.

For each n, apply the weak regularity lemma (Theorem 5.31) repeatedly, to obtain a sequence of partitions

$$\mathcal{P}_{n,1}, \mathcal{P}_{n,2}, \mathcal{P}_{n,3}, \ldots$$

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.

Initially, each partition may be an arbitrary measurable set. However, for each n, we can apply a measure-preserving bijection ϕ to $W_{n,1}$ and $\mathcal{P}_{n,1}$ so that $\mathcal{P}_{n,1}$ is a partition of $[0,1]$ into intervals. For each $k \geq 2$, assuming that $\mathcal{P}_{n,k-1}$ is a partition of $[0,1]$ into intervals, we can apply a measure-preserving bijection to $W_{n,k}$ and $\mathcal{P}_{n,k}$ so that $\mathcal{P}_{n,k}$ is a partition of $[0,1]$ into intervals, and refines $\mathcal{P}_{n,k-1}$. By induction, we therefore have that $\mathcal{P}_{n,k}$ consists of intervals for all n,k. Properties (a) and (b) above still hold. While property (c) may not hold, and it's no longer true that $W_{n,k} = (W_n)_{\mathcal{P}_{n,k}}$, we still know that $\delta_\square(W_n, W_{n,k}) \leq 1/k$ for all n,k. This will suffice for our purposes.

Now, the crux of the proof is a diagonalization argument in countably many steps. Starting with the sequence $W_1, W_2, \ldots$, we will repeatedly pass to a subsequence. In step k, we pick a subsequence $W_{n_1}, W_{n_2}, \ldots$ such that:

1. the endpoints of the parts of $\mathcal{P}_{n_i,k}$ all individually converge as $i \to \infty$, and

2. $W_{n_i,k}$ converges pointwise almost everywhere to some graphon U_k as $i \to \infty$.

There is a subsequence satisfying (1) since each partition $\mathcal{P}_{n,k}$ has exactly m_k parts, and each part has length in $[0,1]$. So consider a subsequence $(W_{a_i})_{i=1}^{\infty}$ satisfying (1). Each $W_{a_i,k}$ can be naturally identified with a function $f_{a_i,k} : [m_k]^2 \to [0,1]$. The space of such functions is bounded, so there is a subsequence $(f_{n_i})_{i=1}^{\infty}$ of $(f_{a_i})_{i=1}^{\infty}$ converging to some $f : [m_k]^2 \to [0,1]$. Now f corresponds to a graphon U_k which is the limit of the subsequence $(W_{n_i})_{i=1}^{\infty}$. Thus, (2) is satisfied as well.

To conclude step k, the subsequence is relabeled as $W_1, W_2, \ldots$ and the discarded terms of the sequence are ignored. The corresponding partitions are also relabeled. Without loss of generality, in step k we pass to a subsequence which contains $W_1, \ldots, W_k$. Thus, the end result of steps $k = 1, 2, \ldots$ is an infinite sequence with the property that $(W_{n,k})_{n=1}^{\infty}$ converges pointwise almost everywhere (a.e.) to U_k for all k:

$$\begin{array}{ccccccl}
 & W_1 & W_2 & W_3 & \cdots & & \\
k=1 & W_{1,1} & W_{2,1} & W_{3,1} & \cdots & \to U_1 & \text{pointwise a.e.} \\
k=2 & W_{1,2} & W_{2,2} & W_{3,2} & \cdots & \to U_2 & \text{pointwise a.e.} \\
k=3 & W_{1,3} & W_{2,3} & W_{3,3} & \cdots & \to U_3 & \text{pointwise a.e.} \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots &
\end{array}$$

Similarly, $(\mathcal{P}_{n,k})_{n=1}^{\infty}$ converges to an interval partition $\mathcal{P}_k$ for all k.

By property (a), each partition $\mathcal{P}_{n,k+1}$ refines $\mathcal{P}_{n,k}$, which implies that $W_{n,k} = (W_{n,k+1})_{\mathcal{P}_{n,k}}$. Taking $n \to \infty$, it follows that $U_k = (U_{k+1})_{\mathcal{P}_k}$ (see Figure 5.6 for an example). Now each U_k can be thought of as a random variable on probability space $[0,1]^2$. From this view, the equalities $U_k = (U_{k+1})_{\mathcal{P}_k}$ exactly imply that the sequence $U_1, U_2, \ldots$ is a martingale.

The range of each U_k is contained in $[0,1]$, so the martingale is bounded. By the martingale convergence theorem (Theorem 5.39), there exists a graphon U such that $U_k \to U$ pointwise almost everywhere as $k \to \infty$.

Recall that our goal was to find a convergent subsequence of $W_1, W_2, \ldots$ under $\delta_\square$. We have passed to a subsequence by the above diagonalization argument, and we claim that it converges to U under

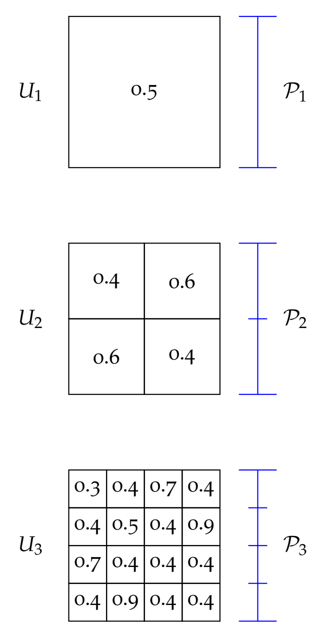

Figure 5.6: An example of possible U_1, U_2, and U_3, each graphon averaging the next.

