Graph on, graph off

Travis, Mathcamp 2024

In class today, we discussed what convergence of a sequence of graphs might mean. There were three possible ideas: A genuine graph on an infinite number of vertices, a “tile” picture as a limit of the adjacency matrices, and convergence in terms of homomorphism densities.

Throughout these notes, F, G, and H are graphs.

Definition 1. A *graph homomorphism* from F to G is a function $f\colon V(F) \to V(G)$ such that $f(x)f(y)$ is an edge of G whenever xy is an edge of F. The number of homomorphsims from F to G is denoted $\hom(F, G)$.

We let $\mathsf{v}(F)$ denote the number of vertices in F.

Definition 2. The *homomorphism density* of F in G is

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

which is the proportion of maps $V(F) \to V(G)$ that are valid homomorphisms.

In other words, $t(F, G)$ is the *probability* that a random map $V(F) \to V(G)$ is a valid homomorphism.

One reason that homomorphism numbers are important is because they mathematically capture the idea of *sampling* from large graphs, which is one of the original motivations of graph limits. Another reason they are important is because they completely determine the graph:

Proposition 3. *If* $\hom(F, G) = \hom(F, H)$ *for every graph* F*, then* $G \cong H$.

The proof uses a lemma. Let $\mathsf{inj}(F, G)$ denote the number of *injective* homomorphisms from F to G.

Lemma 4. *If* $\hom(F, G) = \hom(F, H)$ *for every graph* F*, then* $\mathsf{inj}(F, G) = \mathsf{inj}(F, H)$ *for every graph* F.

The proof of Proposition 3 consists of plugging in $F = G$ and $F = H$ to Lemma 4.

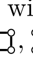
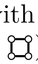
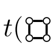
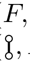
For each graph sequence $(G_1, G_2, G_3, \dots)$ and every graph F, there is a new sequence $\big(t(F, G_1), t(F, G_2), \dots\big)$, which is a sequence of real numbers. We ended class by using this to define convergence:

Definition 5. A sequence of graphs $(G_1, G_2, \dots)$ is *convergent* if $\big(t(F, G_n)\big)_{n \geq 1}$ is a convergent sequence for every graph F.

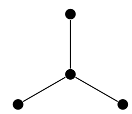
We said that this wasn’t a very satisfactory definition, because it doesn’t tell you anything about what the sequence is converging *to*. We’ll talk about that tomorrow!

PROBLEMS

Problems 1 and 2 are the most important to make sure you’re comfortable with what we talked about in class and are prepared for tomorrow. The rest of the problems are just for fun—work on what seems interesting, or just come talk to me about your questions!

1. Here is some practice with homomorphism density calculations:
   (a) Calculate hom(, ). (Hint: It is more than 8.) What is t(,)?
   (b) What is $\lim_{n\to\infty} t(F, K_n)$? (Work out the solution for every graph F.)
   (c) What is $\lim_{n\to\infty} t$(, P_n), where P_n is the path with n vertices?
   (d) Suppose that F_1 is a subgraph of F_2. Prove that $t(F_1, G) \geq t(F_2, G)$ for every graph G.
   (e) Suppose $(G_1, G_2, \dots)$ is a sequence of graphs where every vertex has degree at most d and $\lim_{n\to\infty} \mathsf{v}(G_n) = +\infty$. Prove that $\lim_{n\to\infty} t(F, G_n) = 0$ for every graph F with at least one edge.
2. For each tile below, find a sequence of graphs that converges to it.

3. Here are some interpretations of homomorphism numbers, to get more used to them.
   (a) Prove that $\hom(P_k, G)$ is the number of walks of length k in G. (P_k is the *path* with k vertices; a *walk* is a sequence of vertices that are connected by an edge.)
   (b) If you paint each vertex of a graph one color, using a palette of r colors, what you get is called an *r-coloring* of the graph. (You don't have to use all r colors.) If every pair of vertices that are connected by an edge have different colors, the r-coloring is called *proper*. Prove that $\hom(G, K_r)$ is the number of proper r-colorings of G.
   (c) The *star graph* S_k has a single vertex connected to k other vertices. The graph S_3 looks like this:



   Prove that
$$\hom(S_k, G) = \sum_{x \in V(G)} \deg(x)^k.$$

4. Here is the "product property" of graph homomorphisms. Given two graphs F_1 and F_2, let $F_1 \sqcup F_2$ denote their *disjoint union*, which is the (disconnected) graph obtained by placing the two graphs side by side.[1] Prove that $t(F_1 \sqcup F_2, G) = t(F_1, G)t(F_2, G)$ for every graph G.
5. This problem relates to Lemma 4. A set $S \subseteq V(F)$ is *independent* if no two vertices in S are adjacent. Let P be a partition of $V(F)$ into independent sets. The graph F/P has vertex set P and the edge XY if and only if there are $x \in X$ and $y \in Y$ with $xy \in E(F)$. That is, F/P is formed by gluing together the vertices in each independent set of P and removing multiedges.
   (a) Prove that
$$\mathsf{hom}(F, G) = \sum_P \mathsf{inj}(F/P, G),$$
   where the sum is over all partitions of $V(F)$ into independent sets.
   (b) Prove Lemma 4.

[1] Formally, the disjoint union of F_1 and F_2 is the graph $F_1 \cup F_2 = F_1F_2$ on vertex set $V(F_1) \sqcup V(F_2)$, where $ij \in E(F_1F_2)$ if and only if $ij \in E(F_1)$ or $ij \in E(F_2)$.

