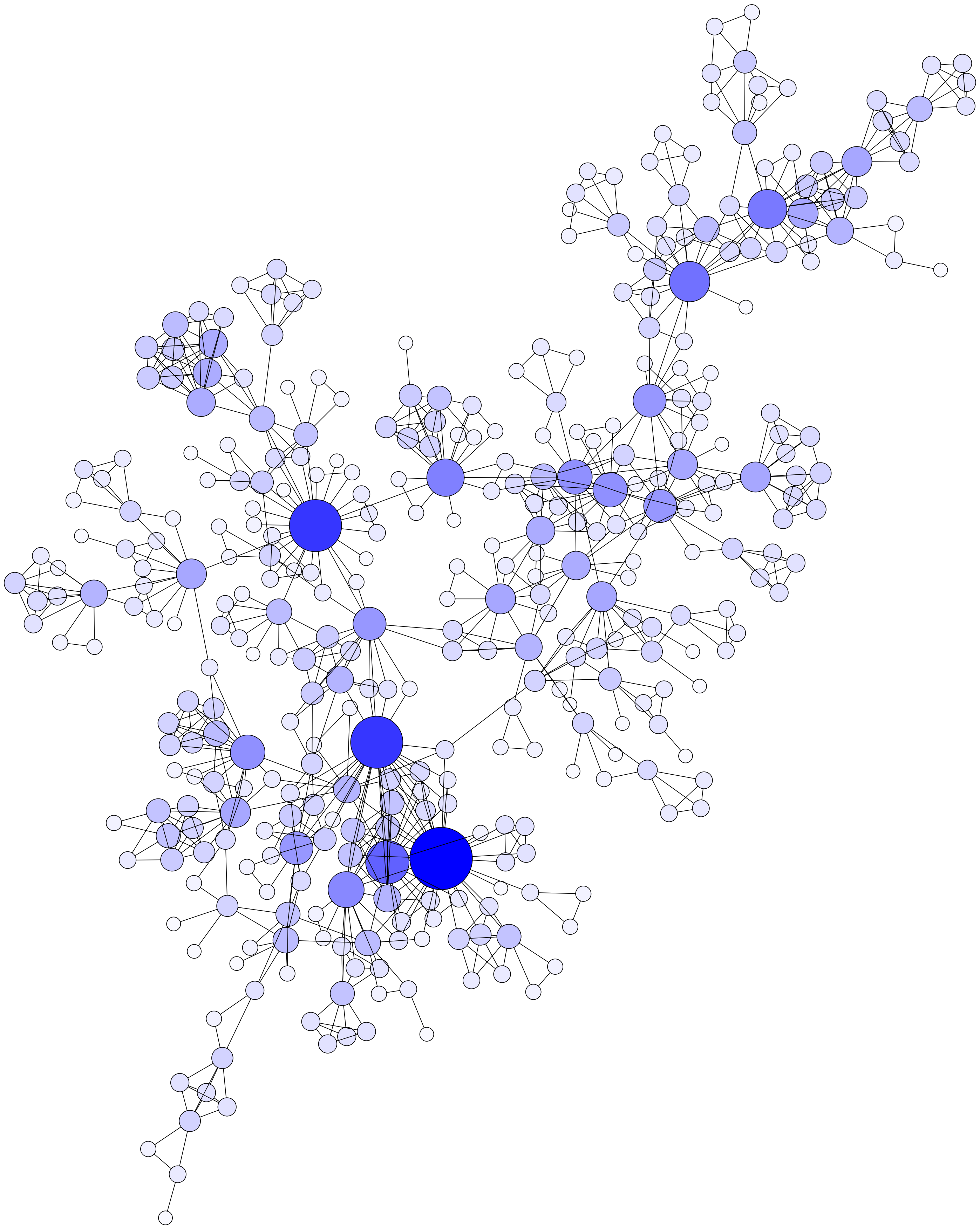
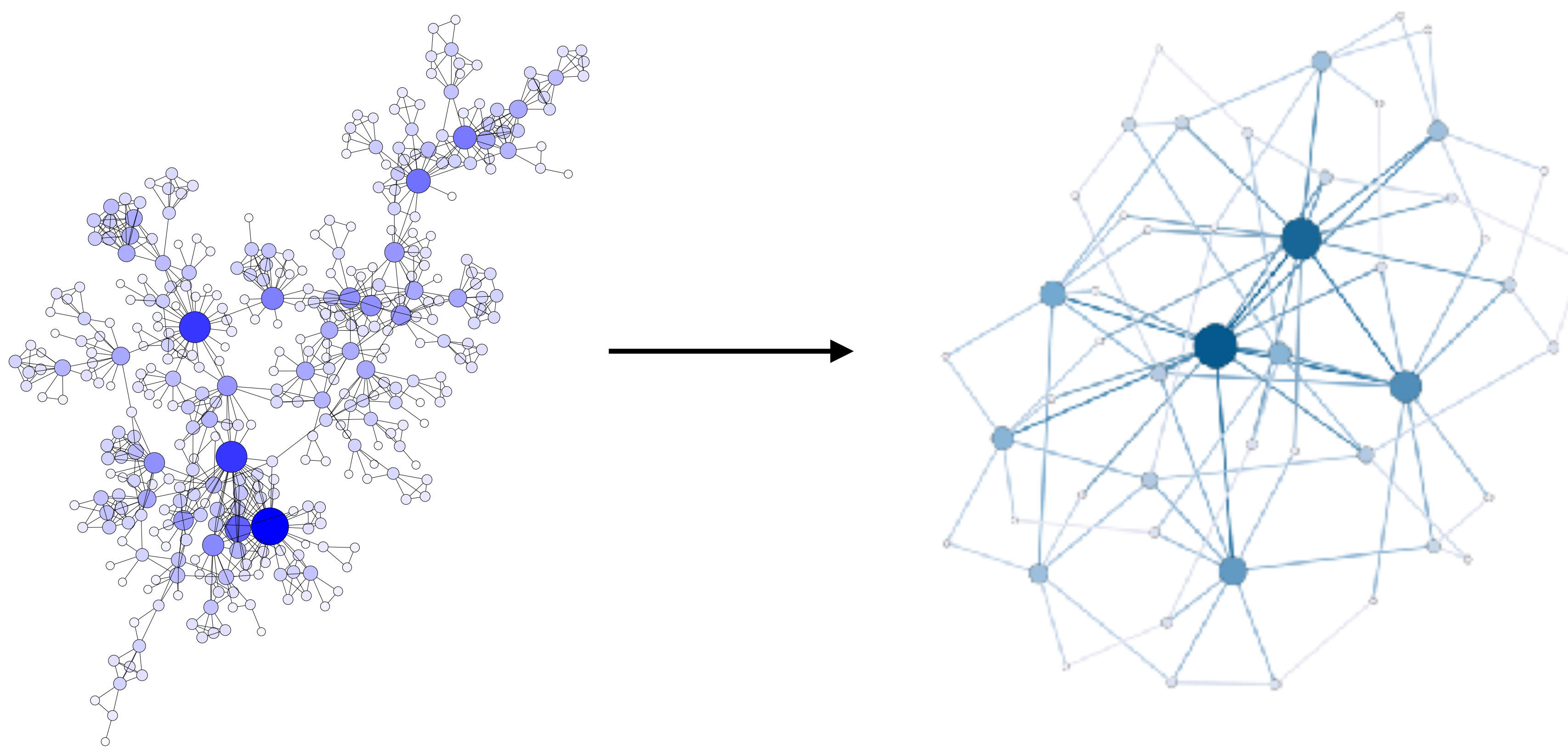


Models of Social Networks

Models of Social Networks

Why do we need models of social network structure?

A model is a way of simplifying something that is very complex…



which allows us to turn the *specific* into the *general*

Models and Modeling

To make that clearer, it helps to talk about models more generally…

Discussion of cows adapted from Miller and Page (2007)
Silly illustrations are my own.

