

**PROFESSOR:** I sent out a survey this morning about how the class is going, what you thought of the problem set. And I would appreciate if you provide me some feedback-- so things you like or don't like about the class or about the problem set that was just due last night. So I can try to adjust to make it more interesting and useful for all of you.

Last time we talked about Szemerédi's graph regularity lemma. So the regularity lemma, as I mentioned, is an extremely powerful tool in modern combinatorics. And last time we saw the statement and the proof of this regularity lemma. Today, I want to show you how to apply the lemma for extremal applications. In particular, we'll see how to prove Roth's theorem that I mentioned in the very first lecture, about subsets of integers lacking three-term arithmetic progressions.

First, let me remind you the regularity lemma. We're always working inside some graph,  $G$ . We say that a pair of subsets of vertices is  $\epsilon$  regular if the following holds-- for all subsets  $A$  of  $X$ ,  $B$  of  $Y$ , neither too small, we have that the edge density between  $A$  and  $B$  is very similar to the edge density between the ambient sets  $X$  and  $Y$ .

So we had this picture from last time. You have two sets. Now, they don't have to be disjoint. They could even be the same set. But for illustration purposes, it's easier to visualize what's going on if I draw them as disjoint subsets.

So there is some edge density. And I say they're  $\epsilon$  regular if they behave random-like in the following sense-- that the edges are somehow distributed in a fairly uniform way so that if I look at some smaller subsets  $A$  and  $B$ , but not too small, then the edge densities between  $A$  and  $B$  is very similar to the ambient edge densities. So by most,  $\epsilon$  difference.

Now I need that  $A$  and  $B$  are not too small because if you allow to take, for example, single vertices, you can easily get densities that are either 0 or 1. So then it's very hard to make any useful statement. So that's why these two conditions are needed.

And here, the edge density is defined to be the number of edges with one endpoint in  $A$ , one endpoint in  $B$ , divided by the product of the sizes of  $A$  and  $B$ . And we say that a partition of the vertex set of the graph is  $\epsilon$  regular if, by summing over all pairs  $i, j$ , such that  $v_i, v_j$  is not  $\epsilon$  regular if we sum up the product of these part sizes, then this sum is at most an  $\epsilon$  fraction of the total number of pairs of vertices.

And the way to think of this is that there are not too many irregular parts. At least in the case when all the parts are equitable. So we should really think about all of them having more or less the same size than saying that at most an epsilon fraction of them are irregular.

And the main theorem from last time was Szemerédi's regularity lemma. And the statement is that for every epsilon, there exists some  $M$ -- so  $M$  depends only on epsilon and not on the graph that we're about to see-- such that every graph has an epsilon regular partition into at most  $M$  parts.

In particular, the number of parts does not depend on the graph. For every epsilon, there is some  $M$ . And no matter how large the graph, there exists a bounded size partition that is epsilon regular.

So the proof last time gave us a bound  $M$  that is quite large as a function of epsilon. So the last time we saw that this  $M$  was a tower of twos of height essentially polynomial in  $1/\epsilon$ . And I mentioned that you basically cannot improve this bound. So this bound is more or less the best possible up to maybe changing the 5.

And so in some sense, the proof that we gave last time for Szemerédi's graph regularity lemma was the right proof. So that was the sequence of steps that were the right things to do. Even though they give a terrible bound, it's somehow the bound that should come out.

What I want to talk about today is, what's a regularity partition good for? So we did all this work to get a regularity partition, and it has all of these nice definitions. But they are useful for something. So what is it useful for?

And here is the intuition. Remember at the beginning of last lecture I mentioned this informal statement of regularity lemma-- namely that there exists a partition of the graph so that most pairs look random-like.

So what does random-like mean? So random-like, there is a specific definition. But the intuition is that in many aspects, especially when it comes to counting small patterns, the graph in the random-like setting looks very similar to what happens in a random graph-- in a genuine random graph.

In particular, if you have three subsets--  $x$ ,  $y$ , and  $z$ -- and suppose that the three pairs are all epsilon regular, then you might be interested in the number of triangles with one vertex in

each set.

Now, if this were a genuine random tripartite graph with specified edge densities, then the number of triangles in such a random graph is pretty easy to calculate. You would expect that it is around the product of the sizes of these vertex sets multiplied by their edge densities.

And what we will see is that in this case, in that of an epsilon regular setting, this is also a true statement. It's a true, deterministic statement. That's one of the consequences of epsilon regularity. Yes, question?

**AUDIENCE:** Why are we only multiplying the sizes [INAUDIBLE]?

**PROFESSOR:** Asking, why are we only multiplying the sizes of  $x$ ,  $y$ , and  $z$ ? So you're asking-- OK. So we're trying to find out how many triangles are there with one vertex in  $x$ , one vertex in  $y$ , and one vertex in  $z$ . So if I put these vertices in there, one by one, then if this were a random graph, I expect that pair to be an edge with probability  $d_{xy}$  and so on.

