

PROFESSOR: So we've been discussing Szemerédi's regularity level for the past couple of lectures. And one of the theorems that we proved was Roth's theorem, which tells us how large can a subset of 1 through N be if it has no 3-term arithmetic progressions. And I mentioned at the end of last time that we can construct fairly large subsets of 1 through N without 3-term arithmetic progressions. I want to begin today's lecture with showing you that example, showing you that construction.

But first, let's try to do something that is somewhat more naive, somewhat more straightforward, to try to just construct greedily 3-AP-free sets of 1 through n . And recall from last time, we showed Roth's theorem that such a set must have size little o of N . So what's one thing you might do?

Well, you can try to greedily construct such a set. It will just make our life a little bit easier if we start with 0 instead of 1 . So you put one element in, and you keep putting in the next integer, as long as it doesn't create a 3-term arithmetic progression. So you keep doing this.

Well, you can't put 2 in, so let's skip 2 . So we go to the next one, 3 . So 4 is OK. So we skip 5 . We have to skip 6 as well because $0, 3, 6$, that's a 3-AP. So we keep going. So what's the next number we can put in?

AUDIENCE: 9.

PROFESSOR: Up. Go to 9 . Then the next one is 10 . What's the next one we can put in? So we can play this game. Find out what is the next number that you can put in. Greedily, if you could put it in, put it in in a way that generates a subset of integers without 3-term arithmetic progressions.

So this actually has a name, so this is called a Stanley sequence. And there's an easier way to see what the sequence is. Namely, if you write the sequence in base 3 , you find that these numbers are $0; 0, 1; 1, 0; 1, 1, 0, 0, 1, 0, 1, 1, 1, 0, 1, 1, 1$.

So these are just numbers whose base 3 representation consists of zeros and ones. So I'll leave it to you as an exercise to figure out why this is the case, if you generate the sequence greedily this way, this is exactly the sequence that you obtain. But once you know that, it's not too hard to find out how many numbers you generate.

So up to-- suppose N is equal to 3 to the k . We get 2 to the k terms, which gives you N raised

to the power of log base 3 of 2. So you can figure out what that number is, but some numbers strictly less than 1.

And actually, for a very long time, people thought this was the best construction. So this construction was known even before Stanley, so it was something that is very natural to come up with. And it was somewhat of a surprise when in the '40s, it was discovered that you can create much larger subsets of the positive integers without arithmetic progressions. So that's the first thing I want to show you today.

So these sets were first discovered by Salem and Spencer back in the '40s. And a few years later, Behrend gave a somewhat improved and simplified version of the Salem-Spencer construction. So these were all back in the '40s. So these days, we usually refer, at least in the additive combinatorics community, to this construction as Behrend's construction. And this, indeed, what I will show you is due to Behrend, but somehow, this Salem-Spencer name has been forgotten and I just wanted to point out that it was Salem and Spencer who first demonstrated that there exists a subset of 1 through N that is 3-AP-free and has size N to the 1 minus little o 1, that there's no power saving.

So there exists examples with no power saving. And this is important because as we saw last time, the proof of Roth's theorem-- and we basically spent two full lectures proving Roth's theorem. And it's somewhat involved, right. It wasn't this one line of inequality you can just use to deduce the result. And that is part of the difficulty. So having such an example is indication that Roth's theorem should not be so easy. That's not a rigorous demonstration, but it's some indication of the difficulty of Roth's theorem.

So let's see this construction due to Behrend. So let me write down the precise statement. There exists a constant, C , such that there exists a subset of 1 through N that is 3-AP-free and has size at least N times e to the minus C root log N .

Perhaps somewhat amazingly enough, this bound is still the current best known. We do not know any constructions that is essentially better than Behrend's construction from the '40s. There were some small recent improvements that clarified what the C could be, but in this form, we do not know anything better than what was known back in the '40s.

And the construction, I will show you-- hopefully you should believe that it's quite simple. So I will show you what the construction is. It's clever, but it's not complicated. And it's a really

interesting direction to figure out is this really the best construction out there. Can you do better?

So let's see. We're going to set some parameters to be decided later, m and d . Let me consider x to be this discrete box in d dimensions. So this is the box of lattice points in d dimensions, 1 through m raised to d .

