[SQUEAKING]

[RUSTLING]

[CLICKING]

PROFESSOR: In this video, we'll look at a beautiful result that provides us with the existence of graphs with high girth and high chromatic number. And it's a very nice demonstration of the power of the probabilistic method.

We'll recall some definitions. Given a graph, G, its chromatic number is the minimum number of colors we need to color the vertices of the graph G so that no edge has both endpoints receiving the same color.

The girth of a graph G is the length of the shortest cycle. For example, the following graph has girth 4 because the shortest cycle in this graph has length 4. And there are a couple of them. Another example-- so this graph here, there are many cycles. But the shortest one has length 3. It is a triangle up here. So the girth is 3.

If the graph is a tree, then it has no cycles, in which case we say that its girth is infinite. OK, so what can we say about girth and chromatic number? Well, suppose your graph contains a clique on k vertices. Then just to color this clique, we need at least k different colors. So then the chromatic number is at least k.

Can we say something in the reverse direction? Suppose we have a graph with high chromatic number. Would it be possible to deduce some local information about the graph? For example, would it be possible to say that if we have very high chromatic number, then there must be a triangle somewhere in the graph, or there's some fairly dense structure that's very local in this graph.

So the main theorem that we'll present in this video gives a definitive no to this question. And it is a classic result due to Paul Erdos from the '50s. And the statement of the theorem is the following. For any positive integers k and l, there exists a graph. There exists a graph with girth l, girth bigger than l, and chromatic number bigger than k.

OK, so let's pause for a moment and think about what this theorem is saying. So given any k and l, no matter how large, we can find a graph. There exists a graph for which there are no short cycles, no cycles of length at most l. And furthermore, to properly color the vertices of the graph, you need at least one. You need strictly more than k colors.

So even though this graph has very high chromatic number, around every vertex, if you only look not too far away, then the graph locally looks like a tree because there are no short cycles. So in particular, having high chromatic number does not give you local information. And Erdos shows that there always exists counterexamples to such a statement if we want some local information from having high chromatic number.

OK, so the rest of this video will concern the proof of Erdos' theorem. It is quite a miraculous result. And certainly before Erdos proved this result, there were other results in graph theory that constructed very explicit graphs that have some large chromatic number and girth, but not for all values of k and l. And those constructions are, in some sense, very hands on. It tells you exactly how to construct these graphs. But they were not able to get arbitrarily large girth and chromatic number.

Eldredge's insight is to use randomness, use the probabilistic method, and show that by appropriately modifying a random graph, one can achieve the desired outcome. So let's see how this works.

We'll begin by taking a random graph, GNP. So this is the Erdos Renyi random graph. And there are n vertices. And the edge probability is p, meaning that we flip a probability p coin for each possible edge and put down an edge independently with probability p between every pair of vertices.

So let G be drawn from this distribution. So G is this random graph. Now let's take a specific choice of p being locked and squared over n. Turns out, other choices of p work as long as p falls in some range. But for concreteness, we'll take this very specific choice of p.

So this is a random graph. And let's construct the random variable x defined to be the number of cycles of length at most l and G. These are short cycles in G. And these are the cycles that we want to avoid because we want to construct a graph with girth strictly larger than l. So we really do not like these cycles.

It would be nice if there were not too many such cycles. So let's compute the expected value of x. For this computation, we use linearity of expectations. And look over all possible cycle lengths from 3 to l.

So I is the target cycle length. And for cycles of length I, there are n choose I different choices of which I vertices the cycles can involve. And once you specify those vertices, they are I minus 1 factorial, different permutations, circular permutations of these vertices. And we divide it by two because a cycle could be counted in two different directions. And they will correspond to the same cycle.

So this quantity here is the number of cycles of length I in the complete graph on n vertices with all edges present. For each such possible cycle, the probability that the cycle appears in G is p to the I. So the second term here is probability that this cycle appears in GNP.

OK, this is some expression. It's a little complicated, but we can simplify it and do some approximations to upper bound this quantity. So you see that the n choose I, if you expand it in terms of factorials, one can upper bound this quantity by n to the I, so incorporating the first two factors, and then leaving p to the I factor intact.

OK, we chose p so that p times n equals 2 log squared of n. And so you can rewrite this quantity here like this. OK, there are l terms here. So let me furthermore upper bound to this expression by l times log n to the 2l. Here, l is a constant. So this quantity here is some growth and the rate of some power of a log and certainly much slower growing compared to the function n. So it's little o of n. It's a fairly crude approximation, but it will be sufficient for our purposes.

So this is a calculation. Conclusion can be read as the expected number of short cycles is fairly small. Well, if there are few short cycles, we can get rid of all of them to get a graph without short cycles. So this is the basic idea of the alteration method in the probabilistic method. So start with this GNP, this random graph.

It's not going to get us what we want right away, but we're going to do something to fix the defects in this graph-- in this case, get rid of short cycles. To that end, let us know that by Markov's inequality, the probability that the number of short cycles-- so short means length at most l-- the probability that the number of short cycles x is at least n over two is upper bounded by the expectation of x divided by n over two.

