

The following content is provided under a Creative Commons license. Your support will help MIT OpenCourseWare continue to offer high quality educational resources for free. To make a donation or view additional materials from hundreds of MIT courses, visit MIT OpenCourseWare at ocw.mit.edu.

PROFESSOR: Today what we're going to do is finish off our discussion about oscillators. In particular, we're going to talk about alternative designs for oscillators. So rather than having these loops that are purely composed of negative interactions, negative feedback, instead we're going to talk about cases where you have both positive and negative interactions.

So in using this kind of combined network structure, you can generate what are known as relaxation oscillators, which have some really wonderful properties. In particular you can get more robust oscillations, relative to the parameters. But also the oscillations become tunable, i.e. you can change the frequency, without compromising, for example, the amplitude of the oscillations. So for both natural and synthetic oscillators these so-called synthetic oscillators are perhaps the way to go.

And then we're going to transition to more of the global structure of some of these networks in the context of transcription networks within cells. And discuss this paper that you guys just read, the Barabasi paper, which is one of the world's most cited papers, I think. And then after thinking about this global structure, of how you might be able to generate these so-called power law structures, we're going to look a little bit more in detail to try to understand something about these network motifs. We've already talked about them a little bit in the context of auto regulatory loops, but now we'll talk about them in a little bit more generality, in particular in the context of feed forward loops. And then on Thursday we will get into some of the possible beneficial features of feed forward loops.

On Thursday we talked about the repressilator. So if you have x inhibiting y inhibiting z coming back and inhibiting x , that it's reasonable to expect that it might

generate oscillations. And indeed in the Elowitz paper that we read, such a synthetic circuit did indeed generate oscillations, but there were perhaps a few problems there, right? So one is that only about 40% of the cells actually oscillated, who knows why not. But also there were other problems that the oscillations seemed rather noisy, there was relatively rapid desynchronization.

Moreover, if you go and you ask, well is it possible, or how easy would be to change the period of the oscillations just by changing something like the degradation rate, what you'll find is that the oscillations are not very tunable. So I'll say the period, or the frequency, is not very tunable, and indeed this is a general feature of oscillatory networks that have purely negative interactions. We talked about a couple of these cases, for example, you can get oscillations just with negative auto regulation. And what is it that's necessary?

AUDIENCE: [INAUDIBLE].

PROFESSOR: What's that?

AUDIENCE: High coordination.

PROFESSOR: High coordination? You me-- oh you're--

AUDIENCE: [INAUDIBLE].

PROFESSOR: Cooperativity in the repression, that I think is necessary, but is it going to be sufficient? Even in this case where I just have, let's say that I say, x dot the rate of production, if this thing is just as a function of x , the sharpest it could be, this is infinite cooperatively, so it's maximal expression. And then when you get above some x critical all of a sudden you fully repress. If I just have this be the formula-- did you guys understand what I'm referring to here? What would this generate? Would this generate oscillations? So it actually doesn't.

In the simple equation, where if we have x dot is equal to this function. So I guess this is a theta, I want to make sure I get this x , less than x critical, that's what this means. So with some [INAUDIBLE] rate beta, minus some alpha x . Does this thing

oscillate? No, and we had a simple argument for why did not oscillate, as well. Yes? Yell it out somebody, I'm sure somebody was here on Thursday.

[LAUGHTER]

AUDIENCE: [INAUDIBLE].

PROFESSOR: That's right. So this is just-- This is just an x dot, there's no x double dot, so that means the derivative of x , the single value is a function of x , that means that we can't get any oscillation here. And then remember we analyzed this model where we explicitly included the mRNA, so then we just had that x comes, and what it does, is it represses expression of this mRNA for x , and then this mRNA comes back and makes x . Right?

And in this model, was this sufficient? Did this give oscillations? No, so this also here, there was no oscillations. Again here, there were no oscillations. But I did tell you that you could do something more to get oscillations, just with a single protein repressing itself. So you need more delays. So if you add delays, then it's possible to get oscillations. So those delays could be in the form of having a model where you explicitly take into account that first mRNA is made, and then that goes, and then you translate that to make some monomer, and then the monomer has to maybe fold, and then the folded protein maybe has to dimerize in order to do a repression.

So if you have a more detailed mechanistic model, that includes all these steps, that kind of introduces some sort of delay, that in principle can lead to oscillations in such a circuit. Or if you wanted to, you could just explicitly put in a delay. So you could say that x dot, instead being a function of x , instead what you can do, is you could say, well its actually a function of x at some time, t minus τ . So instead of having the rate of production of x be a function of x , at that moment in time, instead it could be a function of x at some previous time. Doing that, that's a very explicit form of delay.

And that can also be used to generate oscillations in a simple negative auto

regulatory loop. These are all different kind of approaches for encoding delays into a model and in various approaches will give you oscillations. Yes?

AUDIENCE: Question for the repressilator, when you say the period is not tunable, it's because the mRNA lifeline is very difficult to--

PROFESSOR: All right, when we say--

AUDIENCE: --in the model you can--

PROFESSOR: Yes, that's right. In the model you can, in principle-- So what I mean when I say this is that in this class of model, so you could also have, instead of this repressilator with three, you have the so-called pentalator, where you have five proteins and each is repressing itself. So these all have similar features, so all have these odd numbers of proteins going around and repressing one another. And so you can write down the model with seven, if you want. But in all these cases, it's not tunable. What we mean by that is that, when you tune the frequency, you in general lose the amplitude of the oscillation. So the amplitude will go down.

There was a very nice paper that was written in 2008 on this topic, written by Jim Ferrell at Stanford. So I just want to mention this. So its Ferrell, at Stanford, this is a paper in *Science* 2008, and it's called Robust Tunable Biological Oscillations from Interlinked Positive and Negative Feedback Loops. So nice title, I like titles that say something. So it's sort of the ultimate short version of an abstract, right, if you can do it I recommend it.

Incidentally in graduate school, I once wrote a paper with four words, short words, DNA over-winds when stretched. Nice statement, you may or may not actually know what I mean by that, but it's a nice short, title, it's a statement. I encourage you to think about that when you're writing your papers.