And let me consider an intersection of this box by a sphere of radius \sqrt{L} . So namely, we look at points in this x , such that the sum of their squares is equal to exactly L . You take a bunch of these spheres, they partition your set x , so the smallest possible sum of squares, largest possible sum of squares.

So in particular, there exists some L such that x of L is large, just by pigeonhole principle. You can probably do this step with a bit more finesse, but it's not going to change any of the asymptotics. The intuition here-- and we'll come back to this in a second-- is that xL lies on a sphere. So this is the set of lattice points that lies in a given sphere.

And because you are looking at a sphere, it has no 3-term APs. So we're going to use that property, but right now it's not yet a subset of the integers. So what we're going to do is to take this section of a sphere and then project it to one dimension, so that it is now going to be a subset of integers. And we're going to do this in such a way that it does not affect the presence of 3-term APs.

So let us map x to the integers by base expansion. So this is base $2m$ expansion. All right. So what's the point of this construction here so far? Well, you verified a couple of properties, the first being that this construction here is injective, this map. So if you call this map ϕ , this ϕ is injective.

Well, it's base expansion, so it's injective. But also a somewhat stronger claim is that if you have three points-- so if you have three points in x , that map to a 3-AP in the integers, then the three points originally must be a 3-AP in x . Again, this is not a hard claim. Just think about it.

Here, because we're using base $2m$ expansion and you're only allowed to use digits up to m , you don't have any wrap around effects. You don't have any carryovers when you do the addition. So combining these two observations, we find that the image of xL is a 3-AP-free set off 1 through N . So what is N ?

I can take N to be, for instance, $2m$ raised to power d , so all the numbers up there are less

than this quantity. And the size is the size of x sub L , which is N to the d divided by d^m squared. And now you just need to find the appropriate choices of the parameters m and d to maximize the size of x sub L . And that's an exercise.

So, for instance, if you take m to be e to the root $\log N$ and d to be root $\log N$, then you find that the size here is the bound that we claim. And that finishes the construction of Behrend, giving you a fairly large subset of 1 through N without 3 -term arithmetic progressions. And the idea here is you look at a higher dimensional object, namely, a higher dimensional sphere which has this property of being 3 -AP-free, and then you project it onto the integers. Any questions so far?

OK. So we have some proofs, we have some examples. So now let's go on to variations of Roth's theorem. And I want to show you a higher dimensional version of Roth's theorem. So I mentioned in the very first lecture-- so you have this whole host of theorems and additive combinatorics-- Roth, Szemerédi, multi-dimensional Szemerédi.

So this is, in some sense, the simplest example of multi-dimensional Szemerédi theorem. And this example is known as corners. So what's a corner? A corner is, well, we're working inside two dimensions, so a corner is simply three points, such that the two-- so they're positioned like that-- such that these two segments, they are parallel to the axes and they are the same length. So that's, by definition, what a corner is.

And the question is, if you give me a subset of a grid that is corner-free, how large can this subset be? Here's a theorem. So if A is a subset of 1 through N with no corners, particular, no three points of the form, x comma y ; $x + d$ comma y ; and x comma $y + d$, where d is some positive integer, then the size of A has to be little o of N squared. Question.

AUDIENCE: Do we only care about corners oriented in that direction?

PROFESSOR: OK, good question. So that's one of the first things we will address in this proof. So your question was, do we only care about corners oriented in the positive direction. So you can have a more relaxed version of this problem where you allow d to be negative as well. The first step in the proof is we'll see that that constraint doesn't actually matter. Yes. OK. Great.

Let's get started. So as Michael mentioned, we do have this constraint in this over here that d is positive. And it's somewhat annoying because if you remember in our proof of Roth's theorem, they are positive, negative, they don't play a role. So let's try to find a way to get rid

of this constraint so that we are in a more flexible situation.

So first step is that we'll get rid of this d being positive requirement. So here's a trick. Let's consider the sumset $A + A$. So sumset here means I'm looking at all pairwise sums as a set. So you don't keep track of duplicates, so you'll keep it as a set. And we're living inside this grid, but now somewhat wider in width.

Then there exists an element in this domain of the sumset. By pigeonhole, that's represented in many different ways. So there exists a z represented as a plus b in at least size of $|A|^2$ divided by size of $2N$ squared different ways, just because if you look at how many things come up in that representation, just use pigeonhole.