Graph on, graph off

Travis, Mathcamp 2024

1. Tile homomorphism densities

Yesterday, we said our definition of convergence seemed unsatisfactory because it doesn't say what a sequence converges *to*. The tiles were nicer in this way, so let's see if we can get back to those somehow.

Given that we're defining convergence in terms of homomorphism density, it would make sense to say that $(G_1, G_2, \dots)$ converges to a tile T if $t(F, G_n) \to t(F, T)$ for every graph F. But what is the homomorphism density into a tile?!

To answer that, let's take another look at the definition of homomorphism density for graphs:

$$\begin{aligned} t(\triangle, G) &= \frac{1}{n^3} \sum_{x,y,z \in V(G)} \mathbf{1}\big(xyz \text{ is a } \triangle \text{ in } G\big) \\ &= \frac{1}{n^3} \sum_{x,y,z \in V(G)} A_G(x,y) A_G(y,z) A_G(z,x), \end{aligned}$$

where $A_G(x,y)$ is the corresponding term in the adjacency matrix, meaning $A_G(x,y) = 1$ if xy is an edge and $A_G(x,y) = 0$ otherwise. The term

$$\frac{1}{n^3} \sum_{x,y,z \in V(G)}$$

looks sort of like the Reimann approximation of a function of three variables. And if we think of a tile T as a function $[0,1]^2 \to \{0,1\}$, then we can extend homomorphism density to tiles like this:

$$t(\triangle, T) = \int_0^1 \int_0^1 \int_0^1 T(x,y) T(y,z) T(z,x) \, dx \, dy \, dz.$$

And in fact, this gives an idea of how to extend homomorphism density to tiles for *any* graph F:

Definition 6. If $T \colon [0,1]^2 \to \mathbb{R}$ is symmetric[2] and F is a graph, the *homomorphism density* of F in T is

$$t(F, T) = \int_{x_1, \dots, x_{\mathsf{v}(F)} \in [0,1]} \left(\prod_{ij \in E(F)} T(x_i, x_j) \right) dx_1 \dots dx_{\mathsf{v}(F)}.$$

This formula looks sort of complicated, which is why problem 7 exists: to give you some practice working with this strange thing.

And now we can define convergence *to* something:

Definition 7. A graph sequence $(G_1, G_2, G_3, \dots)$ *converges to* a tile T if $t(F, G_n) \to t(F, T)$ for every graph F.

2. Injective homomorphisms

The kth *falling power* of n is $n^{\underline{k}} = n(n-1)(n-2) \cdots (n-k+1)$. The number of injective functions from $V(F)$ to $V(G)$ is $\mathsf{v}(G)^{\underline{\mathsf{v}(F)}}$.

Definition 8. The *injective homomorphsim density* of F in G is

$$t_{\text{inj}}(F, G) := \frac{\mathsf{inj}(F, G)}{\mathsf{v}(G)^{\underline{\mathsf{v}(F)}}},$$

which is the probability that a random injective function $V(F) \to V(G)$ is a homomorphism.

[2] This means that $T(x,y) = T(y,x)$.

If G is very, very big, then a random function $V(F) \to V(G)$ is probably injective, which suggests that $t(F,G)$ and $t_{\text{inj}}(F,G)$ should be close to each other.

Lemma 9. *If F has k vertices and G has n vertices, then*

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

Thus, for a given graph sequence $(G_1, G_2, \dots)$ where $\mathsf{v}(G_n) \to \infty$, we have $t(F,G_n) - t_{\text{inj}}(F,G_n) \to 0$.

3. Random graphs

Our next focus is to answer a question posed at the end of class yesterday: Does a sequence of random graphs converge?

Definition 10. Given $n \in \mathbb{N}$ and $0 \le p \le 1$, the *Erdős–Rényi random graph* $G(n,p)$ is defined like this: Start with n vertices, and for each pair of vertices, draw an edge between them with probability p.

For example, $G(4, 1/2)$ is the random graph you get by taking four vertices and flipping a coin 6 times, one for each pair of edges: If the coin comes up heads, draw an edge between the corresponding pair of vertices; otherwise, don't draw an edge.

Just to make things easier for now, let's write G_n as shorthand for $G(n, 1/2)$.

Question: Is $(G_1, G_2, G_3, \dots)$ a convergent graph sequence (where $G_n = G(n,1/2)$)?

But wait: G_n isn't a specific graph, it's a *random* one! So it's possible (though very unlikely) that $G_n = K_n$ for every n, in which case it certainly converges. But it's also possible (though again very unlikely) that $G_n = K_n$ if n is even and $G_n = \overline{K_n}$ if n is odd, which does *not* converge. So really, we should ask: What is the *probability* that $(G_1, G_2, G_3, \dots)$ converges?

