

**YUFEI ZHAO:** So the first topic that I want to discuss in this course is extremal graph theory. And in particular, there is a whole class of problems which have to do with what happens if you forbid a specific subgraph. Forbid a specific subgraph. And I ask you, what's the maximum number of edges that can appear in your graph?

In particular, and this is the question that we saw at the end of last lecture, which, now we're going to pretend that's a theorem. Mantel's theorem essentially asks, if you know that your graph has no triangles, what's the maximum number of edges it can have? And Mantel's theorem tells us that the extremal example is when your graph consists of putting half the vertices on one side, half the vertices on the other side, and putting in all the edges between the two sides.

So this is a complete bipartite graph. For these partitions, we denote complete bipartite graphs like that. And Mantel's theorem tells us that this graph, among triangle-free graphs, has the most number of edges.

Triangle-free graph,  $n$  vertices, has at most-- so the number of edges there is  $n$  squared divided by 4. Round down-- that many edges. And from this example, this bound is tight. So Mantel's theorem then gives us a completely satisfactory answer to the question of, what's the maximum number of edges in a graph without triangles?

And I want to begin by showing you a few different proofs of Mantel's theorem. So we're illustrating some different techniques in graph theory. So we'll see quite a few proofs in today's lecture. The first one begins-- well, here's the setup. I have  $G$ , an  $n$  vertex graph. And let me denote the vertices and edges of  $G$  by  $V$  and  $E$ .

If I have an edge in  $G$  between vertices  $x$  and  $y$ , then note that they cannot have any common neighbors. Because if they did, I would see a triangle. So I assume  $G$  is triangle-free. So what can we say about the degrees of the two endpoints of this edge? Well, they cannot add up to?

**AUDIENCE:** They can't have more than  $n$ .

**YUFEI ZHAO:** They cannot add up to more than  $n$ . So in particular-- so exactly. So the degrees of these two endpoints, there's at most  $n$  whenever  $xy$  is an edge. So here I'm using  $d$  to denote degree.

Well, OK. So now let me consider the quantity which is the sum of the squared degrees. On

one hand, I claim that the sum is equal to this quantity here, where I sum over all edges. And the reason is, well, look at the sum. Suppose, imagine writing out all the sum  $n$ 's. How many times does each  $dx$  come up?

So each  $dx$  appears one for each edge  $x$  is in, so appears exactly  $dx$  times. But we saw from up here that each sum  $n$  is at most  $n$ . So this sum here is that most  $mn$ .

On the other hand, let's consider the quantity which is just the sum of the degrees. And you sometimes know it as the handshaking lemma, that the sums of the degrees is just twice the number of edges. Each edge is considered twice in the sum.

Well, now we apply the Cauchy-Schwarz inequality, which, as you'll see many times in this course, although you might think of it as a fairly simple inequality, it's extremely powerful. And it will come up pretty much throughout this course. By the Cauchy-Schwarz inequality, you compare these two quantities. I find that we have this inequality over here relating the sums of the degrees and square of the sum of the degrees.

But we saw, on one hand, the left-hand side is  $4m$  squared. And we also saw that the right-hand side is at most  $mn$  squared. Putting them together, we see that  $m$  is at most  $n$  squared over 4. And of course, because it's an integer, it's at most the floor of this number. And that's a proof of Mantel's theorem.

What can you tell me about the equality case in this proof? So I'll let you think about that. But let me show you some other proofs.

In other words, are there graphs with the same number of edges as the graph shown up there that is also triangle-free? Is that a unique example? So let me show you a different proof of Mantel's theorem.

In this proof, so we begin with a step that seems a little tricky. Let's let  $A$  be a subset of vertices such that  $A$  is a largest independent set of  $G$ . So remember, independent set is a subset of vertices with no edges inside. It may have many independent sets all having the same maximum size. Take one of them.

So why should you do this step? Well, you know, sometimes magic happens. Let's consider some vertex. So consider some vertex, let's say  $x$ , and look at its neighborhood.

The neighborhood must be an independent set. Otherwise, I get a triangle. So every

neighborhood is an independent set. And as a result, the degree of every vertex is at most the size of the largest independent set.

So now, let  $B$  be the complement of  $A$ . So I have this  $A$  and  $B$ . Every edge has to intersect  $B$ . The edge cannot be entirely containing  $A$ , because it has no edges.

So the number of edges of  $G$  is at most-- well, I count over the vertices in  $B$ , the degree. So maybe I overcount. So the edges containing  $B$ , I count twice, but that's OK. So this is an upper bound on the number of edges.