And now let's take A' to be $A \cap (z - A)$. So what's happening here? So you have this set A . And basically what I want to do is look at the-- so suppose A looks like that. So look at $-A$ and then shift $-A$ over so that they intersect in as many elements as you can. And by pigeonhole, I can guarantee that their intersection is fairly large. So because the size of A' , it's essentially the number of ways that z can be represented in the aforementioned manner.

So it suffices to show that A' is little o of N^2 . If you show that, then you automatically show A is little o of N^2 . But now A' is symmetric. A' is symmetric around $z/2$.

So this is centrally symmetric about $z/2$, meaning that $A' = z - A'$. And so you see now we've gotten rid of this d positive requirement because if A' had a positive corner, then it has a negative corner and vice versa. So no corner in A' with d positive implies that no corner with d negative.

So now let's forget about A and A' and just replace A by A' and forget about this d positive condition. So let's forget about this part, but I do want $d \neq 0$. Otherwise, you have trivial corners, and, of course, you always have trivial corners. All right.

So let's remember how the proof of Roth's theorem went. And this relates to the very first lecture where I showed you this connection between additive combinatorics on one hand and graph theory on the other hand, or you take some arithmetic pattern and you try to encode it in a graph so that the patterns in your arithmetic set correspond to patterns in the graph. So we're going to do the same thing here.

So we're going to encode the subset of the grid as a tripartite graph in such a way that the corners correspond to triangles in the graph. So I'll show you how to do this and it will be fairly simple. And in general, sometimes it takes a little bit of ingenuity to figure out what is the right way to set up the graph. So one of an upcoming homework problem will be for you to figure out how to set up a corresponding graph when the pattern is not a corner, but a square. If I add one extra point up there, how would you do that?

So what does this graph look like? Let's build a tripartite graph where-- so this should be somewhat reminiscent of the proof of Roth's theorem-- where I give you three sets, x , y , and z ; and x and y are both going to have N elements and z is now going to have $2N$ elements. x is supposed to enumerate or index all the vertical lines in your grid.

So you have this grid of N by N . So x has size N . Each vertex in x corresponds to a vertical line. y corresponds to the horizontal lines and z corresponds to these negatively sloped diagonal lines-- slope minus 1 lines. And of course you should only take lines that affect your N by N grid. So that's why there are this many of them for each direction.

So what's the graph? I join two vertices if their corresponding lines meet in a point of A . So I might have-- this may be A , a point of A . So I put an edge between those two lines-- one for x , one for z -- because their intersection lies in A . And more explicitly-- I mean, that is pretty explicit-- and alternatively, you could also write this graph by telling us how to put in the edge between x and z . Namely, you do this if x comma z minus x lies in A ; you put an edge between x and y if x comma y lies in A ; and likewise, you put in the final edge if z minus y comma y lies in A .

So two equivalent descriptions of the same graph. Any questions? All right. So the rest of the proof is more or less the same as the proof of Roth's theorem we saw last time. So we need to figure out how many edges are there. Well, every element of A gives you three edges. So that's one of the edges corresponding to that element. The other two pairs with two other edges.

And most importantly, the triangles in this graph, I claim, correspond to corners. So what is a triangle? A triangle corresponds to a horizontal line, a vertical line, and a slope 1 minus line that pairwise intersect in elements of A . If you look at their intersections, that's a corner. And conversely, if you have a corner, you build a triangle.

And because your graph, your set is corner-free, well, you don't have any triangles. Actually, no, that's not true. So like we saw in Roth's theorem, you have some triangles, but corresponding to trivial corners. So triangles correspond to trivial corners because your set A is corner-free. Trivial corners, meaning the same point with d close to 0, so it's not a genuine corner like that.

And in particular, in this graph, every edge is in a unique triangle. And so we're in the same situation. By the corollary of triangle removal lemma, we find that the number of edges must be little o of the number of vertices. The number of edges must be subquadratic, so it must be little o of N squared. And so that implies that the size of A is little o of N squared. So that proves the Corners theorem. Any questions?

So once you set up this graph, the rest is the same as Roth's theorem. So this connection between graph theory and additive combinatorics, well, you need to figure out how to set up this graph. And then sometimes it's clear, but sometimes you have to really think hard about how to do this.