So if all the edge densities were one half, then I expect one eighth of these triples to be actual triangles. And what we're saying is that in an epsilon regular setting, that is approximately a true statement. So let me formalize this intuition into an actual statement.

And this type of statements are known as counting lemmas in literature. And in particular, let's look at the triangle counting lemma. In the triangle counting lemma-- so we're using the same picture over there-- I have three vertex subsets of some given graph. Again, they don't have to be disjoint. They could overlap, but it's fine to think about that picture over there.

And suppose that these three pairs of subsets-- so these three subsets-- they are mutually epsilon regular. Then, for abbreviation, let me write the sub  $xy$  to be the edge density between  $x$  and  $y$ , and so on for the other two pairs.

The conclusion is that the number of triangles-- where I'm looking at triangles and only counting triangles with one specified vertex in  $x$ , one in  $y$ , and one in  $z$ -- is at least some quantity. So there is a small potential error loss but otherwise the product, as I mentioned earlier.

So it is at least this quantity I mentioned earlier up to a potential small error because we're looking at epsilon regularity. So there could be some fluctuations in both directions. A similar statement is also true as an upper bound. But the lower bound will be more useful, so I will

show you the proof of the lower bound. But you can figure out how to do the upper bound. And later on we'll see a general proof what happens, instead of triangles, if you have other subgraphs that you wish to count.

So here's the intuition. So you have a random-like setting, and we'll formalize it in the setting of epsilon regular pairs. Yeah?

**AUDIENCE:** Where does the  $1 - 2\epsilon$  come from?

**PROFESSOR:** OK. The question is, where does  $1 - 2\epsilon$  come from? You'll see in the proof. But you should think of this as essentially a negligible factor. Any more questions? All right.

So here's how this proof is going to go. Let's look at  $x$  and think about its relationship to  $y$ . It's epsilon regular. And they claim, as a result of them being epsilon regular, fewer than epsilon fraction of  $x$ . So fewer than this many vertices in  $x$  have very small number of neighbors in  $y$ .

Because if this were not the case, then you can violate the condition of absolute regularity. So if not, then let's look at this subset, which has size at least epsilon  $x$ . And all of them have fewer than that number of neighbors to  $y$ . So these two sets-- so this set,  $x'$  and  $y$ , would witness non-epsilon regularity.

So you cannot have too many vertices with small degrees going to  $x$ -- going to  $y$ . OK. Great. And likewise, fewer than epsilon  $x$  vertices have a small number of neighbors to  $z$ .

So what does the picture now look like? So you have this  $x$  and then these two other sets,  $y$  and  $z$  where I'm going to throw out a small proportion of  $x$  less than  $2\epsilon$  fraction of  $x$  that have the wrong kinds of degrees. And everything else in here have lots of neighbors in both  $y$  and in  $z$ .

And in particular, for all  $x$  up here it has lots of neighbors to  $y$ , lots of neighbors to  $z$ . How many? Well, we have at least  $d_{xy} - \epsilon y$  neighbors to  $y$  and at least  $d_{xz} - \epsilon z$  neighbors to  $z$ .

OK. So now I realize I'm missing a hypothesis in the counting lemma. Let me assume that none of these edge densities are too small. They're all at least  $2\epsilon$ .

So now these guys are at least epsilon fractions of  $y$  and  $z$ . So I can apply the definition of epsilon regularity to the pair  $yz$  to deduce that there are lots of edges between these two sets.

So over here, the number of edges is at least the products of the sizes multiplied by the edge density between them. And by the definition of epsilon regularity, the edge density between these two small or these two red sets is at least  $d$  of  $yz$  minus epsilon.

So putting everything now together, we find that the total number of triangles, looking at all the possible places where  $x$  can go-- so at least  $1 - 2\epsilon$  times the size of  $x$ . And then multiply by this factor over here. And so we find the statement up there.

So this calculation formalizes the intuition that if you have epsilon regular pairs, then they behave like random settings when it comes to counting small patterns-- namely that of a triangle. So what can we use this for? The next statement I want to show you is called a triangle removal lemma.

So this is a somewhat innocuous looking statement that is surprisingly tricky to prove. And part of the development of this regularity lemma was to prove Szemerédi's-- to the triangle removal lemma. This was one of the early applications of the regularity lemma. So it's due to Ruzsa and Szemerédi back in the '70s.

Here's the statement. For every epsilon there exists a delta, such that every graph of  $n$  vertices with a small number of triangles-- so a small number of triangles means a negligible fraction of all the possible triples of vertices are actual triangles. So fewer than  $\delta n^3$  triangles.

