

**YUFEI ZHAO:** So we've been discussing graph limits for a couple of lectures now. In the first lecture on graph limits, so two lectures ago, I stated a number of main theorems. And today, we will prove these theorems using some of the tools that we developed last time, namely the regularity lemma. And also we proved this Martingale convergence theorem, which will also come in.

So let me recall what were the three main theorems that we stated at the end of two lectures ago. So one of them was the equivalence of convergence. On one hand, we defined a notion of convergence where we say that  $W_n$  approaches  $W$ , by definition, if the  $F$  densities converge. We can say convergence even without a limit in mind, where we see a sequence converges if all of these  $F$  densities converge. So the first main theorem was that the two notions of convergence are equivalent, so one notion being convergence in terms of  $F$  densities, and the second notion being convergence in the sense of the cut norm, the cut distance.

There was a second term that tells us that limits always exist. If you have a convergent sequence, then you can represent a limit by a graphon. And the third statement was about compactness of the space of graphons.

So we're actually going to prove these theorems in reverse order. We're going to start with the compactness and work backwards. So this is not how these theorems were originally proved, or not in that order, but it will be helpful for us to do this by first considering the compactness statement.

So remember we're compactness says. I start with this space  $\tilde{W}$ , which is the space of graphons, where I identify graphons that have distance 0. So if they have cut distance 0, then I refer to them as the same point. So this is now a metric space. And the theorem is that this space is compact.

So I think this, it's a really nice theorem. It's a beautiful theorem that encapsulates a lot of what we've been talking about so far with similarities, regularity, and what not, in a qualitatively succinct way, just that this space of graphons is compact. You may not have some intuition about what the space looks like at the moment, but we'll see the proof and hopefully that will give you some more intuition

I first learned about this theorem when Laszlo Lovasz, who was one of the pioneers in the subject, when he came to MIT to give a talk when I was a graduate student. And he said that analysts thought that they pretty much knew all the naturally occurring compact spaces out there. So there are lots of spaces that occur in analysis and topology that are compact.

I mean, the first one you learn in analysis undergraduate is probably that an interval is compact. But there are also many other spaces. But this one here doesn't seem to be any of these classical notions of compactness. So it's, in some sense, a new compact space.

So let's see how the proof goes. Now, because we are working in a metric space, it suffices to show, due to the equivalence between compactness in the sense of finite open covers and sequential compactness in the metric space, so it suffices to show sequential compactness, that every sequence of graphons has a convergent subsequence with respect to this cut metric and also will produce a limit as a convergence subsequence with a limit point.

So that's what we'll do. I give you an arbitrary sequence of graphons. I want to construct by taking subsequences a convergent sequence. And I will tell you what that limit is.

So here is what we're going to do, given this sequence. As I hinted before, it has to do with the regularity lemma. So we're going to apply the regularity lemma in the above form, which we did last time. So apply the weak regularity lemma, which will tell us that for each  $W_n$ , there exists a partition, in fact, a sequence of partitions, each one refining the next.

So what's going to happen is, I'm going to start with  $W_n$  and starting with a trivial partition, apply that lemma, and obtain a partition  $P_{n,1}$ . And then starting with that as my  $P_0$ , I'm going to apply regularity lemma again and obtain a refinement. I will have this sequence of partitions, each one refining the next.

So all of these are going to be partitions of the  $[0, 1]$  interval. And as I mentioned last time, everything's going to be measurable. I'm not going to even mention measurability. Everything will be measurable such that they satisfy the following conditions.

So the first one is what I mentioned earlier, is that you have a sequence of refinements. So each  $P_{n,k+1}$  refines the previous one for all  $n$  and  $k$ .

And the second condition, as given by the regularity lemma, you get to control the number of parts. So I will say in the third part what the error of approximation is. But you get to control the

number of parts. So in particular, I can make sure that this number here, the number of parts in the  $k$ 'th partition depends only on  $k$ .

Now, you might complain somewhat, because the regularity lemma only tells you an upper bound on the number of parts. But that's OK. I can allow empty parts. So now I make sure that the  $k$ 'th partition has exactly  $nk$  parts.

And the third one has to do with the area of approximation. OK, so suppose we write  $W_{nk}$  as the graphon obtained by applying the stepping operator. So this is the averaging operator on this partition, corresponding to the  $k$ 'th partition. I apply that partition, do a stepping averaging operator on the  $n$ 'th graphon. I get  $W_{nk}$ .