But I also know that every vertex in the graph has a degree at most the size of  $A$ . So each sum  $n$  is at most the size of  $A$ . And I have  $B$  terms. So I have that.

Now, by the AMGM inequality, you'll have that. And the sizes of  $A$  and  $B$  add up to the entire vertex set,  $n$  squared over 4. So that gives you another proof of Mantel's theorem. What does this proof tell us about the equality case?

Now, something I always want you to keep in mind when we do proofs is, especially if we have a tight example like that-- and later on in the course, we'll almost never have good examples like that. So this is still early on in the course. And we're still very clean in the examples-- is to keep the extremal example in mind. And every step in your proof, it should be tight for that example. Otherwise, something went wrong with your proof.

Let's think about the inequalities. At equality, we must have that-- so looking at this inequality. So there are no edges in  $B$ . We also have that-- so by that, so every vertex in  $B$  is complete to  $A$ . And finally,  $A$  and  $B$  should have the same size.

So it is exactly the configuration shown up there. Now, when  $n$  is odd, we're rounding down by 1. So you can lose a little bit. But you can check that actually what we described up there is also the unique example. So that graph, the complete bipartite graph with two equal parts, is the unique maximal example, number of edges in a triangle-free graph.

Great. So once we know what the answer is for triangle-free, of course, we should ask further questions. Instead of forbidding a triangle, what if we forbid other graphs? And what are some natural next steps to take?

Well, say, instead of triangles, we can ask, what about if we forbid a  $K_4$ , a clique of four vertices? Or in general, what if we forbid a clique on the fixed number of vertices? So what is

the maximum number of edges in a  $K_r$  plus 1-free-- there's a good reason why index is  $r$  plus 1-- graph on  $n$  vertices.

So for example, if we're interested in  $K_4$ -free-- so what might be a good candidate for a graph with lots of edges that has no  $K_4$ 's? I mean, certainly, that example we saw, it does not have  $K_4$ 's, because it doesn't have any triangles. But we can do even better.

Instead of taking two parts, you can take more parts. For  $K_4$ , if I take three parts, each with  $n/3$  number of vertices, of course, if  $n$  is not divisible by 3, round up and round down. And putting all the edges between different parts, you can see this graph here has no  $K_4$ . So that's an example of a  $K_4$ -free graph. Well, does it have the maximum possible number of edges? So is this the best that we can do?

So it turns out the answer is yes. And that's the next theorem that we'll see. But just to give it a name, so we're going to call graphs like these Turán graphs. So the next theorem is proved by Turán. It's Turán's theorem. So a Turán graph, so we'll denote  $T_{n,r}$ . It is a complete  $r$ -partite graph such that there are  $n$  vertices whose part sizes are all nearly the same, up to at most 1 difference.

So this is an example here. But in general, maybe you have  $r$  parts. And I put in all the edges between different parts.

So it's not too hard to calculate the number of edges in such a graph. And Turán's theorem tells us that that is the extremal example. You cannot do better in terms of getting more edges. So if  $G$  is an  $n$  vertex, try a  $K_{r+1}$ -free graph. Then it has at most the number of edges of the Turán graph.

It's a generalization of Mantel's theorem. And well, you can think about if the proofs that we did for Mantel's generalizes to Turán's theorem. And it's actually not entirely clear how to do it. So let me present for you three different proofs of Turán's theorem.

So some of them, you can think about, are they related to the proofs of Mantel's theorem that we did? And they all are going to look somewhat different, but maybe superficially.

The first proof, we will use induction on the number of vertices. So actually, this is one of the very few times in this course where we will see induction. So of course, induction is a powerful technique in combinatorics. But for almost the rest of the course, we're not going to have clean

examples. And when we do have clean examples to work with, somehow increasing  $n$  by 1 doesn't buy you all that much. Here, there are very clean examples, very clean answers, and induction works out quite well.

When  $n$  is small, of course, you should always address that. You can come up with many funny proofs if you don't address when  $n$  is small. So when  $n$  is small, this problem is basically trivial.  $n$  is almost  $r$ . You could have the complete graph on  $r$  vertices. And then we're good.

So let's assume that we're not in this case. And also, by induction hypothesis, let's assume that it is true for all graphs fewer than  $n$  vertices. And let  $G$  be a graph that is  $K_{r+1}$ -free on  $n$  vertices.

And also, let's assume a maximum example to begin with. So let's assume that the  $G$  that you chose has already the maximum possible number of edges. There are only finite of any such examples. So pick one that already has the maximum number of edges. And let's think about what properties this graph  $G$  has.