So if you have a graph with a small number of triangles, the question is, can you make it triangle free by getting rid of a small number of edges? So actually, there was already a problem on the first homework set that is in that spirit.

So if you compare what I'm doing here to the homework set, you'll see that there are different scales. So fewer than  $\delta n^3$  triangles can be made triangle free by removing  $\epsilon n^2$  edges.

So if you have a small number of triangles, you can get rid of all the triangles by removing a small number of edges. If I put it that way, it actually sounds kind of trivial. You just get rid of all the triangles. But if you look at the scales it's not trivial at all, because there are only a subcubic number of triangles.

So if you take out one edge from each triangle, maybe you got rid of a lot of edges. So this is a

very innocent looking statement, but it's actually incredibly deep and tricky. Before jumping to the proof, let me first show you an equivalent reformulation of the statement that also helps you to think about what this statement is trying to say.

So the triangle removal lemma can be equivalently stated as saying that every  $n$  vertex graph with a subcubic number of triangles-- so little  $o$  of  $n$  cubed triangles-- can be made triangle free by removing a subquadratic-- namely, little  $o$  of  $n$  squared-- number of edges.

So this is an equivalent statement to what I wrote above, although it actually takes some thought to figure out what this is even saying because everybody loves using asymptotic notation, but there is also ambiguity with, what do you mean by asymptotic notation, especially if it appears in the hypothesis of a claim?

So what do you think this statement means? Can you write out more of a full form? I think of this as a lazy version of trying to say something. So what do you mean by having little  $o$  of  $n$  cubed triangles? Yes.

**AUDIENCE:** The sequence of the graph. [INAUDIBLE].

**AUDIENCE:** [INAUDIBLE] function has  $n$  and only  $n$ . [INAUDIBLE].

**PROFESSOR:** OK. Great. So I have a sequence of graphs. And also, we can put some functions in. So I'll write down the statement here, but that's kind of the idea. We're looking at not just a single graph, but we're looking at a sequence.

Another way to say this is that for every function  $f_n$ , that is subcubic. So for example, if  $f$  of  $n$  is  $n$  cubed divided about  $\log n$ , there exists some function  $g$ , which is subquadratic, such that if you replace the first one by  $f$  of  $n$  and the second thing by  $g$  of  $n$ , then this is a true statement. And I'll leave it to you as an exercise in quantified elimination, let's say, to explain why these two statements are equivalent to each other.

I want to explain a recipe for applying Szemerédi's regularity lemma. How does one use the regularity lemma to prove, well, statements in graph theory? The most standard applications of regularity lemma generally have the following steps.

Let me call this a recipe. And we'll see it a few times. The first step is we apply Szemerédi's regularity lemma to obtain a partition. So let me call the first step partition.

In the second step, we look at the partition that we obtained, and we clean it up. So in the partition, you have some irregular pairs that are undesirable to work with. And there are some other pairs that we'll see. So in particular, if your pair involves edges that are fairly sparse or subsets of vertices that are fairly small, then maybe we don't want to touch them because they're kind of not so good to deal with.

So we're going to clean the graph by removing edges in irregular pairs and low density pairs. And unless you're using the version of regularity lemma that allows you to have equitable parts, you also want to get rid of edges where one of the parts is too small.

And the third step, I'll call this count. Once you've cleaned up the regularity partition, say, well, let's try to find some patterns. If you find one pattern in the cleaned graph-- and we can use the counting lemma to find lots of patterns. Here, for the purpose of triangle removal lemma and what we've been doing so far, pattern just means a triangle.

So we're going to use the triangle counting lemma to find us lots of triangles. So we'll see the details in a bit. But if we run through the strategy-- you give me a graph. Let's say, starting from the triangle removal lemma, it has a small number of triangles. You apply the partition, clean it up, and I claim this cleaning removes a small number of edges.

And it should result in a triangle free graph because if it did not result in a triangle free graph, then there's some triangle. And from that triangle I can apply the triangle counting lemma to get lots of triangles. And that would violate the hypothesis of the triangle removal lemma. So that's how the proof is going to go.

So I want to take a very quick break. And then when we come back, we'll see the details of how to apply the irregularity lemma. Are there any questions so far? Yeah?

**AUDIENCE:** So when we're removing edges in one of the [INAUDIBLE], is that too small? Can we do that for every vertex, or is it too small?

**PROFESSOR:** So you're asking about what happens when we remove vertexes that are too small. You will see in the details of the proof. So hold on to that question for a bit. More questions.

OK. So let's see the proof of the triangle removal lemma. So the first step is to apply Szemerédi's regularity lemma and find a partition. So we'll find a partition that's  $\epsilon$  over  $4$  regular. So here,  $\epsilon$  is the same  $\epsilon$  in the statement-- in the top statement-- of the triangle removal lemma.