The third condition is that the  $k$ 'th partition approximates-- it's a good approximation in the cut norm up to error  $1/k$ . So  $1/k$  is some arbitrary sequence going to 0 as  $k$  goes to infinity.

So I obtained a sequence of partitions, so by applying the regularity lemma to each to each graphon in the sequence. Now, these graphons, I mean, they each have their own vertex set. And so far, they're not related to each other.

But to make the visualization easier and also in order to do the next step in the proof, I am going to do some measure-preserving bisection. So think of this as permuting the vertex labels. So by replacing each  $W_n$  by some  $W_n \circ \phi$ , where  $\phi$  is a measure-preserving bisection, we can assume that all these partitions are partitions into intervals.

So initially, you might have a partition into arbitrary measurable sets. Well, what I can do is to push over the first set to the left, and so on, so do a measure-preserving bisection in a way so that I can maintain that all the partitions are visually chopping up into intervals.

Yeah?

**AUDIENCE:** So at some point, we need just one measure for the projection, like all of them be in  $k$ ?

**YUFEI ZHAO:** OK, so the question is, it may be the case that, for a given  $k$ , I can do this arrangement, but it's not clear to you at the moment why you can do this uniformly for all  $k$ . So one way to get around this is, for now, just think of for each given  $k$ . And then you'll see at the end that that's already enough. OK, any more questions?

So now assume all of these  $P_{nk}$ 's are intervals. So in fact, what you said may be a better way to go. But to make our life a little bit easier, let's just assume for now that you can do this.

OK, and what's going to happen next is some kind of a diagonalization argument. We're going to be picking subsequences. So I'm going to be picking subsequences so that they are going to have very nice convergence properties.

And so I'm going to repeatedly throw out a lot of the sequence. So this is a diagonalization argument. And basically what happens is that, by passing two subsequences-- and we're going to do this repeatedly, many times-- we can assume, first, that the end points of  $P_{n1}$ , they converge as  $n$  goes to infinity. So each  $P_{n1}$  is some partition of interval into some fixed number of parts. So by passing to a subsequence, I make sure that the division points all converge.

And now, by passing one more time, so by passing to subsequence one more time, let's assume that also,  $W_{n1}$  converges to some function, some graphon  $u_1$ , point-wise. So initially, I have these graphons. Each one of them is an  $m$  by  $n$  block. They have various division points.

By passing to a subsequence, I assume that the points of division, they converge. And now by passing to an additional subsequence, I can make sure the individual values, they converge. So as a result,  $W_{n,1}$  converges to  $W_1$ -- converges to some graphon,  $u_1$ , point-wise, almost everywhere.

And we repeat for  $W_{nk}$  for each  $k$ . So do this sequentially. So we just did it for  $k$  equals to 1. Now do it for 2, 3, 4, and so on. So this is a diagonalization argument. We do this countably many times.

At the end, what do we get? We pass down to the following subsequence. And just to make my life a bit more convenient, instead of labeling the indices of the subsequence, I'm going to relabel the sequence so that it's still labeled by 1, 2, 3, 4, and so on.

So we now pass to a sequence  $W_1, W_2, W_3$ , and so on, such that if you look at the first partition, the first weak regularity partition, they produce  $W_{1,1}, W_{2,1}, W_{3,1}$ , and so on. And these guys, they converge to  $u_1$ , point-wise. The second level,  $W_{2,1}$ -- sorry,  $W_{1,2}, W_{2,2}, W_{3,2}$ , they converge to  $u_2$ , point-wise, and so on.

OK, so far so good? Question?

**AUDIENCE:** Sorry, earlier, why did that converge to  $u_1$  point-wise?

**YUFEI ZHAO:** OK, so the question is, why is this true? Why is  $W_{n,1}$  converge to  $u_1$  point-wise? Initially, it might not. But what I'm saying is, you can pass to a subsequence.

**AUDIENCE:** Yes.

**YUFEI ZHAO:** You can pass to subsequence, because there are only  $n_1$  parts. So it's an  $n_1$  by  $m_1$  matrix of real numbers. And so you only have finite bounded many of them. So you can pick a subsequence so that they converge. Yeah?

**AUDIENCE:** So how do you make sure that your subsequence is not empty at the end-- like, could you fix the first  $k$ ?

**YUFEI ZHAO:** OK, so you're asking, if we do this slightly not so carefully, we might end up with an empty sequence. So this is why I say, you have to do a diagonalization argument. Each step, you keep the first term, the sequence, so that you always maintain some sequence. You have to be slightly careful with diagonalization. Any more questions?