So he wrote this paper where, he said, all right, well, oscillations are really important. Thinking about context of heart rhythms, or cell cycle, or this or that. Oscillations are important, but if you go and you look at the circuits that are generating oscillations in biology, they often have so-called interlinked positive and

negative feedback loops.

There are many cases where you have, some x that actually is positively, it's kind of activating itself. And this is very much something that will not lead to oscillations on its own. It might be bistable, which is interesting, but not oscillations on its own. But then there's also maybe a negative feedback loop through another protein.

And the idea is that this one somehow operates-- this one's fast, and this one's slow. And the key feature of these relaxation oscillators is they are two time scales. And it's the slow time scale that specifies the period of the oscillation, and this fast one kind of locks the system into these alternative states. And this helps maintain the amplitude, because it has this nature being bistable, right, so it's on or off. So this helps you maintain amplitude, so this is kind of in charge of amplitude, and this one over here is in charge of the period.

So what you can imagine that by changing this time scale, you change the period of the oscillation, whereas this loop allows you to maintain the amplitude. And what Jim's group did computationally in this paper, is they analyzed many different circuit designs that can lead to oscillations, and they showed that for the loops that are made of purely negative interactions like this, if you change a parameter in order to change the period, you'll also in general make the amplitude of the oscillations drop dramatically.

So that's the sense in which they're not tunable. Whereas if you have this kind of design, you can actually tune over, in some cases a very wide range, but maintain the amplitude of the oscillation. And in addition to being tunable, these things also end up being robust in various ways. The oscillation is maintained subject to various kinds of-- If you twiddle with the parameters, you double this, you have that, you still get nice oscillations here, whereas in those designs you tend to lose the oscillations more easily.

So they claim that based on that, that these might be more evolvable. So even in cases where you don't need to tune the period, maybe you still end up evolving towards this design, just because it's robust to stochastic fluctuations in the

concentrations of things, but also it might be easier to evolve these sorts of oscillations.

Are there any questions about the kind of intuition behind this for now?

There's a nice kind of circuit analogy that people often talk about in the context of this. So if you imagine you have some battery, with some voltage, v , well, we'll say v battery, some capacitor over here, but over here you have something that will spark at some voltage, some v_t , you get a spark. Now the question is, well, what happens over time if the threshold is less than v battery? We maybe should have a resistor in here.

So the threshold is less than v battery, then this can generate nice oscillations in the voltage say across the capacitor as a function of time, that are tunable. Because if you plot as a function of time. This is the voltage across the capacitor, where up here we might have the v , the battery, here we might have v threshold. Now in the absence of-- This thing's that's going to short periodically, we're just going to charge up the capacitor. So in principle, there's going to be this standard r, c time constant, coming up to here, but before we get there, we get the spark. So then we discharge across here and this drops. So you get something that looks like this.

Now you can imagine by changing, for example, the resistor, you can change the rate that this thing, the capacitor, will charge up. But the amplitude of the oscillations stay constant, because that's set by the voltage threshold across this-- where it shorts. This is capturing this dynamic of the separation of time scales. So there's a slow time scale, which is this r, c time constant, and then there's the rapid time scale is where this shorts out. So you can imagine that this is an example of an oscillatory signal that we can tune the frequency without sacrificing the amplitude.

What we've said so far is that there are engineering analogs to these sorts of relaxation oscillators. We can model various synthetic circuits, or we can look at natural oscillatory networks, in order to get a sense of what's going on. But of course, a major goal of this kind of system synthetic approach to the field, is that if all this stuff is really true, we should be able to build it.

And there's a very nice demonstration of this, also in 2008, by Jeff Hasty's group. So Jeff Hasty was actually trained as a high energy theorist, and then I think it was during his postdoc, maybe he switched into experimental biology. Went and did his postdoc, I think, with Jim Collins. And then eventually now has his own group doing systems synthetic biology.

In this paper, it was a *Nature* paper in 2008, it's called A Fast Robust and Tunable Synthetic Gene Oscillator. It's a nice statement, tells you what he's about to do. The data here, this is again using this basic insight of having both interlinked positive negative feedback loops in *E. coli*. He demonstrated that he can get really beautiful oscillations, in essentially all the cells, and that they're tunable, enter n period, by a factor of three, or four, or so, by a fair amount. And indeed as fast as 13 minutes, the oscillatory period. Which is pretty nice, right? So I encourage you to check out this paper.

This paper was also an example of how it was in principle possible to get oscillations just by doing negative auto regulation. Right, so this was a case where they designed a gene network that they could tune and had this wonderful property. But then after they did that they noticed that in their model at least, they could get oscillations in some parameter regime, just by having the negative auto-regulatory loop. And as a result of all these intermediate processes, of protein maturation, and so forth, and then they went and they constructed that network, and they showed that that could also oscillate.

So again this is an example of the interplay between modeling, experiment theory, modeling, And Jeff Hasty has gone on to write another several, really beautiful papers looking at these sorts of oscillations, looking at how you can get synchronization of oscillators, and you get period doubling ideas. It's really a whole string of wonderful, wonderful papers. So I encourage you to, if you're interested in oscillations, to look at Jeff Hasty's work over the years.

If you want a quick introduction to these papers, I also wrote a news and views in

Nature on these two papers. So you can read that, it's only a page. Although I guess you won't hear anything that you haven't already heard probably.

Any other questions about this idea of how we can use both positive and negative feedback in order to get some nice oscillatory properties? OK, then let's move on.

What did you guys think of this paper, the Barabasi paper? Good, bad, difficult, easy?

AUDIENCE: Why does it have so many citations?

PROFESSOR: Why does it have so many citations? All right that's an inter-- and you should look at how many-- according to Google Scholar, I haven't checked this year, but it's probably 20,000 citations. I mean it's--

AUDIENCE: Is it a cult thing?

PROFESSOR: It's a cult thing. Well, I don't know. That might be exaggerating.

AUDIENCE: I mean it's a nice paper.