I claim that  $G$  must already contain a clique on  $r$  vertices. So think about that. If  $G$  does not contain a clique on  $r$  vertices, then I can add in more edges. And I can still maintain the property of being  $K_{r+1}$ -free. So I can assume that  $G$  must contain a  $K_r$ . So let's look at one of these  $K_r$ 's.

So let  $n$  be the vertices of some  $r$  clique in  $G$ . So we have some  $A$ . And the complement to that, let me call that  $B$ .

Look at the vertex in  $B$ . How many neighbors can have in  $A$ ? It cannot be complete to  $A$ , otherwise I would have an  $r+1$  clique. So every vertex in  $B$  has at most  $r-1$  neighbors in  $A$ .

So let's count all the edges. The number of edges in  $G$  is that most-- well, first, we should account for the edges inside  $A$ . And there are-- or choose two of them. And then the edges between  $A$  and  $B$ , for every vertex in  $B$ , there are at most  $r-1$  vertices going to  $A$  for each vertex in  $B$ . And finally, the edge set of  $B$ .

Well, we can say something more about these quantities. We know that the size of  $B$  is exactly  $n - r$ . But what can we say about the number of edges in  $B$ ? We can use induction hypothesis. So  $B$  is also  $r+1$  clique-free. So the number of edges is at most the number of edges in a corresponding Turán graph.

Now, at this point, you can do a calculation. Well, I mean, you should expect that the answer we're looking for is-- I mean, this should be equal to the number of edges in the Turán graph. So you can either do a calculation to figure this out, or remember what I said earlier, that keep the tight example in mind, and everything should check out for the tight example.

So in particular, if you are in the situation of a complete multipartite graph with equal size or nearly equal-size parts, what is  $A$ ? So  $A$  is one vertex from each part. So you take out one vertex from each part. And read off this calculation for this graph over here. And then you see that that is indeed equality. So it should check out. You don't need to do any actual calculations.

Any questions about the proofs we've done so far? All right. Let me show you another proof of Turán's theorem.

So this proof has a name. So it's called Zykov's symmetrization. So Zykov has the unfortunate honor of having a name that's hard to beat in terms of alphabetical order. I think, if he and I were to write a paper, I wouldn't be the last author. So what's this about?

So let  $G$  be the graph. Again, as before, we're going to take a maximal example. So be the  $n$ -vertex graph that is free of cliques of  $r$  plus 1 vertices and has already the maximum number of edges.

So here's the property I want to prove about this graph. I claim that if you look at the complement of the graph, of this extremal example, it must be an equivalence relation. So more precisely, claim that if  $xy$  is a non-edge, and  $yz$  is a non-edge, then  $xz$  must also be a non-edge. So in other words, non-edges form equivalence relation.

And again, you should always think about the extremal example. And it is true for the extremal example. Because the complement are a bunch of cliques. So it is an equivalence relation.

So let's prove this claim. So let's assume that the conclusion is not true. So suppose, for the contrary, that we have  $xy$  and  $yz$  being non-edges, but  $xz$  is an edge.

So let's think about what happens to the degrees of non-adjacent vertices. So I claim that if the degree of  $y$  were smaller than the degree of  $x$ , then I can do something to the graph that violates the maximality of the number of edges. Namely, I can replace  $y$  by a clone of  $x$ .

So I chop off  $y$  from the graph. And I look at  $x$ . And I clone. So cloning it means taking some  $x$  prime. And I join  $x$  prime to all the same neighbors as  $x$ . So clone  $x$  into some other vertex,  $x$  prime.

Now, when you do this, I claim that you also get a graph that is free of  $K_{r+1}$ . So we also obtain a  $K_{r+1}$ -free graph. So why is that?

So if you were to have an  $r+1$  clique, well, it shouldn't have both  $x$  and  $x$  clone. Because there's no edge between them. So it only has one of  $x$  and  $x$  clone. But then that would have been a clique in the original graph.

However, what about the number of edges that we obtain after this transformation? If  $x$  had more outgoing edges, a higher degree than  $y$ , then cloning  $x$  to replace  $y$  increases the number of edges. We obtain a graph that is also  $K_{r+1}$ -free and has more edges than  $G$ , which is not possible, because we assumed that  $G$  started with the maximum number of edges.

So therefore, the degree of  $y$  is at most the degree  $x$ , basically for every non-edge  $xy$ . Likewise, we also have that there is no edge between  $y$  and  $z$ . So the degree of  $y$  is-- thus the degree of  $y$  is at least the degree of  $x$  now. So degree of  $y$  is at least the degree of  $z$ . Now,  $y$  has a lot of outgoing edges. So now let's replace both  $x$  and  $z$  by clones of  $y$ .