So by passing to a subsequence, we obtain this very nice sequence, this nice subsequence, such that each row corresponding to each level of regularization converges point-wise to some  $u$ . So what do this  $u$ 's look like? So they are step graphons.

So let's explore the structure of  $u$  a bit more. OK, so since we have that each-- OK, so we have the property that each partition refines the previous partition. And as a result, if you look at the  $k$  plus 1'th stepping, and I step it by the previous partition in the sequence, I should get back, I should go back one in the sequence.

So this was this graphon obtained by averaging over the  $k$ 'th partition. And this is the graphon obtained by averaging over the  $k$ 'th plus 1st partition. So if I go back one more, I should go back in the sequence.

And since the  $u$ 's are the point-wise limit of these  $W$ 's, the same relationships should also hold for the  $u$ 's, namely that  $u_{sub k}$  should equal to  $u_{sub k plus 1}$  if I step it with  $P_k$ , where  $P_k$  is the-- so if you look at, all these endpoints converge. And these partitions, they converge to  $P_1$ . So if you look at the partitions that correspond to  $P_{1,1}$ ,  $P_{2,1}$ , and so on, I want these partitions to converge to  $P_1$ . I want these partitions to converge to  $P_2$ .

So all these partitions, they are partitions into intervals. So I'm just saying, if you look at where the intervals, where the divisions of intervals go, they converge. And then I'm calling the limit partition  $P_{k+1}$ . And here we're using that  $P_{k+1}$  refines  $P_k$ , because the same is true for each end. So in the limit, the same must be true as well.

So you have this column of  $u$ 's. So let me draw you a picture of what these  $u$ 's could look like. So here is an illustration that may be helpful. So what could these  $u$ 's look like? Each one of them is represented by values on the unit square. And I write this in matrix notation so that inversion is in the top left corner.

Well, maybe  $P_1$  is just the trivial partition, in which case  $u_1$  is going to be a constant graphon. Let's say it has value 0.5.  $u_2$  came from  $u_1$  by some partitioning. And suppose just for the sake of illustration, there was only a partitioning into two parts.

And OK, so it doesn't have to be at the origin. It doesn't have to be at midpoint. But just for illustration, suppose the division were at the midpoint. Because  $u_1$  needs to have-- so this 0.5 value should be the average value in all of these four squares.

So for instance, the points may be like 0.6, 0.6, 0.4, 0.4, so for example. And in  $u_3$ , the partition, the  $P_3$  partition-- so here are the partition is  $P_1$ . The partition is  $P_2$ . There's two parts. And suppose  $P_3$  three now has four parts. And again, for illustration's sake, suppose it is equally dividing the interval into four intervals.

It could be that now each of these parts is split up into four different values in a way that so you can obtain the original numbers by averaging. So that's one possible example. Likewise, you can have something like that. Sorry, 4, 7-- so and I should maintain symmetry in this matrix.

And the last one, I'm just going to be lazy and say that it's still 0.4 throughout. OK, so this is what the sequence of  $u$ 's are going to look like. Each one of them splits up a box in the previous  $u$  in such way that the local averages, the step averages are preserved. Any questions so far?

All right, so now we get to Martingales. So I claim that this is basically a Martingale. So and suppose you let  $x, y$  be a uniform point in the unit square. And consider this sequence. So this is now a random sequence, because  $x, y$  are random. I evaluate these  $u$ 's on this uniform random point,  $x, y$ . So this is a random sequence.

And the main observation is that this is a Martingale. So remember the definition of a Martingale from last time. Martingale is one where, if you look at the value of  $u_{sub k}$  conditioned on the previous values, the expectation is just the previous term. And I claim this is true for the sequence, because of the way we constructed it, it's splitting up each box in an averaging-preserving way.

A different way to see this, and for those of you who actually know what the definition of a random variable is in the sense of probability theory, is that you should view this  $[0, 1]$  squared as the probability space, in which case  $u$  itself is the random variable. And this partitioning gives you a filtration of the space. It's a sequence of sigma algebras dividing up the space into finer and finer pieces.

So this is really what a martingale is. So we have a Martingale. It's bounded because the values take place in  $[0, 1]$ . So by the Martingale convergence theorem, which we proved last time, we find that this sequence must converge to some limit. So this sequence of Martingale converges, which means, so if you think about the interpretation up there, so there exists a  $u$  which is a graphon such that  $u_k$  converges to  $u$  point-wise almost everywhere as  $k$  goes to infinity.