Theorem 11. *The graph sequence $(G_1, G_2, G_3, \dots)$ where $G_n = G(n, 1/2)$ converges with probability 1.*

According to Definition 5, to prove that the sequence converges, we should consider the sequence of real numbers $t(F, G_n)$ for specific graphs F. Let's start by considering $t(\bullet\!\!-\!\!\bullet, G_n)$. On the one hand, we expect G_n to have approximately half its edges, so

$$t(\bullet\!\!-\!\!\bullet, G_n) \approx \frac{1}{2}.$$

The way to make this formal is to use linearity of expectation:

$$\mathbb{E}\big[t(\bullet\!\!-\!\!\bullet, G_n)\big] = \frac{\mathbb{E}\big[\hom(\bullet\!\!-\!\!\bullet, G_n)\big]}{n^2} = \frac{\frac{1}{2}n(n-1)}{n^2} = \frac{1}{2} - \frac{1}{2n}.$$

So $t(\bullet\!\!-\!\!\bullet, G_n)$ *averages* close to $\frac{1}{2}$, but it might *often* be far away: It's possible that one particular instance of the random graphs has $t(\bullet\!\!-\!\!\bullet, G_n) > 2/3$ for every n. The question we have to answer is: How (un)likely is that?

For this, we need a *concentration inequality*, which is a type of inequality that shows that a random variable is often close to a single value.

Definition 12. The *variance* of a random variable X with $\mathbb{E}[X] = \mu$ is

$$\operatorname{Var}(X) := \mathbb{E}[X^2] - \big(\mathbb{E}[X]\big)^2.$$

Lemma 13 (Chebyshev's inequality). *If X is a random variable with variance ν and expected value μ, then*

$$\mathbb{P}\big(|X - \mu| > \varepsilon\big) \le \frac{\nu}{\varepsilon^2}.$$

Now let's take $X_n = t(\overset{\circ}{\underset{\circ}{|}}, G_n)$. To use Chebyshev's inequality, we need to calculate the variance of $t(\overset{\circ}{\underset{\circ}{|}}, G_n)$. Using problem 4, we have

$$X_n^2 = t(\overset{\circ}{\underset{\circ}{|}}, G_n)^2 = t(\overset{\circ}{\underset{\circ}{|}}\overset{\circ}{\underset{\circ}{|}}, G_n).$$

This would be a bit tedious to calculate directly, but $t_{\mathrm{inj}}(\overset{\circ}{\underset{\circ}{|}}\overset{\circ}{\underset{\circ}{|}}, G_n)$ is easy:

$$\mathbb{E}\big[t_{\mathrm{inj}}(\overset{\circ}{\underset{\circ}{|}}\overset{\circ}{\underset{\circ}{|}}, G_n)\big] = \frac{1}{4}.$$

Therefore, if we use Lemma 9, we get

$$\begin{aligned}\mathrm{Var}(X_n) &\le \mathbb{E}\big[t_{\mathrm{inj}}(\overset{\circ}{\underset{\circ}{|}}\overset{\circ}{\underset{\circ}{|}}, G_n)\big] - \mathbb{E}\big[t(\overset{\circ}{\underset{\circ}{|}}, G_n)\big]^2 + \frac{1}{n}\binom{4}{2} \\ &= \frac{1}{4} - \left(\frac{1}{2} - \frac{1}{2n}\right)^2 + \frac{6}{n} \\ &< \frac{7}{n}.\end{aligned}$$

We can now apply Chebyshev's inequality to conclude that

$$\mathbb{P}\big(|t(\overset{\circ}{\underset{\circ}{|}}, G_n) - \mu_n| \ge \varepsilon\big) \le \frac{7}{n\varepsilon^2},$$

where $\mu_n = \frac{1}{2} - \frac{1}{2n}$ is the expected value of X_n. So for large n, the homomorphism density $t(\overset{\circ}{\underset{\circ}{|}}, G_n)$ will be close to $1/2$. But how likely is it that *all* of them are simultaneously close to $1/2$? For that, we need

Lemma 14 (Borel–Cantelli). *Let $A_1, A_2, A_3, \ldots$ be a sequence of events such that $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$. Then the probability that infinitely many A_n occur is 0.*

Let A_n be the event that $|t(\overset{\circ}{\underset{\circ}{|}}, G_n) - \mu_n| > \varepsilon_n$. Then $\mathbb{P}(A_n) \le 7(n\varepsilon_n^2)^{-1}$. So

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) \le \sum_{n=1}^{\infty} \frac{7}{n\varepsilon_n^2}.$$

But if $\varepsilon_n \to 0$, then this sum diverges, so we can't use the Borel–Cantelli lemma! So we change our goal:

Proposition 15. *The sequence $\big(t(\overset{\circ}{\underset{\circ}{|}}, G_{n^2})\big)$ converges to $1/2$ with probability 1.*

Proof. Let B_n be the event that $|t(\overset{\circ}{\underset{\circ}{|}}, G_n) - \mu_n| > n^{-1/3}$. Then

$$\sum_{n=1}^{\infty} \mathbb{P}(B_n) \le \sum_{n=1}^{\infty} \frac{7}{n^2 \cdot n^{-2/3}} = 7\sum_{n=1}^{\infty} \frac{1}{n^{4/3}} < \infty$$

by the integral test.[3] The Borel–Cantelli lemma implies that, with probability 1, only finitely many of the events B_n occur. And if this is the case, then

$$\lim_{n\to\infty} t(\overset{\circ}{\underset{\circ}{|}}, G_{n^2}) = \frac{1}{2}.$$

□

In fact, this whole proof works for *all* graphs F (see problem 9), meaning we can prove that

Proposition 16. *For every graph F, the sequence $t(F, G_{n^2})$ converges to $2^{-\mathsf{e}(F)}$ with probability 1.*

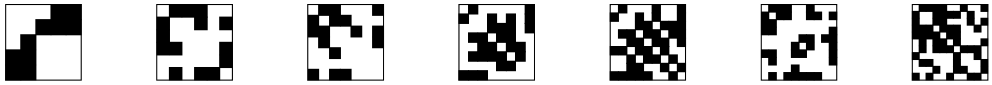
Theorem 11 is true, it's just that to prove it, we need to use more complicated tools from probability. The idea of the proof is exactly the same.