In the second step, let's clean the graph by removing all edges in-- so we are going to get rid of edges between-- OK. So let me do it this way. So all edges between the  $v_i$  and the  $v_j$  whenever  $v_i$  and  $v_j$  is not  $\epsilon$  regular. Get rid of the edges between irregular parts.

**AUDIENCE:**  $\epsilon$  over 4 regular.

**PROFESSOR:** Sorry?

**AUDIENCE:**  $\epsilon$  over 4 regular.

**PROFESSOR:**  $\epsilon$  over 4 regular. Thank you. Also, between parts where the edge density is too small-- if the edge density is less than  $\epsilon/2$ , get rid of those edges. And if one of the two vertex sets has size too small-- and here, too small means  $\epsilon/4M$  times the size of  $n$ .

So here-- OK. So let me use big  $M$  for the number of parts. So that's the  $M$  that comes out of Szemerédi's regularity lemma. If you like, some of the vertex sets can be empty. It doesn't change the proof. And  $n$  is the number of vertices in the graph.

And this step, you don't really need the step if your regular partition is equitable. So let's see how many vertices-- how many edges have we gotten rid of. We want to show that we're not deleting too many edges.

In the first step-- so the number of deleted edges. In the first step, you see that the number of edges deleted is at most the sum of product of  $v_i v_j$  when you sum over  $ij$  such that this pair is not  $\epsilon$  regular or  $\epsilon/4$  regular.  $\epsilon$  over 4 regular. By the definition of an  $\epsilon$  regular partition, the sum here is at most  $\epsilon/4$  times  $n$  squared.

In the second step, I'm getting rid of low density pairs. By the virtue of them being low density, I'm not removing so many edges. So at most  $\epsilon/2$  times  $n$  squared edges I'm getting rid of.

In the third part, you see every time I take a very small piece, every vertex here is adjacent to at most  $n$  vertices. So the number of such things, such edges I'm getting rid of in the last step is at most this number times a number of parts  $M$  then times  $n$ . So it's at most  $\epsilon/4$  times  $n$  squared.

So here I'm telling you how many edges I've deleted in each step. And in total, putting them together, we see that we get rid of at most  $\epsilon n$  squared edges from this graph. So that's

the cleaning step. So we cleaned up the graph by getting rid of low density pairs, getting rid of irregular pairs, and small vertex sets.

Now suppose, after this cleaning, some triangles still remain. So we're now onto the third step. So suppose some triangle remains. So where could this triangle sit? Has to be between three parts--  $v_i$ ,  $v_j$ , and  $v_k$ .  $i$ ,  $j$ , and  $k$ , they don't have to be distinct. So the argument will be OK if some of them are the same, but it's easier to draw if they're all different.

So I have some triangle, like that. Because these edges have not yet been deleted in the cleaning step, I know that the vertex sets are not too small, the edge densities are not too small, and they are all regular with each other. So here, each pair in  $v_i$ ,  $v_j$ ,  $v_k$  is  $\epsilon$  over 4 regular and have edge density at least  $\epsilon$  over 2.

And now we apply the triangle counting lemma, and we find that the number of triangles with one vertex in  $v_i$ , one vertex in  $v_j$ , one vertex in the  $v_k$  is at least this quantity here. So that's a correction factor. So  $1$  over this  $2\epsilon$ .

And then a bunch of densities-- so densities are not too small. So I have at least  $\epsilon$  over 4  $n$  cubed multiplied by the sizes of the vertex sets. Now I know that-- use the fact that these part sizes are not too small. So I have that.

Just in case, if  $i$ ,  $j$ , and  $k$  happen to be the same, or two of them happen to be the same, I might overcount the number of triangles a little bit. But at most, you overcount by a factor of 6. So that's OK. So if you're worried about that, put the  $1$  over 6 factor in, just in case  $i$ ,  $j$ ,  $k$  not distinct.

Or if you like, in the cleaning step, you can-- if you apply the equitable version of the regularity lemma, you can also get rid of edges inside the parts. But there are many ways to do this. It's not an important step.

Now, this quantity, let me set it to be  $\delta$ . You see,  $\delta$  is a function of  $\epsilon$  because  $M$  is a function of  $\epsilon$ . So now, looking back at the statement, you see for every  $\epsilon$  there exists a  $\delta$ , such that if your graph has fewer than  $\delta n^3$  triangles, then let me get rid of all those edges. I've gotten rid of fewer than  $\epsilon n^2$  edges, and the remaining graph should be triangle free.

Because if it were not triangle free, then I can find some triangle. And that will lead to a lot

more triangles. So for example, if you set this as  $\delta$  over 2, then this will give you  $2\delta n^3$  triangles. Therefore, it would contradict the hypothesis.