That's the limit. So this is the limit. And we're going to show that it is indeed a limit. But you see, this is a construction of the limit, where we took regularity, got all these nice pieces, found convergent subsequences, and then applied the martingale convergence theorem to produce for us this candidate for the limit, this  $u$ .

So now let us show that it is indeed the limit that we're looking for in the subsequence. So again, I've tossed out all the terms which we removed in passing to subsequences. So in the remaining subsequence, I want to show that the  $W_n$ 's indeed converge to  $u$ .

And this is now a fairly straightforward three epsilons argument, the standard analysis type argument. But OK, so let's carry it through. So for every epsilon bigger than 0, suppose you pick a sufficiently large  $k$ . There exists a sufficiently large  $k$ . And we make sure  $k$  is large enough such that  $u$  differs from  $u_{sub k}$  in  $L^1$  norm by, at most,  $\epsilon/3$ , because the  $u_k$ 's, they converge to  $u$  point-wise almost everywhere.

So we find this  $k$ . So let's fix this  $k$ . Then there exists an  $n_0$  such that if you look at this  $u_{sub k}$ , it does not-- it is very close to  $W_{sub nk}$  for all  $n$  large enough because of what happened up

there.

So we can now compute the difference between-- in fact, let's do it this way. So now let's compute the difference, the cut norm of the difference between the term in the sequence  $W_{\text{sub } n}$  and  $u$ . So by triangle inequality, we have that the following is true.

The cut norm is upperbounded by the  $l_1$  norm. Look at the definitions. So I'm going to replace the first couple of these cut norms by  $l_1$  norms and leave the last one in tact.

The first term, I claim is, at most,  $\epsilon/3$ , because up there. The second term is going to be at least  $\epsilon/3$ , because over here. And the third term is going to be also, at most,  $\epsilon/3$ , because well, from the regularity approximation, I know that it is, at most,  $1/k$ .

And I chose  $k$  large enough so that there is also, at most,  $\epsilon/3$ . Put everything together, we find that these two are-- they differ by, at most,  $\epsilon$  if  $n$  is large enough. But now, since  $\epsilon$  can be arbitrarily small, we find that you indeed have convergence, as claimed.

And this finishes the proof of compactness. So there are a few components. One is passing to-- so applying regularity, passing to subsequences, and obtaining this limit from the regularity approximations, these  $u$ 's. And then we observe that these  $u$ 's, they form a Martingale.

So we can apply the Martingale convergence theorem to get us a candidate for the limit. And then the rest is fairly straightforward, because all the steps are good approximations. You put them together, you prove the limit. Any questions?

All right, so you may ask, well, now we have compactness. What is compactness good for? So it may seem like a somewhat abstract concept.

So in the second half of today's lecture, I want to show you how to use this compactness claim combined with the first definition of compactness that you've seen, namely every open cover contains a finite sub cover, and to use that to prove many consequences about the space of graphons. And some things that we had to work a bit hard at, but they turn out to fall from the compactness statement. So let's take a quick break.

In the first part of this lecture, we proved that a space of graphons is compact. So now let me show you what we can reap as consequences from the compactness result. So I want to show

you how to apply compactness and prove some consequences.

As I mentioned earlier, the compactness result is related to regularity. And in fact, many of the results I'm going to state, you can prove maybe with some more work using the regularity lemma. But I also want to show you how to deduce them directly from compactness. In fact, we'll deduce the regularity lemma from compactness. So these two ideas, compactness and regularity, they go hand-in-hand.

So first, more so as a warm up, but as also an interesting result, statement, an interesting statement on its own, so let me prove the following. So here is a statement that we can deduce from compactness. So for every  $\epsilon$ , there exists some number  $N$  which depend only on  $\epsilon$ , such that for every  $W$  graphon, there exists a graph  $G$  with  $N$  vertices, such that the cut distance between  $G$  and  $W$  is, at most,  $\epsilon$ .

So think about what this says. So for every  $\epsilon$ , there is some bound  $N$  such that every graphon-- so a graphon is some real-value function, so taking values between 0 and 1. You can approximate it in the distance that we care about by a graph with a bounded number of vertices. This is kind of like regularity lemma. If you are allowed edge weights on this graph  $G$ , then it immediately follows from the weak regularity lemma that we already proved. And from that weak regularity lemma which allows you to get some  $G$  with edge weights, you can think about how you might turn an edge-weighted graph into an unweighted graph. So that can also be done.

But I want to show you a completely different way of proving this result that follows from compactness. And so I say it's a warm up, because it's really a warm up for the next thing we're going to do. This is an easier example showing you how to use compactness.