As before, we obtain some graph that we'll call  $G$  prime. And  $G$  prime, same reason as before, is  $K_{r+1}$ -free. If you had a  $K_{r+1}$ , then the same clique should have shown up in  $G$ . So  $G$  prime does not have  $r+1$  cliques either. But what about the number of edges in  $G$  prime?

So the number of edges in  $G$  prime-- well, we started with  $G$ . We deleted  $x$  and  $z$ . So we deleted this many edges. Here, we're crucially using the fact that there is an edge between  $x$  and  $z$ .

But now we also added back in a bunch of edges coming from the clone of  $y$ . And they are  $2$  times  $d_y$  edges added back in. Now, you see, because  $y$  has degree at least that of both  $x$  and  $z$ , we obtain that this number of edges is strictly bigger than that of  $G$ , which contradicts the maximality of  $G$ .

So at this point, we know that the complement of  $G$  must be an equivalence relation. So the complement is a bunch of cliques. And that's a lot of information. So that's almost the best

thing we can hope for is the structural information.

We're not quite done yet. And why is that? Well, we're almost there. But we're not quite done yet.

**AUDIENCE:** [INAUDIBLE] you can figure out the sizes [INAUDIBLE].

**YUFEI ZHAO:** OK. So we need to figure out the sizes of the individual parts. So we need to figure out the sizes of individual parts. And also, note that you cannot have too many parts.

So at this point, after the claim, we know that  $G$  is a complete multipartite graph. It has at most  $r$  parts. If it had  $r$  plus 1 parts, I would see a clique in  $r$  plus 1 vertices.

So then, finally, I need to show-- I mean, the rest is fairly routine. I need to show what the part sizes have to be to maximize the number of edges. And basically, if you had some two parts-- so consider exactly  $r$  parts. They're all empty parts. And some two parts have the number of vertices differing by more than 1.

Then what I can do is-- so if you had one part much bigger than the other part, you can imagine moving one vertex from one part to the other part. And you should convince yourself that this operation should strictly increase the number of edges. And then moving vertices should strictly increase the number of edges of  $G$ .

And putting this all together, we see that the extremal example necessarily has to be the current graph, namely a complete  $r$ -partite graph, where all the parts have the same size, up to almost 1 difference. Great. Any questions? Yep?

**AUDIENCE:** Can the Zykov symmetrization technique be used for any other type of problem?

**YUFEI ZHAO:** So the question is, can the Zykov symmetrization technique be used for any other types of problem? I do have something else in mind. But I don't want to discuss it now. Any more questions?

Great. And you see that in both proofs, if you look at this proof as well, you see that the Turán graph, it's the unique extremizer. I want to give you a third proof that has a somewhat different flavor. So these two proofs, they're both somewhat combinatorial in flavor. So I'm doing some arguments either looking at, in this case, a clique and arguing what happens outside of it. And over there, again, I'm looking at maximal example.

The third proof I want to show is more of a probabilistic proof. So this highlights an important method in combinatorics, and almost a probabilistic method, where we start with a problem that comes with no randomness. But we introduce some randomness to make the problem amenable. It's a very pretty idea.

So we start, again, with  $G$  being a  $n$ -uniform, so  $n$ -vertex,  $K_{r+1}$ -free graph on  $m$  edges. And what I want to do is to randomly sort the vertex set. Consider a random order of the vertices. So I put all the vertices on the line, but the vertices are chosen at random order. And you see some of the edges like that.

Let me show you how to find a clique. And essentially, we do it in a not-so-smart way. Namely, I basically pick all the vertices in some greedy-like manner, where I include the vertex in my set if it is adjacent to all earlier vertices. So if all of its neighbors-- so I include  $v$ , if  $v$  is the earliest vertex among its neighbors. So earliest means the leftmost, so if I go from left to right.

So in this case up here, so I look at the first vertex. Well, the first vertex should always be in your set. So this vertex has one neighbor, just to the right. So that's OK. And hold on. It's-- no. Sorry. That's not what I want to do. Sorry.

So I actually do mean what I was going to write down initially. So if  $v$  is adjacent to all the earlier vertices in this order-- so for example, this vertex, it's OK. And I'm also going to include this vertex here, because both of its earlier vertices are in-- both the earlier vertices are adjacent to the third vertex. And I think that's it.

Now, this set  $x$ , I claim two things. One, that  $x$  has to be a clique. So claim that  $x$  has to be a clique. Because every vertex in the  $x$  is adjacent to all of the [INAUDIBLE] vertices in particular. All the vertices in  $x$  are joined to each other.