So with that, let's return to the old question:

> Question: What does $(G_1, G_2, \ldots)$ converge *to*?

The tile pictures for $G(n, 1/2)$ look like this:

[3] That is, since $\int_1^{\infty} n^{-4/3} < \infty$, the sum is finite, too.



which looks like it doesn't converge to *anything*. 😢

But maybe the limit just isn't what we've been thinking of as a tile. Let's take a step back: Is there any kind of "tile" T such that $t(F,T) = 2^{-\mathsf{e}(F)}$ for every graph F, where $t(F,T)$ is defined according to Definition 6?

Yes! The tile that has value 1/2 everywhere! That one works perfectly, and it turns out that no tile whose values are only 0 or 1 will work. So we have to extend the possible set of limit objects include real numbers, not just 0 and 1.

Definition 17. A *graphon* is a symmetric function $W\colon [0,1]^2 \to [0,1]$.

If we let W_p denote the graphon defined by $W(x,y) = p$ for every x, y, what we've proved today is that $G(n^2, 1/2) \to W_{1/2}$ with probability 1. (And, in fact, $G(n^2, p) \to W_p$ for every $p \in [0,1]$.)

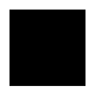
These graphons will be the true limit objects of graphs.[4]

PROBLEMS

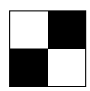
GRAPHON HOMOMORPHISMS

The first two problems are meant to give some practice dealing with homomorphism densities for graphons.

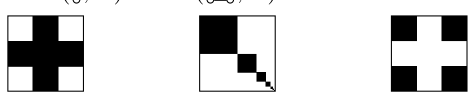
6. (a) Show that $(K_1, K_2, \dots)$ converges to the graphon



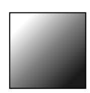
   (b) Show that the sequence $(K_{1,1}, K_{2,2}, K_{3,3}, \dots)$ of complete bipartite graphs converges to



7. (a) For each tile T below, calculate $t(\;, T)$ and $t(\;, T)$.

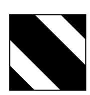


   (b) Calculate $t(\;, W)$ for this graphon:



   which has formula $W(x,y) = \frac{1}{2}(x+y)$. (Remember that for pictures of graphons, the $(0,0)$ point is in the upper left corner while the $(1,1)$ point is in the lower right.)

   (c) (Bonus problem: the only difference is the calculation of the integral is a little different.) Calculate $t(\;, T)$ for this tile:



MORE GRAPH HOMOMORPHISMS

Some more challenging problems about homomorphisms.

8. Prove Lemma 9. [Hint: Split $\hom(F,G)$ into the numbers of injective and non-injective homomorphisms.]
9. Prove Proposition 16 by (more or less) copying the proof for $F =$.

[4] And now you understand that the title of this class is a pun.

PROBABILITY

These problems go through proofs of the probability results that we used in class.

10. This problem is a proof of Chebyshev's inequality.
    (a) (Markov's inequality) Suppose that X is a random variable whose values are always nonnegative. Prove that $\mathbb{P}(X \geq a) \leq \frac{1}{a}\mathbb{E}[X]$. [Hint: Multiply both sides by a.]
    (b) Prove Chebyshev's inequality using Markov's inequality. [Hint: Prove that $\mathbb{E}\big[(X-\mu)^2\big] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.]
11. (Proof of Borel–Cantelli) Let p be the probability that infinitely many A_i occur. The following line is a sketch of a proof of the Borel–Cantelli lemma:

$$p \leq \mathbb{P}\left(\bigcup_{i=N}^{\infty} A_i\right) \leq \sum_{i=N}^{\infty} \mathbb{P}(A_i) \xrightarrow{N\to\infty} 0.$$

    Using this, write out a full proof.

Graph on, graph off

Travis, Mathcamp 2024

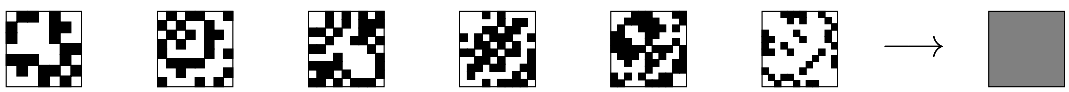
4. More problems and some solutions

Since we have defined $t(F, W)$ for any graphon W, we can also talk about convergence of graphons!

Definition 18. A sequence $W_1, W_2, W_3, \dots$ of graphons *converges to* a graphon W if $t(F, W_n) \to t(F, W)$ for every graph F.

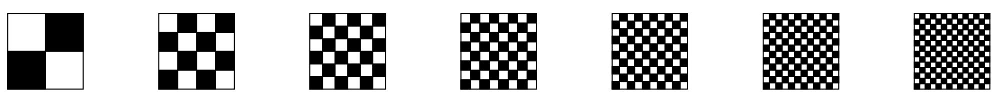
The tile of any graph G is a graphon W_G, and $t(F, G) = t(F, W_G)$, so convergence of graphs is a special case of convergence of graphons. But there are some strange things with this definition. First: We have a space of things (graphons), but in order to talk about whether a sequence of them converges, we have to travel *outside* that space to talk about homomorphism densities. Typically, we think of a convergent sequence as getting closer and closer to a specific object, but we don't have a way to talk about that here.

