[SQUEAKING]

[RUSTLING]

[CLICKING]

**YUFEI ZHAO:** The probabilistic method in combinatorics is a powerful way to demonstrate the existence of certain special configurations in combinatorial objects by introducing randomness. In this video, we will see one of the earliest examples of an application of this method and its use to prove lower bounds to Ramsey numbers.

Now, Ramsey numbers are interesting objects in combinatorics, and they are defined as follows. The Ramsey number Rkl is defined to be the smallest integer n, such that no matter how we color the edges of a complete graph on n vertices, we'll denote such objects by kn. So a complete graph is n vertices with all the edges, all n choose two edges available. And if we color all such edges, each one of them, either with red or green, so with one of two colors, then there always exists a red kl, red clique on k vertices or green clique on l vertices.

So the Ramsey number is the smallest n such that this is possible. So to be more concrete, let me demonstrate an example. So R 3, 3 equals to 6, and what this means is the following. First, if we have six vertices and we color all the possible 6 choose two edges, each one with one of two colors, red or green.

So for example, maybe these edges are colored green and the remaining edges are colored red, then no matter how we do these colorings, there always exists a triangle, which is completely red or completely green. In this example here, you see this. There's a triangle that is completely red.

On the other hand, if we start with only five vertices, then it is possible to color the edges with one of two colors. So color each edge with one of two colors so that there is no monochromatic triangle. So this is what it means for R 3, 3 to be 6.

An important result in combinatorics, due to Ramsey, so this is known as Ramsey's theorem, and this was proved back in the 1920s, says that this number, Rkl is always finite. It is always a well-defined number. So as long as n is large enough, no matter how we color the edges of a complete graph on n vertices, there always exists a large monochromatic, either a large red k clique or a large green L clique.

What I want to prove in this video is the following theorem due to Erdos from the 1940s. An Erdos showed that Rkk is bigger than 2 to the k over 2 for every integer k bigger than 3. So this is a lower bound to the Ramsey numbers. And what this says that if you have too few vertices, namely if you have 2 to the k over 2 vertices, then there is always a way to color the edges with two colors so that there is no monochromatic clique on k vertices.

Erdos's theorem is an important foundational result and is also one of the earliest demonstrations of the probabilistic method. So we will see this proof in this lecture, in this video. In fact, we will prove a slightly more general statement. We will show that if two integers, n and k, satisfy the following inequality, n choose k times 2 to the 1 minus k choose two, if this quantity is less than 1. So right now, this is just some formula. It will come out of the proof. So let's not worry too much about it. Then Rkk is strictly larger than n.

We will prove the second theorem and, by a calculation that is fairly straightforward, which I will admit, one can use the second result to imply the first result. OK, so let us prove the second theorem over here. So what is this saying? If n satisfies certain inequalities, so the n is not too large, then it is possible to color the edges of a complete graph on n vertices so that there is no monochromatic clique on k vertices.

How do we find such a coloring? Well, let's do it randomly. So let's color the edges of a complete graph on k vertices using two colors uniformly at random, so meaning that for each edge, we flip a coin. If it comes up heads, we color it red. If it comes up tails, we color it green. And the goal is to show that we can do this in such a way that avoids having a monochromatic clique of size k.

So there are things that we want to avoid, and let us encode these bad events as follows. For every k vertex subset s of vertices, there are n vertices in total. So for every k vertex subset s, we'll use A sub s to denote the following bad event that we're trying to avoid. This is the event that s induces a monochromatic clique. OK, it induces a monochromatic clique in this coloring.

Let's calculate the probability that this occurs. So what's the probability that in this s, which let's say k is 4 for purpose of this illustration, of the six edges, all of them have the same color. Well, there are two possible colors, and for each of those colors, the probability that all of these edges are of that color is this quantity, 2 to the minus k choose two. So this is the probability 2 times 2 to the minus k choose two. This is the probability that s, all the edges in s are entirely of the same color.

Well, what about the probability that there is some clique of size k in the graph that's monochromatic? That's really the quantity that we care about. So let's consider the probability that there is some monochromatic clique of size k in this graph.

Well, if there is some monochromatic clique of size k, then it must be one of the S's. So we can do what's called a union bound to upper bound this probability by the sum of the individual event probabilities. So summing over all subsets s of the vertices with size exactly k of the probability that A sub s this event occurs.

In this summation, there are n choose k terms, and each term is this probability that we calculated earlier. And you see this expression. It was the expression in the hypothesis of the theorem. And we assume that it is strictly less than 1. And this is why we had this expression in the theorem.