$\delta_\square$. That is, we want to show that $\delta(W_n, U)_\square \to 0$ as $n \to \infty$. This follows from a standard "3-epsilons argument": let $\epsilon > 0$. Then there exists some $k > 3/\epsilon$ such that $\|U - U_k\|_1 < \epsilon/3$, by pointwise convergence and the dominated convergence theorem. Since $W_{n,k} \to U_k$ pointwise almost everywhere (and by another application of the dominated convergence theorem), there exists some $n_0 \in \mathbb{N}$ such that $\left\|U_k - W_{n,k}\right\|_1 < \epsilon/3$ for all $n > n_0$. Finally, since we chose $k > 3/\epsilon$, we already know that $\delta(W_n, W_{n,k})_\square < \epsilon/3$ for all n. We conclude that

$$\begin{aligned}\delta(U, W_n)_\square &\leq \delta(U, U_k)_\square + \delta(U_k, W_{n,k})_\square + \delta(W_{n,k}, W_n)_\square \\ &\leq \|U - U_k\|_1 + \left\|U_k - W_{n,k}\right\|_1 + \delta(W_{n,k}, W_n)_\square \\ &\leq \epsilon.\end{aligned}$$

The second inequality uses the general bound that

$$\delta(W_1, W_2)_\square \leq \|W_1 - W_2\|_\square \leq \|W_1 - W_2\|_1$$

for graphons W_1, W_2. $\square$

5.5 *Applications of compactness*

We will now use the compactness of $(\widetilde{W}_0, \delta_\square)$ to prove several results, notably the strong regularity lemma for graphons, the equivalence of the convergence criteria defined by graph homomorphism densities and by the cut norm, and the existence of a graphon limit for every sequence of graphons with convergent homomorphism densities.

As a warm-up, we will prove that graphons can be uniformly approximated by graphs under the cut distance. The following lemma expresses what we could easily prove without compactness:

Lemma 5.42. *For every $\epsilon > 0$ and every graphon W, there exists some graph G such that $\delta_\square(G, W) < \epsilon$.*

Proof. By a well-known fact from measure theory, there is a step function U such that $\|W - U\|_1 < \epsilon/2$. For any constant graphon p there is a graph G such that $\|G - p\|_\square < \epsilon/2$; in fact, a random graph $G(n, p)$ satisfies this bound with high probability, for sufficiently large n. Thus, we can find a graph G such that $\|G - U\|_\square < \epsilon/2$ by piecing together random graphs of various densities. So

$$\delta_\square(G, W) \leq \|W - U\|_1 + \|U - G\|_\square < \epsilon$$

as desired. $\square$

However, in the above lemma, the size of the graph may depend on W. This can be remedied via compactness.

Proposition 5.43. *For every $\epsilon > 0$ there is some $N \in \mathbb{N}$ such that for any graphon W, there is a graph G with N vertices such that $\delta_\square(G, W) < \epsilon$.*

Proof. For a graph G, define the ϵ-ball around G by $B_\epsilon(G) = \{W \in \widetilde{W}_0 : \delta_\square(G, W) < \epsilon\}$.

As G ranges over all graphs, the balls $B_\epsilon(G)$ form an open cover of $\widetilde{W}_0$, by Lemma 5.42. By compactness, this open cover has a finite subcover. So there is a finite set of graphs $G_1, \ldots, G_k$ such that $B_\epsilon(G_1), \ldots, B_\epsilon(G_k)$ cover $\widetilde{W}_0$. Let N be the least common multiple of the vertex sizes of $G_1, \ldots, G_k$. Then for each G_i there is some N-vertex graph G_i' with $\delta_\square(G_i, G_i') = 0$, obtained by replacing each vertex of G_i with $N/|V(G_i)|$ vertices. But now W is contained in an ϵ-ball around some N-vertex graph. $\square$

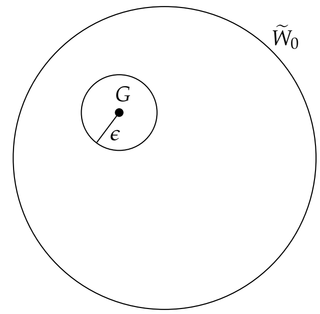

Figure 5.7: Cover of $\widetilde{W}_0$ by open balls

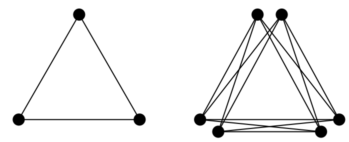

Figure 5.8: A K_3 and its 2-blowup. Note that the graphs define equal graphons.

Remark 5.44. Unfortunately, the above proof gives no information about the dependence of N on ϵ. This is a byproduct of applying compactness. One can use regularity to find an alternate proof which gives a bound.