And if we *do* think about things that way, it's still strange. We proved that $G(n^2, 1/2)$ converges to $W_{1/2}$ as $n \to \infty$, which looks like this:



which . . . isn't super believable. One justification for it is that any "average" over a portion of the tile will be 1/2, so the "smooth" limit should be 1/2.

But! Imagine you take the complete bipartite graphs again, but relabel the vertices. Then the tiles look like this:



to which the same argument applies. But you can check that $K_{n,n}$ does *not* converge to $W_{1/2}$ (this is problem 1).

So we have some problems to resolve here: What does it mean for two graphons to be close? And what's up with relabelling?

In the rest of the notes, we'll need to talk about functions with any real value:

Definition 19. A *kernel* is a symmetric function $W\colon [0,1]^2 \to \mathbb{R}$.

Here is a formal definition of relabelling for graphons:

Definition 20. Let $\varphi\colon [0,1] \to [0,1]$ be a measure-preserving map and $W\colon [0,1]^2 \to \mathbb{R}$ a kernel. The kernel W^φ is defined by

$$W^\varphi(x, y) := W\big(\varphi(x), \varphi(y)\big).$$

Two graphons U, W are *weakly isomorphic* if there are measure-preserving maps φ, ψ so that $U^\varphi = W^\psi$ (up to a set of measure 0).

If the terms "measure preserving" and "set of measure 0" don't mean anything to you, that's okay—these are terms from analysis which make everything work, but they're somewhat technical, and you don't need to know the precise definitions to see how everything works.

The important part is that because φ is measure-preserving,

$$t(F, W) = t(F, W^\varphi)$$

for every graph F. (Just as you would expect in a relabelling of the vertices of a graph.) The "proper" or precise way to think of a graphon is as the equivalence class of weakly isomorphic graphons.

This is all very nice, but it doesn't solve the problem of what it means for two graphons to be close. That's what we turn to next.

5. Cut distance

Definition 21. The *cut norm* of a kernel $W\colon [0,1]^2 \to \mathbb{R}$ is

$$\|W\|_\square = \max_{S,T\subseteq[0,1]} \left| \iint_{S\times T} W(x,y)\,dx\,dy \right|.$$

To define an actual distance on the set of graphons, we should also consider relabellings:

Definition 22. The *cut distance* between two kernels U and W is

$$\delta_\square(U,W) = \min_{\varphi,\psi} \|U^\varphi - W^\psi\|_\square$$

The cut distance is a true distance: If U, V, W are graphons, then

$$\delta_\square(U,W) \le \delta_\square(U,V) + \delta_\square(V,W).$$

You can prove this in problem 2. Also, it turns out that $\delta_\square(U,W) = 0$ if and only if U and W are weakly isomorphic.

The most important part of the cut distance is that it actually captures the idea of convergence:

Theorem 23. *A sequence of graphons $(W_1, W_2, W_3, \dots)$ converges to a graphon W if and only if $\delta_\square(W_n, W) \to 0$.*

Rejoice! We now finally have a way of thinking about graph limits that's very similar to limits of real numbers. We just need to prove it.

6. The counting lemma

Theorem 24 (Counting lemma). *Let F be a finite simple graph. For any two graphons U and W,*

$$|t(F,U) - t(F,W)| \le \mathsf{e}(F)\delta_\square(U,W).$$

This proves one direction of Theorem 23: If $\delta_\square(W_n, W) \to 0$, then $t(F, W_n) \to t(F,W)$.

To prove the Counting lemma, we need another lemma, which is another way to define the cut norm.

Lemma 25. *For any kernel W,*

$$\|W\|_\square = \max_{f,g\colon [0,1]\to[0,1]} \left| \iint_{[0,1]^2} f(x)W(x,y)g(y) \right|.$$

You can find a proof of this equality at the end of these notes.

Proof of Theorem 24. We will prove that $|t(F,U) - t(F,W)| \le \mathsf{e}(F)\|U - W\|_\square$. Since $t(F,W) = t(F,W^\varphi)$ for any measure-preserving bijection $\varphi\colon [0,1] \to [0,1]$, replacing W by W^φ and taking the infimum over φ yields the stronger inequality in the theorem statement.

Let's focus on the case $F = \triangle$.

The left-hand side is

$$|t(\triangle, U) - t(\triangle, W)| = \left| \iiint \Big(U(x,y)U(y,z)U(z,x) - W(x,y)W(y,z)W(z,x) \Big)\,dx\,dy\,dz \right|.$$

To work with this expression, we pull a trick from analysis: adding "ghost terms". In proving that $\lim a_n b_n = ab$ for convergent real sequences $a_n \to a$ and $b_n \to b$, we need to bound the inequality $|a_n b_n - ab|$, but all we know is that $|a_n - a| \to 0$ and $|b_n - b| \to 0$. The solution is to add $a_n b - a_n b$ and use the triangle inequality:

$$|a_n b_n - ab| \le |a_n||b_n - b| + |b||a_n - a| \longrightarrow 0.$$

The same trick works here; we just need to introduce more ghost terms:

$$\Big| \iiint \Big(U(x,y)U(y,z)U(z,x) - U(x,y)U(y,z)W(z,x) \Big) + \Big(U(x,y)U(y,z)W(z,x) \\ - U(x,y)W(y,z)W(z,x) \Big) + \Big(U(x,y)W(y,z)W(z,x) - W(x,y)W(y,z)W(z,x) \Big) \Big|.$$