OK, great. So the probability that there is some monochromatic clique of size k is strictly less than 1, and thus with positive probability a random coloring has no monochromatic clique of size k.

And that's basically, what we are trying to do. We're trying to demonstrate the existence of a coloring with no monochromatic clique of size k. And we show that, in fact, by doing everything at random, the random coloring succeeds with positive probability. And therefore, such an object must exist. Thus such a coloring exists. And this finishes the proof of this theorem over here that we were trying to demonstrate.

And as I mentioned earlier, by a more routine calculation, which we will not do here, one can deduce Erdos's result from this theorem. And in fact, by a more careful calculation that, again, I will not do, one can show an even more precise lower bound, that Rkk is bigger than the following asymptotic 1 over the following constant e times root 2. So here e is the natural constant, around 2.71, plus a term that goes to 0 as k goes to infinity times k 2 to the k over 2. So this is a much more precise bound that one can get out of analyzing the consequence of this theorem that we just proved.

In the next part of this video, we will refine or come up with additional techniques that allows us to improve this lower bound. But for now, let me make a comment about this technique. What this theorem shows is that by doing a random coloring there is a positive probability that this coloring gets us what we want. And in fact, the probability that this coloring doesn't get us what we want is this quantity here, which is not only less than 1, but it is actually quite small. It is very, very small.

So if you, for example, on a computer flip these random coins and come up with this coloring, with very high probability, it will succeed. However, it will be very difficult to verify that you have succeeded because the number of cliques that one needs to check is n choose k, which grows very quickly as a function of k. So this is a phenomenon that is sometimes called finding hay in a haystack.

A random coloring works with overwhelming probability. And so if you just pick something at random, it will work. But one, it is very hard to check that what you found actually works. And two, it is an active research direction and what seems to be a very difficult problem to find other ways that does not involve randomness but maybe even involving randomness in some other ways to guarantee that what you found works. And this finding hay in a haystack describes this phenomenon, which is quite counterintuitive, that-- and also demonstrates the power of the probabilistic method, that by just doing everything in random, it works really, really well. But we don't understand this process as much as we would really like to.

So now you've seen a proof of Erdos's seminal result that gives a lower bound to Ramsey numbers. In the next two segments of this video, we will see a couple of refinements to this technique that allows us to get slightly better lower bounds by introducing new ideas to the probabilistic method.

Now, let us introduce an additional idea to the probabilistic method to prove a slightly better lower bound on Ramsey numbers. And this is a method of alterations called the alteration method. What we'll prove is the following. So let me write down the statement. It's slightly technical. The statement itself is not as important as the ideas I will introduce, so bear with me for a second.

So theorem says that for any positive integers k and n, the Ramsey number, Rkk, is bigger than the following quantity, n minus n choose k times 2 to the 1 minus k choose two. OK, so that's the theorem. It gives you a lower bound on Ramsey numbers. So, by optimizing over the value of k as a value of n, as a function of k, we can deduce the following corollary.

And I will not show this deduction here because it's a more routine calculation. So the corollary is that Rkk is bigger than the following quantity, 1 over e-- e is the natural constant 2.71, and so on-- plus little 1-- a term that goes to 0 as k goes to infinity-- times k times 2 to the k over 2, OK? So that's the asymptotic lower bound that one obtains by taking this theorem and optimizing the value of n as a function of k.

We see that it is slightly better than the bound that we got previously, is slightly better by a factor of root 2. OK, so let us prove this theorem here. Like before, we will start with a random coloring but then make some adjustments. So in fact, this construction will have two steps. Construct a random coloring in two steps.

In the first step, We'll. Do what we did previously just to randomly color each edge with one of two colors uniformly at random. OK, so flip a coin for each edge, and color it with one of the two colors.

Now, this coloring might have some monochromatic k cliques, in which case, we will delete a vertex from each such k clique and to destroy all the monochromatic k cliques. So the second step is to delete a vertex from each monochromatic k clique.

OK, so after this process, see, what we are left with is an edge coloring of some complete graph where there are no monochromatic k cliques left because we've destroyed all such monochromatic k cliques in the second step of this process. How many vertices are left? So we started with initially n vertices, but in the second step we deleted some. So we don't have as many vertices as we started with initially. So let's ask ourselves, how many vertices are left at the end of this process?