All right. What does corner have to do with Roth's theorem, other than that they have very similar looking proofs? Well, actually, you can deduce Roth's theorem from the Corners theorem. And to show you this precisely, so let me use $r_{3,N}$ to denote the size of the largest 3-AP-free set of 1 through N .

So this notation, the $r_{3,N}$ is actually fairly standard. The next one's not so standard, but let's just do that. So that's not an L , but that's a corner. So that is the size of the largest corner-free subset of N squared. So we gave bounds for both quantities, but they are actually related to each other through this fairly simple proposition that if you have an upper bound for corners, then you have an upper bound for Roth's theorem.

Indeed, given A a 3-AP-free subset of 1 through N , let me build for you a corner-free subset of the grid that is a fairly large subset of that grid. I can form B by setting it to be the set of pairs inside this grid whose difference, x minus y , lies in A . So what does this look like?

This is the grid of size $2N$. And if I start with A that is 3-AP-free, what I can do-- so over here-- is look at the lines like that, putting all of those points. And you see that this set of points should not have any corners because if they have corners, then the corners would project down to a 3-AP. So B is corner-free.

To recap, if we have upper bound on corners, we have upper bound on Roth's theorem. But you also know that if you have lower bound on Roth's theorem, then you have lower bound on corners. So the Behrend construction we saw at the beginning of today extends to, you know, through exactly this way to a fairly large corner-free subset. And that's more or less the best thing that we know how to do.

In fact, there aren't that many constructions in additive combinatorics that's known. Almost everything that I know how to construct that's fairly large come from Behrend's construction or some variant of Behrend's construction. So it looks pretty simple. It comes from the '40s, yet we don't really have too many new ideas besides playing and massaging Behrend's construction.

Let me tell you what is the best known upper bound on the Corners theorem. This is due to Shkredov. So the proof using triangle removal lemma, it goes to Szemerédi's regularity lemma. It gives you pretty horrible bounds, but using Fourier analytic methods-- so you see if you have upper bound on Roth, it doesn't give you an upper bound on corner, so you need to do something extra. And so the best known bound so far is of the form, N^2 divided by polylog $\log N$, so $\log \log N$ raised to some small constant, C . Any questions?

Last time, we discussed the triangle counting lemma and the triangle removal lemma. Well, it shouldn't be a surprise to you that if we can do it for triangles, we may be able to do it for other subgraphs. So that's the next thing I want to discuss-- how to generalize the techniques and results that we obtain for triangles to other graphs, and what are some of the implications if you combine it with Szemerédi's regularity lemma.

So let's generalize the triangle counting lemma. So the strategy for the triangle counting lemma, let me remind you, was that we embedded the vertices one by one. Putting a vertex and a typical vertex here should have many neighbors to both vertex sets. So these two guys should have sizes typically roughly the same as a fraction of them corresponding to the edge density between the vertex parts.

And if they're not too small, then from the epsilon regularity of these two sets, you can deduce the number of edges between them. So that was the strategy for the triangle counting lemma. So you can try to extend the same strategy for other graphs. So let me show you how this would have been done, but I don't want to give too many details because it does get somewhat hairy if you try to execute it.

So the first strategy is to embed the vertices of H one at a time. So my H now is going to be, let's say, a K_4 . And I wish to embed this H in this setting where you have these four parts in the regularity partition, and they are pairwise epsilon regular with edge densities that are not too small.

Well, what you can try to do, mimicking the strategy over there, is to first find a typical image for the top vertex. And a typical image over here, minus some small bad exceptions, will have many neighbors to each of the three parts. Next, I need to figure out where this vertex can go.

I'm going to embed this vertex somewhere here. Again, a typical place for this vertex, modulo some small fraction, which I'm going to throw away. So now you see you need somewhat stronger hypotheses on the epsilon regularity, but they are still all polynomial dependent, so you just have to choose your parameters correctly.

So this typical green vertex should have lots of neighbors over here. So you just keep embedding. So it's almost this greedy strategy. You keep embedding each vertex, but you have to guarantee that there are still lots of options left. So embed vertices one at a time.

And I want to embed each vertex so that the yet to be embedded vertices have many choices left. So epsilon regularity guarantees that you can always do this. And you do this all the way until the end and you arrive at some statement. And depending on how you do this, the exact formulation of the statement will be somewhat different, but let me give you one statement which we will not prove, but you can deduce using this strategy.