So the idea is, I have this compact space. I'm going to cover this space by open sets, by open balls. So the open balls are going to be this  $B_{\epsilon}(G)$ . So for each graph,  $G$ , I'm going to consider the set of graphons that are within  $\epsilon$  of  $G$ . So this is you have some topological space or some metric space. I have a point  $G$ . And I look at its open ball. This is the ball. So I claim that these open balls, they form an open cover, of the space. Where is that? So I want to show every point  $W$  is covered.

So this follows from the claim that every  $W$  is the limit of some sequence of graphs. So we didn't technically actually prove this claim. I said that if you take  $W$  random graphs, you get this. So we didn't technically prove that. But OK, so it turns out to be true. There are easier

ways to establish it as well by taking  $l_1$  approximations.

But the point is that, if you use this claim here, you do not get a bound on the number of vertices. It could be that for very bizarre-looking  $W$ 's, you might require much more number of vertices. And a priori, you do not know that it is bounded as a function of  $\epsilon$ .

But now we have this open cover. So by compactness of this space of graphons, we can find an open cover using a finite subset of these graphs, so  $G_1$  to  $G_k$ , so a finite subset to do an open cover. And now we let  $N$  to be the least-common multiple of all of these vertex set sizes.

So all of these graphs, they are within-- so for each of these graphs, I can replace it by a graph on exactly  $N$  vertices. There exists a graph  $G_i$  prime of exactly  $N$  vertices, such that they represent the same point in the space of graphons. So why is this? Think about the representation of a graphon using from a graph.

If I start with  $G$  and I blow up each vertex into some  $k$  vertices, then it turns out-- I mean, you should think about why this is true. But it's really not hard to see if you draw the picture. So remember, this black and white picture, that actually, they're the same point. They are represented by the same graphon.

OK, and that's it. So we found these  $G$ 's. All of them have exactly  $N$  vertices such that their  $\epsilon$  open balls form an open cover of the space of graphons. So every graphon can be approximated by one of these graphs. So you get that from compactness.

The statement says, for every  $\epsilon$ , there exists an  $N$ . So  $N$  is a function of  $\epsilon$ . What's the function? This proof doesn't tell you anything about that. So this proof gives no information about the dependence of  $N$  on  $\epsilon$ .

So in some sense, it's even worse than some of the things we've seen in the earlier discussion on Szemerédi's regularity lemma where there were tower or Wowzer-types. Here there is no information, because it comes from a compactness statement. So you just know there exists a finite open cover, no bounds.

OK, any questions about this warm-up application? So it feels a bit magical. So you have compactness. And then you have all of these consequences.

So now let me show you how you can deduce the regularity lemma itself from compactness. In

fact, in the proof of the existence, in the proof of compactness, we only used weak regularity. And now let me show you how you can use the weak regularity consequence of namely compactness to bootstrap itself to strong regularity.

So we saw a version of strong regularity in the earlier chapter when we discussed Szemerédi's regularity lemma. So let me state it in a somewhat different-looking form, but that turns out to be morally equivalent. Suppose I have a vector of epsilons. So all of these are positive real numbers.

The claim is that there exists an  $M$  which only depends on this vector such that for every graphon  $W$ , one can-- so every graphon  $W$  can be written as the following, decomposing the following way. We write  $W$  as a sum of a structured part, a pseudo-random part, and a small part, where the structured part is a step function with  $k$  parts, but  $k$  is, at most,  $M$ , this claimed bound  $M$ .

The pseudo-random part has a very small cut norm, so its cut norm, very small, even compared to the number of parts. And finally, the small part has  $L_1$  norm bounded by  $\epsilon$ . So that's the claim. You can always-- there exists some bound  $M$  in terms of these error parameters so that you have this decomposition.

So we saw some version of this earlier when we discussed the spectral proof of regularity lemma. And I don't want to go into details of how these two things are related, but just to comment that depending on your choice of the epsilon parameters, it relates to some of the different versions of regularity lemma that we've seen before. So for example, if  $\epsilon_k$  is roughly  $\epsilon$ , some fixed  $\epsilon$  over  $k$  squared, then this is basically the same as Szemerédi's regularity lemma, whereas if all the  $\epsilon_k$ 's are the same  $\epsilon$ , then this is roughly the same as the weak regularity lemma.

All right, so how to prove this claim? We're going to use compactness again. So first, there always exists an  $L_1$  approximation so that every  $W$  has some step function  $u$  associated to it such that the  $L_1$  distance between  $W$  and  $u$  is, at most,  $\epsilon$ .