Using the triangle inequality, we get an upper bound of

$$\Big| \iiint U(x,y)U(y,z)U(z,x) - U(x,y)U(y,z)W(z,x) \Big| \\ + \Big| \iiint U(x,y)U(y,z)W(z,x) - U(x,y)W(y,z)W(z,x) \Big| \\ + \Big| \iiint U(x,y)W(y,z)W(z,x) - W(x,y)W(y,z)W(z,x) \Big|.$$

To finish the proof, we show that each term of this sum is at most $\|U - W\|_\square$.

Take the first term as an example. If we define $f_y(x) = U(x,y)$ and $g_y(z) = U(y,z)$, the first term becomes

$$\Big| \iiint U(x,y)U(y,z)U(z,x) - U(x,y)U(y,z)W(z,x)\,dx\,dy\,dz \Big| \\ \le \int_0^1 \Big| \int_0^1 \int_0^1 f_y(x) \Big(U(z,x) - W(z,x) \Big) g_y(z)\,dx\,dz \Big| dy.$$

By Lemma 25, the inner integral is at most $\|U - W\|_\square$. Each term can be bounded the same way, so we finally obtain

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

The proof for other graphs follows exactly the same pattern. The only difference is that the notation becomes more complicated because you need to keep track of a more complicated graph. $\square$

proof of Lemma 25

Given a set $S \subseteq [0,1]$, define the characteristic function $\chi_S \colon [0,1] \to [0,1]$ by

$$\chi_S(x) = \begin{cases} 1 & \text{if } x \in S \\ 0 & \text{if } x \notin S. \end{cases}$$

Suppose that S and T maximize $\big| \int_{S\times T} W \big|$; then

$$\|W\|_\square = \Big| \iint_{S\times T} W \Big| = \Big| \iint \chi_S(x) W(x,y) \chi_T(y)\,dx\,dy \Big| \le \max_{f,g \colon [0,1]\to[0,1]} \Big| \iint f(x)W(x,y)g(y) \Big|.$$

To prove the other inequality, we will show that we can choose the maximal choices of f and g to be $\{0,1\}$-valued. Let $H(f,g) = \Big| \iint f(x)W(x,y)g(y)\,dx\,dy \Big|$. Choose functions f and g so that $H(f,g)$ is maximal and define

$$f_1(x) = \begin{cases} 1 & \text{if } f(x) \ge \frac{1}{2} \\ 0 & \text{if } f(x) < \frac{1}{2} \end{cases} \qquad \text{and} \qquad f_2(x) = \begin{cases} 2f(x) - 1 & \text{if } f(x) \ge \frac{1}{2} \\ 2f(x) & \text{if } f(x) < \frac{1}{2}. \end{cases}$$

Then $f = \frac{1}{2}f_1 + \frac{1}{2}f_2$. Using the triangle inequality and the maximality of $H(f,g)$, we have

$$H(f,g) \le H\left(\frac{1}{2}f_1 + \frac{1}{2}f_2, g\right) \le \frac{1}{2}H(f_1,g) + \frac{1}{2}H(f_2,g) \le H(f,g).$$

Equality must hold in each step, so in particular $H(f_1,g) = H(f,g)$ is maximal and f_1 is $\{0,1\}$-valued. Repeat for g. Then $H(f,g) = H(f_1,g_1)$, and both f_1 and g_1 are characteristic functions of sets. This shows that $H(f,g) \le \|W\|_\square$. $\square$

This last step of using the triangle inequality might seem strange and magical, but it's a common trick in the field of convex optimization; I'm happy to talk about it if you'd like to know more!

PROBLEMS

1. Prove that $K_{1,1}, K_{2,2}, K_{3,3}, \dots$ does not converge to $W_{1/2}$.
2. (a) Prove that for any two kernels U, W, we have
$$\|U + W\|_\square \leq \|U\|_\square + \|W\|_\square.$$
   (b) Prove the triangle inequality for $\delta_\square$.
3. (a) The L_1-norm of a kernel W is
$$\|W\|_1 = \int_0^1 \int_0^1 \big|W(x,y)\big| \, dx \, dy.$$
   Why is $\|W\|_\square \leq \|W\|_1$?
   (b) Let T_n denote the *threshold graph* on the vertex set $[n]$, with $ij \in E(T_n)$ if and only if $i + j \leq n$. Define the *threshold graphon* by
$$W(x,y) = \begin{cases} 1 & \text{if } x + y \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$
   Show that $\delta_\square(T_n, W) \to 0$ as $n \to \infty$.
   (c) Let H_n be the *half-graph* with vertex set $V(H_n) = \{a_i : 1 \leq i \leq n\} \cup \{b_i : 1 \leq i \leq n\}$ and an edge from a_i to b_j if $i \leq j$. Find a graphon W so that $H_1, H_2, H_3, \dots \to W$, and prove it by showing that $\delta_\square(H_n, W) \to 0$.
4. Define a modified cut norm by
$$\|W\|_\blacksquare = \max_{S \subseteq [0,1]} \left| \iint\limits_{S \times S} W(x,y) \, dx \, dy \right|.$$
   Prove that $\|W\|_\blacksquare \leq \|W\|_\square \leq 2\|W\|_\blacksquare$.
5. For a graphon W, define
$$\delta(W) = \min_{x \in [0,1]} \int_0^1 W(x,y) \, dy \qquad \text{and} \qquad \Delta(W) = \max_{x \in [0,1]} \int_0^1 W(x,y) \, dy.$$
   Prove that for any tree T on k vertices, $\delta(W)^{k-1} \leq t(T, W) \leq \Delta(W)^{k-1}$.