And this is known as a graph embedding lemma. And again, as I mentioned when I started discussing Szemerédi's regularity lemma, the exact statements, they are fairly robust and they are not as important as the spirit of the ideas. So if you have some application in mind, you might have to go into the proof, tweak a thing here and there, but you get what you want.

So the graph embedding lemma says, for example, that if H is bipartite, and with maximum degree and most Δ -- so the H maximum degrees at most Δ -- and suppose you have vertex sets V_1 through V_r , such that each vertex set is not too small; and if these vertex sets are pairwise epsilon regular and the density is not too small. So here I'm assuming some lower bound on the density which depends on your epsilon. Then the conclusion is that G contains a copy of H .

I just want to give a few remarks on the statement of this theorem. Again, we will not discuss

the proof. So what is this hypothesis on H being r -partite have to do with anything? So here, as an example, when r equals to 4, instead of this K_4 , maybe I also care about that graph over there.

Well, maybe some more vertices, some more edges, but if it's 4-partite. And the point is that I want to embed the vertices in such a way that that top vertex goes to the part that it's supposed to go into. So I'm embedding this configuration in the same way that corresponds to the proper coloring of H .

So if you do this, there's enough room still to go, as long as you have some lower bound on the edge density between the individual parts, and you only depend really on not the number of edges of H , but the maximum degree. Because if you look at how many times each vertex, its possibilities can be shrunk, it's in most δ times. Now, this graph embedding lemma-- so I give you this statement here, but it's a fairly robust statement. And if you want to get, for example, not just a single copy of H , if you want to get many copies of H , you can tweak the hypotheses, you can tweak the proofs somewhat to get you what you want, again, following what we did for the triangle counting lemma. Question.

AUDIENCE: Is the bound on the H edge density between partitions correct? So if the maximum degree increases, the lower bound decreases?

PROFESSOR: If maximum degree increases, this number goes up.

AUDIENCE: Oh, OK.

PROFESSOR: I want to show you a different way to do counting that does not go through this embedding vertices one by one, but instead we will try to analyze what happens if you take out an edge of H one by one. And that's an alternative approach which I like more. It's somewhat less intuitive if you are not used to thinking about it, but the execution works out to be much cleaner. And it's also in line with some of the techniques that we'll see later on when we discuss graph limits. Let's take a quick break.

Any questions so far? We will see a second strategy for proving a graph counting lemma. And the second strategy is more analytic in nature, which is to embed, well, to analytically somehow we'll analyze what happens when we pick out one edge of H at a time.

So let me give you the statement first. So the graph counting lemma says that if you have a graph H with vertex set elements of 1 through k -- I also have an epsilon parameter and I have

a graph G and subsets V_1 through V_k of G , such that $V_i V_j$ is ϵ regular whenever ij is an edge of H .

So here, the setup is slightly different from the picture I drew up there. So what's going on here is that you have, let's say, H being this graph. And suppose you are in a situation where I know that some of these-- so I know that five of these pairs are ϵ regular, and what I really want to do is embed this H into this configuration.

So $1, 2; V_1, V_2$; and so on. And I want to know how many ways can you embed this way. The conclusion is that the number of tuples, the number of such embeddings, such that V_i, V_j is an edge of G for all ij being the edge of H -- so exactly as showing that picture, the number of such embeddings, the little v 's, is within a small error, which is the number of edges of H ϵ times the total, the product of these vertex set sizes of this number, which is what you would predict the number of embeddings to be if all of your bipartite graphs were actually random.

So like in the triangle counting lemma, if you look at the edge densities in this configuration and predict how many copies of H you would get, that's the number you should write down. And this counting lemma tells you that the truth is not so far from the prediction. Any questions?

So we will prove the graph counting lemma in an analytic manner. It helps to-- so it will be convenient for me to rephrase the result just a little bit in this probabilistic form. So it has the equivalent to show that if you have uniformly randomly chosen vertices, little v_1 and big V_1 , and so on-- so little v_k and big V_k . So they're independent, uniformly, and random, then the probability-- so basically, I am putting down a potential image for each vertex of H and asking what's the probability that you actually have an embedding of H .

So the probability that little v_i , little v_j is an actual edge of G for all ij being an edge of H , this number here, we're saying that it differs from the prediction, which is simply multiplying all the edge densities together. So the difference between the actual and the predicted values is fairly small. So I haven't done anything, just rephrasing the problem. Instead of counting, now we're looking at probabilities.