So again, this is one of these more measured theoretic technicalities I don't want to get into, but so it's not hard to prove. So roughly speaking, you have some function. You can approximate it using steps.

So similar to what we did just now, if you just do that, the number of steps might not be a function of  $\epsilon$ , so you might need much more steps just doing that if your  $W$  looks more pathological. So now what we're going to do is consider the following function,  $k$  of  $W$ . And I define it to be the minimum  $k$  such that there exists a  $k$  step graphon  $u$  such that  $\|u - W\|$  is, at most,  $\epsilon$ .

So among all the step function approximations, pick one that has the minimum number of steps and call the number of steps  $k$  of  $W$ . So now, as before, we're going to come up with an open cover of the space of graphons. So the open cover is going to be consisting of the cut norm balls of-- actually, what notation did I use over there? So this is a ball centered around  $W$  with radius  $\epsilon$ . This is an open cover of the space of graphons as  $W$  ranges over all graphons.

So I'm literally looking at every point in the space and putting an open ball around it. So obviously, this is an open cover. And because of compactness, there exists a finite sub cover. So there exists a finite set, we write curly  $s$ , of graphons such that these balls, as I range over  $W$  and curly  $s$ , they cover the space of graphons.

Now the goal is, given the  $W$ , I want to approximate it in some way. So having a finite set of things to work with allows us to do some kind of approximations. So thus, for every  $W$  graphon, there exists a  $W'$  in  $s$  whose ball in that collection covers the point  $W$ , such that  $W$  is contained in this ball.

And OK, so given this  $W'$ , because of this definition over here, so there exists a  $u$  which is a  $k$  step graphon with  $k$ , at most, the maximum over all such possible number of steps, such that  $W$  and  $W'$ , they are close in cut norm because you have this open cover. And furthermore,  $W'$  is close to a graphon with a small number of steps. So suppose we now write  $W$  as  $u$  plus  $W - W'$  and then plus  $W' - u$ .

We find that this is the decomposition that we are looking for because the  $u$ -- so this is the structural component-- has  $k$  steps, where  $k$  is less than this quantity here. And that quantity there is just some function of  $\epsilon$ . So it's, at most, some function of the  $\epsilon$ . It doesn't depend on the specific choice of  $W$ .

The second term, this is this pseudo-random piece, because its cut norm is small, so what we have here. Yeah, so this entire thing should be subscript. And finally, the third term here is the small term, because its  $L_1$  norm is small.

So putting them together, we get the regularity lemma. So again, the proof gives you no information whatsoever about the bound  $M$  as a function of the input parameters, the epsilons. So it turns out you can use a different method to get the bounds. Namely, we actually more or less did this proof when we discussed regularity lemma, the strong regularity lemma.

So we did a different proof where we iterated an energy increment argument. And that gave you some concrete bounds, some bounds which iterates on these epsilons. But here is a different proof. It gives you less information, but it elegantly uses this compactness feature of the space of graphons. Any questions?

OK, so we proved compactness. So now let's go on to the other two claims, namely the existence of the limit and that equivalences of convergence. The existence of the limit more or less is a consequence of compactness.

So you have this sequence of graphons,  $W_1, W_2$ , and so on. And the claim is that, if this sequence of  $F$  densities converges for each  $F$ , then there exists some limit  $W$  such that all of these sequences of  $F$  densities converge to the limit density. So that was the claim, so nothing about cut norms in at least as far as the statement goes.

Well OK, from compactness, we know that you can produce always a subsequential limit. So by compactness or sequential compactness, there exists some limit point which we call  $W$ . And this  $W$  has the property that, for some subsequence, the cut distance from the subsequence converges to  $W$ . So for some subsequence  $n_0$  as  $n_i$  going to infinity.

But now, by the counting lemma, the sequence of  $F$  densities-- so the counting Lemma tells you, if you have cut distance going to 0, then the  $F$  density should also go to 0. So indeed, that's what we have here. So this is so far just for the subsequence.

But we assumed already that the entire sequence converges in respect to every  $F$  densities. So it must be the same limit. And that finishes the proof of convergence, so proof of the existence of the limit. So we obtain this limit from compactness.

Next, let's prove the equivalence of convergence. And this one is somewhat trickier. So what happens here is that we would like to show that these two notions of convergence, one having to do with  $F$  densities and another having to do with cut distance, that these two notions are equivalent to each other.

So the goal here is to show that this  $F$  density convergence is equivalent to the statement that  $W$  sub  $n$  is Cauchy with respect to the cut distance. All right, claim one of the directions is easy. Which direction is that?