Intuitively, the compactness theorem has a similar flavor to the regularity lemma; both are statements that the space of graphs is in some sense very small. As a more explicit connection, we used the weak regularity lemma in our proof of compactness, and the strong regularity lemma follows from compactness straightforwardly.

Theorem 5.45 (Strong regularity lemma for graphons)**.** *Let $\epsilon = (\epsilon_1, \epsilon_2, \ldots)$ be a sequence of positive real numbers. Then there is some $M = M(\epsilon)$ such that every graph W can be written*

$$W = W_{str} + W_{psr} + W_{sml}$$

where

- *W_{str} is a step function with $k \le M$ parts,*
- *$\|W_{psr}\|_\square \le \epsilon_k$,*
- *$\|W_{sml}\|_1 \le \epsilon_1$.*

Lovász and Szegedy (2007)

If $\epsilon_k = \epsilon/k^2$, then this theorem approximately recovers Szemerédi's Regularity Lemma. If $\epsilon_k = \epsilon$, then it approximately recovers the Weak Regularity Lemma.

Proof. It is a well-known fact from measure theory that any measurable function can be approximated arbitrarily well by a step function. Thus, for every graphon W there is some step function U such that $\|W - U\|_1 \le \epsilon_1$. Unfortunately, the number of steps may depend on W; this is where we will use compactness.

For graphon W, let $k(W)$ be the minimum k such that some k-step graphon U satisfies $\|W - U\|_1 \le \epsilon_1$. Then $\{B_{\epsilon_{k(W)}}\}_{W \in \widetilde{W}_0}$ is clearly an open cover of $\widetilde{W}_0$, and by compactness there is a finite set of graphons $\mathcal{S} \subset \widetilde{W}_0$ such that $\{B_{\epsilon_{k(W)}}(W)\}_{W \in \mathcal{S}}$ covers $\widetilde{W}_0$.

Let $M = \max_{W \in \mathcal{S}} k(W)$. Then for every graphon W, there is some $W' \in \mathcal{S}$ such that $\delta_\square(W, W') \le \epsilon_{k(W')}$. Furthermore, there is a k-step graphon U with $k = k(W') \le M$ such that $\|W' - U\|_1 \le \epsilon_1$. Hence,

$$W = U + (W - W') + (W' - U)$$

is the desired decomposition, with $W_{\text{str}} = U$, $W_{\text{psr}} = W - W'$, and $W_{\text{sml}} = W' - U$. $\square$

Earlier we defined convergence of a sequence of graphons in terms of the sequences of F-densities. However, up until now we did not know that the limiting F-densities of a convergent sequence of graphons are achievable by a single graphon. Without completing the space of graphs to include graphons, this is in fact not true, as we saw in the setting of quasirandom graphs. Nonetheless in the space of graphons, the result is true, and follows swiftly from compactness.

Theorem 5.46 (Existence of limit)**.** *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. Then the sequence of graphons converges to some W. That is, there exists a graphon W such that $t(F, W_n) \to t(F, W)$ for every F.*

Lovász and Szegedy (2006)

Proof. By sequential compactness, 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 a graph F. By Theorem 5.29, it follows that $t(F, W_{n_i}) \to t(F, W)$. But by assumption, the sequence $\{t(F, W_n)\}_n$ converges, so all subsequences have the same limit. Therefore $t(F, W_n) \to t(F, W)$. $\square$

The last main result of graph limits is the equivalence of the two notions of convergence which we had defined previously.

Theorem 5.47 (Equivalence of convergence)**.** *Convergence of F-densities is equivalent to convergence under the cut norm. That is, let $W_1, W_2, \ldots$ be a sequence of graphons. Then the following are equivalent:*

Borgs, Chayes, Lovász, Sós, and Vesztergombi (2008)

- *The sequence of F-densities $\{t(F, W_n)\}_n$ converges for all graphs F*
- *The sequence $\{W_n\}_n$ is Cauchy with respect to $\delta_\square$.*

Proof. One direction follows immediately from Theorem 5.29, the counting lemma: if the sequence $\{W_n\}_n$ is Cauchy with respect to $\delta_\square$, then the counting lemma implies that for every graph F, the sequence of F-densities is Cauchy, and therefore convergent.

For the reverse direction, suppose that the sequence of F-densities converges for all graphs F. Let W and U be limit points of $\{W_n\}_n$ (i.e. limits of convergent subsequences). We want to show that $W = U$.

Let $(n_i)_{i=1}^\infty$ be the 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.

By the subsequent lemma, this implies that $U = W$. $\square$

Lemma 5.48 (Moment lemma)**.** *Let U and W be graphons such that $t(F, W) = t(F, U)$ for all F. Then $\delta_\square(U, W) = 0$.*

This lemma is named in analogy with the moment lemma from probability, which states that if two random variable have the same moments (and are sufficiently well-behaved) then they are in fact identically distributed.