PROFESSOR: Yeah, right. So this is interesting, and I think that the basic answer is that there are networks that are relevant in many, many, many fields, which they allude to. And there are many researchers that have been excited about studying those networks in many, many fields, and many, many, many of the networks that are observed in nature or social science, the web, everywhere, they have these power law structures. And this is the first clear simple mechanism to generate it. My understanding is that actually a mathematician decades before, actually did demonstrate that this kind of thing could be constructed, that would lead to this, but that paper doesn't have 20,000 citations.

I mean like it's a lot of these things, you have to be the right time, right place, and have the right idea.

AUDIENCE: Yeah, I guess my main thought about the paper is exactly that, the interesting thing about it was, it came out at about the time that data on large networks was readily

available.

PROFESSOR: That's right. There's a reason that this paper was published at this time, and of course if Barabasi didn't do it here, someone else would have done it a year or two later. But it was really that the data were available everywhere, and we were seeing these power law distributions, and it's really crying out for an explanation. I think it's-- You know sometimes people complain, that they say, oh yeah, you know I could have come up with this idea, it's not that deep. And maybe you could have, but you didn't.

[LAUGHTER]

PROFESSOR: And also I'd say that Barabasi has a record of doing interesting things, and being the first to point out a simple idea. If you can reliably be the first to point out a simple explanation for important things, then that's another kind of genius, right? I mean-- and it's the kind of genius that I aspire to, because I know that I'm not going to reach the other kind of genius. I mean there are some things you look at, oh well, I would never be able to do that, right? And everyone agrees that that's hard.

But I think that there is something about being able to see what the scientific opportunities are at a given time, and you don't have to come up with a really complicated model or proof in order to have really important impact. And this paper is way beyond, in terms of number of people that have read it, cited it, and so forth, it's way beyond, probably any other paper you'll likely read in your life. Yes?

AUDIENCE: Just more thoughts.

PROFESSOR: More thoughts. Yeah, that's fine.

AUDIENCE: I guess, what's interesting is, it starts a conversation, so we can analyze these networks, and the feature that works the best [INAUDIBLE] starts that conversation, there's still a lot more that we can do with it [INAUDIBLE]. I think that's why I like it.

PROFESSOR: That's a totally-- so this is the Barabasi and Reka Albert. So Barabasi is a professor over at Northeastern now, Reka Albert is a professor at Penn State, and I think

they've both gone on to do what, I think, are really very interesting things in this network space, and more generally. So I you encourage you to check out what each of them have been doing over the years.

All right, so this model, what are the two key ingredients of this model?

AUDIENCE: Growth and preferential attachment.

PROFESSOR: Right, so the two things you should be able to recapitulate on an exam is that there are two assumptions here, there's growth, and there's preferential attachment. And we'll talk about the degree to which we think each of those things might be necessary, but what's the key-- Can somebody be a little more explicit than what we've been so far about what the key observation is that we're trying to explain?

AUDIENCE: There are nodes that, inactive nodes, that have more edges than you expect, either by random or--

PROFESSOR: Right. Perfect, right. Observation-- So some nodes have lots of edges.

AUDIENCE: I mean it is sort of a meta irony that this paper is now very widely cited.

[LAUGHTER]

PROFESSOR: Yes.

AUDIENCE: Every time we talk about it--

PROFESSOR: Yes. Indeed it is ironic, and we'll talk about how the scaling of the citation networks go in a moment. Right, so some nodes have lots of edges, and you want to be very clear about this, this is what you're trying to explain. So it's a power law distribution, so in particular, you quantify this thing. That the probability of having k nodes, or I'm sorry, the probability that a node has k edges, falls off as a power law, 1 over k to some power $2, 3, 4$.

So the probability of having k edges, goes as 1 over k to some power α . Where α is maybe between 2 and 4 , for a lot of these. Now it's important to make sure

that you keep this qualitative statement in mind, because it's true that it falls off, and sort of rapidly. Right, 1 over k squared, k cubed, or k to the fourth, you'd say, oh, that's a pretty rapid fall off. Right? But it's not rapid compared to what?

AUDIENCE: [INAUDIBLE].

PROFESSOR: --exponent. Right. So for these other models then, it falls off exponentially. Right? So even faster. So it's easy to look at 1 over k to the fourth, and think, oh, that's a fast fall off. We have to remember that it's slow compared to some other things. So in particular, if you look at the data for real networks, and you see that the probability distribution in many cases goes over orders of magnitude in terms of this probability. You think oh, that's a big range. And it is a big range, but the fact is that you actually see some nodes with the thousand ideas or whatnot, which is something that you would just never see, if it were a random network, or if it were not a power law distributed network.

And I think that this is also highlighting another statement, which is that a powerful way to make a difference, for example, if you're going to write down a model, or you're going to do a theory, is that it's nice if there's a clear observation that needs to be explained. Because you can always write down a model of something, and maybe you'll find something interesting. But a way to massively increase the probability that you're going to discover something interesting is if you already know there's something interesting there and that you're trying to explain it.

And I think that this is an example of that, right there, it was already an observation, it was already known. It's not that he was the first person to make those plots. There are other plots of citation networks before. So Sid Redner, for example, had already done some analyses of citation networks, he's a theoretical statistical physicist over at BU, but just now, I guess, moving over to the Santa Fe Institute. But it's not that he was the first person to make that observation, but he knew there was something interesting that needed to be explained. So I'd say that for any of you that are thinking about doing theory, or writing down models, I would say, whenever possible start with an interesting observation. So can somebody-- maybe you guys could just

throw out, what are some examples of nodes and edges that were given there or elsewhere?

AUDIENCE: Web pages and links.

PROFESSOR: Right, web pages and links. And is this a directed or undirected?

AUDIENCE: Directed.

PROFESSOR: So this is indeed directed. Some others?

AUDIENCE: Movie stars and movies.

PROFESSOR: Movie stars and movies. This one's a funny one, rig-- So movie stars and then this is like being in a movie together, right? So co-starring or so. Others?

AUDIENCE: Articles and Citations.

PROFESSOR: Articles and citations. And this is again directed, and this is not directed, right? And we can maybe even try to remind ourselves, this fell off as alpha was equal to what? I guess it was 3, I think they said. Actors work around 2.3, I guess they said. The web was 2.1. Just because it's a power law doesn't mean that it's always going to have the same alpha right?