On the other hand, how big is  $x$ , at least in expectation? So for every vertex, I want to understand the probability that this  $v$  is included in  $x$ . So  $v$  has a bunch of non-neighbors. And the property of  $v$  being included in  $x$  is that all of its non-neighbors appear after  $v$ .

So the probability that  $v$ , little  $v$ , is in the set  $x$  is equal to the probability that  $v$  appears before all its non-neighbors. All  $v$  and its non-neighbors, they're sorted uniformly. So the probability that this occurs is exactly  $1$  over  $1$  plus the number of non-neighbors, which is  $n$  minus the degree of.

Well, we know that this graph  $G$  has no cliques of size  $r + 1$ . So if we consider the expected number of the size of  $x$ , on one hand, it is at most  $r$ . Because  $G$  is  $K_{r+1}$ -free. On the other hand, by linearity of expectations, each vertex is included with some probability. So the size of  $x$  in expectation is just a sum of all of these individual inclusion probabilities, which individually we've computed above like that.

Now, by convexity, I can conclude that this quantity, this sum here, is at least the quantity that would have been obtained if all the degrees were equal to each other. And if you rearrange this equation, this inequality, we obtain that  $m$  is at most  $\frac{1}{r} n^2$ . And if you compare this number to the number of edges in the Turán graph, so you see that this is basically the number of edges, and this gives you a proof if  $n$  is divisible by  $r$ .

And in fact, the number of edges in the Turán graph is exactly this number here, if this divisibility condition is true. If it's not, you need to do a little bit more work. Because we were a little bit lost here in this step. So you should see that this quantity here is minimized when all the degrees are roughly equal to each other, and you make them as close to each other as possible.

So with a little bit more work, you can get the exact version of Turán's theorem up there. But at least what we've shown is-- so from here, we've shown that the number of edges is at most this quantity, which, for most purposes, is basically as good as Turán's theorem. Any questions about this proof here? And so it's a probabilistic method of proof. So we're introducing some randomness into the problem that originally had no randomness.

Great. So let's take a very quick 2-minute break. And then when we come back, I want to show you that even though we've shown so many different proofs of Turán's theorem and Mantel's theorem-- you might think, OK, this is a pretty simple thing-- even if I tweak the problem just a little bit, there are so many things that we do not understand, and many important open problems in combinatorics that are variants of Turán's theorem. So let's take a quick break.

So so far we've been talking about Turán's theorem, or generally the problem of, if you forbid a certain structure, forbid a certain subgraph, what is the maximum number of edges? We're going to spend the next few lectures discussing more problems of that form. And it turns out, for the answers we've just seen, they are deceptively simple. And for almost any other situation, we don't really know the exact answer. And for many questions, we don't know anything close to the truth.

And in particular, I want to show you one variant of this problem, namely what happens, instead of looking at graphs, if you look at hypergraphs. And there, that's a major open problem in combinatorics, what the truth should be. So here is an open problem, which is, what happens to Turán's theorem for 3-uniform hypergraphs?

So don't be scared by the word hypergraph. So whereas you think of graphs as having edges consisting of pairs of vertices, a hypergraph is simply a structure where, for a 3-uniform hypergraph, the edges are triples, triples of vertices. So the question is, what is the maximum number of triples in 3-uniform hypergraph without-- well, for Mantel's theorem, we asked what happens without a triangle. For a 3-uniform, the basic question you can ask is, what about forbidding a tetrahedron?

So suppose you do have four vertices such that every triple inside the four vertices is a niche. What's the maximum number of edges you can have in an  $n$ -vertex, 3-uniform hypergraph. Already, it's not so easy to come up with good examples. So Turán's theorem says take a bipartite graph, complete bipartite graph. Well, here, it's not so easy to come up with examples, but there are some examples.

And in fact, Turán suggested the following construction. Namely, you divide the set of vertices, as before, to three roughly equal-sized parts. And let me take all triples that look like one of the following forms, either three vertices, one in each vertex set, or a triple, looking like that-- two vertex in one set, one in the next set, or going cyclically, like that. So I include all triples, one of these forms.

And you should check that this construction has no tetrahedron. If it had a tetrahedron, you would have had at least two vertices in one part, also then where the other two vertices can be. If you check that, it cannot happen.

So then, how many edges does it have? The exact number is not so important. But what's important is that the edge density-- of all the possible triples, the fraction of edges that are containing this construction is  $5/9$ . And it is conjectured that this is optimal.