So which direction is the easy direction? So which way, left, going left, going right? OK, so going left? So I claim that this is easy, because it follows from counting lemma.

Counting lemma, remember the spirit of the counting lemma, at least qualitatively, is that if you have two graphons that are close in cut distance, then they are close in  $F$  densities. So if you have Cauchy with respect to cut distance, then they are Cauchy, and hence, convergent in  $F$  densities.

And it's the other direction that will require some work. And this one is actually genuinely tricky. So and it's almost kind of a miraculous statement, that somehow if you only knew the  $F$  densities-- so somebody gives you this very large sequence of graphs and only tells you that the triangle densities, the  $C_4$  densities, all of these graph densities, they converge. Somehow from these small statistics, you conclude that the graphs globally look very similar to each other. That's actually, if you think about, this is an amazing statement.

OK, so let's see the proof. The proof method here is somewhat representative of these graph-limit-type arguments. So it's worth paying attention to see how this one goes. So by compactness, if-- OK, we're going to set up by contradiction.

If the sequence is not Cauchy, then there exists two limit points. So there exists at least two distinct limit points. And call them  $u$  and  $W$ , and such that-- so because you have two separate limit points, you must have that this sequence, at least along a subsequence that converges to  $W$ , converges in  $F$  densities to  $W$ . So initially, this is true along subsequence. But the left-hand side is convergent, so this is true along the sequence.

But  $u$  is also a limit point. So the same is true for  $u$ . And therefore, the  $F$  density in  $W$  must equal to the  $F$  density in  $u$  for all  $F$ .

So we would be done if we can prove that the  $F$  densities, the collection of all these  $F$  densities, they determine the graphon. And that's indeed the case. And so this is the next claim. So it's what I will call a moment lemma, is that if  $u$  and  $W$  are graphons such that the  $F$  densities agree for all  $F$ , then the cut distance between  $u$  and  $W$  is equal to 0.

Somehow the local statistics tells you globally that these two graphons must agree with each other. Does anyone know why I call it a moment lemma? There is something else which this should remind you of. So there are some classical results in probability that tells you, if you have two probability distributions, both, assume are nice enough, then if they have the same  $k$ 'th moment for every  $k$ , so first moment, second moment, third moment, if all the moments agree, that these two probability distributions should agree.

And this is some graphical version of that. So instead of looking at the probability distribution, we're looking at graphons, which are two-dimensional. These are two-dimensional objects. And this moments lemma tells you that in these two-dimensional, in the corresponding two-dimensional moments, namely these  $F$  moments, if they agree, then the two graphons must agree. So it's the analog of the probability theory statement about moments.

The proof is actually somewhat tricky. So I'm only going to give you a sketch. And the key here is to consider the  $W$  random graph, which we saw last lecture. So this is  $W$  random graph with  $k$  vertices sampled using the graphon  $W$ .

So a key observation here is that, for every  $F$ , the probability that the sampled  $W$  random graph agrees with  $F$ — and here, there is a bit of a technicality. I want them to agree as labeled graphs. So the vertices of  $W$  are a priori labeled 1 through  $k$ . And this  $kW$  random graph is generated with vertices labeled 1 through  $k$ . They agree with some probability that is completely determined by the  $F$  densities. Yes?

**AUDIENCE:** Is  $k$  the number of vertices of  $F$ ?

**PROFESSOR:** Yeah, so  $k$  is the number of vertices of  $F$ . And the specific formula is not so important. Let me just write it down. But the point is that, if you know all the  $F$  densities, then you have all the information about the distribution of this  $W$  random graph.

And the way you can calculate the actual probability is via an inclusion exclusion. And the reason we have to do this inclusion exclusion is just because this is more like counting induced subgraphs. And this is counting actual subgraphs. So there is an extra step. But the point is that, if you knew this data, the moment's data then you immediately know the distribution of the  $W$  random graphs.

OK, so if I have two graphons for which I know that their  $F$  densities agree, then I should be able to conclude that the corresponding  $W$  random graphs also have the same distribution, in

particular, this, the  $W$  random graph and the  $u$  random graph have the same distribution. I am going to create a variant of the  $W$  random graph which is something called an  $H$  random graph. It's kind of like the  $W$  random graph except I forget the very last step.

So I only keep a weighted-- think of it as, so think this is an edge-weighted graph where you sample  $x_1$  through  $x_k$  uniformly between 0 and 1. And I put edge weight between  $i$  and  $j$  to be  $W$  of  $x_i x_j$ . So the difference between this  $H$  version and the  $G$  version is that the  $G$  version is obtained by turning this weight into an actual edge with that probability, but if I don't do the last step, I obtain this intermediate object.