But for example, what this means is that for every paper that has say 200 citations, there are going to be roughly 10 papers that have 100 citations. If you increase k by a factor of 2, you get almost an order of magnitude in terms of the probability distribution.

So this is an interesting observation, and where Barabasi came in and said, well, what would be a model that would recapitulate this? And what are the models that did not recapitulate it?

AUDIENCE: [INAUDIBLE].

PROFESSOR: Right. So the Erdos Renyi-- so other models, there's the E R, other models, there's the Erdos Renyi network, random network, and that's because here the degree

distribution is peaked around something and then falls off exponentially as you go above that. And this is actually where I think the equations are wrong in this paper. Because if you look at the paper, page 510, where they say the Erdos Renyi, you connect the edges of probability p , and then they say you get a poisson distribution, p of k , where λ the mean is something, but then they say, oh λ is equal to some binomial of something of k , and so forth. So I think this is all not true, but rather that you can approximate the binomial with a poisson in the limit of small P s. So be aware if you're looking at that.

I know-- there was another network that-- Do you have a question?

AUDIENCE: Oh, no, no. I was--

PROFESSOR: So we're going to spend a lot of time talking about probability distributions in the coming weeks, but I just wanted to highlight that there, as far as I tell, that is not true what they say. But there was one other model for a network that they talk about, or they mention. Does anybody--

AUDIENCE: Small world.

PROFESSOR: The so-called small world network, right? And this is-- small world network, and this is based on a paper by Strogatz-- Watts and Strogatz, small world. That's Watts and Strogatz, and this was a paper where they demonstrated that there was a very simple mechanism. Just by rewiring a network that you could get this so-called small world phenomenon.

Where the Kevin Bacon thing, where you can take any-- You're right, from Kevin Bacon, and this is actually the actor network, so you could say, starting with Kevin Bacon can you construct a list of actors that costarred with each person that gets you to any given actor. And the statement is that you are supposed to be able to do that from a path of six. So that all the actors are supposed to be connected to Kevin Bacon by six. Although maybe you guys don't even remember who Kevin Bacon is anymore. Oh, you do? OK.

This rule works for anybody so just insert your favorite actor into that sentence. And

it's important, just to mention that just because something is a small world network, does not mean that it has power law distributions. It may be the case that many power law networks also have this small world character, and I'd say maybe even most of them, because some of those highly connected nodes are going to be useful for connecting anybody to anybody else. But that's not required to get the small world character. Any questions about that statement?

AUDIENCE: So you can go from this small world statement to any sort of strong statement concerning connectivity?

PROFESSOR: Well stron-- I guess that the strong statement is that this property does not imply that property.

AUDIENCE: You're not saying that the universe is true [INAUDIBLE] because it seems like, at least the examples we've listed, ought to be small world.

PROFESSOR: Yeah. I agree. I think that this small world property, that's why I saying that, it's-- What I do not know, it's whether it would be possible to construct a power law distributed network that does not have the small world property, but I would say is that the ones that I'm aware of would have the small world property arm. Any other questions about where we are?

So there's interesting properties of networks that we would like to explain. And I would say that what this paper does, I think kind of convincingly, is that they demonstrate that at least this model, and we'll get into the assumptions, does lead to a power law distributed network.

The answer to the reading questions about whether both of these is strictly necessary, I think was an interesting one, and I'd say that this gets into the wider issue of there's a observation that is maybe interesting. And then we want to understand why that might be, and then what you can do is you can write down a model that leads to that behavior. We've already talked about. Does that prove that the assumptions of the model are correct? No.

In this case, these are pretty generic features of lots and lots of the network. So

when you read it you kind of believe that this is a dominant mechanism, but it very much does not prove that these are the only, this is not at all the only way to get a power law distributed network. I'd say that some of the language in the paper might kind of lead you to believe that that is the case, and I think this is a standard logical fallacy that we have to be careful of, and something I think that some the language is a little bit dangerous.

The development of the power law scaling the model indicates that growth and preferential attachment play an important role in networ-- I'd say that it's quite true, but once again this question of-- This is certainly not a proof, that those assumptions are relevant for any given network. Of course, in all of these cases, the network does grow, and there is preferential attachment. But there are other things that are also true, that may be important, for example, in determining exactly what alpha is or in other things. And I think that as indicated that there are other ways of getting power law networks without making the exact assumptions that are here.

But its, in my mind, it's probably a or d dominant mechanism in a lot of these networks. I think it's a fine paper, but just remember that it doesn't prove that those are the only two important things. Yes?

AUDIENCE: Just above the preferential attachments, I think you mentioned that you tried different ways, and only the linearly one was

AUDIENCE: [INAUDIBLE].

PROFESSOR: Troubling.

AUDIENCE: [INAUDIBLE].

PROFESSOR: I agree. I agree. And what they assume in the model here is that the preferential attachment goes linearly with the number of existing edges. And I would say that I very much believe that preferential attachment is present in all those things, but I'm sure that if you go and you measure it you're not going to find that its linear with the number of edges. It's going to-- actually, I don't know what you'll find in each of

those cases, but there's no reason to believe it has to be linear.

That being said it may be, the question is how strong of a deviation from linearity is there? And then how sensitive is the power law behavior to that? And that's the kind of thing that I'm sure that one of the 20,000 papers that have cited this paper in the last 15 years address this issue. Yeah, but I mean, this is also why there are so many papers that have cite-- It's like you read this paper , like oh, you know, it would be really interesting to do this, tha-- and people have been following that interest.

Let's go and-- I think that the derivation is a little bit tricky, and so I think it's worth just walking through it. Especially since some people apparently couldn't even get the equations, which is going to be a problem.

Maybe while we're on this question of preferential attachment-- How do you guys feel about this question of networks within, say the transcriptional network of E. coli or other cells? I mean do you think that these properties are relevant in the cell or--

So what would growth mean?

AUDIENCE: [INAUDIBLE].

PROFESSOR: So growth would correspond to adding a new gene. Does that ever happen?