and cows…

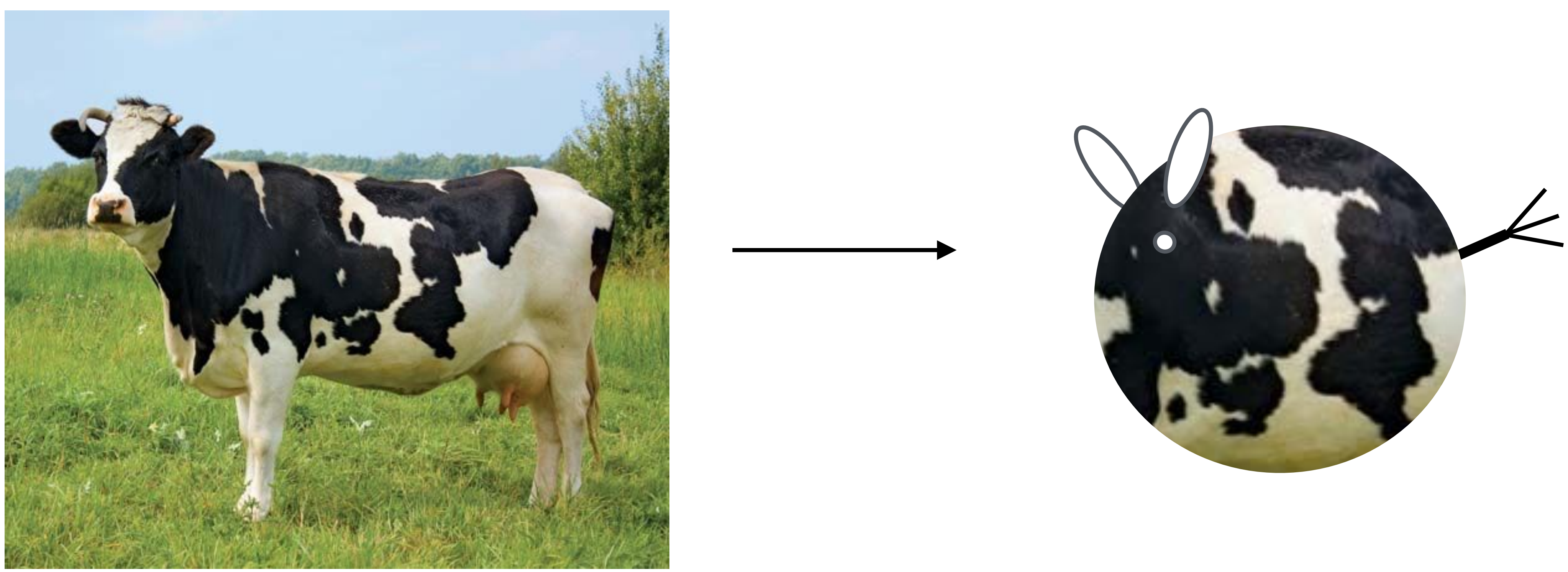
Models and Modeling

Cows are surprisingly complex creatures…

Models and Modeling

A *model* of the cow simplifies it by leaving out some of the details.

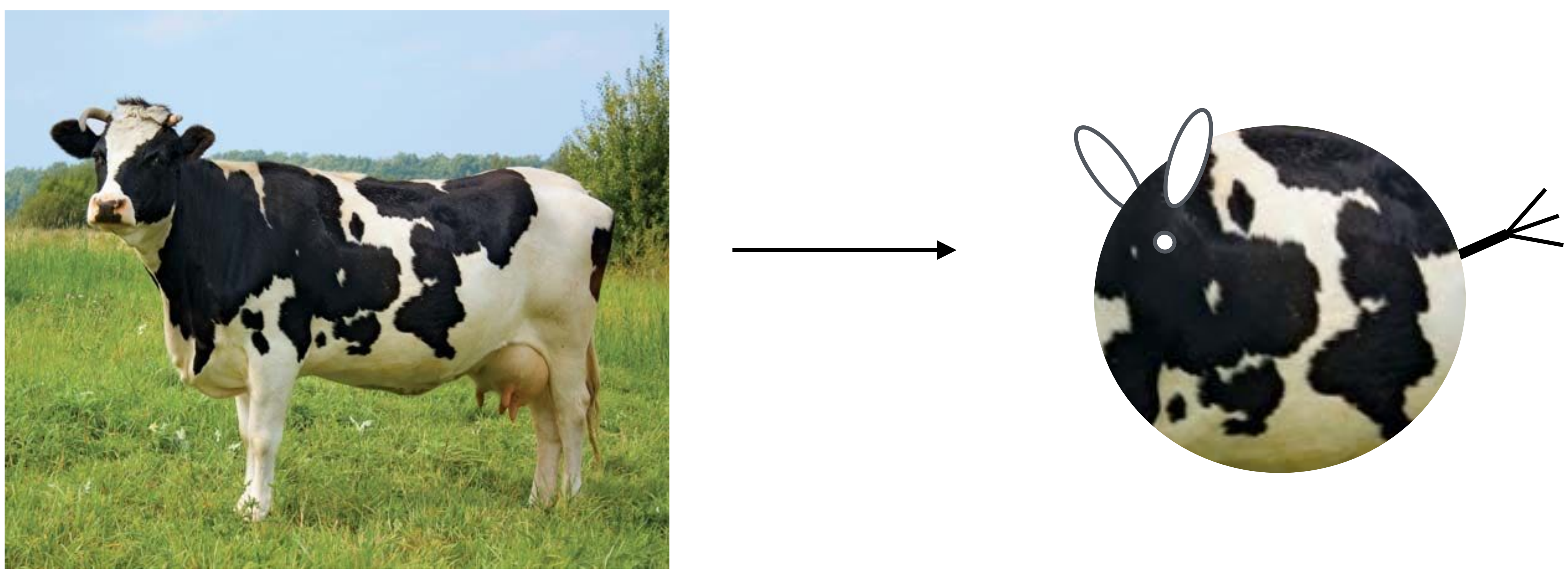
Example: a spherical cow



Again: discussion of cows adapted from Miller and Page (2007)
Silly illustrations are my own.

Models and Modeling

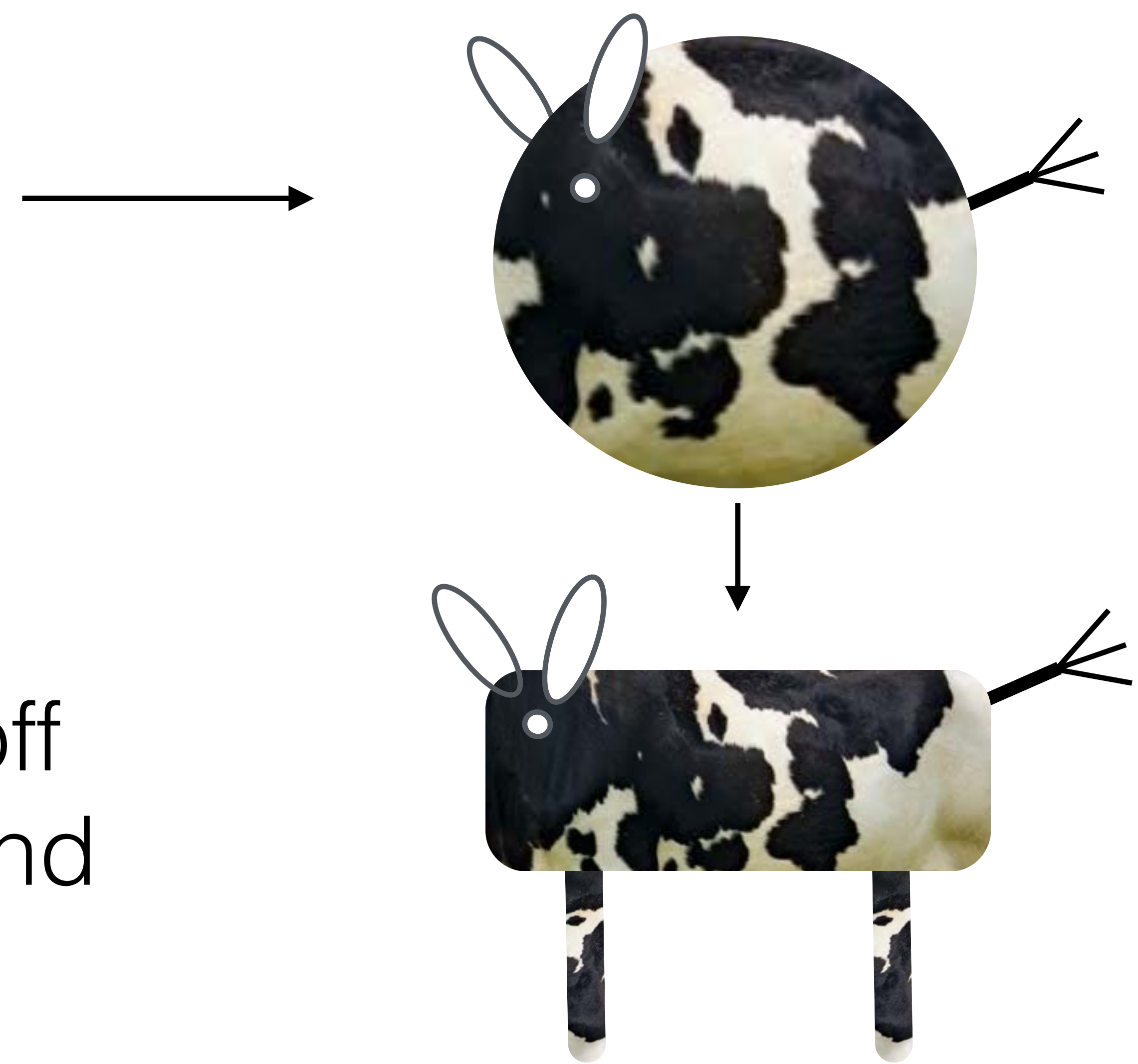
Building a model of a cow lets us answer general questions about cows, because it leaves out the details that vary from one cow to another, but don't matter for the question at hand.



Any given model will leave some things out, but a good model will leave the right things in!