However, we're quite far from proving this number. So the best upper bound that is currently available-- and it's found quite recently using this fairly new method in graph theory called flag algebras, essentially a computerized way to try to prove such inequalities. And the best upper bound is something like 0.562. And there's a major open problem to either prove or disprove that this construction here is the optimal one.

So you see, even though I presented so many proofs of Mantel's theorem and Turán's theorem, at this point, hopefully they should all be-- they seem quite simple in retrospect. It's deceptive. And even if I changed and tweak the problem just a little bit, going to 3-uniform hypergraph instead of graphs, we really have no idea what's going on. Yes? Question?

**AUDIENCE:** So basically, how do you-- why you say that this is really bad compared to a two-number? It doesn't seem like that big a difference.

**YUFEI ZHAO:** OK. So great. So the question is, why do I say it's a pretty big gap. So we know that there has to be some proportion of the total number of triples. So, well, you have two numbers. And well, I mean, to me, they seem pretty far apart. It's not some lower-order gap. So this is a first-order gap. Any more questions?

But it's true that later in the course, we'll see gaps that are much bigger. We'll see gaps where there's a polynomial on one side and a power of exponentials on the other side. And there, I agree, that's much worse. The gap is much bigger. Here, it's just two numbers. But in the worst case, there can be two numbers anywhere. Any more questions?

So throughout this course, I will try to bring out some open problems. And there are lots. So I will tell you what we do understand. But most things, we do not understand. And hopefully, some of you will go out and try to understand them better so the next time I teach the course, I could have something new to present.

So now that we've addressed the question of what's the maximum number of edges if you forbid a triangle or a clique, the next natural thing to ask is, what about if you forbid a general  $H$ ? So give me a graph  $H$ . And what's the maximum number of edges if you forbid that  $H$ ?

It will be helpful to do some notation. So the extremal number, which we'll denote by  $ex$ , and this is also called the Turán number sometimes because of Turán's theorem. So I'll probably call it the extremal number more frequently. So this number here is defined to be the maximum number of edges in an  $n$ -vertex graph containing no copy of  $H$  as a subgraph.

I want to just clarify a piece of notation. So when I say a subgraph, so what do I mean? So there are several notions of different kinds of subgraphs. And a couple that come up somewhat frequently, one is subgraph. And then there's something called induced subgraph. So it's probably easiest if I just show you an example.

So suppose my  $H$  is the four cycle. So in the example of  $H$  being a subgraph, suppose I have this graph here. So  $H$  is a subgraph of this graph here in many different ways, but in particular, like that. But you see there are some more edges that are among the vertices. But that's OK. So subgraph only requires you to have a subset of vertices and a subset of edges, whereas induced subgraphs-- this is not an example of induced subgraph.

And so induced subgraph means that you take this set of vertices and you look at all the edges in the big graph among your set of vertices. So that's an induced subgraph. So here, the four cycle is an induced subgraph, but not induced subgraph over here. So I just want to make that distinction clear. And for now, in this chapter, we'll only talk about subgraphs-- so not induced, necessarily.

All right. So let's recap what happened for Turán's theorem. So Turán's theorem told us that the extremal number of these cliques, well, we know very precisely to be the number of edges in the Turán graph. And in particular, we saw from the last proof, but also, similarly, if you just do a calculation, that it is at most this quantity over here. And it's basically that quantity.

So the number of edges in this Turán graph is asymptotically that quantity up to a lower-order error; if you like  $n$  go to infinity,  $r$  fixed. And the basic question is, what about general  $H$ ? And so if I give you some arbitrary graph  $H$ , what can you tell me about the maximum number of edges in the graph forbidding this  $H$  as a subgraph?

And it turns out, for most  $H$ 's, we have a pretty good understanding, and perhaps quite surprisingly, because you can imagine this problem looks like it might get quite complicated. All the proofs that we've done are very specific to cliques. But it turns out that we already understand a lot. And the critical parameter that governs how this quantity behaves is the chromatic number of  $H$ .

So if you call me the chromatic number of  $H$ , I can already tell you quite a lot. So just to remind you, the chromatic number of a graph is the minimum number of colors you need to properly color this graph. Chromatic number of  $H$ , denoted  $\chi$  of  $H$ , is the minimum number of colors needed to color the vertices of  $H$  so that no two adjacent vertices have the same color.

So for instance, if I give you a clique of  $r$  plus 1 vertices, so all the vertices must receive different colors, because every pair of vertices is adjacent-- so the chromatic number is  $r$  plus 1. The chromatic number of the Turán graph is what? It's  $r$ . So the chromatic number of the

Turán graph is  $r$ . I can color each part in this complete bipartite graph using a different color. And that's the best I can do.