Towards this goal, let me introduce a random variable x. And x is the number of monochromatic k cliques at the end of the first step. Well, the expectation of x can be computed as follows, which is very similar calculation to what we did earlier. There are n choose k possible k cliques, and each one of them is monochromatic with probability 2 to the 1 minus k choose two. So that's the expected number of monochromatic k cliques.

Now, in the second step of this process, we delete some number of vertices and one vertex for each monochromatic k clique. Although, some vertices may be used to destroy more than one monochromatic k clique. So the number of vertices that we delete is not necessarily x, but it is no more than x. So delete at most the size of x many vertices.

And thus, the final graph that we get at the end of this two-step process has at least n minus the size of x vertices. Let me remind you that this is a random process. The final graph and this coloring that we get is random, and x is also a random quantity.

This is a number of vertices, and this quantity here has expectation n minus expectation of the size of x. Oh, so x is already a size. So let me-- I don't have to use the absolute value bars. So the expectation is n minus n choose k 2 to the 1 minus k choose two.

OK, so the final graph has at least this many vertices in expectation. Well, that's the average number through this random process, and thus, there must always be, with some positive probability, some graph that beats this average.

So thus, or therefore, with positive probability the remaining graph has at least this many vertices. And furthermore, it has no monochromatic clique on k vertices because we destroyed all such cliques in the second step of this process.

OK, this finishes the proof of this theorem, which, as you see in the corollary, gives you a slightly better lower bound compared to what we saw last time. In the third segment of this video, we'll see yet another technique that gets us even further along to prove an even better lower bound on the Ramsey numbers.

Earlier, we saw two different approaches to using the probabilistic method to lower bound Ramsey numbers. The first method was by taking a union bound to show that the sum of all the probabilities of bad events is fairly small. And the second method is through alteration, where we start by constructing some random coloring and then fix the blemishes. So we show that by removing the bad parts, removing the vertices that contribute to monochromatic cliques, we can get a coloring with no monochromatic cliques.

Now let me introduce a more advanced method that can get us even further. So the general motivation, that we already saw a couple of times, is that we wish to avoid a certain collection of bad events, which we'll denote E1 through EL. So these are bad events. In our case, the bad events correspond to having a monochromatic clique in the random coloring.

There are some extreme situations that are easier to handle. And these extreme situations are one of two types. The first is when the events collectively have very small probability, even when you just sum the probabilities of these events.

So for example, if the sum of the probabilities of these bad events is less than 1, then we can apply the union bound to deduce that with positive probability none of the bad events occur. So that's an easy situation to handle.

Another easy situation to handle is if all the events, so all the bad events are independent. So if they are independent events, then the probability that none of them occur is equal to the product of the probabilities that the individual events do not occur. And this is positive, provided that all the individual event probabilities are strictly less than 1.

Now, this situation typically is not the case in a lot of applications, including the one that we're considering. For example, having two triangles, if they intersect at an edge, if they overlap by an edge, then being-- or if you have two cliques that intersect in some large portion, then having them individually being monochromatic are not independent events.

So what we're often dealing with is situations where there is some dependencies, so some event dependencies. And often, these are the hardest and most interesting situations that one has to deal with in the probabilistic method. A powerful tool that we will see is the Lovász local lemma, OK? So let me explain what is a Lovász local lemma and then apply it to the problem of lower bounds to Ramsey numbers.

So here is a version of the statement of the Lovász local lemma and specifically applied to a setting known as the random variable model, random variables model. The setting is as follows. Suppose we have x1 through xv, and these are random variables. And importantly, they are independent random variables. So think of them as independent coin tosses, the outcomes of independent random coin tosses.

We'll have B1 through Bm denote index sets. So they are subsets of numbers 1 through v. And for each i ranging from 1 through m, let Ei be some event that depends only on the variables indexed by Bi, namely, the variables x sub j as j ranges over the elements of Bi.

So this is a situation that captures a lot of applications where you have bad events, for example, cliques being monochromatic in a coloring of a large graph. And some of these events may have dependencies because the events are based on some underlying coin tosses, underlying independent random variables. And the different events may involve overlapping variables. And so this is capturing that kind of setup.

Now, let me introduce some hypotheses that captures the notion of a weak amount of dependence, a small amount of dependence. So suppose for every i from 1 to m, the set, the index set Bi is disjoint from all but at most d other Bj's. So d here is some other parameter in the theorem. So each Bi has overlapping variables with a small number of other Bj's.

And furthermore, another hypothesis is that each bad event occurs with some probability which is not too large, so 1 over d plus 1 times a constant E. Here, E is the natural constant, 2.71 so on. And this is true for every i. So let me just write it a little bit more clearly like this.