Models and Modeling

More accurate models may give you more insights, but they are also more complex, and thus harder to understand

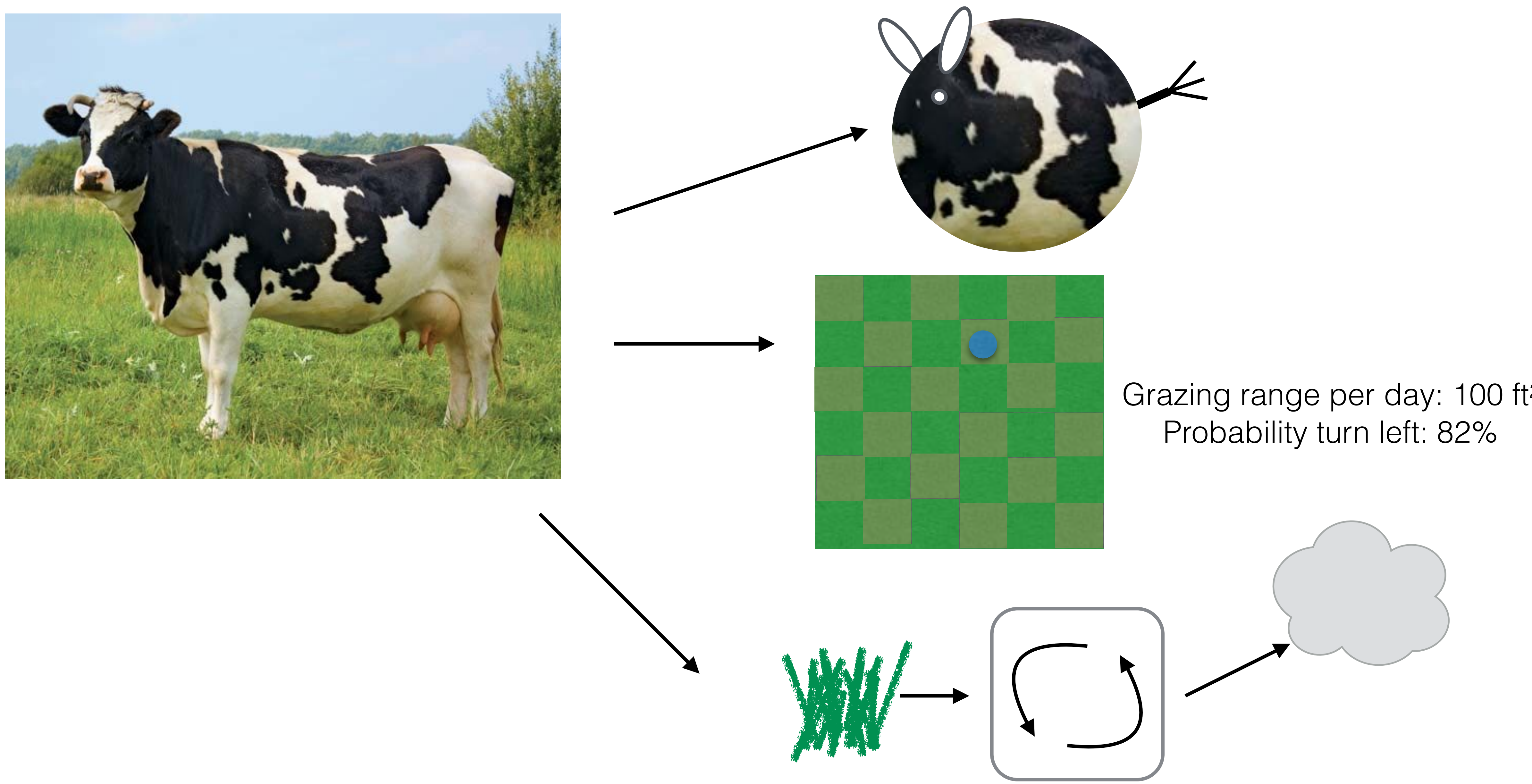


There is often a tradeoff between complexity and understanding

Models and Modeling

Not all models are suited for answering all questions…

Different questions require different simplifications!

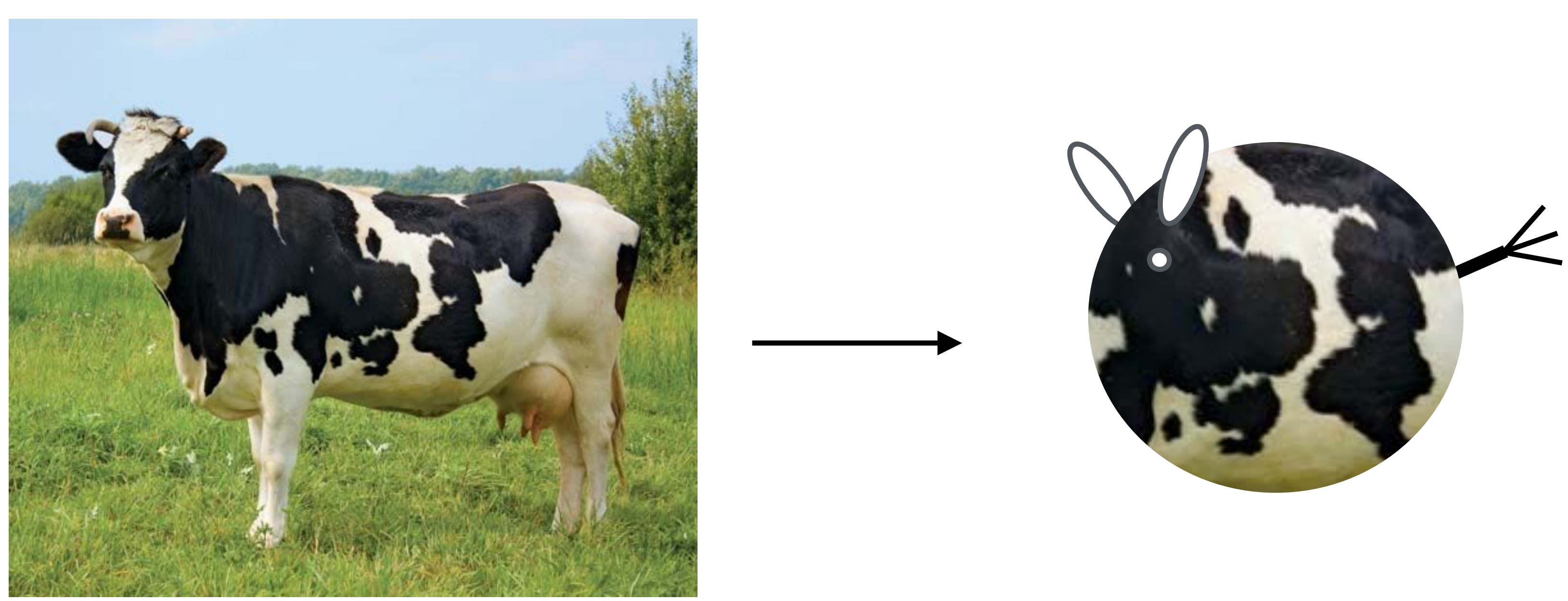


A Brief Soap Box Moment…

All social science inquiry (including empirical work) involves modeling: explicitly or implicitly.

Being a responsible consumer of social science research means understanding how to be critical of models

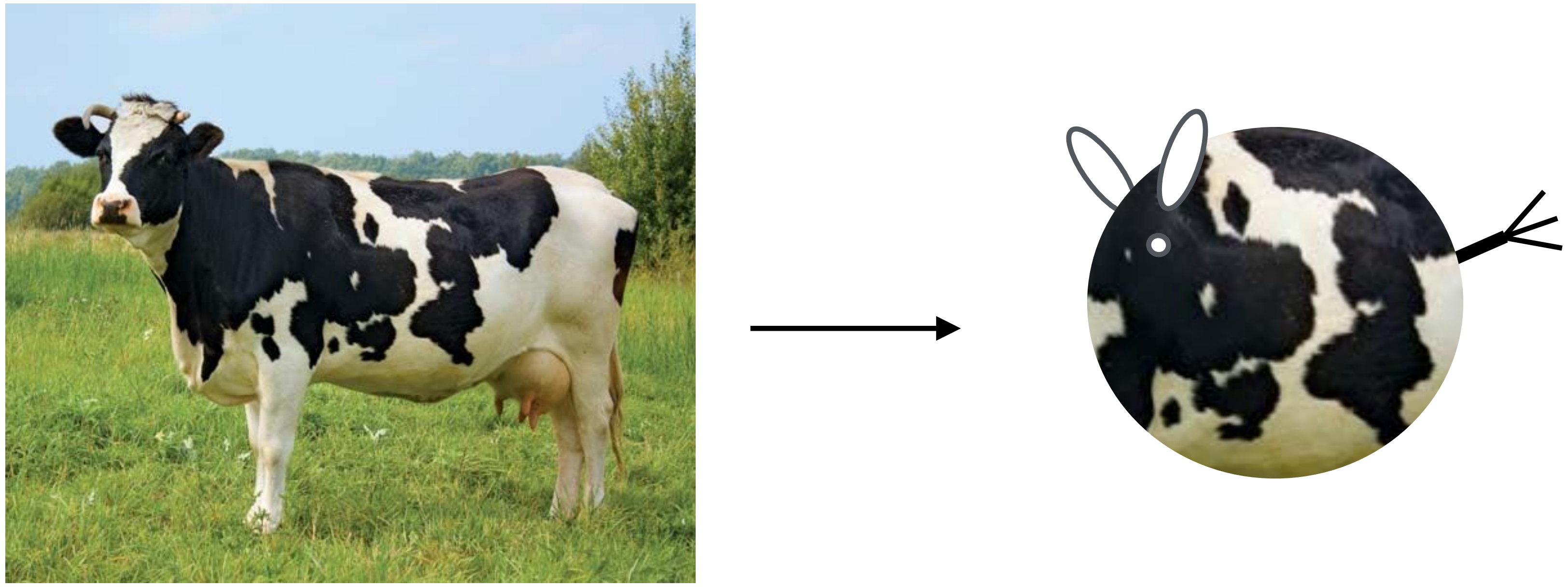
This may be one of the most important things you learn in your time at CMU!