And that finishes the proof of the triangle removal lemma, saying that thus the resulting graph is triangle free. So that's the proof of the triangle removal lemma. So let me recap.

We start with a graph, apply Szemerédi's regularity lemma, and clean up the regularity partition by getting rid of low density pairs, getting rid of irregular pairs, and getting rid of edges touching a very small vertex set. And I claim that the resulting graph, after cleaning up, should be triangle free.

Because if it were not triangle free and I find some triangle, then I should be able to use that triple of vertex sets, combined with a triangle counting lemma, to produce a lot more triangles and. That would violate the hypothesis of the theorem.

Any questions? Yeah.

**AUDIENCE:** Where are you using that there exists a triangle?

**PROFESSOR:** Ah, great. So question is, where am I using there exists a triangle? If there were no triangles, then we're done. So the purpose of the triangle-- the claim in the triangle removal lemma is that you can get rid of all triangles by removing at most  $\epsilon n^2$  edges.

**AUDIENCE:** So say we did that, and now-- why does this not prove that we still have triangles?

**PROFESSOR:** Can you say your question again?

**AUDIENCE:** So say we've removed everything by our cleaning step, and we've removed  $\epsilon n^2$  edges, why does this logic not prove that we still have  $\delta n^3$  triangles.

**PROFESSOR:** OK. So let me try to answer your question. So why does this proof show that you still have  $\delta n^3$  triangles? So I only set  $\delta$  at the end. But of course, you can also set  $\delta$  in the beginning of this proof. So I'm saying that you do the step. You get rid of  $\epsilon n^2$  edges. And now I claim, after the step-- so I claim the remaining graph is triangle free.

If it were not triangle free, then, well, it has some triangle. Then the triangle counting lemma would tell me there are lots of triangles. And that would contradict the hypothesis where we assume that this graph  $G$  has a small number of triangles.

**AUDIENCE:** So if there is no triangle, then we've removed edges between  $v_i, v_j$ , or  $v_i, v_k$ , or  $v_j, v_k$  for any three  $i, j, k$ .

**PROFESSOR:** That's correct. So we're saying, if you do not have any triangles-- well, after the cleaning step, we have gotten rid of all the edges between the bad pairs. And I'm claiming that there is no configuration like this left. And this is the proof because if you have some configuration where you did not delete the edges between these three parts, then you should be able to get a lot more triangles from the triangle counting lemma. Yeah.

**AUDIENCE:** What if there were lots of triangles inside each individual  $v_i, v_j, v_k$ ?

**PROFESSOR:** You asked me, what happens if there were a lot of triangles inside each  $v_i, v_j, v_k$ ? So that is fine. If you find some triangle-- so this picture,  $i, j$ , or  $k$ , they do not have to be distinct. So the same proof works if  $i, j$ , and  $k$ , some of them are equal to each other. Yep.

**AUDIENCE:** [INAUDIBLE] but, I don't really understand why-- isn't  $\delta > 2$  there?

**PROFESSOR:** So you're asking, why did I put the  $\delta > 2$ ? Just because I put less than or equal to  $\delta$ . If I put strictly less than  $\delta$ , then I don't need a  $\delta > 2$ .

**AUDIENCE:** [INAUDIBLE]  $\delta > 2$  or  $2\delta$ .

**PROFESSOR:** OK. Don't worry about it. Yes.

**AUDIENCE:** Is there a way to generalize the triangle counting lemma to a general graph?

**PROFESSOR:** OK. You're asking, is there a way to generalize the triangle counting lemma to a general graph? So yes. We will see that not today but I think next time. Any more questions? Great.

So why do people care about the triangle removal lemma? So it's a nice, maybe somewhat unintuitive statement. But there was a very good reason why the statement was formulated, and it's because you can use it to prove Roth's theorem. So that's what I want to explain, how to connect this graph theoretic statement to a statement about three-term AP-- three AP-free subsets of the integers.

This goes back to the very connection between graph theory and additive combinatorics that I highlighted in the first lecture. First, let me state a corollary of the triangle removal lemma-- namely, that if you have an  $n$  vertex graph  $G$ , where-- so if  $G$  is  $n$  vertex, and every edge is in exactly one triangle, then the number of edges of  $G$  is little  $o$  of  $n^2$ .

These are actually kind of strange graphs. Every edge is in exactly one triangle. OK. Well, the number of triangles in  $G$ -- every edge is in exactly one triangle. So the number of triangles in  $G$  is the number of edges divided by 3. The number of edges is at most  $n^2$ .

So this quantity is at most quadratic order, which in particular is little  $o$  of  $n^3$ . And thus the triangle removal lemma tells us that  $G$  can be made triangle free by removing little  $o$  of  $n^2$  edges.