Definition 26. A set X and a distance d form a *metric space* if d is a sensible notion of distance:
- $d(x,y)=0$ if and only if $x=y$,
- $d(x,y)=d(y,x)$ for all $x,y\in X$, and
- $d(x,y)\le d(x,z)+d(z,x)$ (triangle inequality).

The set of graphons together with the cut distance $\delta_\square$ forms a metric space.

Definition 27. A metric space (X,d) is *compact* if every sequence in X has a convergent subsequence.

Theorem 28 (The Big Theorem). *The space of graphons is compact under the $\delta_\square$ distance.*

Theorem 28 is equivalent to a Very Important Theorem in structural graph theory called Szemerédi's Regularity Lemma, first proved in 1978. This theorem is a superstar in the graph theory community and has applications throughout graph theory. The original proof was a monster: The original paper is widely regarded as nearly impossible to read, and in order to even understand the structure of the proof, Szemerédi had to include a diagram of the logical connections (Figure 1). Even stating Szemerédi's Regularity Lemma would take 20 minutes or so, but in terms of graphons, it's simple to state. And we could prove it if we had two more days, but we don't have two more days.

So instead, we'll focus on applications. Using a combination of Theorem 11 and a bit of analysis, we can prove that

Lemma 29. *For every graphon W, there is a sequence of graphs $(G_1,G_2,G_3,\dots)$ that converges to W.*

and this lemma plus Theorem 28 tells us:

Theorem 30.
- *Every convergent sequence of graphs $(G_1,G_2,G_3,\dots)$ converges to a graphon.*
- *For every $\varepsilon>0$, there is a finite set $\mathcal{G}$ of graphs such that for every graphon W there is a graph $G\in\mathcal{G}$ such that $\delta_\square(W,G)<\varepsilon$.*

In fact, we can squeeze a bit more from this theorem:

Definition 31. Let $\mathcal{P}=\{V_1,\dots,V_k\}$ be a partition of $[0,1]$ into k sets. The *stepping* of a kernel W with respect to $\mathcal{P}$ is kernel $W_\mathcal{P}$ defined by

$$W_\mathcal{P}(x,y)=\frac{1}{m(V_i)m(V_j)}\int\limits_{V_i\times V_j} W(x,y)\,dxdy \qquad \text{when } x\in V_i\times V_j.$$

In other words, the partition $\mathcal{P}$ divides $[0,1]^2$ into k^2 regions, and to get $W_\mathcal{P}$, just average W over each region.

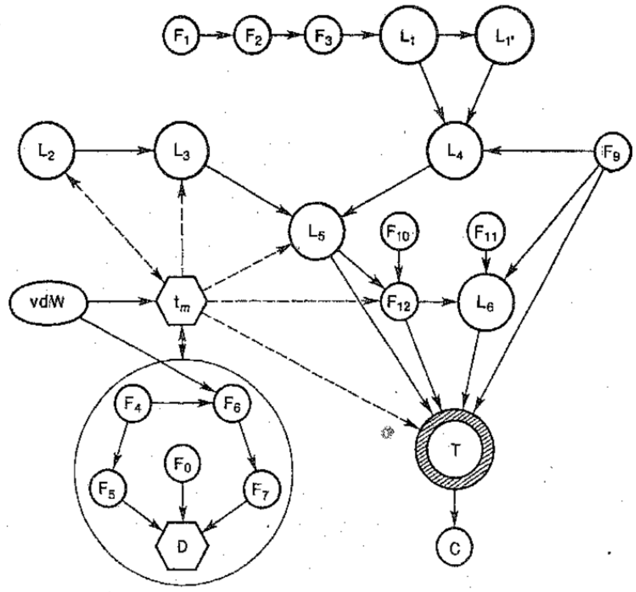
Proposition 32. *For every $\varepsilon>0$, there is a k such that: For every kernel W, there is a partition $\mathcal{P}$ into k sets, each with measure $1/k$, such that*

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

You can think of $W_\mathcal{P}$ as the adjacency matrix of a weighted graph with k vertices.

7. Applications of the Big Theorem

Theorem 33 (Triangle removal lemma). *For every $\varepsilon>0$, there is a $\delta>0$ so that any graph with at most δn^3 triangles can be made triangle-free by removing at most εn^2 edges.*



The diagram represents an approximate flow chart for the accompanying proof of Szemerédi's theorem. The various symbols have the following meanings: F_k = Fact k, L_k = Lemma k, T = Theorem, C = Corollary, D = Definitions of B, S, P, α, β, etc., t_m = Definition of t_m, vdW = van der Waerden's theorem, F_0 = "If $f: \mathbf{R}^+ \to \mathbf{R}^+$ is subadditive then $\lim_{n\to\infty} \frac{f(n)}{n}$ exists".