A Brief Soap Box Moment…

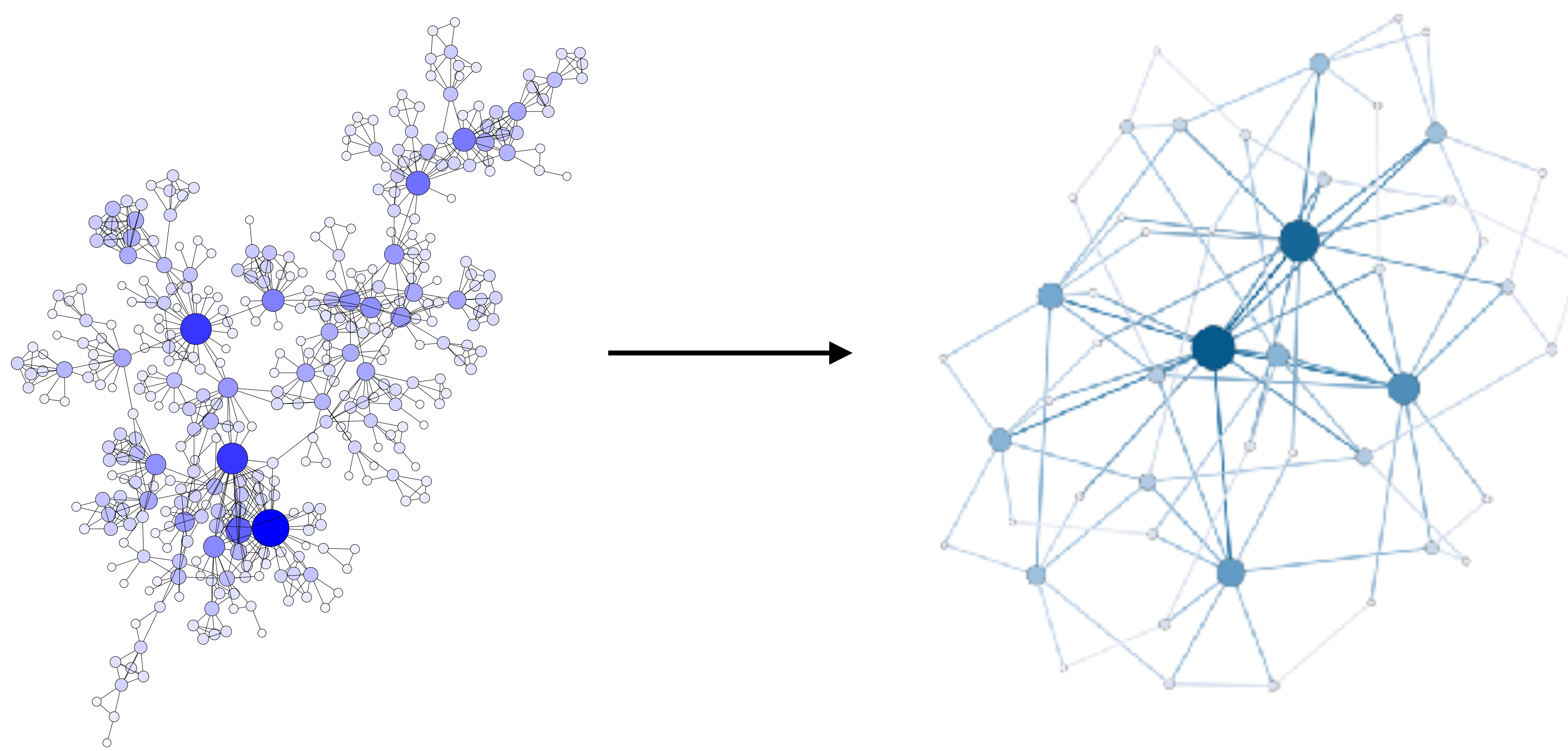
How to be critical of a (social science) model in three simple questions:

- What simplifying assumptions have been made?
- Are those assumptions reasonable in the context of the question being explored?
- How would things be different if those assumptions were changed?



Models of Social Networks

This is all particularly important in the context of social networks: we often have only one truly independent observation of any given network. Models allow us to consider what might make that network look the way it does. They also potentially let us generalize from what we observe in one context to other contexts.

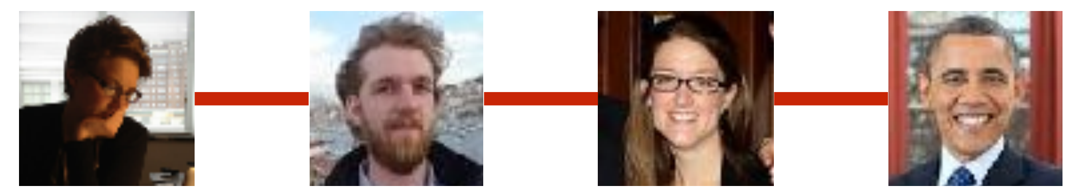


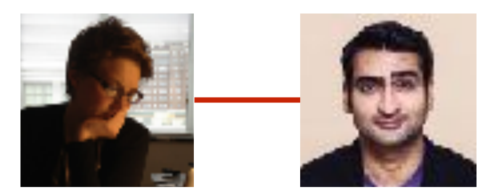
Networks: The General

So what are some things we generally observe to be true in social networks?

- Small World (low average distance between nodes)

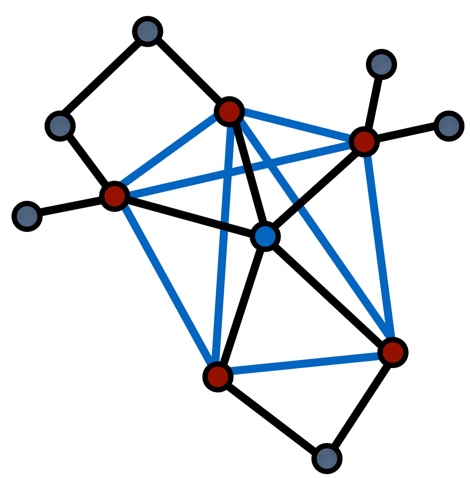
Strong links



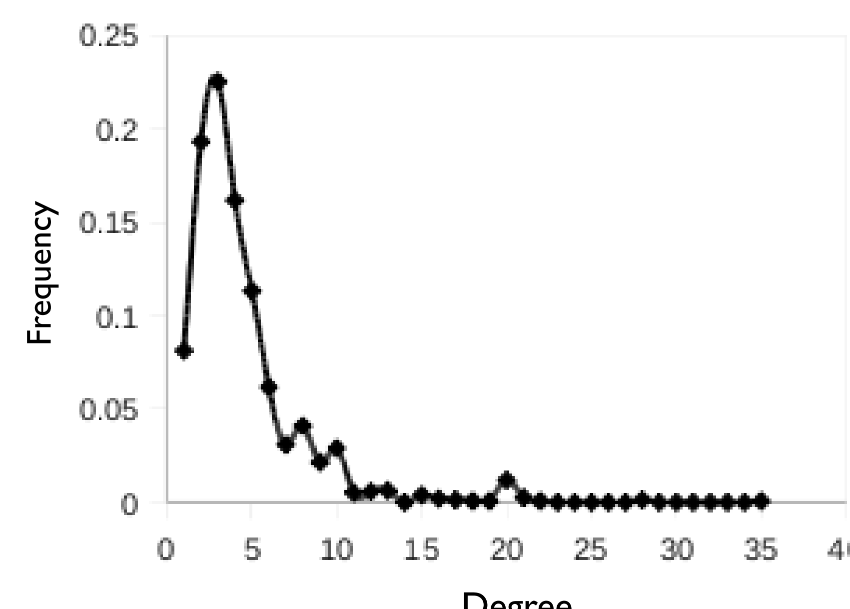


Weak links

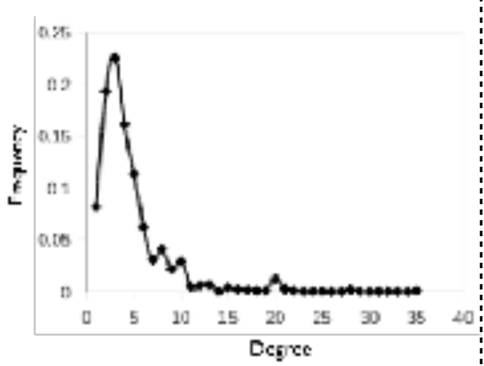
- High Clustering (triadic closure)



- Long-tailed degree distribution (skewed right)



Network Models Summary

	Empirical	Erdös-Renyi	Watts-Strogatz	Preferential Attachment
Average Distance	Low			
Clustering	High			
Degree Distribution				

Models of Social Networks

So! What are some ways that people have modeled social networks?