AUDIENCE: Yes.

PROFESSOR: Can some given a possible mechanism by which a new gene is added to the genome?

AUDIENCE: Duplication.

PROFESSOR: For example, duplication is common, right? eg. duplication. So what does this mean for preferential attachment?

AUDIENCE: --duplicate the gene and it will probably also duplicate the promoter region, which means--

PROFESSOR: Right. So this, I think, is very interesting. So duplication, in general you'll duplicate both the coding region makes protein, but also maybe the promoter region that specifies the regulation. So if you imagine you have some x here that is-- And we can remind ourselves, are both the incoming and outgoing edges power law distributed in transcription networks? No, I know this was in the pre-class reading, but just in case.

So what you find is that some transcription factors regulate many genes, but we don't have any proteins that are regulated by 200 genes, so in that sense typically we have the things that are regulated, there's maybe some x_1 , x_2 , x_3 . And there might be a few incoming edges, so the expression a gene is typically specified by a few transcription factors. Whereas some transcription factors might have 100 outgoing edges.

So it's the outgoing edges that are power law distributed, and the ingoing are closer to being plus on or so. So you can imagine that this guy might have 100 or so, whereas over here some y transcription factor that is just regulating two genes, say y_1 , and y_2 . Now, question is, if gene duplication occurs kind of randomly throughout the genome, which transcription factor x or y is more likely to have a target that's duplicated?

AUDIENCE: x .

PROFESSOR: x , all right. Interestingly, how does that scale with the number of targets?

AUDIENCE: Linear?

PROFESSOR: This actually is linear, right? So I'd say that gene duplication does give growth and preferential attachment that is basically linear with a number of targets. It's interesting I'd say I find this kind of observation quite interesting, and compelling, and makes me feel kind of comfortable about this as a mechanism for some of the global properties. I mean there's no selection, there's no way to explain the interesting network motifs and so forth here, but I'd say just in terms of some general properties I think it's interesting.

Of course, once again not a proof. Evolution can do whatever it wants with these gene duplication events, but also I would say not everybody finds this argument very, very compelling. But I'd say I think it's kind of-- I get a warm fuzzy feeling inside.

AUDIENCE: We're talking about transcription network, it's different from the other networks you were talking about in that you also lose genes, and so is there any discussion--

PROFESSOR: Well you know, you could lose web pages, you can--

AUDIENCE: Are you losing them nearly as fast as you're adding them?

PROFESSOR: Yeah, I don't know. I find that lots of links to my web pages just disappear over time, and I-- It's a reasonable question. I don't-- In some of these you say, oh well right, so with the web has been growing a lot recently, and so then we'd say the birth dominates over death there. Where if you talk about genome sizes along different lineages, it certainly is not growing exponentially the way the web pag--

I think that that's fair and true, but we haven't really actually specified or made clear, within a model what happens if you allow for birth and death. But I think that you could introduce death and recapitulate these behaviors, so it's not-- I think just because some nodes disappear, doesn't mean that we have to throw the whole idea out the window. But in the presence of evolution this is all very complicated, right? So you can't carry this argument too far.

AUDIENCE: So it's [INAUDIBLE].

PROFESSOR: Well what we're assuming is that there is some segment of DNA that's in front of the gene that specifies-- gives instructions of when to transcribe the gene. So the linearity is really just assuming that genes have the same rate of being duplicated on average. And this is a very global property, so I think that it's kind of roughly-- I would say it's the middle model that you would use, if you're had to write an old model.

AUDIENCE: Is there anything in looking for evidence to support [INAUDIBLE].

PROFESSOR: That's an interesting question. It's hard to know what it would even mean to collect the evidence to support it in the sense that-- You're saying along different evolutionary lineages, could we say that it's more likely to grow. Of course the other thing to say is that, the rate of death would also scale linearly. In the sense that a gene being stochastically removed from the genome should also scale linearly, so it's not that you don't actually then expect there to be any systematic change.

I mean it's not as simple as just saying, oh the number of targets of a transcription factor with many targets should grow faster. It's really that the expectation is that it should be changing faster because both duplication and removal would both be increasing. So I think the signature is not totally obvious in that sense.

So how many people actually tried to piece this derivation apart? Anybody? All right, and were you happy with it at the end of your--

AUDIENCE: I think that--

PROFESSOR: --permissions?

AUDIENCE: --that I was a little bit iffy about.

PROFESSOR: There is like a crux of the climb at the end. So let's make sure that we can understand what happened there. It's worth-- since we read the paper it's worth trying to figure it out. So what we're going to assume is that we start with m_0 nodes. So they're going to be here, and the idea is it doesn't really matter how we start this thing.

They might start out being unconnected, or they might be connected. But over time the signature how we start is not supposed to be that important. What we're going to do is at each time point we're going to add one more node. And as we do that we're going to add m edges as well. So we then have the number of, we'll say, nodes, N , as a function of time, is going to be equal to what?

[INTERPOSING VOICES]

PROFESSOR: Right. This is just going to be-- we're going to start at m_0 and we're going to add 1 each time, m_0 plus 2. Number of edges is just going to be equal to the number that we add each time point, times the time. So here we're assuming that we start out with these nodes being unconnected.

Now we're given the assumption that there's preferential attachment, so that means that the probability of connecting to some i -th node that has k edges is going to be k to the i divided by the sum over all the edges. Yes?

AUDIENCE: Why is [INAUDIBLE]?

PROFESSOR: All right, so the assumption is at each time point we add a new node, let's say this node, and with that we bring in some number, n , of new edges. So this could be 3, and then we go randomly to 3 of the existing nodes. So each time point we add m edges.

AUDIENCE: How do we necessarily add them to the new node? Like [INAUDIBLE].

PROFESSOR: I'm sorry I don't understa-- oh yeah right, so the assumption is that the new node is indeed being connected to-- that all m edges that we're adding are to this new node. So this is the linear preferential attachment that we were talking about.

So what we want to know first, is how after a node is connected, how is it that number of edges will grow over time. What we know is that when it's first added it has it exactly m edges, right? But then as new nodes come, then we'll maybe get some more and then it'll grow.