And because expectation of x is little o of n, this final quantity is little o of 1 in the case to 0 as n goes to infinity. OK, so this is a good sign. So it means that, typically, the graph does not have very many short cycles, has fewer than n over 2. And then eventually, we'll be able to get rid of one edge from each cycle to get us a graph with still fairly large number of vertices. But there are no short cycles left in this after the alteration.

Now, let's think about chromatic number. So we want to ensure that the graph that we end up having has high chromatic number. To that end, we'll recall the following fact. If we give you a graph, G, then the chromatic number of G-- so this is chi of G. So that's the chromatic number.

Well, the chromatic number is the minimum number of colors that one needs to color the vertices of the graph so that no edge has the same color on both sides. In particular every color class-- so if you look at a single color like red, and look at the vertices colored by that color, these vertices have no edges directly between them, and so they form an independent set.

So each color class is an independent set. And so the number of color classes you need, then, should be at least the number of vertices divided by the independence number. This is the size of the largest independent set. So this alpha of G. The independence number is the size of the largest independent set.

Indeed, each color class is an independent set. And each independent set has size at most alpha of G. And therefore, the number of colors you need should be at least the number of vertices divided by the independence number.

OK, now for every h, the probability that the independence number-- so after here, we see that it makes sense to think about the independence number. So the probability that independence number is bigger than h is at most the following. So here we'll use union bound.

For each h vertex subset, of which there are n choose h, such h vertex subsets, the probability that those h vertices form an independent set is the probability that there are no edges appearing among these h vertices. And that's 1 minus p raised to the power h choose 2.

So this is some expression. And we can do some manipulations to simplify this expression, yielding an upper bound of n to the h times. So the latter expression, we can simplify it as upper bounded by e to the minus ph, h minus 1 over 2. So 1 minus p is less than e to the minus p.

Simplifying even further, we see that this expression can be rewritten this way. And now let's make a choice for what age I want to set. So let's set age to be the quantity, which, for now seems a little mysterious. But it turns out to be a natural thing to set after we see what the calculation looks like.

So we set age to be this quantity, 3 times log n over p. And knowing what p was from earlier, this quantity is 3 n over log n. If we set h to be this quantity here, then you see that here, in this expression, the numerator here can be simplified. And the result is that this expression decays according to some rate, which is n to something that is a minus constant, something on a constant word.

In particular, it goes to 0 as n goes to infinity. And that's really the only thing that we need, that we need out of this calculation. So if we set h to be of an appropriate quantity, then typically, one does not have independent sets of size, at least h in the graph.

OK, so let's regroup. So we proved a couple of things. First, we showed that typically-- so when I say typically, I mean with probability approaching 1 as n goes to infinity. So typically, this graph does not have short cycles, meaning cycles of length at most l. And also, typically, this graph does not have large independent sets.

OK, now let us put everything together. So by choosing n sufficiently large, we can simultaneously ensure that the probability of x, the number of short cycles, exceeds n over 2, is strictly less than 1/2. And the probability that the independence number exceeds h is also strictly less than half. So both of these are quantities that decay to 0 as n goes to infinity.

So by choosing sufficiently r, h, n, we can make sure that both probabilities are strictly less than 1/2. And therefore, there is some possibility that lie outside both events. Thus, there exists a graph G with fewer than n over 2 cycles of length at most l.

And furthermore, its independence number, the size of the largest independent set, is, at most, h, which we set earlier to be 3 n over log n. OK, so we have such a graph G that has these nice properties.

Well, we're looking for a graph with high girth. Here, G has not too many short cycles. So we can get rid of one vertex from each short cycle to remove all the short cycles, so remove a vertex from each short cycle, each cycle of length at most l.

So remove one vertex from each such cycle. And doing so results in a subgraph in G prime. And then we know that the girth of G prime is strictly larger than l, because we got rid of all the short cycles. And furthermore, the chromatic number of g prime is by what we saw earlier, at least the number of vertices of g prime divided by its independence number.

While the number of vertices is at least n over 2 because we removed at most n over 2 vertices from G to obtain G prime. The independence number of G prime-- well, the independence number of G prime is, at most, the independence number of G because any independent set in G prime is automatically an independent set in G.

And here we saw that the independence number of G prime is at most h, which is 3 log n over 3 n over log n. So we have, in the denominator, 3 n over log n.

And whole expression simplifies to log n over 6, which is bigger than the constant that we were given, k, as long as n is sufficiently large. And thus, G prime satisfies the desired requirements.

This finishes the proof of this theorem of Erdos that there exists graphs with arbitrarily large girth and arbitrarily large chromatic number. To review some of the key ideas in this proof, we use the probabilistic method with alterations to construct first a random graph with various parameters so that this random graph typically has very few short cycles so that we can then remove one vertex from each short cycle and get rid of all the short cycles. And this is a way to ensure large girth.

To obtain high chromatic number, we note that this graph typically has-- typically does not have large independent sets. And so its independence number is typically small. And therefore, the independence number of the subgraph must be typically small, as well. And having small independent number implies high chromatic number.

OK, so it's a beautiful illustration of the probabilistic method with alterations that allows us to deduce this highly counterintuitive result, which was quite surprising because previously, various researchers have tried to construct such examples of graphs by hand explicitly, but unsuccessfully. And Erdos brought this beautiful insight that introducing randomness by looking at random graphs, one can, indeed, prove the existence of graphs with highly counterintuitive properties.