As I mentioned, we'll take out one edge at a time. So relabeling if necessary, let's assume that $1, 2$ is an edge of H . So now we will show the following plane, so I'll denote star. That, if you

look at this quantity over here compared to if you take out just the edge density between V_1 and V_2 , but now you put in a similar quantity where I'm considering all of the edges of H , except for 1, 2. I claim that this difference is, at most, ϵ .

So you can think of this quantity here as the same quantity as the green one, except not on H , but on H minus the edge 1, 2. To show this star claim, let us couple the two random processes choosing these little v_i 's. By that, I mean here you have the random little v_i 's and here you have the random little v_i 's, but use the same little v_i 's in both probabilities.

So in both, there are two different random events, but you use the same little v_i 's, it suffices to show this inequality star with-- so what's this process, this random process? You pick V_1 , you pick V_2 , you pick V_3 , all of them independently uniformly at random. But if I show you the inequality under a further constraint of arbitrary V_3 through V_N , then that's even better. So with V_3 through V_k , fixed arbitrary, and only little v_1 and little v_2 random.

So you can phrase this in terms of conditional probabilities, if you like. You're comparing these two probabilities. Now I fix the V_3 's through V_k 's, and I just let V_1 and V_2 be random. And if condition on V_3 through V_k , you have this inequality, then letting V_3 through V_k go, letting them be random as well, by triangle inequality, you obtain the bound that we're looking for.

Any questions about this step? If you're confused about what's happening here, another way to bypass all the probability language is to go back to the counting language. So in other words, we're trying to count embeddings. And I'm asking if you arbitrarily fix V_3 through V_k , how many different choices are there to V_1 and V_2 in the two different settings.

OK. So let A_1 be the set of places where little v_1 can go, if you already knew what V_3 through V_k should be. OK. So you look at all the neighbors of V_1 , neighbors of 1 and H , except for 2. And I want to make sure that V_1, v_i , as i ranges over all such neighbors, is indeed a valid H in G .

I'll draw a picture in a second. And A_2 , likewise, is the same quantity, but with 2 instead of 1. So for example, if you're trying to embed a K_4 , what's happening here is that you have this V_1 , this V_2 , and somebody already arbitrarily fixed the locations where V_3 and V_4 are embedded. And you're asking, well, how many different choices are now left for little v_1 . It's the common neighborhood of V_3 and V_4 . So that's for A_1 . And V_3 and V_4 , also their common neighborhood in B_2 is A_2 .

OK. So with that notation, what is it that we're trying to show over here? So if you rewrite this inequality with the V_3 's through V_k 's fixed, you find that what we're trying to show is the following. So we claim-- and this claim implies the star, that if you look at what the first term should be-- so this is the number of edges between A_1 and A_2 as a fraction of the product of V_1 and V_2 .

And then the second guy here is if you use the prediction d of V_1 and V_2 . So this is each of them, each of these two factors, there's a probability that little v_1 lies in A_1 , little v_2 lies in A_2 , and then you tack on this extra constant, namely, this constant here. So we're trying to show that this difference is small. And the claim is that this difference is, indeed, always small. There's always, at most, ϵ for every A_1 in V_1 and A_2 in V_2 .

And here, in particular, so this statement looks somewhat like the definition of ϵ -regularity, but there's no restrictions on the sizes of A_1 and A_2 , and they don't have to be big. And as you can imagine, this statement, we're not really using all that much. All we're assuming is the ϵ -regularity between B_1 and B_2 . So we will deduce this inequality from the hypotheses of ϵ -regularity between B_1 and B_2 . So let's check.

So we know that B_1 and B_2 is ϵ -regular by hypotheses. So if either A_1 or A_2 is too small, so if A_1 is too small or A_2 is too small, then we see that both of these terms here are, at most, ϵ . So if the A 's are too small, then neither of these terms can be too large. Here it's bounded by-- if you took the product of A_1 and

A_2 and likewise over there.

So in this case, their difference is, at most, ϵ , and we're good to go. Otherwise, if A_1 and A_2 are both at least an ϵ -fraction of their [INAUDIBLE] sets, then we find that-- so what happens?