On the other hand, since every edge is in exactly one triangle, well, how many edges do you need to remove to get rid of all the triangles? Well, I need to remove at least a third of the edges. I need to remove at least a third of edges to make  $G$  triangle free. Putting these two claims together, we see that the number of edges of  $G$  must be little  $o$  of  $n^2$ .

Any questions?

**AUDIENCE:** Are there not more elementary ways to prove this?

**PROFESSOR:** Great. Question is, are there not more elementary ways to prove this? Let me make some comments about that. So the short answer is, yes but not really. And really, the answer is no.

[LAUGHTER]

So you can ask, what about quantitative bounds? Because what is more elementary, what is less elementary is kind of subjective. But quantitative bounds, something that is very concrete. It's hard to argue.

So if you look at the triangle removal lemma, you can ask, how is the dependence of  $\delta$  on  $\epsilon$ ? So what does the proof give you? Where's the bottleneck? The bottleneck is always in the application of Szemerédi's regularity lemma-- namely in this  $M$ . So none of the other  $\epsilon$ 's really matter. It's this  $M$  that kills you in terms of quantitative bounds.

So in triangle removal lemma, this proof gives  $1/\delta$ . So you can take  $1/\delta$  being a tower of twos of height at most polynomial in  $1/\epsilon$ . So that is your different proof.

Well, the best known bound due to Fox is that you can replace this height by a different height that is at most essentially logarithmic in  $1/\epsilon$ . Still a tower of twos. So we've changed some really big number to another, but slightly smaller, really big number. So this is still an astronomical number for any reasonable  $\epsilon$ .

And in terms of that corollary, basically the only known proof goes through the triangle removal lemma. Currently, we do not know any other approach to this problem. And you'll see later on that, well, what's the best thing that we can hope for? So it is quite possible that there are other proofs that are yet to be found.

So that's actually-- people believe this, that this is not the right proof, that maybe there's some other way to do this. And the best lower bound, which we'll see either later today or next time, shows that we cannot do better than  $1/\epsilon$  being essentially just a little bit more than polynomial in  $\epsilon$ .

So  $\epsilon$  raised to something that is logarithmic in  $1/\epsilon$ . So you can think of this as very-- it's a little bit bigger than polynomial in  $1/\epsilon$  but not that much bigger than polynomial in  $1/\epsilon$ . So there is a very big gap in our knowledge on what is the right dependence between  $\epsilon$  and  $\delta$  in the triangle removal lemma. And that's one of the-- it's a major open problem in extremal combinatorics to close this gap.

Other questions? All right. So let's prove Roth's theorem. So let me remind you that Roth's theorem, which we saw in the very first lecture, says that if you have a subset of  $1$  through  $n$  that is free of three-term arithmetic progressions, then the size of the set must be sublinear.

So what does this have to do with a triangle removal lemma? So if you remember the first lecture, maybe the connection shouldn't be so surprising. What we will do is we will set up a graph, starting from some arithmetic sets such that the graph encodes some arithmetic information-- in particular, the three-term APs in your graph, in the set, correspond to the triangles in the graph.

So let's set up this graph. It will be helpful to view  $A$  not as a subset of the integers. It'll just be more convenient to view it as a subset of a cyclic group. Because I don't have to worry about edge cases so much when you're working a cyclic group. Here I take  $M$  to be  $2N + 1$ . So having it odd makes my life a bit simpler.

Then if  $A$  is three AP free subset of  $1$  through  $n$ , then I claim that  $A$  now sitting inside this cyclic group is also three AP free. So it's a subset of  $\mathbb{Z} \bmod n$ . And what we will do is that we will set up a certain graph. So we will set up a tripartite graph,  $x, y, z$ .

And here,  $x, y$ , and  $z$  are going to be  $M$  elements whose vertices are represented by elements

of  $Z \text{ mod } n$ . And I need to tell you what are the edges of this graph. So here are the edges. I'm putting an edge between vertex  $x$  and  $y$  if and only if  $y - x$  is an element of  $A$ .

So it's a rule for how to put in the edges. And this is basically a Cayley graph, a bipartite variant of a Cayley graph. Likewise, I put an edge between  $x$  and  $z$ . So let me put  $x$  down here and  $y$  up there. So let me put in the edge between  $y$  and  $z$  if and only if  $z - y$  is an element of  $A$ .

And for the very last pair, it's similar but slightly different. I'm putting that edge if and only if  $z - x$  divided by 2 is an element of  $A$ . Because we're in an odd cyclic group I can divide by 2.