- Models we'll explore today:
    - Erdös-Renyi Random Graph
    - Watts-Strogatz Small World Network (Watts and Strogatz (1998))
    - Preferential Attachment (Barabasi and Albert (1999))

Models of Social Networks: Erdös-Renyi Random Graphs

Oldest model of a network (1959)

Procedure:

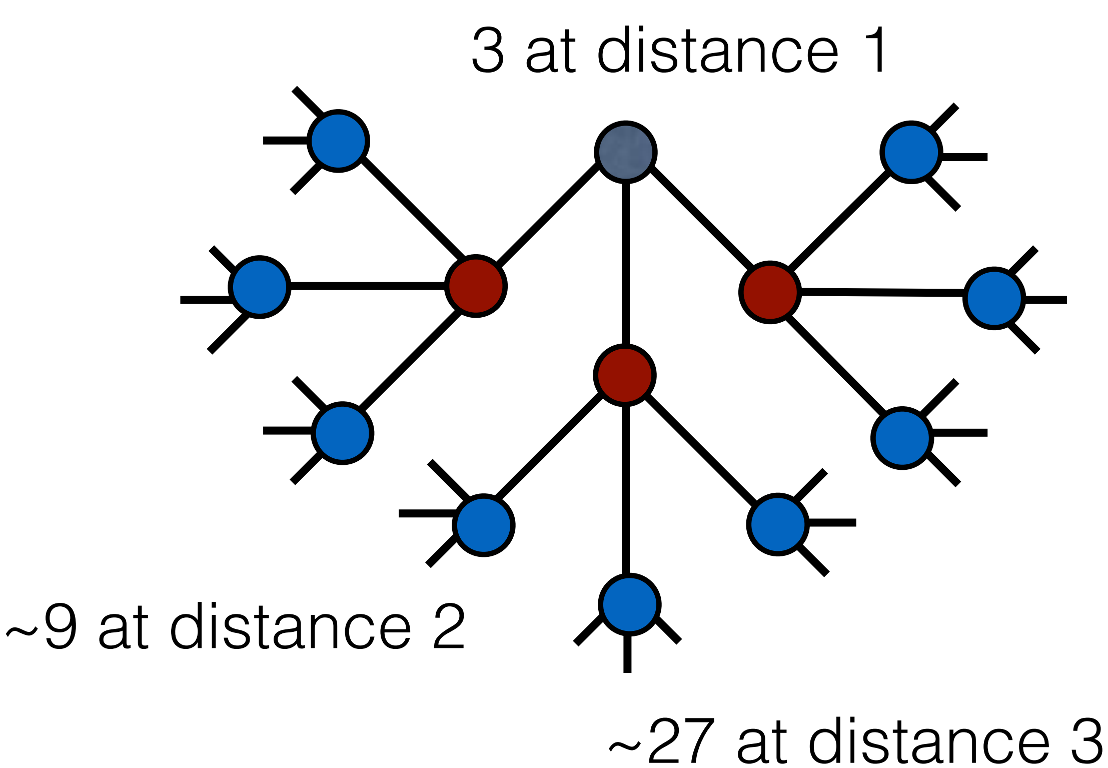
- Select a node, i
- For every other node in the network, connect to node i with probability p
- Repeat for all nodes in the network

In the end, every pair of nodes is connected with probability p

Result is called an Erdös-Renyi random graph (or simply a random graph)

Characteristics of Erdös-Renyi Random Graphs

Erdös-Renyi Random Graphs have a low average distance:



→you aren't very far from anyone!

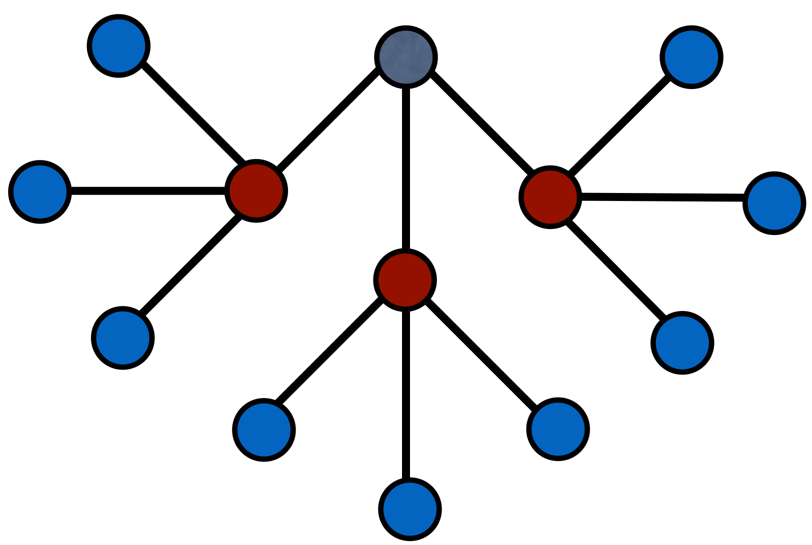
Suppose you are linked to three people on campus at random.

If those people also know three random people, then chances are, they are all different

The number of people who are x hops out from you is $\approx d^x$ where d is the average degree of a person in the network

Characteristics of Erdös-Renyi Random Graphs

Erdös-Renyi Random Graphs have a low clustering coefficient

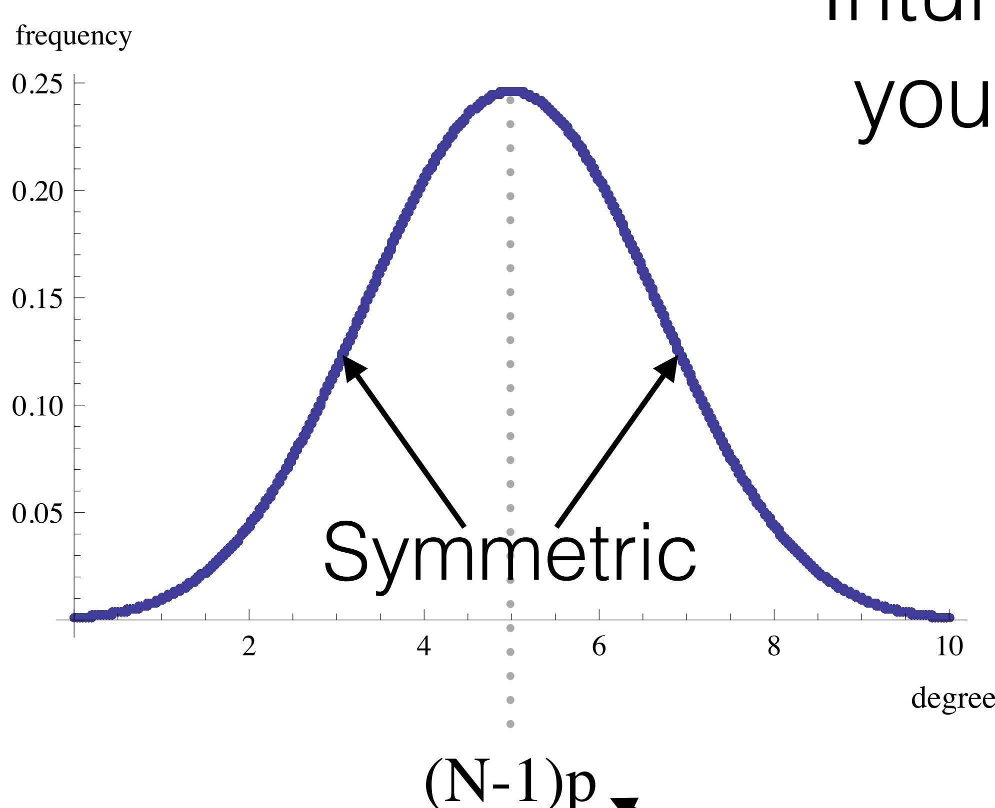


Again, if everyone knows three random people, then chances are that none of the people you know know each other.