So here, so by the hypothesis of ϵ -regularity, we find that d of V_1 and V_2 differs from the number of edges between A_1 and A_2 , divided by the product of their sizes. So that's just d of A_1 and A_2 . So this difference is, at most, ϵ , which then implies the inequality up there. So here we're using that the size of A is, at most, the size of V .

So we have this claim. And that claim proves this inequality in star. And basically, what we've done is we showed that if you took out a single edge, you change the desired quantity by, at most, an ϵ , essentially. So now you do this for every edge of H . Alternatively, you can do

induction on the number of edges of H .

So to complete a proof of the counting lemma, so we do induction on the number of edges of H . And when H has exactly one edge, well, that's pretty easy. But now if you have more edges, well, you apply induction hypothesis to the graph, which is H minus the edge e . And you find that this quantity here differs from the predicted quantity by the number of edges of H minus 1 times ϵ .

In other words, you run this prove that we just did one edge at a time. So each time you take out an edge, you use ϵ regularity to show that the effect of taking that edge out from H does not have too big of an effect on the actual number of embeddings. Do this for one edge at a time, and eventually you prove the graph counting lemma.

So this is one of those proofs which may be less intuitive compared to the one I showed earlier, in the sense that there's not as nice of a story you can tell about putting in one vertex at a time. But on the other hand, if you were to carry out this proof to bound each time how big the sets have to be, it gets much hairier over here. And here, the execution is much cleaner, but maybe less intuitive, unless you're willing to be comfortable with these calculations. And it's really not so bad.

And the strength of these two results are somewhat different. So again, it's not so much the exact statements that matter, but the spirit of these statements, which is that if you have a bunch of ϵ regular pairs, then you can embed and kind of pretending that everything behaved roughly like random. Any questions?

So now we have Szemerédi's graph regularity lemma, we have the graph counting lemma, embedding lemmas, we can use it to derive some additional applications that don't just involve triangles. So when we only had a triangle counting lemma, we can only do the triangle removal lemma, but now we can do other removal lemmas.

So in particular, there's the graph removal lemma, which generalizes the triangle removal lemma. So in the graph removal lemma, the statement is that for every H and ϵ , there exists a δ , such that every N vertex graph with fewer than δN^2 edges has fewer than ϵN^2 copies of H -- so it has very few copies of H -- such graph can be made H -free by removing a fairly small number of edges. All right, so same statement as the triangle removal lemma, except now for general graph H .

And as you expect, the proof is more or less the same as that of a triangle removal lemma, once we have the H counting lemma. So let me remind you how this goes. So it's really the same proof as triangle removal, where there was this recipe for applying the regularity lemma from last time. So what is it? You first-- so what's the first step when you do regularity?

You partition. So let's do partition. OK. So apply the regularity lemma to do partitioning. And what's the second step? So we clean this graph. And you do the same cleaning procedure as in the triangle removal lemma, except maybe you have to adjust the parameters somewhat, so remove edges, remove low density pairs, irregular pairs, and small vertex sets.

And the last step-- OK, so what do we do now? OK. So you can count. So if there were any H left, then the counting lemma shows you that you must have lots of copies of H left.

So now let me show you how to use the strategy. Now that we have this general graph counting lemma, we'll prove the Erdos-Stone-Simonovits theorem, which we omitted, the proof that we omitted from the first part of the course. So remind you, the Erdos-Stone-Simonovits theorem says that if you have a graph H, then the extremal number of H is equal to this quantity which depends only on the chromatic number of H.

The lower bound comes from taking the Turan graph. If you take the Turan graph, you get this lower bound, so it's really the upper bound that we need to think about. All right. So what's the strategy here? The statement really is that if you have a graph G that's N vertex whose number of edges is at least that much-- OK, so I fixed an epsilon bigger than zero, fixed a positive epsilon. So the claim, what we're trying to show with Erdos-Stone-Simonovits, is that if you have a graph G with too many edges-- too many meaning this many edges-- then G contains a copy of H if N is sufficiently large.

OK. So let's use the regularity method, so applying this three-step recipe. First, we partition. So partition the vertex set of G into m pieces, and in such a way that it is ϵ regular-- and for some parameter ϵ that we'll decide later.