And in particular we want to get-- We're told that it's going to grow as this differential equation, so we want to kind of get to this. And the way to think about this is that, all right well, how is it that the number of edges will change at each time point, so Δk_i . Well the expected number of edges that will be attached to some node, well that's going to be m , this is the number of edges that were attached by this incoming node, times this probability of attaching to this node. So this is the probability of k_i .

Now this is in one time step. So this is really a Δk_i . If we want, we could say over some Δt , which is 1. So from that standpoint, we can actually then write it as differential equation, where you say the change in this number of edges with respect to time is indeed going to be equal to m times this guy here, which is the number of edges that that node has at this time, divided by this sum over all those edges. This is just kind of the expected number of edges to be added to that node at each time point.

What does this thing-- What does that thing equal to? Yes?

AUDIENCE: --that equation, because it seemed like you just wrote the same equation on the line above that line. You just substituted it--

PROFESSOR: I did.

AUDIENCE: OK, but [INAUDIBLE] wrote it as [INAUDIBLE].

PROFESSOR: Yeah, so this is kind of the discrete version of this differential equation.

AUDIENCE: Oh.

PROFESSOR: Right. Yeah that's right, that's right. And of course the beginning could be highly stochastic but we're just thinking about in the limit of if it's deterministic. What is this thing in terms of-- from here this is just a normalization constant, right? Because each edge has to be attached somewhere, we're assuming it's linear with respect to the number of edges at each node, right? And that means that for normalization we have to divide by the sum over all those edges, the edges that each of the nodes might have.

What is this thing equal to in terms of something else that we might have on the board? Yeah?

AUDIENCE: These have edges with respect to [INAUDIBLE].

PROFESSOR: Right. So I guess the question is this, can we write this? Where E is a function of time? Is that correct? So we're getting some shakes.

AUDIENCE: Isn't it $2E$?

PROFESSOR: Right. So it's actually $2E$. Because what you notice here is that this is the sum over all of the edges that each of the nodes have. But each edge is connecting 2 nodes. So the sum over all these edge distributions is twice the number of edges. Now I would say as a physicist, working in biology, my general attitude is that a factor of 2 here, factor of 2 there, doesn't really matter. But this factor of 2 actually is relevant because it ends up determining the scaling over time. So not all factors of 2 are created equal, and this is one that is worth paying attention to.

Does everyone here understand why this is 2 times the number of edges? k_1 is equal to 1, k_2 is equal to 1, number of edges is equal to 1. Yeah.

AUDIENCE: So that means we're in an undirected network, if we were in a directed network, then we would not have that factor of 2.

PROFESSOR: Yes. So we are indeed in an undirected, and I'd say in a directed network you have to then be more careful about what you-- you have to specify the k 's in and k 's out. So actually, already just by writing this we've already assumed it's undirected, because we haven't specified what we mean by k .

We're here, but very conveniently we already know how many edges there are as a function of time. This is just equal to m times t . So we get something that's very convenient k_i divided by $2t$. From here we can solve the differential equation. This is what we want to show.

The fact that we're doing partials doesn't really matter, because it's just time here. So it's really-- so we have $d k_i$ over k_i , is equal to dt over $2t$. This 2, really again, is going to make a difference, because when we go and we integrate, we get the logs and so forth. And so we get that k_i as a function of time is going to grow with time, with some constant c , proportionality to the square root of time. So if we didn't have the half it would just be linear with time.

Now how do we know what c -- in general how do we get constants of integration in

life?

AUDIENCE: Boundary conditions.

PROFESSOR: Yeah, boundary conditions, in this case, the initial condition. And what is it that we know?

AUDIENCE: k_i .

PROFESSOR: Right. So what we know is that k_i , so this i -th node, when it's added at time t_i , it should be equal to what?

AUDIENCE: m .

PROFESSOR: Yeah. It's equal to m . So when it's first added, at some time t_i , its number of edges is equal to m . Because that's what we've assumed, is that we add a node and we connect it randomly and other things, so it has m edges initially. So from this kot, this is then equal to m times the square root of t divided by t initial. Where t_i is the time that i -th node was added to the network.

Are there any questions about how we got there?

So I think that this is relatively straightforward. The part that gets confusing is this later part about the probabilities and keeping everything straight. And so what Barabasi did next, is he said, all right, well, what we're going to do, is we're going to talk about the probability, P . Now this is an actual honest to goodness probability. The big P is actually a probability, and that's as compared to a probability distribution, little p .

And I'll put in a little curly here thing, so it's a little p . This is saying if you want to get an actual probability here, then you have to multiply that probability distribution times some range Δk . If you want to know that the probability that some node has between some number and some number of edges, then you multiply it by that range. Right?

Probability distribution, this is an actual probability. And as befits an actual

probability, we're going to say, OK the probability that the i -th node has k edges, that are less than some value k . And remember this thing is actually a function of time.

But we have an expression for k_i as a function of time, it's equal to this. So we can solve when we show that this probability is also the same as this other probability. That the i -th node was added after some time t that can be written as this. So this is saying, the probability that some random, say i -th node, has fewer than k edges, is the same as saying it's the probability that the i -th node was added after some time, t , which is this thing. Because the number of edges will grow over time for each of these nodes.

Do you understand that kind of conceptual statement that was made there? Yes?
Any questions?

All right, so the probability that this i -th node was added after this time, is also of course 1 minus the probability that it was added before that time. Whereas time, little t here, this is at the time that you're actually looking. So this is saying, oh well, if little t is 100 , for example, it's saying all right, at that time point after I got 100 nodes, we want to say, all right, what's the probability that some random i -th node was added before this quantity. And this is just again some other kind of time, if you'd like.

I think this is the part that it is especially kind of weird. So this is also equal to this thing. And I think reasonable people can argue about exactly what you should write here, but let's figure out the basic argument first. So there's this probability is equal to this thing.

So this statement is really that at some time t we have how many nodes? We have m_0 plus t nodes, right? So this is something here. And of course there are edges going around doing things. And what we want to know is, what's the probability if I grab one of them, we're going to call that the i -th node. What's the probability if I grab one of them that it was added before sometime here. And it's useful to just imagine this is as just being some time t , just so that we don't get confused by all the symbols.