Now, if I have one graph being a subgraph of another graph, what can you tell me about the relationships between their chromatic numbers? So if  $H$  is a subgraph of  $G$ , so what can you tell me about relationship between their chromatic numbers?

**AUDIENCE:** Chi of  $H$  is less than the other chi.

**YUFEI ZHAO:** So OK. You're telling me that the chi of  $H$  is at most chi of  $G$ . And why is that?

**AUDIENCE:** Because if  $G$  can be colored with a certain number of colors, then that also has [INAUDIBLE].

**YUFEI ZHAO:** Great. Whatever coloring you do for  $G$ , you use the same coloring, and that's a proper coloring for  $H$  as well. So the chromatic number of  $H$  might be smaller, but certainly cannot be bigger than that of  $G$ . So in particular, if you have a graph  $H$  with chromatic number  $r + 1$ , then this Turán on graph is always  $H$ -free. So if  $H$  requires four colors, it cannot be embedded into the complete multipartite graph of three parts.

So the Turán graph is also an example of an  $H$ -free graph with lots of edges. And this tells us that the extremal number of  $H$  is at least that of this Turán number, where  $r$  is defined as the chromatic number minus 1. So that's some lower-bound construction.

So as we saw earlier, and we know what the asymptotic is like for the number of edges, this  $n$  goes to infinity. Namely, it's like that. And now the question is, is this the right answer? Is it possible that we completely missed some construction that might produce a lot more edges?

And it turns out-- and I think this should be somewhat surprising. Because so far, I feel like I haven't told you anything all that surprising yet. This seems like a fairly mysterious problem. But it turns out that this is more or less the right answer. So you cannot do much better than the Turán graph.

And there's the theorem of Erdos, Stone, and Simonovits rich that for every graph  $H$ -- so if I fixed  $H$ , then the limit as  $n$  goes to infinity of the extremal number is a fraction of total number of pairs. So here, this is the edge density. So this quantity here is equal to  $1 - \frac{1}{\chi(H)}$ . So  $1 - \frac{1}{\chi(H)}$ .

So the chromatic number, in some sense, completely determines how big the extremal

number should be. And you see that so far everything we've proved, Turán's theorem, Mantel's theorem agree with this formula here. But if I give you some  $H$ , maybe quite complicated, and those previous proofs don't work, well, still you know the first order asymptotics. But there's still more to say. But first, let me just run through some examples for a sanity check.

So if  $H$  is a triangle-- so if  $H$ , the chromatic number, and this limit-- so if  $H$  is the triangle, chromatic number is 3. So then this limit is  $1/2$ . And that's indeed the case. So that's what we did with Mantel's theorem.

If  $H$  is a clique of four vertices, then the chromatic number is 4. And the answer is  $2/3$ . And also, that agrees with Turán's theorem. It agrees with Turán's theorem.

But I can give you some fairly complicated-looking  $H$ . So for example,  $H$  might be the Petersen graph. So every good graph theory course should feature a Petersen graph at least once, somewhere. So that's the Petersen graph. What's the chromatic number of the Petersen graph?

Actually, that's kind of a tricky question. The last time I taught this course, I even got the answer wrong. So it turns out you can three-color the Petersen graph.

So it's completely not obvious how to do this. But there are only so many vertices. If you stare at it long enough, you can see what happens. But let me just show you a three-coloring of the Petersen graph.

Then the third color is like that. So that's a three-coloring of the Petersen graph. That's chromatic number 3. It's not 2. So why is it not 2?

**AUDIENCE:** It's not bipartite.

**YUFEI ZHAO:** It's not bipartite. It's a five cycle, which cannot be two-colored. So then the limit is  $1/2$ . And-- no. That's 3. Great. So the limit is  $1/2$ .

And here, I think that we can apply this theorem of Erdos-Stone-Simonovits. And I think this should be somewhat surprising. Because the Petersen graph looks quite complicated. If you try to forbid this graph in some big  $G$ , it seems like it's kind of hard to do. But it turns out the chromatic number completely governs the behavior of the extremal numbers.

But it turns out that's not the entire story. Because while this is quite a good theorem, and I said, it gives you the first-order asymptotics, actually, that's a lie. It doesn't always give you the first-order asymptotics. And when is Erdos-Stone-Simonovits not effective? Or rather, it's not the complete-- it's not the final answer. Well, you can say, well, what about this little  $\epsilon$ ?