Proof. We will sketch the proof. Let $\mathbb{G}(k, W)$ denote the W-random graph on k vertices (see Definition 5.26). It can be shown that for any k-vertex graph F,

$$\Pr[\mathbb{G}(k, W) \cong F \text{ as labelled graph}] = \sum_{F' \supseteq F} (-1)^{E(F')-E(F)} t(F', W).$$

In particular, this implies that the distribution of W-random graphs is entirely determined by F-densities. So $\mathbb{G}(k, W)$ and $\mathbb{G}(k, U)$ have the same distributions.

Let $\mathbb{H}(k, W)$ be an edge-weighted W-random graph on vertex set $[k]$, with edge weights sampled as follows. Let $x_1, \ldots, x_k \sim \text{Unif}([0, 1])$ be independent random variables. Set the edge-weight of (i, j) to be $W(x_i, x_j)$.

We claim two facts, whose proofs we omit

- $\delta_\square(\mathbb{H}(k, W), \mathbb{G}(k, W)) \to 0$ as $k \to \infty$ with probability 1, and
- $\delta_1(\mathbb{H}(k, W), W) \to 0$ as $k \to \infty$ with probability 1.

Since $\mathbb{G}(k, W)$ and $\mathbb{G}(k, U)$ have the same distribution, it follows from the above facts and the triangle inequality that $\delta_\square(W, U) = 0$. $\square$

A consequence of compactness and the moment lemma is that the "inverse" of the graphon counting lemma also holds: a bound on F-densities implies a bound on the cut distance. The proof is left as an exercise.

Corollary 5.49 (Inverse counting lemma)**.** *For every $\epsilon > 0$ there is some $\eta > 0$ and integer $k > 0$ such that if U and W are graphons with*

$$|t(F, U) - t(F, W)| \leq \eta$$

for every graph F on at most k vertices, then $\delta_\square(U, W) \leq \epsilon$.

Remark 5.50*.* The moment lemma implies that a graphon can be recovered by its F-densities. We might ask whether all F-densities are necessary, or whether a graphon can be recovered from, say, finitely many densities. For example, we have seen that if W is the pseudorandom graphon with density p, then $t(K_2, W) = p$ and $t(C_4, W) = p^4$; furthermore, it is uniquely determined by these densities. If the equalities hold then $\delta_\square(W, p) = 0$.

The graphons which can be recovered from finitely many F-densities in this way are called "finitely forcible graphons". Among the graphons known to be finitely forcible are any step function and the half graphon $W(x,y) = \mathbf{1}_{x+y\geq 1}$. More generally, $W(x,y) = \mathbf{1}_{p(x,y)\geq 0}$ is finitely forcible for any symmetric polynomial $p \in \mathbb{R}[x,y]$ which is monotone decreasing on $[0,1]$.

Lovász and Sós (2008)

Lovász and Szegedy (2011)

5.6 *Inequalities between subgraph densities*

One of the motivations for studying graph limits is that they provide an efficient language with which to think about graph inequalities. For instance, we could be able to answer questions such as the following:

Question 5.51. If $t(K_2, G) = 1/2$, what is the minimum possible value of $t(C_4, G)$?

We know the answer to this question; as discussed previously, by Theorem 4.1 we can consider a sequence of quasirandom graphs; their limit is a graphon W such that $t(K_2, W) = 2^{-4}$.

In this section we work on these kind of problems; specifically, we are interested in homomorphism density inequalities. Two graph inequaities have been discussed previously in this book; Mantel's theorem (Theorem 2.2) and Turán's theorem (Theorem 2.6):

Theorem 5.52 (Mantel's Theorem)**.** *Let* $W : [0,1]^2 \to [0,1]$ *be a graphon. If* $t(K_3, W) = 0$*, then* $t(K_2, W) \leq 1/2$.

Theorem 5.53 (Turán's theorem)**.** *Let* $W : [0,1]^2 \to [0,1]$ *be a graphon. If* $t(K_{r+1}, W) = 0$*, then* $t(K_2, W) \leq 1 - 1/r$.

Our goal in this section is to determine the set of all feasible ***edge density, triangle density*** pairs for a graphon W, which can be formally written as

$$D_{2,3} = \{(t(K_2, W), t(K_3, W)) \ : \ W \text{ graphon } \} \subseteq [0,1]^2.$$

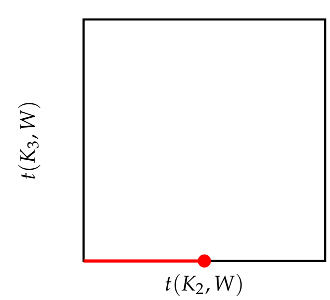

Figure 5.9: Mantel's Theorem implication in the plot of $D_{2,3}$ (red line)

We know that the limit point of a sequence of graphs is a graphon (Theorem 5.23), hence the region $D_{2,3}$ is closed. Moreover, Mantel's Theorem (Theorem 5.52) tells us that the horizontal section of this region when triangle density is zero extends at most until the point $(1/2, 0) \in [0,1]^2$ (see Figure 5.9).