OK, so these are all the hypotheses. So there's a lot, but roughly speaking, it's saying that every bad event is related to only a small number of other bad events. And each event occurs with not too high probability, which depends not on the total number of events, but only on this d, which is the number of other events that it is associated with.

The conclusion is that the probability that none of the bad events-- so the bad events are the events indexed. So none of the bad events you went through Em occurs is strictly positive. So with positive probability, none of these bad events occur. OK, so that's the statement of the Lovász local lemma.

And now let us apply this local lemma to prove a even better lower bound than what we saw earlier for Ramsey numbers. We will not be able to prove the local lemma in this video. So its proof is, although not long, quite intricate. And I refer you to the references for a proof.

So let us now see a lower bound to Ramsey numbers and this is a result due to Spencer, Joe Spencer from the '70s. And such that if-- if k choose 2 times n choose 2, or n minus 2 choose k minus 2 plus 1 times 2 to the 1 minus k choose 2 is less than 1 over E, then the Ramsey number, Rkk, is strictly bigger than n.

So, as with earlier, this is some mysterious-looking formula. And I want you to not worry too much about it. It will come out of the proof. What may be more illuminating is the consequence of this theorem. So as a corollary, Rkk has the lower bound 1-- so a lower bound-- the following asymptotic root 2 divided by E plus little 1 k times 2 to the k over 2. And this yet beats the bound that we saw earlier by another factor of root 2.

This may seem like a small improvement, and quantitatively, indeed, it's only a constant factor more. Whereas, the remaining asymptotics, we do not know how far it is away from the actual truth. But yet, this result proved in the '70s is the best known lower bound to date for diagonal Ramsey numbers, these Rkk's.

OK, so let us now see the proof of this theorem of Spencer, which applies the Lovász local lemma. As earlier, we will begin by coloring the edges of the complete graph on n vertices with two colors uniformly at random.

And also, similar to the first proof that we saw, for every k vertex subset s, let us define a bad event, E sub s to be the event that s induces a monochromatic clique on the vertices in s. And we also saw earlier that the probability of this bad event is 2 raised to 1 minus k choose two.

OK, let us see how this setup corresponds to this random variable setup earlier. The variables correspond to the coloring of the edges. So in other words, these are indexed by the n choose two edges of the complete graph, Kn. So there are n choose two such variables.

Well, what about this condition here? So Bi's, so each Bi is a set of size k choose two corresponding to the edges of this complete graph inside this clique of size k. So when do the Bi's overlap?

So if s and s prime, so both k vertex subsets of vertices, so if these are both k vertex subsets, then their cliques overlap in edges if and only if the sets of vertices s and s prime overlapping at least two vertices. So as an illustration, if you have this clique of size 4 and another clique of size 4, then they overlap in at least one edge, if and only if the vertex sets overlap in at least two vertices.

So to apply the local lemma, we need to check how many other s primes intersect a given s in terms of their edge set. And this is a calculation that we can do as follows. For each k vertex s, the number of k vertex s primes that has intersection size at least two with s is at most following.

Well, I need to pick a pair of vertices of s where there is this overlap and then choose n choose-- and then select the remaining vertices of s prime from the remaining n choose two vertices of the entire clique. There are some possible overcount here because some s primes may be counted in more than one way through this formula here, but it is certainly an upper bound. And now we can apply the local lemma.

By applying the local lemma, one checks that as long as this hypothesis inequality is satisfied, this hypothesis in the statement of the local lemma is satisfied as well. So this inequality here corresponds to this inequality in the hypothesis. So this, it is satisfied, and so we can apply the local lemma to deduce that with positive probability none of the bad events, so none of the bad events Es occur for ranging over all k vertex subsets s in the graph.

And therefore, with positive probability, none of the k vertex subsets induce a monochromatic clique. And that's exactly what we're trying to show with this result over-- to this claim over here. And this concludes the proof of Spencer's theorem.

So you see that this is a demonstration of the powerful Lovász local lemma. And it is a very interesting method that was pioneered by Lovász and has had important applications subsequently in the development of the probabilistic method. Although we did not see the proof of the local lemma here today, I hope that you can at least appreciate its application. And throughout this lecture, throughout this video, we saw three different approaches to lower bounding, the Ramsey number, the diagonal Ramsey numbers using different ideas from the probabilistic method. It is a powerful method, and it has lots of other applications and many beautiful applications to various areas of combinatorics and beyond.