So we're still trying to understand, there's a limit. And you can understand what is the-- how quickly does it converge to this number here? So that's certainly a valid question. But more importantly, though, if your graph  $H$  is bipartite, if  $\chi(H) = 2$ , i.e. bipartite, then all this theorem tells you is that the limit is equal to 0, which somehow is not the most satisfying answer. You want to know the first-order asymptotics. I mean, this still tells you something.

So the Erdos-Stone-Simonovits theorem tells you that the extremal number is little  $\epsilon$  of  $n$  squared. But of course, no. You are a curious mathematician. And you want to know, really, what is the asymptotics? Is it like  $n$  to the  $3/2$ ? Is it like  $n$  to the  $4/3$ ? You know, is this is not satisfying.

And it turns out, from most bipartite graphs  $H$ , it is a very difficult problem that we still do not know the answer to, even what is the order of the asymptotics. So the next few lectures, what I want to do is to show you some techniques that will allow you to prove some upper bounds for this extremal number in the case when  $H$  is bipartite that shows you that the exponent can be less than 2. And I will also show you some constructions that sometimes, but in the very few cases, matches the upper bound.

So there are very few examples of graphs  $H$  for which we know the first-order asymptotics. And for most graph  $H$ , they are major open problems in combinatorics. And there are some really old ones for which any solution may be quite exciting.

I will not show you a proof of Erdos-Stone-Simonovits now. We will see that later in the term, once we have developed some more machinery. Although, later on in the term, we'll see a proof using the so-called Szemerédi's regularity lemma, which I've mentioned a few times in the first lecture. Although, you don't actually need such heavy machinery.

So the original proofs of Erdos and Stone, and also by Simonovits-- so Erdos and Stone first proved this result for  $H$  being a complete multipartite graph. And Simonovits observed that knowing  $H$  for a complete multipartite graph actually implies this result in general. So you will not find a paper with all three of them being authors. But this is what the theorem is called.

So later on in the course, once we've developed the machinery of graph regularity lemmas, I will show you how you can deduce Erdos-Stone-Simonovits from Turán's theorem. So somehow you use Turán's theorem and bootstrap it to Erdos-Stone-Simonovits. So that should seem somewhat magical.

So on one hand, you have cliques. On the other hand, you have things that somehow don't have cliques in them. They don't really look like cliques. But you can still bootstrap one to the other. And so after we develop some machinery, we'll do that.

But there is a combinatorial proof which doesn't use any of this heavy machinery. I won't discuss it in lecture, but you can look it up. And it's quite nice. It has some combinatorial techniques and some double-counting arguments.

So going forward, I want to show you some questions, mostly questions, but some answers as well-- so questions such as, what is the extremal number of-- well, so what are some of the basic bipartite graphs? One of them is just a complete bipartite graph.

So this has a name. It's called the Zarankiewicz problem. And for some values of  $s$  and  $t$ , we know the answer. But for most values, we have no idea what the answer is. For example for  $K_{4,4}$ , we do not know what is the correct exponent on the  $n$ . So even fairly small cases, it is very much open.

So this is all to show you that these simple proofs that we did today, they are perhaps too deceptively simple. Because even if you change the question a little bit, we don't understand a lot. So questions like these will occupy the next several lectures. And I'll also show you some constructions. And there are some constructions that use nice algebraic ideas and some probabilistic ideas. So we'll see ideas coming from many different sources.

I want to close off by just a cultural remark. So in this course, especially the first half, we'll encounter a lot of Hungarian names. So we already saw a couple of them. And I just want to give you a very quick tutorial on how to pronounce Hungarian names, just for cultural purposes now.

In the past, when I took this CLASS there were no Hungarian speakers. But I think we do have at least one Hungarian speaker in the room. So I want you-- but you are a native Hungarian speaker. So you should tell us. So can you help us pronounce these names?

**AUDIENCE:** Erdos.

**YUFEI ZHAO:** Erdos. And this one?

**AUDIENCE:** That doesn't seem like a Hungarian name.

[LAUGHTER]

**YUFEI ZHAO:** So this, Hungarian, it's Simonovits. So I'll just tell you two things about Hungarian names. One of them is that the S--

**AUDIENCE:** It's /sh/.

**YUFEI ZHAO:** --is pronounced like /sh/. And another thing that comes up is S-Z, which we saw in Szemerédi. So this is pronounced like /s/. Forget the Z.

So Erdos, another thing about Erdos that you should know is that-- what is this accent? It is not a double dot. So in LaTeX you would type it as slash H, and in particular, not like that. So just a few cultural remarks about names. Great. So we'll end today. And then next time, we'll start looking at other extremal numbers for more bipartite graphs.