A way in which we can describe $D_{2,3}$ is by its cross sections. A simple argument below shows that each vertical cross section of $D_{2,3}$ is a line segment:

Proposition 5.54. *For every* $0 \leq r \leq 1$*, the set* $D_{2,3} \cap [0,1] \times \{r\}$ *is a line segment with no gaps.*

Proof. Consider two graphons W_0, W_1 with the same edge density; then, we can consider

$$W_t = (1-t)W_0 + tW_1,$$

which is a graphon; moreover, its triangle density is mapped continuously as t varies from 0 to 1. Its initial and final values are $t(K_3, W_0)$ and $r(K_3, W_1)$, respectively, so every triangle density between these values can be achieved. $\square$

Then, in order to better understand the shape of $D_{2,3}$, we would like to determine the minimum and maximum subgraph densities that can be achieved given a fixed edge density. We begin by addressing this question:

Question 5.55. What is the maximum number of triangles in an n-vertex m-edge graph?

An intuitive answer would be that the edges should be arranged so as to form a clique. This turns out to be the correct answer: a result known as the Kruskal–Katona theorem implies that a graph with $\binom{k}{2}$ has at most $\binom{k}{3}$ triangles. Here we prove an slightly weaker version of this bound.

The Kruskal–Katona theorem can be proved using a "compression argument": we repeatedly "push" the edges towards the clique and show that number of triangles can never decrease in the process.

Theorem 5.56. *For every graphon* $W : [0,1]^2 \to [0,1]$,

$$t(K_3, W) \leq t(K_2, W)^{3/2}.$$

Remark 5.57. This upper bound is achieved by a graphon like the one shown in Figure 5.10, which is a limit graphon of a sequence of cliques in G; for each of these graphons, edge and triangle densities are, respectively,

$$t(K_2, W) = a^2, \quad t(K_3, W) = a^3.$$

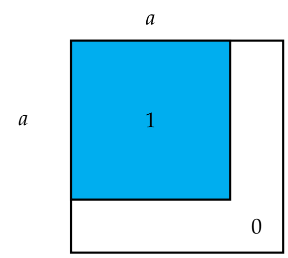

Figure 5.10: Graphon which achieves upper boundary of $D_{2,3}$: $t(K_2, W) = a^2$ and $t(K_3, W) = a^3$

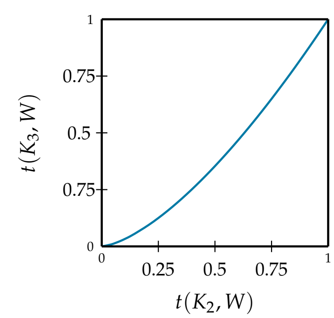

Figure 5.11: Plot of upper boundary of $D_{2,3}$, given by the curve $y = x^{3/2}$ in $[0,1]^2$

Therefore, The upper boundary of the region $D_{3,2}$ is given by the curve $y = x^{3/2}$, as shown by Figure 5.11.

Proof of Theorem 5.56. It suffices to prove the following inequality for every graph G:

$$t(K_3, G) \leq t(K_2, G)^{3/2}.$$

Let us look at $\hom(K_3, G)$ and $\hom(K_2, G)$; these count the number of closed walks in the graph of length 3 and 2, respectively. These values correspond to the second and third moments of the spectrum of the graph G:

$$\hom(K_3, G) = \sum_{i=1}^{k} \lambda_i^3 \quad \text{and} \quad \hom(K_2, G) = \sum_{i=1}^{k} \lambda_i^2$$

Where $\{\lambda_i\}_{i=1}^n$ are the eigenvalues of the adjacency matrix A_G. We then have that

$$\hom(K_3, G) = \sum_{i=1}^n \lambda_i^3 \le \left(\sum_{i=1}^n \lambda_i^2\right)^{3/2} = \hom(K_2, G)^{3/2}. \tag{5.1}$$

After dividing by $|V(G)|^3$ on both sides, the result follows. □

Note that in the last proof, we used the following useful inequality, with $a_i = \lambda_i^2$ and $t = 3/2$:

Claim 5.58. Let $t > 1$, and $a_1, \cdots, a_n \ge 0$. Then,

$$a_1^t + \cdots + a_n^t \le (a_1 + \cdots + a_n)^t$$

Proof. This inequality is homogeneous with respect to the variables a_i, so we can normalize and assume that $\sum a_i = 1$; therefore, each of the $a_i \in [0,1]$, so that $a_i^t \le a_i$ for each i. Therefore,

$$LHS = a_1^t + \cdots + a_n^t \le a_1 + \cdots + a_n = 1 = 1^t = RHS. \qquad \square$$

The reader might wonder whether there is a way to prove this without using eigenvalues of the graph G. We have following result, whose proof does not require spectral graph theory:

Theorem 5.59. *For every $W : [0,1]^2 \to \mathbb{R}$ which is symmetric,*

$$t(K_3, W) \le t(K_2, W^2)^{3/2}$$

where W^2 corresponds to the graphon W, squared pointwise.

Note that above, $t(K_2, W)^{3/2}$ falls in between these two terms when W is a graphon because all the terms would be bounded between 0 and 1; therefore, the above result is stronger than that of Theorem 5.56. The proof of this result follows from applying the Cauchy–Schwarz inequality three times; one corresponding to each edge of a triangle K_3.