So this is a graph. So starting with a set  $A$  I give you this rule for constructing this tripartite graph. And the question now is, what are the triangles in this graph? If the vertices  $x, y, z$  is a triangle, then these three numbers by definition, because of the edges-- because they're all edges in this graph, these three numbers, they all lie in  $A$ .

But now notice that these three numbers, they form a three-term arithmetic progression because the middle element is the average of the two others. But we said that  $A$  is a set that is three AP free. Has no three-term arithmetic progression. So what must be the case?

So  $A$  is 3 AP free. But you can still have three APs using the same element three times. So all the three-term arithmetic progressions must be of that form. So these three numbers must then equal to each other. And in particular, you see that if you select  $x$  and  $y$ , it determines  $z$ . This equality here is the same as saying that  $x, y$ , and  $z$  they themselves form a three AP in  $Z \text{ mod } n$ .

So this is precisely the description of all the triangles in the graph. So all the triangles in the graph  $G$  are precisely  $x, y, z$ , where  $x, y$ , and  $z$  form a three-term arithmetic progression. And in particular, every edge of  $G$  lies in exactly one triangle. You give me an edge-- for example,  $xy$ -- I complete it to a three AP,  $x, y, z$ . And that's the triangle.

And that's the unique triangle that the edge sits in. And likewise, if you give me  $xz$  or  $yz$ , I can produce for you a unique triangle. So we have this graph. It has this property that every edge lies in exactly one triangle, so we can apply the corollary up there to deduce a bound on the total number of edges.

Well, how many edges are there? On one hand, we see that because it's a Cayley graph,

each of the three parts-- there are three parts here. Each of the three parts, if I start with any vertex, I have  $A$  edges coming out of that vertex to the next part by the construction.

On the other hand, by the corollary up there, the number of edges has to be little  $o$  of  $M$  squared. And because  $M$  is essentially twice  $n$ , we obtain that the size of  $A$  is little  $o$  of  $M$ .

And that proves Roth's theorem. Yeah?

**AUDIENCE:** Could you explain one more time why every edge is in exactly one triangle?

**PROFESSOR:** OK. So the question is, why is every edge in exactly one triangle? So you know what all the edges are. So this is a description of what all the edges are. And what are all the triangles. Well,  $x, y, z$  is a triangle precisely when these three expressions all lie in  $A$ . But note that these three expressions, they form a three AP because the middle term is the average of the two others.

So  $x, y, z$  is the triangle if and only if this equation is true. And this equation is true if and only if  $x, y, z$  form a three AP in  $\mathbb{Z} \bmod n$ . So if you just read out this equation, I give you  $x$  and  $y$ . So what is  $z$ ? So all the triangles in  $x, y, z$  are precisely given by three APs, where one of the differences  $y$  minus  $x$  is in  $A$ .

OK. So I give you an edge. For example,  $xy$ , such that  $y$  minus  $x$  is in  $A$ . And I claim there's a unique  $z$  that completes this edge to a triangle. Well, it tells you what that  $z$  is.  $z$  has to be the element in  $\mathbb{Z} \bmod m$  that completes  $x$  and  $y$  to a three AP. Namely,  $z$  is the solution to this equation.

No other  $z$  can work. And you can check that  $z$  indeed works and that all the remaining pairs are edges. So it's something you can check. Any more questions?

So starting with the set  $A$  that is three AP free, we set up this graph with a property that every edge lies in exactly one triangle. And the one triangle basically corresponds to the fact that you always have these trivial three APs repeating the same element three times. And then, by applying this corollary of the triangle removal lemma, we deduce that the number of edges in the graph must be subquadratic. So then the size of  $A$  must be sublinear. And that proves Roth's theorem.

So we did quite a bit of work in proving this theorem-- Szemerédi's regularity lemma, counting lemma, removal lemma, and then we set up this graph. So it's not an easy theorem. Later in

the course, we'll see a different proof of Roth's theorem that goes through Fourier analysis. That will look somewhat different, but it will have similar themes.

So we'll also have this theme comparing structure and pseudorandomness, which comes up in the proof-- in the statement and proof of Szemerédi's graph regularity lemma. So there, it's really about understanding what is the structure of the graph in terms of decomposition into parts that look pseudorandom. Yeah.

**AUDIENCE:** You called the graph the Cayley graph. Why?

**PROFESSOR:** OK. So question is, why do I call this graph the Cayley graph? So usually the Cayley graph refers to a graph where I give you a group, and I give you a subset of the group, and I connect two elements if, let's say, their difference lies in my subset. This basically has that form. So it's not exactly what people mean by a Cayley graph, but it has that spirit. Any more questions?

OK. So earlier I talked about bounds for triangle removal lemma. So what about bounds for Roth's theorem? We do know somewhat better bounds for Roth's theorem compared to this proof. Somehow it's a nice proof, it's a nice graph, theoretic proof, but it doesn't give you very good bounds. It gives you bounds that decay very poorly as a function of  $n$ .