→clustering coefficient very close to 0

Characteristics of Erdös-Renyi Random Graphs

Erdös-Renyi random graphs have a binomial degree distribution

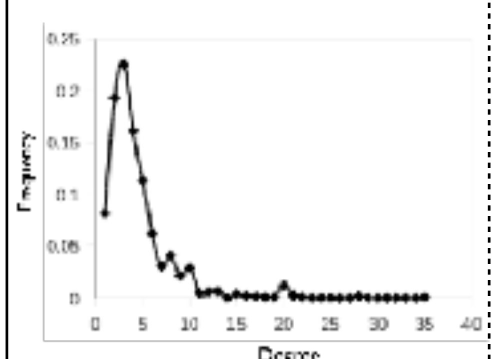
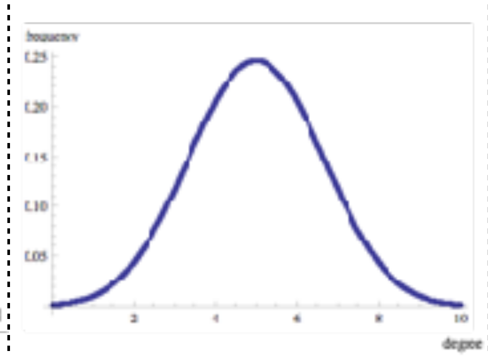


Intuition: For each node, it is as if you do (N-1) Bernouli trials with probability p of success

The distribution of outcomes will be binomial

Average degree: (N-1)p

Network Models Summary

	Empirical	Erdös-Renyi	Watts-Strogatz	Preferential Attachment
Average Distance	Low	Low		
Clustering	High	Low		
Degree Distribution				

Models of Social Networks: The Giant Component in Random Networks

The Erdös-Renyi Model does not match many of the elements of real-world social networks…but it still gives us insight!

In particular, it can tell us about what happens as people become more connected to each other…

Models of Social Networks: The Giant Component in Random Networks

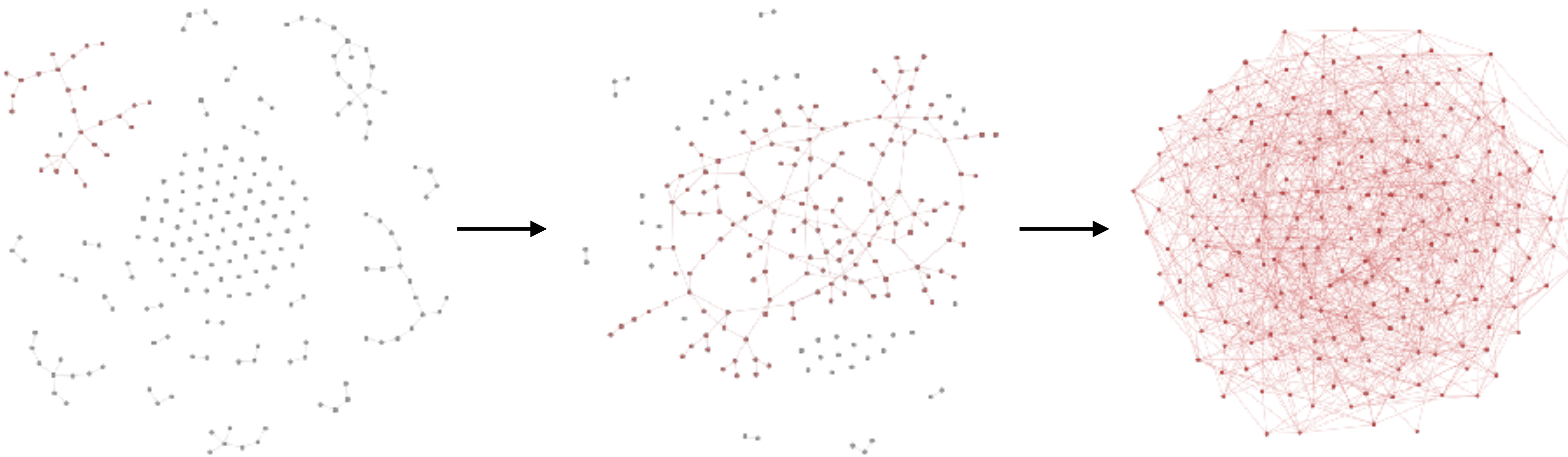
Recall…

- On an undirected network, a *connected component* is a group of nodes that can all be reached from each other via a path
- The *largest connected component* is…um…the largest one.
- If the largest connected component remains proportional to number of nodes in the network, we call it the *giant component*

  → as the name suggests, this component is often substantially larger than the rest. Hence…gigantic.

Models of Social Networks: The Giant Component in Random Networks

A question: what do you expect to happen to the size of the largest connected component in a random graph as we increase the number of connections?



When does it become a giant component? Does it happen gradually? Or all at once?

Models of Social Networks: The Giant Component in Random Networks

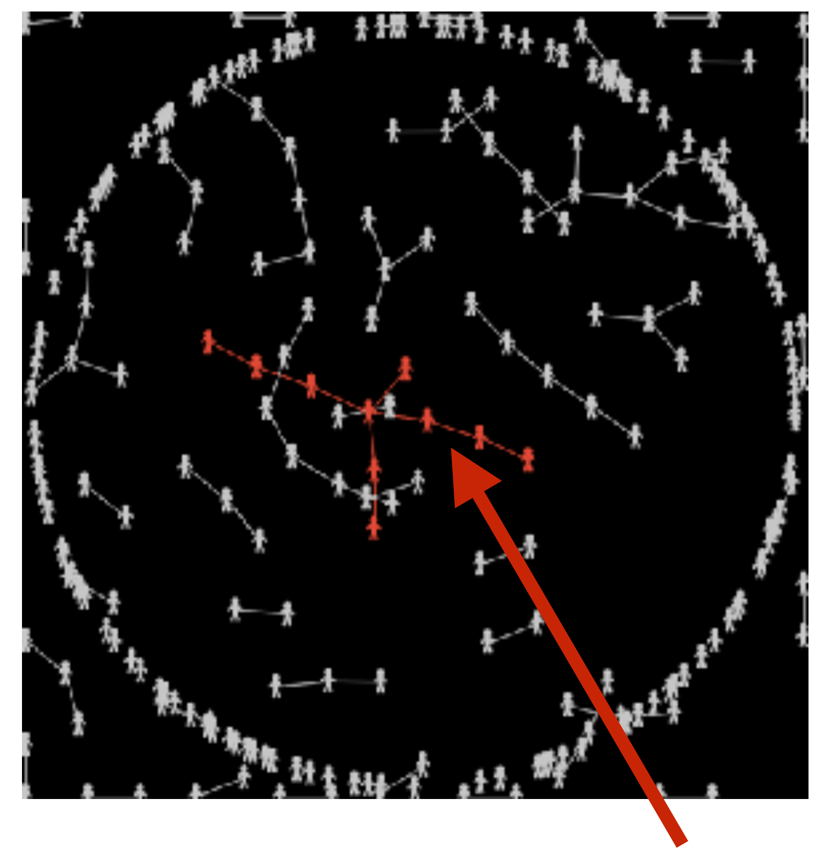
An Agent Based Model (ABM):

- Put down *N* agents (people)
- Every time period, link two random people

As the number of links increases, average degree increases and density increases

Load the net logo program:

Giant Component ER random.nlogo



largest connected component (red)

An ABM of how the giant component grows in an ER random network

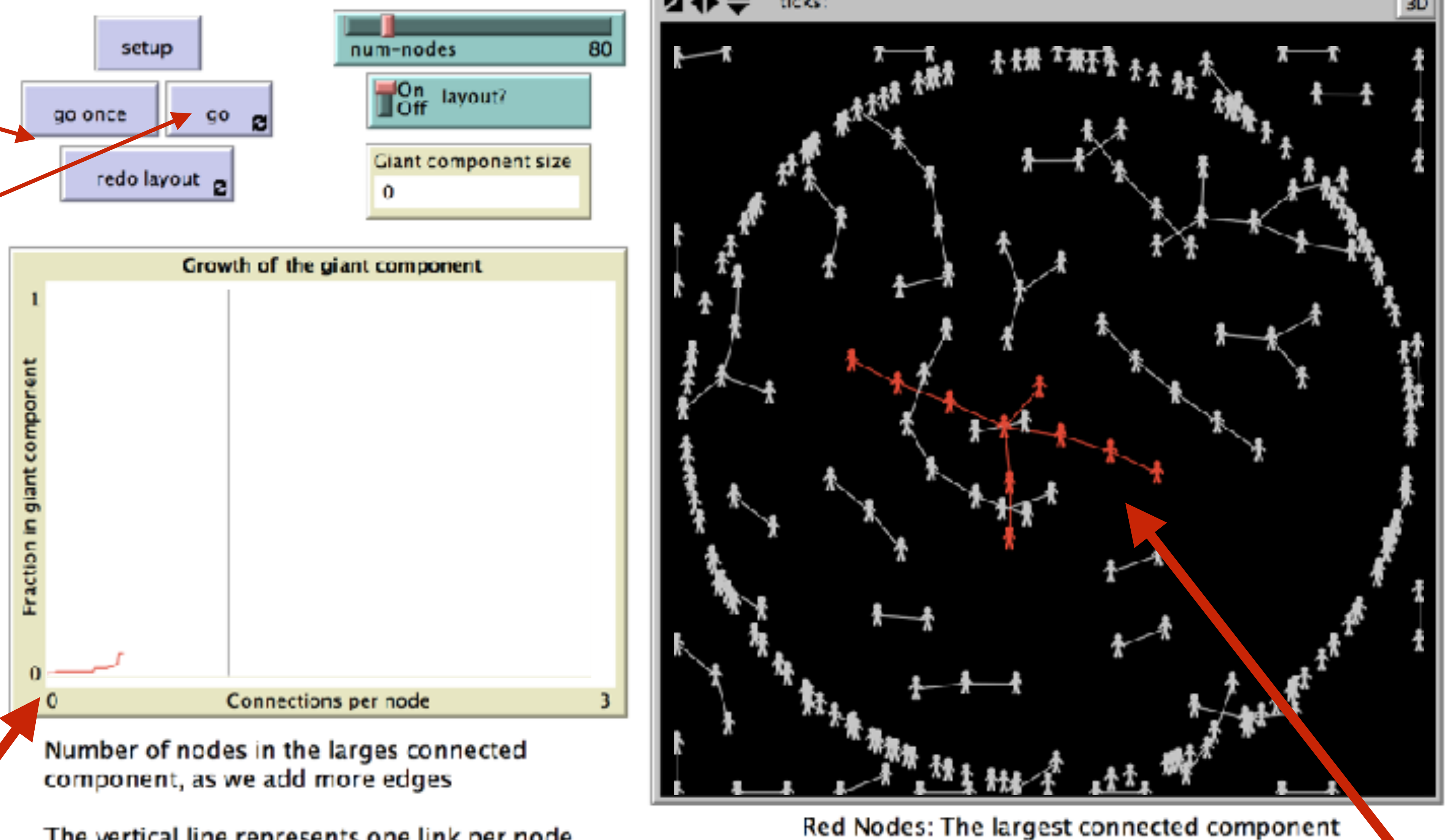
To run:
1) press “setup”
2) press “go once” or “go”