Proof. We have

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

From now on, we drop the notation for our intervals of integration. We can apply the Cauchy–Schwarz inequality on the following integral; first with respect to the variable dx, and subsequently with respect to the variables dy, dz, each time holding the other two variables constant:

$$\begin{aligned}
t(K_3, W) &= \int W(x,y)W(x,z)W(y,z)dxdydz \\
&\leq \int \left(\int W(x,y)^2 dx\right)^{1/2} \left(\int W(x,z)^2 dx\right)^{1/2} W(y,z)dydz \\
&\leq \int \left(\int W(x,y)^2 dxdy\right)^{1/2} \left(\int W(x,z)^2 dx\right)^{1/2} \left(\int W(y,z)^2 dy\right)^{1/2} dz \\
&\leq \left(\int W(x,y)^2 dxdy\right)^{1/2} \left(\int W(x,z)^2 dxdz\right)^{1/2} \left(\int W(y,z)^2 dydz\right)^{1/2} \\
&= \|W\|_2^3 \\
&= t(K_2, W)^{3/2},
\end{aligned}$$

completing the proof. □

Remark 5.60. If we did not have the condition that W is symmetric, we could still use Hölder's inequality; however, we would obtain a weaker statement. In this situation, Hölder's inequality would imply that

$$\int_{[0,1]^3} f(x,y)g(x,z)h(y,z)dxdydz \leq \|f\|_3 \|g\|_3 \|h\|_3,$$

and by setting $f = g = h = W$, we could derive a weaker bound than the one obtained in the proof of Theorem 5.59 because, in general, $\|W\|_2 \leq \|W\|_3$.

The next theorem allows us to prove linear inequalities between clique densities.

Theorem 5.61 (Bollobás). *Let $c_1, \cdots, c_n \in \mathbb{R}$. The inequality*

Bollobás (1986)

$$\sum_{r=1}^{n} c_r t(K_r, G) \geq 0$$

holds for every graph G if and only if it holds for every $G = K_m$ with $m \geq 1$. More explicitly, the inequality holds for all graphs G if and only if

$$\sum_{r=1}^{n} c_r \cdot \frac{m(m-1)\cdots(m-r+1)}{m^r} \geq 0$$

for every $m \geq 1$.

Proof. One direction follows immediately because the set of clique graphs is a subset of the set of all graphs.

We now prove the other direction. The inequality holds for all graphs if and only if it holds for all graphons, again since the set of graphs is dense in $\widetilde{W_0}$ with respect to the cut distance metric. In particular, let us consider the set $\mathcal{S}$ of node-weighted simple graphs, with a normalization $\sum a_i = 1$.

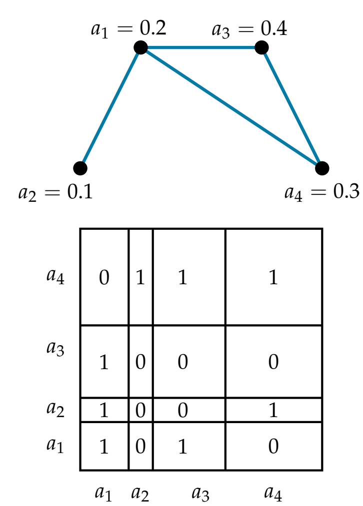

Figure 5.12: Example of a node weighted graph on four vertices, whose weights sum to 1, and its corresponding graphon

As Figure 5.12 shows, each node weighted graph can be represented by a graphon. The set $\mathcal{S}$ is dense in $\widetilde{W}_0$, because this set contains the set of unweighted simple graphs. Then, it suffices to prove this inequality for graphs in $\mathcal{S}$.

Suppose for the sake of contradiction that there exists a node weighted simple graph H such that

$$f(H) := \sum_{r=1}^{n} c_r t(K_r, H) < 0$$

Among all such H, we choose one with smallest possible number m of nodes. We choose node weights $a_1, \cdots, a_m$ with sum equal to 1 such that $f(H)$ is minimized. We can find such H because we have a finite number of parameters, and f is a continuous function over a compact set.

We have that $a_i > 0$ without loss of generality; otherwise we would have a contradiction because we could delete that node and decrease the quantity $|V(H)|$, while $f(H) < 0$ would still hold.

Moreover, H is a complete graph; otherwise there exist i, j such that $ij \notin E(H)$. Note that the clique density is a polynomial in terms of the node weights; this polynomial would not have an a_i^2 term because the set of graphons $\mathcal{S}$ corresponds to simple graphs, and the vertex i would not be adjacent to itself. This polynomial does not have an $a_i a_j$ term either, because i and j are not adjacent. Therefore, $f(H)$ is multilinear in the variables a_i and a_j.

Fixing all of the other node weights and considering a_i, a_j as our variables of the multilinear function $f(H)$, this function would be minimized by setting $a_i = 0$ or $a_j = 0$. If one of these weights were set to zero, this would imply a decrease in the number of nodes, while $a_i + a_j$ would be preserved, hence not increasing $f(H)$. This is a contradiction to the minimality of number of nodes in H such that $f(H) < 0$.