You say, oh well, that probability is really just the probability-- well how many nodes total do we have here, m_0 plus t . How many nodes were there that were added before this time t ? Well that's going to be t , you might want to say t plus m_0 . There's a question of whether you include those nodes that started there or not. Given the equations that Barabasi wrote down, he kind of assumes that we're only counting the nodes that were added later.

So I'd say if you want, you could either add an m_0 up there, or get rid of this m_0 , depending on what you like. But broadly there's this idea that we have this many nodes, and this many of them were added for some time t . And that's how we get this $m^2 t$ over k^2 was just that time t divided by the total number of nodes.

And this whole discussion about whether you count the initial m_0 nodes or not, it doesn't matter because we're going to take the limit as t goes to infinity, and that all goes away. Are there questions about this? There is something kind of mind twisting about this argument, even though we're really just picking big T objects out of essentially little t objects, but somehow something funny goes on there. Any questions about that?

AUDIENCE: Could you just go through the argument one more time?

PROFESSOR: Yeah, sure, sure Right so I think that what's confusing about it is the fact that we're asking whether the i -th node was added before some time t . And this time t is equal to something that's funny based on what we've just done. But it's useful to just ask, if at time little t you look at this network and I ask you, all right, was it added before this time, big T . Let's just for concreteness say m_0 is equal to-- we start with 10 nodes. And we say, OK, at time t equal to 100, I ask you, what's the probability that if I grab a random node, what's the probability it was added before some time big T equal 10.

Well you would say, very roughly actually. We can say let's actually, we can even if you'd like, say we're not going to count-- we're not going to count those m_0 initial

nodes. So we're just going to be looking at nodes that we added later, if you'd like. And then when you would say, all right well, at time $t = 100$, we've added 100 nodes.

And I'm asking, if I grab one of the nodes, what's the probability that the node I grab was added in the first 10 time steps. Well you'd say, it's going to be 10%, because there were 10 nodes that were added before time T , and we added 100, so it's really just this divided by this. And with the question of whether you want to include m_0 's or not.

So I think that that argument is surprisingly straightforward, but somehow it gets really confusing is that the time t we're referring it's depending on the k 's and t 's and so forth. But that's a way of keeping track of how are things scaling as a function of time. But if you boil the argument down to this, then it makes sense, but then of course then you look back at this and you get confused you again. Which is how I feel every year when I prepare this lecture, but I think it all does make sense if you--

Any questions about this argument or that argument or any part of it? Yes?

AUDIENCE: So the t_i 's are very important [INAUDIBLE]?

PROFESSOR: Yes. So this is just saying that if I pick some random node, we're calling it the i -th node. I'm asking what's the probability that the time that was added was before something. So this is not one of the variables, and you'll see the t_i doesn't appear down here. Because this is just saying-- I'm asking you, if I grab some random node, the i -th node. I'm asking you, what's the probability that it was added before some other time, which is all this. And what you can see is that it's a function of the time that we look, because if I go to longer times you know then indeed this probability should it go-- What should it do?

AUDIENCE: [INAUDIBLE].

PROFESSOR: OK, but it depends on k 's as well, right? What do I want to say?

Ultimately what we see here is that as time goes infinity, so after a long time, then we reach this stationary distribution where the base structure of the network is not

changing anymore. And that's because there's a t in both the numerator and denominator. So then the only thing that is left is this behavior as a function of k .

And this is really saying that the probability that some node was added before some time, is kind of the same as saying that, well, that you have a lot of edges. And that's how we got here to begin with, because the nodes that were added early end up with a lot of edges. This is the so-called rich get richer phenomenon. So if you're sitting on a manuscript, and you're not submitting it for publication you should get on it because the earlier that it's published the more citations it's going to get.

But this is saying that the probability that some random node has a small number of edges is the same as that the probability that the node was added late. And that makes sense, because if it's added late it doesn't have very many edges, hasn't had time to grow. And then from those calculations you get it at this degree distribution. Yes?

AUDIENCE: So for this analytical [INAUDIBLE] we're assuming the links could be [INAUDIBLE].

PROFESSOR: Yes. So we're taking, in principle it's a discrete problem and converting it into a differential equation. And it's an interesting question of I don't know how big of an error this ends up making, and of course this expression doesn't actually end up having integers. But this is a way of making it so that the errors don't grow or so, right? I think that it basically works. If you'd like you could actually do the simulation with all the discrete-- I think that is actually going to be the stochastic dynamics that end up being more relevant than the integer kind of issue, but I haven't actually looked into that though. Any other questions about that so far? Yes?

AUDIENCE: [INAUDIBLE].

PROFESSOR: So there's no loss of edges, no loss of nodes, strictly verboten. I spent a lot of time trying to plan an upcoming trip to Germany last night so German is on my mind.

So are we done yet incidentally? Nearly right? Because we have-- What we really wanted is the degree distribution, not this probability. So we have to take a derivative still, but as t goes to infinity, regardless of how you treat the m_0 's, actually

what we-- maybe we'll take the derivative first. So this probability density is going to be the derivative with respect to k of the actual probability here.

So we take a derivative, this one derivative that nothing happens, case squared, it's going to turn into a k cubed. So we get $2m$ squared t over k cubed, we still have the t plus m_0 , but when we let t go to infinity, so after this thing has reached its stationary distribution, then we end up just getting $2m$ squared over k cubed. I just want to be clear this is to the k .

The key feature here is that the probability distribution goes as 1 over k cubed. What is interesting is that when I first read the paper I actually thought that this exponent here would be a function of the linearity of the preferential attachment. So I actually-- and of course they say that it's not true, but when I was halfway through the paper I thought, oh well, if you just let this go as some power to the beta, or so, that you would maybe get something like this was 2 plus beta-- I thought something like that, but apparently it's not true.