Figure 1: Diagram of the logical structure of Szemerédi's original proof

The important point of this theorem is that δ and ε are *independent* of n. While you might think a theorem like this could be proven by, for example, linearity of expectation, that's not true: The original proof gives a bound on δ of

$$\frac{1}{\delta} = 2^{2^{\cdot^{\cdot^{2}}}},$$

where the number of 2s is approximately $1/\varepsilon$. The best bound we have right now is a tower of height at least $\log(1/\varepsilon)$, which is still verrrrrry large. For a proof of the Triangle Removal Lemma, see Lemma 11.64 in the book *Large Networks and Graph Limits* by László Lovász.

Nevertheless, the triangle removal lemma can be used to prove many results, not just in graph theory, but also in number theory.

Definition 34. A sequence $a, a+b, a+2b, \ldots, a+(k-1)b$ with $b > 0$ is called a *k-term arithmetic progression*, or k-AP.

How big can a subset of $\{1, 2, \ldots, n\}$ be if it does not contain a k-AP? This is a question in arithmetic Ramsey theory, which is like Ramsey theory for graphs (if you've heard of such things), but for the integers. The principle of all of Ramsey theory is that "big sets must contain ordered structure". In this case, the structure is a k-AP.

The answer to the previous question is that there is any set with a "positive density" must contain a 3-AP:

Theorem 35 (Roth). *Let $f(n)$ be the size of the largest set $A \subseteq \{1, 2, \ldots, n\}$ that contains no 3-term arithmetic progression. Then*

$$\frac{f(n)}{n} \to 0$$

as $n \to \infty$.

Proof sketch. Suppose that A contains no 3-AP. The key point of the proof is to define a graph G that transforms 3-APs in A to triangles in G. The vertex set of G is $X \cup Y \cup Z$, where
- $X = \{x_1, \ldots, x_{2n+1}\}$,
- $Y = \{y_1, \ldots, y_{2n+1}\}$, and
- $Z = \{z_1, \ldots, z_{2n+1}\}$.

The edges of G are formed as follows:
- $x_i y_j \in E(G)$ if $j - i \in A \mod 2n+1$,
- $y_j z_k \in E(G)$ if $k - j \in A \mod 2n+1$, and
- $x_i z_k \in E(G)$ if $(k-i)/2 \in A \mod 2n+1$.

Note that $x_i y_j z_k$ is a triangle in G if and only if $j-i, (k-i)/2, k-j$ is an arithmetic progression in A (with $a = j - i$ and $b = (k+i)/2 - j$), or if all three terms are equal. Since A has no 3-AP, every triangle in G corresponds to the latter category, where all terms are equal. Therefore, every edge in G is contained in exactly one triangle. (This is the reason we use modular arithmetic.)

There are exactly $3(2n+1)|A| = 3(2n+1)f(n)$ edges in G. The triangle removal lemma implies that

$$\frac{3(2n+1)f(n)}{n^2} \to 0$$

as $n \to \infty$ (see problem 4). Cancelling the factors of n proves the theorem. □

Take a step back for a moment to see what we've outlined: We proved a theorem in **number theory** using a result from **graph theory** that we proved using **analysis**. This mix of areas of mathematics is one of the reasons I find graph limits so amazing.

Let $f_k(n)$ be the size of the largest set $A \subseteq \{1, 2, \ldots, n\}$ that contains no k-term arithmetic progression. Forbidding a 4-AP is a weaker condition than forbidding a 3-AP, so $f_4(n) \geq f_3(n)$. Nevertheless, the same result is true for all k:

Theorem 36 (Szemerédi's theorem on arithmetic progressions). *For every $k \geq 3$,*

$$\frac{f_k(n)}{n} \to 0$$

as $n \to \infty$.

The proof also uses the regularity lemma, but in a much more complicated way.

PROBLEMS

1. Is it true that $t(F, W) = t(F, W_{\mathcal{P}})$?
2. Let $\mathcal{P} = \{V_1, \ldots, V_k\}$ be a partition of $[0,1]$ into k sets. Show that for any graphon W that is constant on the sets $V_i \times V_j$,

$$\|W\|_\square = \max\left\{\left|\int_{S\times T} W(x,y)\,dxdy\right| : S, T \text{ are unions of elements of } \mathcal{P}\right\}.$$

   Use this to show that $\|W_{\mathcal{P}}\|_\square \leq \|W\|_\square$.
3. (a) The inner product of two graphons U, W is defined by

$$\langle U, W\rangle = \int_0^1 \int_0^1 U(x,y)W(x,y)\,dx\,dy.$$

   Prove that $\langle U, W_{\mathcal{P}}\rangle = \langle U_{\mathcal{P}}, W_{\mathcal{P}}\rangle$ for any two graphons U, W.

   (b) The L^2-norm of a graphon W is

$$\|W\|_2 = \int_0^1 \int_0^1 W(x,y)^2\,dx\,dy$$

   Use this to prove that $\|W_{\mathcal{P}}\|_2 \leq \|W\|_2$ for every graphon W. [Hint: What is $\|W - W_{\mathcal{P}}\|_2^2$?]

4. (Challenge) Let $g(n)$ be the maximum number of edges a graph with n vertices can have if every edge is contained in exactly one triangle. Use the triangle removal lemma to prove that

$$\frac{g(n)}{n^2} \to 0$$

as $n \to \infty$.