“go once” takes one step (adds one link)

“go” continues to take steps (add links) until you press it again

(yes, you hit “go” to stop)

adjust the number of nodes in the network

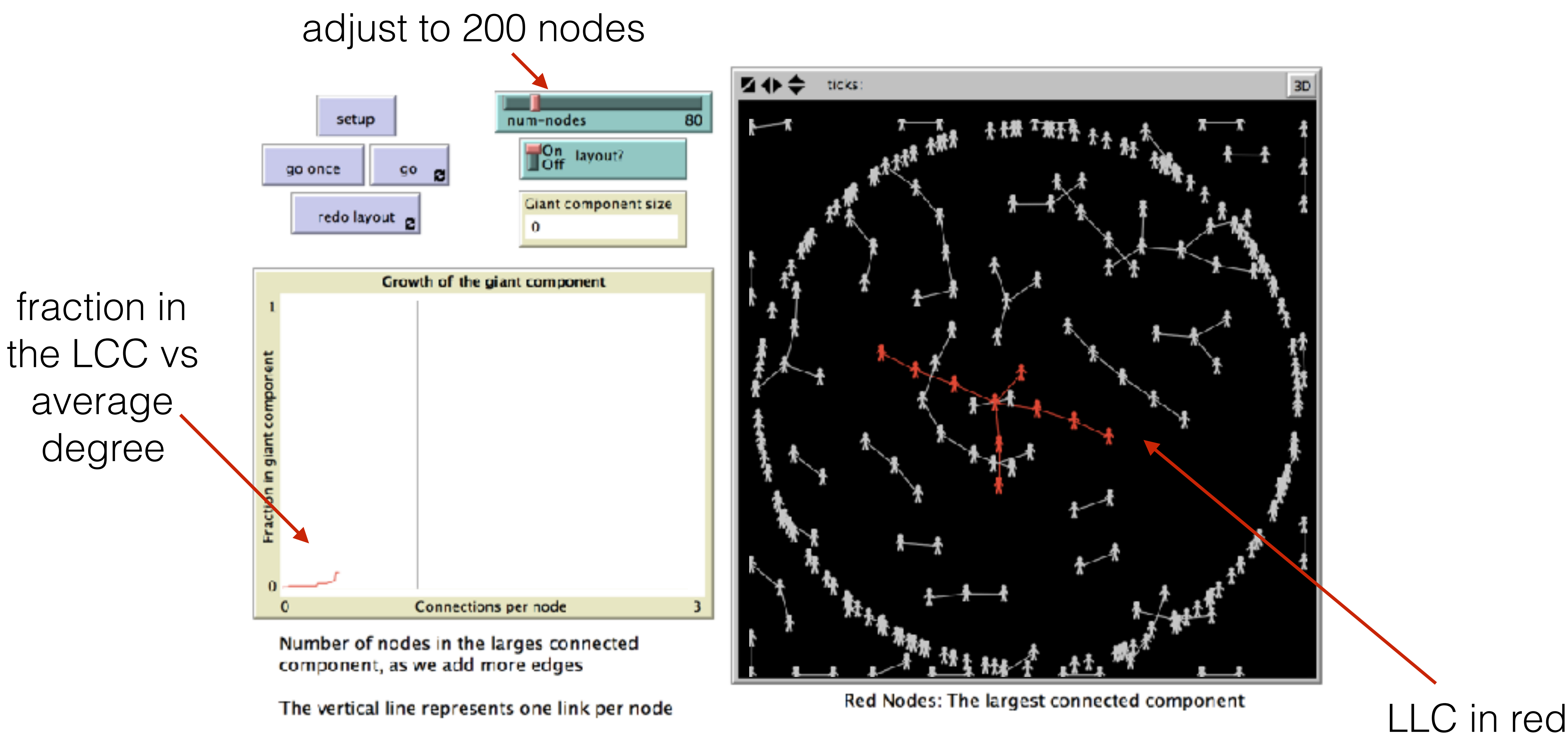


plot: fraction of nodes in the LCC vs average degree (updated with every step)

LLC in red

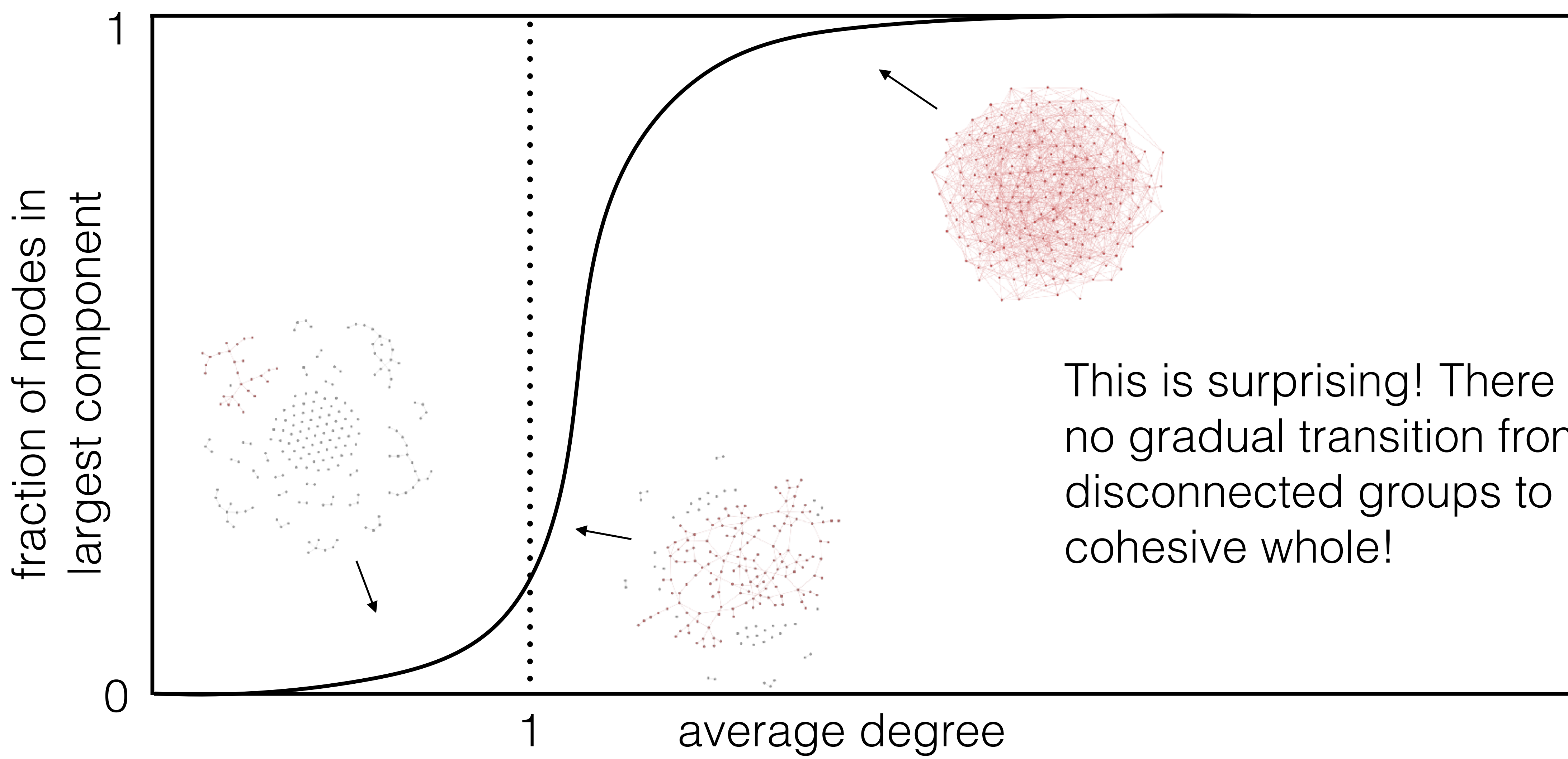
Now take a look at what happens:

1) Set up the nodes (no links) — 200 is a good number
2) Go one step at a time. Watch the LCC. How does it grow?
3) When you get bored, hit “run”. What happens to the LCC? Is there a gradual change, or does it change all at once?
4) Try it again. Do you get the same thing? What if you change the number of starting nodes? Do you have any idea what is going on?