Actually, what does it give you as a function of  $n$ ? If you were to replace this little  $o$  by a function of  $n$  according to this proof, what would you get? I'm basically asking, what is the inverse of the function where you input some number and it gives you a tower of exponentials of height with that input?

It's called a log star. So the log star-- so this is essentially  $N$  over the log star of  $N$ . So the log star basically is the number of times you have to take the logarithm to get you below 1. So that's the log star.

And there's a saying that the log star, we know that it grows to infinity, but it has never been observed to do so. It's extremely slowly growing function. Any more questions?

So I want to-- so next time I want to show you a construction that gives you a-- so next time I will show you a construction that gives you a subset  $A$  of  $n$  that is fairly large. So you might ask, OK, so you have this upper bound, but what should the truth be? And here's more or less the state of knowledge.

So best bounds of Roth's theorem. Basically, the best bounds have the form  $N$  divided by

basically  $\log N$  raised to power  $1 + \text{little } o(1)$ . The precise bounds are of the form  $N$  over  $\log N$ , and then there's some extra log-log factors. But let's not worry about that.

The best lower bounds-- so we'll see this next time. So there exists subsets of  $1$  through  $N$  such that the size of  $A$  is at least  $e$  to the-- so  $N$  times-- so first, let me say it's pretty close to-- the exponent is as close to  $1$  as you wish.

So there exists as  $A$  such that the size of  $A$  is  $N$  to the  $1$  minus little  $o(1)$ . And already, this fact is an indication of the difficulty of the problem because if you could prove Roth's theorem through some fairly elementary techniques, like using a Cauchy-Schwarz a bunch of times for instance, then experience tells us that you probably expect some bound that's power saving, replacing the  $1$  by some smaller number.

But that's not the case. And the fact that that's not the case is already indication of the difficulty of this upper bound of Roth's theorem, even getting a little  $o$ . So you don't expect there to be simple proofs getting the little  $o$ . The bound that we'll see next time-- so we'll see a construction which gives you a bound that is of this form.

So it's maybe a little bit hard to think about how quickly this function grows, but I'll let you think about it. Now, how does this-- so let's look at this corollary here. Can you see a way to construct a graph which has lots of edges, such that every edge lies in exactly one triangle?

So we did this connection showing how to use this corollary to prove Roth's theorem. But you can run the same connection. So starting from this three AP free  $A$ , we can use that construction to build a graph  $n$ , such that a graph of  $n$  vertices with essentially order of  $n$  times the size of  $A$  number of edges, such that every edge lies in exactly one triangle.

So you run the same construction. And this is actually more or less the only way that we know how to construct such graphs that are fairly dense. So on one hand-- basically what I said earlier. On one hand, you have this upper bound, which is given by the proof of using Szemerédi's regularity lemma that gives you a tower in the upper bound of  $1$  over  $\delta$ .

And if you use this construction here of three AP free set to construct the graph, you get this lower bound on  $\delta$ , which is quasipolynomial. And that's more or less that we know. And there's a major open problem to close these two gaps.

Any more questions? So I want to give you a plan on what's coming up ahead. So today we

saw one application of Szemerédi's regularity lemma-- namely, the triangle removal lemma, which has this application to Roth's theorem. So we've seen our first proof of Roth's theorem. And next lecture, and the next couple lectures, I want to show you a few extensions and applications of Szemerédi's regularity lemma.

So one of the questions today was, we knew how to count the triangles, but what about other graphs? And as you can imagine, if you can count triangles, then the other graphs should also be doable using the same ideas. And we'll do that. So we'll see how to count other graphs.

And we'll give you a-- well, I'll give you a proof of the Erdos-Stone-Simonovits theorem that we did not prove in the first part of this course. So it gives you an upper bound on the extremal number of a graph  $H$  that depends only on the chromatic number of  $H$ . So we'll do that.

And then I'll also mention, although not prove, some extensions of the regularity lemma to other settings, such as to hypergraphs. And what that's useful for is that it will allow us to deduce generalizations of Roth's theorem to longer arithmetic progressions. Proving Szemerédi's theorem.

So one way to deduce Szemerédi's theorem is to use a hypergraph removal lemma-- the hypergraph extension of the graph removal lemma, the triangle removal lemma that we saw today. It would also let us derive higher dimensional generalizations of these theorems.

So it's a very powerful tool. And actually, the hypergraph removal lemma, as mentioned in the very first lecture, it's a very difficult extension of the graph removal lemma. And the hypergraph regularity lemma, which can be used to prove the hypergraph removal lemma, is a difficult extension of the graph regularity lemma.

So we'll see that in the next few lectures.