The second step is cleaning. The cleaning step, again, it's the same kind of cleaning as we've done before. So let's remove an edge from V_i cross V_j , if any of the following hold-- if V_i V_j is not ϵ regular, if the density is too small, or either the two sets is too small. So same cleaning as before. And we can check that the number of edges removed is not too small.

So in the first case, so again, it's the same calculation as last time. In the first case, the

number of edges removed is, at most, ϵN^2 . And we'll choose ϵ to be less than $\epsilon/8$, although it will be actually significantly smaller, as you will see in a second. In the second step, same as what happened in the triangle removal stage, the number of edges removed in the second type is, at most, that amount, still very small number.

And the third one here is also a very small number. So the third type, they start with one of these sets, the m possible-- so it's a very small number. And so the total is, at most, an $\epsilon/2 N^2$. So maybe I want $\epsilon/2$ -- so let's say $\epsilon/2 N^2$ number of edges.

And I would like that to be strictly bigger than. So now after removing these edges from G , we have this G' , which has strictly more than $(1 - \epsilon/2) N^2$ edges. So now what do we do?

So we knew from Turan's theorem that if your graph has strictly more than this number of edges, you must have a K_{r+1} . So even after deleting all these edges, G' still has lots of edges left, in particular, Turan's theorem implies that G' contains a clique on $r+1$ vertices. So here I should say that r is chromatic number of H minus 1.

So I find one copy of this clique, but what does that copy look like? I find this one copy of a clique. Let's say r equals to 4. And the point, now, is that the counting lemma will allow me to amplify that clique into H . So it will allow me to amplify this clique into a copy of H . So, for example, if H were this graph over here, so then you would find a copy of H in G , which is what we want.

So why does the counting lemma allow you to do this amplification? So it's this point, the ideas are all there, but there's a slight wrinkle in the calculations. I mean, there's, in the executions, I just want to point out, just in case some of the vertices of H end up in the same vertex in G . But that turns out not to be an issue. So by counting lemma, the number of homomorphisms from H to G' , where I'm really only considering homomorphisms that map each vertex of H to its assigned part.

It's at least this quantity where I'm looking at the predicted density of such homomorphisms, and all of these edge densities are at least $\epsilon/8$. So it's at least that amount minus a small error that comes from the counting lemma. And, well, all of the vertex parts are quite large, so all of the vertex parts are of size like that. So that's the result of the counting lemma combined with information about the densities of the parts and the sizes of the parts that came

out of cleaning.

So setting ϵ to be an appropriate value, we see that for sufficiently large N , this quantity here, is on the order of N to the number of vertices of H . But I'm only counting homomorphisms, and so it could be that some of the vertices of H end up in the same vertex of G . And those would not be genuine subgraphs, so I shouldn't consider those as subgraphs. Because otherwise, if you were to allow those, then if you found this K_4 , then you found all four chromatic graphs.

So you shouldn't consider copies that are degenerate, but that's OK because the number of maps from the vertex set of H to the vertex set of G that are non-injective is of a lower order. The number of non-injective maps from the vertex set, well, you have to pick two vertices of H to map to the same vertex, and then the number of choices, you have one order less. So there are negligible fraction of these homomorphisms.

And the conclusion, then, is that G prime contains a copy of H , which is what we're looking for. If G prime contains copy of H , then G contains a copy of H , and that proves the Erdos-Stone-Simonovits theorem. You get a bit more out of this proof. So you see that not only does G contain one copy of H , but the counting lemma actually shows you it contains many copies of H .

And this is a phenomenon known as supersaturation, which you already saw in the first problem set, that often when you are beyond a certain threshold, an extremal threshold, you don't just gain one extra copy, but you often gain many copies. And you see this in this proof here. So to summarize, we've seen this proof of Erdos-Stone-Simonovits, which comes from applying regularity and then finding a single copy of a clique from Turan's theorem, and then using counting lemma to boost that copy from Turan's theorem into an actual copy of H .

So in the second homework, one of the problems is to come up with a different proof of Erdos-Stone-Simonovits that is more similar to the proof of Kovari-Sos-Turan, more through double-counting like arguments. And that is closer in spirit, although not exactly the same as the original proof in Erdos-Stone. So this regularity proof, I think it's more conceptual. You get to see how to do this boosting, but it gets a terrible bound. And the other proof that you see in the homework gives you a much more reasonable bound on the dependence between how N grows versus how quickly this little ϵ has to go to zero.