In other words, H must be a complete graph; further, the polynomial $f(H)$ on the variables a_i has to be symmetric:

$$f(H) = \sum_{r=1}^{n} c_r r! s_r,$$

where each s_r is an elementary symmetric polynomial of degree r

$$s_r = \sum_{i_1 < \cdots < i_r} a_{i_1} \cdots a_{i_r}.$$

In particular, by making constant all variables but a_1, a_2, the polynomial $f(H)$ can be written as

$$f(H) = A + B_1 a_1 + B_2 a_2 + C a_1 a_2,$$

where A, B_1, B_2, C are constants; by symmetry, we have $B_1 = B_2$; also, since $\sum a_i = 1$, we have that $a_1 + a_2$ is constant, so that

$$f(H) = A' + Ca_1a_2.$$

If $C > 0$ then f would be minimized when $a_1 = 0$ or $a_2 = 0$; this cannot occur because of the minimality of the number of nodes in H. If $C = 0$ then any value of a_1, a_2 would yield the same minimum value of $f(H)$; in particular we could set $a_1 = 0$, again contradicting minimality on the number of nodes. Therefore, the constant C must be negative,implying that $f(H)$ would be minimized when $a_1 = a_2$. Then, all of the a_i have to be equal, and H can also be regarded as an unweighted graph.

In other words, if the inequality of interest fails for some graph H, then it must fail for some unweighted clique H; this completes the proof. □

Remark 5.62. In the proof above, we only considered clique densities; an inequality over other kinds of graphs would not necessarily hold.

Thanks to the theorem above, it is relatively simple to test linear inequalities between densities, since we just have to verify them for cliques. We have the following corollary:

Corollary 5.63. *For each n, the extremal points of the convex hull of*

$$\{((t(K_2, W), t(K_3, W), \cdots, t(K_n, W)) \ : \ W \text{ graphon}\} \subset [0,1]^{n-1}$$

are given by $W = K_m$ for all $m \geq 1$.

Note that the above claim implies Turán's theorem, because by Theorem 5.61, the extrema of the set above are given in terms of clique densities, which can be understood by taking W to be a clique. Thus, if $t(K_{r+1}, W) = 0$, then this cross section on the higher dimensional cube $[0,1]^r$ will be bounded by the value $t(K_2, W) = 1 - \frac{1}{r}$.

In the particular case that we want to find the extremal points in the convex hull of $D_{2,3} \subset [0,1]^2$, they correspond to

$$p_m = \left(\frac{m-1}{m}, \frac{(m-1)(m-2)}{m^2} \right)$$

All of the points of these form in fact fall into the curve given by $y = x(2x-1)$, which is the dotted red curve in Section 5.6.

Because the region $D_{2,3}$ is contained in the convex hull of the red points $\{p_m\}_{m \geq 0}$, it also lies above the curve $y = x(2x-1)$. We can moreover draw line segments between the convex hull points, so as to obtain a polygonal region that bounds $D_{2,3}$.

The region $D_{2,3}$ was determined by Razborov, who developed the theory of ***flag algebras***, which have provided a useful framework in

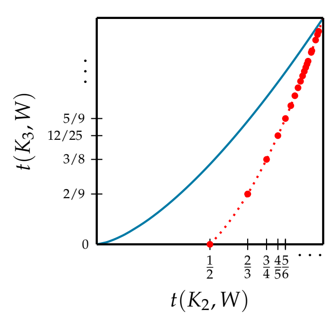

Figure 5.13: Set of lower boundary points of $D_{2,3}$, all found in the curve given by $y = x(2x-1)$
Razborov (2007)

which to set up sum of squares inequalities, e.g., large systematic applications of the Cauchy–Schwarz inequalities, that could be used in order to prove graph density inequalities.

Theorem 5.64 (Razborov). *For a fixed edge density $t(K_2, W)$, which falls into the following interval, for some $k \in \mathbb{N}$*

$$t(K_2, W) \in \left[1 - \frac{1}{k-1}, 1 - \frac{1}{k}\right],$$

the minimum feasible $t(K_3, W)$ is attained by a unique step function graphon corresponding to a k-clique with node weights $a_1, a_2, \cdots, a_k$ with sum equal to 1, and such that $a_1 = \cdots = a_{k-1} \geq a_k$.

Razborov (2008)

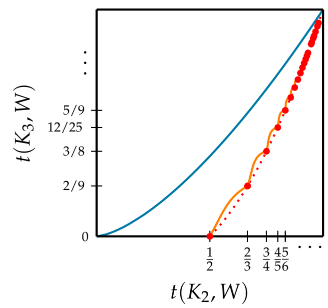

Figure 5.14: Complete description of the region $D_{2,3} \subset [0,1]^2$

α_3	0	1	1
α_2	1	0	1
α_1	1	1	0
	α_1	α_2	α_3

Figure 5.15: A non unique optimal graphon in the case $k = 3$.

The region $D_{2,3}$ is illustrated on the right in Section 5.6. We have exaggerated the drawwings of the concave "scallops" in the lower boundary of the region for better visual effects.

Note that in Turán's theorem, the construction for the graphs which correspond to extrema value (Chapter 2, definition 2.5) are unique; however, in all of the intermediate values $t(t_2, W) \neq 1 - 1/k$, this theorem provides us with non-unique constructions.