That if you do not have linear attachment here then you just don't get power law distributions. They suggest other ways that you could maybe get different exponents, which is very relevant given the fact that different real networks indeed have different exponents. But I'd say that their proffered explanation, which is to include directed edges, feels unsatisfying because not all networks are directed. And this network here is not directed, it has next exponents closer to 2 . So you really want to have other mechanisms. But this is as we mentioned, it's a thriving field and people have explored many different aspects of this problem.

Are there any other questions about this derivation, how we got there, how convincing maybe you think it should be or not be?

So I want to just spend the last five minutes of the class kind of setting up the discussion of how we should be searching for network motifs. In particular there's a natural question which is, we have to decide what the right null model is, in terms of deciding what the expected frequency of a network motif, like a feed forward loop might be.

So first of all, why is it that we maybe should not use an Erdos Renyi network? Yes?

AUDIENCE: Because it's not very good for handling directed networks?

PROFESSOR: Right. So you'd say, oh, not very good-- I can maybe make-- there's a clear analog to it-- you could take a random undirected ER network and say put arrows randomly on each-- I mean I think that there's a natural ER version of a directed network.

AUDIENCE: There are constraints.

PROFESSOR: Like what?

AUDIENCE: Like when you [INAUDIBLE] duplication, you don't randomly assign the edge.

PROFESSOR: That's right. OK, so one thing is that it may be that biologically there are constraints, but that should manifest itself somehow. In the sense that if, you know all that may be well and good, it may be true, what you're saying, but if we go and we look at a transcription network, if it looks like an ER network, then I would say it just doesn't matter. The fact that there's microscopic things going on, I mean if at the end of the day it looks like an ER network, then maybe it's fine anyways, right?

AUDIENCE: Hum.

PROFESSOR: Or maybe not. You can argue either way.

AUDIENCE: It depends on what you want. If a particular motif occurs a lot it might be because it's selected for it, but it's not what you were-- --it's for some other reason.

PROFESSOR: That's right. So this is an important point, that I would say that in Erdos approach, he basically says if we see a network motif more frequently than we would expect based on some null model, some null network, then it's kind of prima facie evidence that maybe evolution was selecting for it for some reason. And what you're saying is that it could be there's a microscopic mechanism that just leads to those things happening, and so it doesn't have to be selection, it could be just due to the mechanistic processes below. And I think that's a fair concern.

And it's related to a lot of these other things, in that just for example, duplication will naturally lead to something-- if you start out with x regulating Y, and Y is duplicated then now you have x regulating some Y1 and also some Y2. And this is the beginnings of a network motif, and so it's a reasonable thing to worry about but maybe we can correct for at least a majority of this by using the proper null model. At least that would be the hope.

AUDIENCE: Well, that's why you don't want necessarily--

PROFESSOR: OK, that's fair. But then the question is, what you null model should we be using? Yeah?

AUDIENCE: So you feel like having the microscopic constraints does not necessarily need to be in the null model. I feel we can have a null model but without using the microscopic constraints and then just say, oh well that's another possibility for why we might have these divergences. I don't think they need to be in the null model.

AUDIENCE: Yeah, it's just that then you can't say anything about evolution.

AUDIENCE: Well fair, but I don't should have to-- I don't think you have to say something about evolution afterwards necessarily.

PROFESSOR: Yeah, and I think that this question about how strongly you can argue that evolution, selective or something, and this is a little bit of a judgment call, because most of these evolutionary arguments are not ironclad, it's more a matter of making you feel kind of comfortable with looking for what the evolutionary explanation might have been. This is just the nature of looking at historical science, right? I mean, you can speculate about what would have happened if Napoleon had done something else, or whatever. But it's a speculation.

Of course the hope is that we can collect multiple pieces of evidence that make us more and more comfortable with it and in some cases we can do laboratory evolution to get more comfort, but laboratory evolution doesn't prove that that's what happened a million years ago either. But I'd say it's more the accumulation of evidence to make you feel comfortable with an argument.

But you know, let's first make sure we understand what the null model is, and then on Thursday we'll decide, well we won't decide, we'll discuss what we think that means about evolution. Yeah?

AUDIENCE: So I think what we the other part of the appendix that we read about the in and out distributions is important for the null model.

PROFESSOR: Yes.

AUDIENCE: Because it seems to me that the Erdos Renyi network might be a good model for the in distributions, but not for the out distributions.

PROFESSOR: That's right. And I think this is really important. I think that it's clear that the actual transcription network of *E. coli*, for example, is not well described as an Erdos Renyi random network, but then it does beg the question of what should we be using. And you could say, well, we just make a power law network, but then you say, oh, but there's the in degree, and the out degree. How much do you want to keep track of that? And I think that there is a fairly strong argument that what you should do is what they call this degree preserving network.

In particular what that means is that you take the real network, so you take the actual network that you're going to be analyzing, and there is some actual degree distribution. So there's 1 node has-- so k_1 might be 106, k_2 might be 73, dot, dot, dot, dot, up to k_n which is equal to 1. And of course I'm not even talking about it being directed, but you do the same thing with directed.

But then what you do, is you kind of mix things up. So you start with a real network and then you do something to randomize it. And it's a rather clever scheme, I'm just going to describe it briefly here and then we'll talk more about it on Thursday.

What you do is you take all of the actual-- so let's say we have x_1 , some x_2 and here we have a Y_1 , Y_2 , Y_3 , now let's say that these guys are regulating something like this. What you do is you take two edges randomly, we'll pick this one that one, and what we do is we swap the targets. So what we do is we make this guy come

over here, and then this one comes over here. So now what we do is we erase this, and we erase this, now we have a new network, but intriguingly, the degree distributions for both incoming and outgoing edges are identical to what we had before this.

Every guy has the outgoing edges, incoming edges, but they're just different targets. So if you just do this procedure many, many times then what you do is you achieve some randomized version of the real network. And then what you can do is you can ask how many feed forward loops are there. How many, this, that--

And so there's a fair argument that this is in some ways the proper null model to be asking the question in. And indeed, for example, there are many more feed forward loops than there would be in an Erdos Renyi, but still what you see is that you lose many feed forward loop. So this then the argument for feed forward loops being selected for. We'll talk about this and we'll quantify it on Thursday, but I'm available for the next half hour if anybody has any questions.