So the following are true. And this is where I'm going to skip the proofs. If I look at this  $H$  random graph and the  $G$  random graph, they are very close in cut distance. You can think of this as the claim that  $G$  and  $P$  is very close to  $G$  in cut distance. So they are very close in cut distance. As  $k$  going to infinity with probability 1-- so now I'm going to do the proof. But it's some kind of a concentration argument.

And the second claim is that the  $H$  random graph is actually very close to the original graphon  $W$ , as well. This is also little  $l_1$  in distance as  $k$  goes to infinity. So this one is, again, not so obvious. But it's easier in the case when  $W$  itself is a step function, in which case, the produced  $H$  is almost the same as  $W$ , except the boundaries are slightly shifted, perhaps. And so you first approximate  $W$  by a step function, and prove this up to an epsilon approximation, and then let the steps go to infinity.

So if you have these two claims, so then we see that this one here is identically distributed as  $u$ . So it should follow that the corresponding  $H$  random graph for  $u$ , if you place the same inequalities by the  $u$  versions, it should also be true. So because these two are the same distribution, if you follow this chain, you obtain that the cut distance between  $u$  and  $w$  is equal to 0.

I want to close by mentioning, in some sense-- so here, you have two graphons that have exactly the same  $F$  moments. But what if I give you two graphons which have very similar moments to each other? Can you conclude that the two graphons are close to each other? And that will be some kind of an inverse counting lemma. And in fact, it does follow as a corollary.

And the statement is that, for every epsilon, there exists  $k$  and  $\eta$  such that if the two

graphons  $u$  and  $W$  are such that the  $F$  densities do not differ by more than  $\eta$  for every  $F$  on, at most,  $k$  vertices, then the cut distance between  $u$  and  $W$  is, at most,  $\epsilon$ . So the counting lemma tells you, if the cut distance is small, then all the  $F$  moments are close to each other. And the inverse tells you this converse. So it tells you this, if you have similar  $F$  moments, up to a certain point, then this is small.

You can deduce the inverse counting lemma from the moments lemma via a compactness argument similar to the one that we did in class today. And I want to give you a chance to practice with that argument. So this will be on the homework, for the next homework. I'll give you some practice with using these compactness arguments.

But you see, just with the other compactness statements, it doesn't tell you anything about the  $k$  and the  $\epsilon$  as a function of-- the  $k$  and  $\eta$  as a function of  $\epsilon$ . So there are other proofs that gives you concrete bounds, but this proof here is much simpler if you assume the corresponding results about compactness.

And finally, I want to mention that in the moments lemma, in order to deduce that  $u$  and  $w$  have the same-- that they are basically the same graphon, we need to consider  $F$  moments for all  $F$ 's. So you might ask, could it be the case that we only need some finite set of  $F$ 's to deduce-- to recover the graphon? Is it the case that you can recover  $W$  from only a finite number of  $F$  moments?

And this is, it's actually a very interesting problem for which we already saw one instance. Namely, when we discussed quasi-random graphs, we saw that if you know that the  $k_2$  moment is  $p$  and also the  $C_4$  moment is  $p$  to the 4, then we can deduce that the graphon must be the constant graphon,  $p$ . OK, so we didn't do it in this language, but that's what the proof does.

And likewise, you can use this to deduce a qualitative version where you have an extra slack and an extra slack over here. So you might ask, except for the constant graphons, are there other graphons for which you can similarly deduce-- recover this graphon from just a finite amount of moments data? And such graphons are known as finitely forcible. So finitely forcible graphons  $W$  such that a finite number of moments can uniquely recover-- can uniquely identify this graphon,  $W$ .

And a very interesting question is, what is the set of all finitely forcible graphons? And it turns out, this is not at all obvious. And let me just give you some examples, highly non-trivial, that

turned out to be finitely forcible.

For example, anything which is a step graphon is finitely forcible. The half graphon which corresponds to the limit of a sequence of half graphs is finitely forcible. I mean, already, I think neither of these two examples are easy at all.

And this example here can be generalized where you have any polynomial curve. I think this has to be-- so if it's a polynomial curve, it's also finitely forcible. But turns out finitely forcible graphons can get quite complicated. And there is still rather quite a bit of mystery around them.

OK, so next time, I want to discuss some inequalities that come out of-- you can state between different F densities. OK, great. That's all for today.