To illustrate why these constructions are not unique, the graphon in Figure 5.15, which is a minimizer for triangle density when $t(t_2, W) = 2/3$ can be modified by replacing the highlighted region by any graphon with the same edge density.

Non-uniqueness of graphons that minimize $t(K_3, W)$ implies that this optimization problem is actually difficult.

The problem of minimizing the K_r-density in a graph of given edge density was solved for $r = 4$ by Nikiforov and all r by Reiher., respectively.

Nikiforov (2011)

Reiher (2016)

More generally, given some inequality between various subgraph densities, can we decide if the inequality holds for all graphons?

For polynomial inequalities between homomorphism densities, it suffices to only consider linear densities, since $t(H, W)t(H', W) = t(H \sqcup H', W)$.

Let us further motivate with a related, more classical question regarding nonnegativity of polynomials:

Question 5.65. Given a multivariable polynomial $p \in \mathbb{R}[x_1, x_2, \cdots, x_n]$, is $p(x) \geq 0$ for every $x = (x_1, \cdots, x_n)$?

This problem is decidable, due to a classic result of Tarski that every the first-order theory of the reals is decidable. In fact, we have the following characterization of nonnegative real polynomials.

Theorem 5.66 (Artin). *A polynomial $p \in \mathbb{R}[x_1, x_2, \cdots, x_n]$ is nonnegative if and only if it can be written as a sum of squares of rational functions.*

However, when we turn our interest into the set of lattice points, the landscape changes:

Question 5.67. Given a multivariable polynomial $p \in \mathbb{R}[x_1, x_2, \cdots, x_n]$, can it be determined whether $p(x_1, \cdots, x_n) \geq 0$ for all $x \in \mathbb{Z}^n$?

The answer to the above question is no. This is related to the fact that one cannot solve diophantine equations, or even tell whether there is a solution:

Theorem 5.68 (Matiyasevich; Hilbert's 10th problem)**.** *Given a general diophantine equation is an undecidable problem to find its solutions, or even to determine whether integer solutions exist.*

Matiyasevich (2011)

Turning back to our original question of interest, we want to know whether the following question is decidable

Question 5.69. For a given set of graphs $\{H_i\}_{i\in[k]}$ and $a_1, \cdots, a_k \in \mathbb{R}$, is $\sum_{i=1}^{k} a_i t(H_i, G) \geq 0$ true for every graph G?

The following theorem provides an answer to this question:

Theorem 5.70 (Hatami - Norine)**.** *Given a set of graphs $\{H_i\}_{i\in[k]}$ and $a_1, \cdots, a_k \in \mathbb{R}$, whether the inequality*

$$\sum_{i=1}^{k} a_i t(H_i, G) \geq 0$$

is true for every graph G is undecidable.

Hatami and Norine (2011)

A rough intuition for why the above theorem is true is that we actually have a discrete set of points along the lower boundary of $D_{2,3}$; one could reduce the above problem into proving the same inequalities along the points in the intersection of the red curve and the region. The set of points in this intersection forms a discrete set, and the idea is to encode integer inequalities (which are undecidable) into graph inequalities by using the special points on the lower boundary of $D_{2,3}$.

Another kind of interesting question is to ask whether specific inequalities are true; there are several open problems of that type. Here is an important conjecture in extremal graph theory:

Conjecture 5.71 (Sidorenko's Conjecture)**.** *If H is a bipartite graph then*

$$t(H, W) \geq t(K_2, W)^{e(H)}.$$

Sidorenko (1993)

We worked recently with an instance of the above inequality, when $H = C_4$, when we were discussing quasirandomness. However, the above problem is open. Let us consider the Möebius strip graph

- which consists in removing a 10-cycle from a complete bipartite graph $K_{5,5}$ (Section 5.6).

The name of this graph comes from its realization as a face-vertex incidence graph of the usual simplicial complex of the Möebius strip. The graph above is the first one for which this inequality remains an open problem.

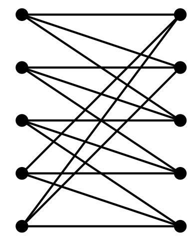

Figure 5.16: The Möebius strip graph.

Even if nonnegativeness of a general linear graph inequalities is undecidable, if one wants to decide whether they are true up to an ε-error, the problem becomes more accessible:

Theorem 5.72. *There exists an algorithm that, for every $\varepsilon > 0$ decides correctly that*

$$\sum_{i=1}^{n} c_i t(H_i, G) \geq -\varepsilon$$

for all graphs G, or outputs a graph G such that

$$\sum_{i=1}^{n} c_i t(H_i, G) < 0.$$

Proof sketch. As a result of weak regularity lemma, we can take a weakly ε-regular partition. All the information regarding edge densities can be represented by this partition; in other words, one would only have to test a bounded number of possibilities on weighted node graphs with $\leq M(\varepsilon)$ parts whose edge weights are multiples of ε. If the estimate for the corresponding weighted sum of graph densities is true for the auxiliary graph one gets from weak regularity lemma, then it is also true for the original graph up to an ε-error; otherwise, we can output a counterexample. $\square$

18.217 Graph Theory and Additive Combinatorics
Fall 2019