Models of Social Networks: The Giant Component in Random Networks

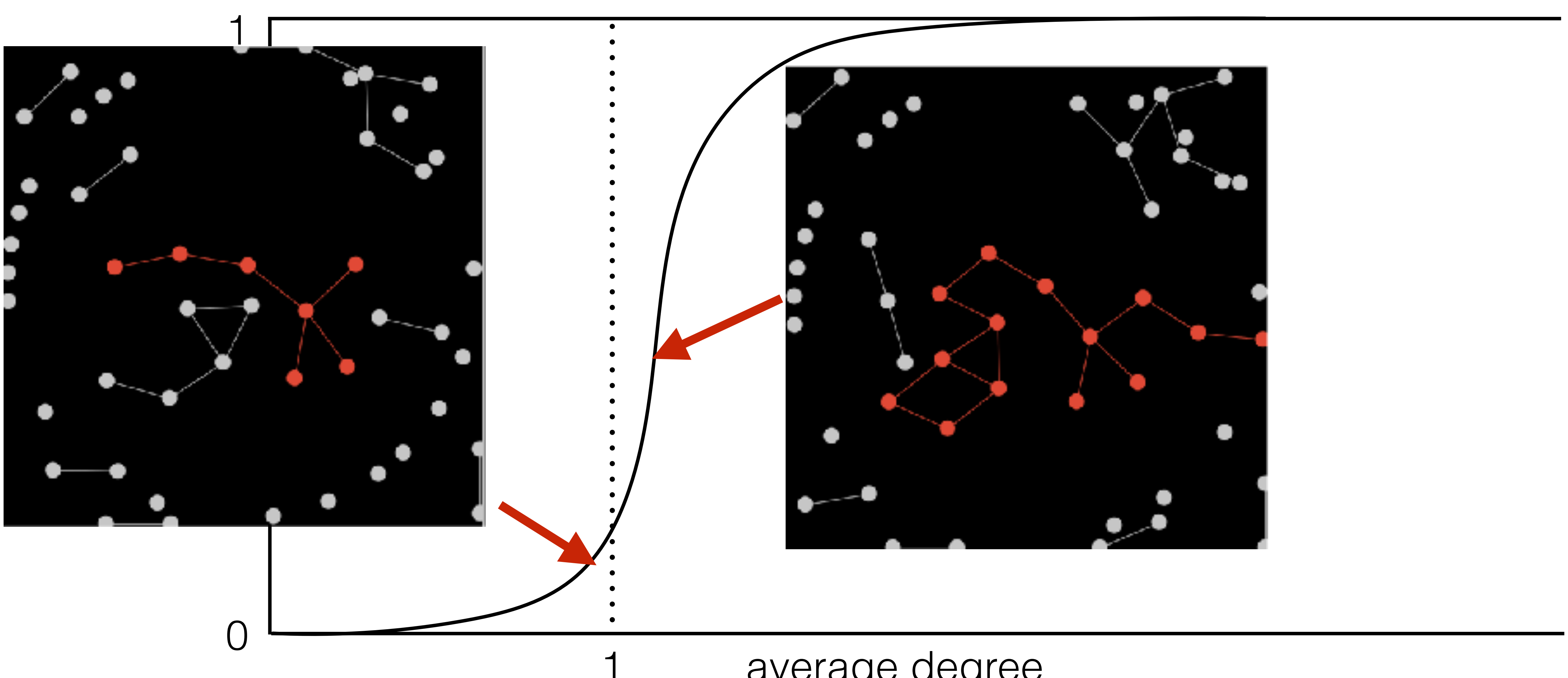
In a random network, the largest connected component remains small until the average degree (*Np*) reaches 1. Then a giant component *suddenly* emerges.



This is surprising! There is no gradual transition from disconnected groups to cohesive whole!

Models of Social Networks: The Giant Component in Random Networks

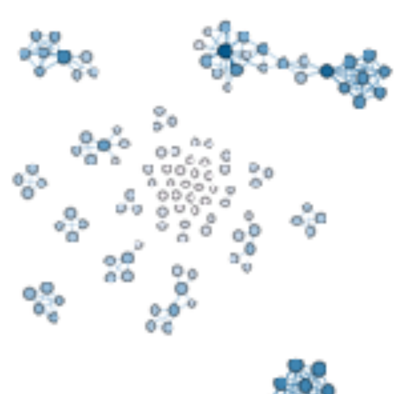
The intuition: CCs grow individually, then two or more connect together. That component jumps in size, and then has a much higher probability of getting links to new nodes.



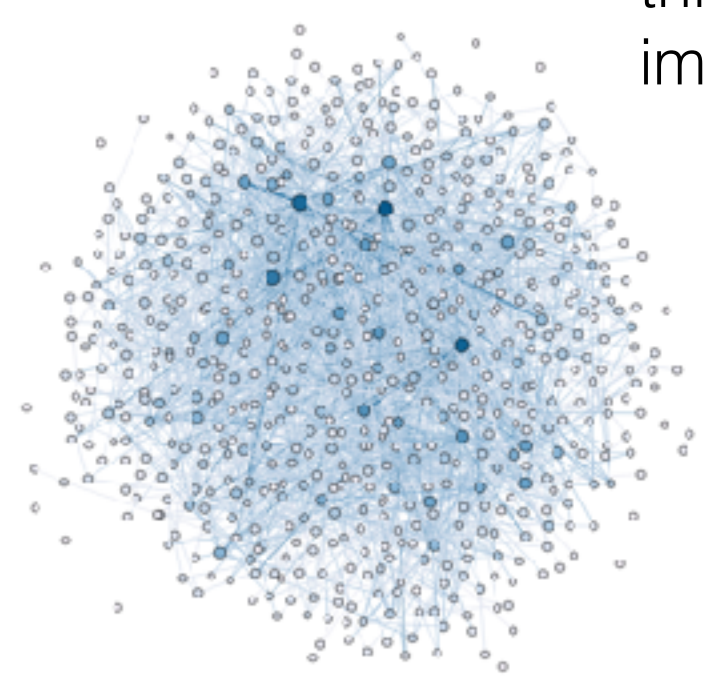
Models of Social Networks: The Giant Component in Random Networks

This is backed up empirically in my own work: in this emerging collaborative community, there was a transition point, where isolated groups transitioned to a single, cohesive whole.

network of authors on 80s papers



network of authors on 90s papers



network of authors on 2000s papers

So despite being inaccurate on a number of dimensions, this model gives us some important insights!

Models of Social Networks: The Small World

But what about clustering?

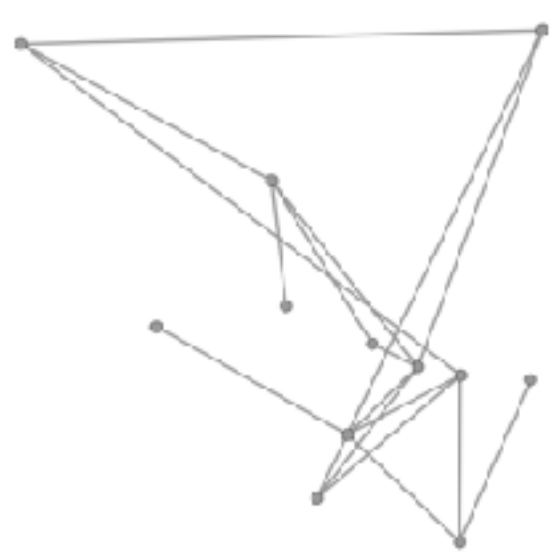
Empirically, we have a bit of a puzzle. Social networks have high clustering, giving them a highly local structure.

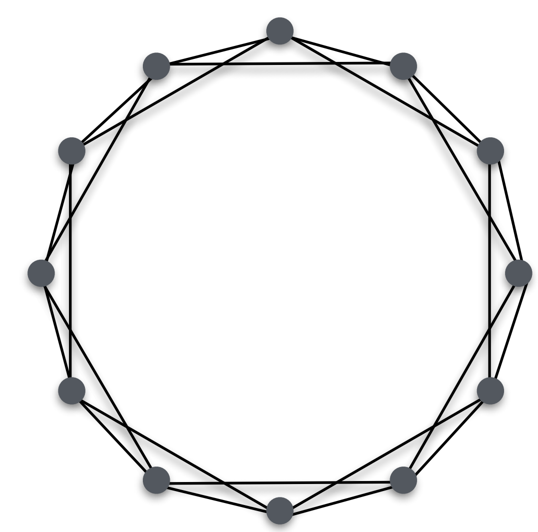
But they also have a low average distance, making them globally very small…

So how is it possible for networks to be both local and global at the same time?

Models of Social Networks: The Small World

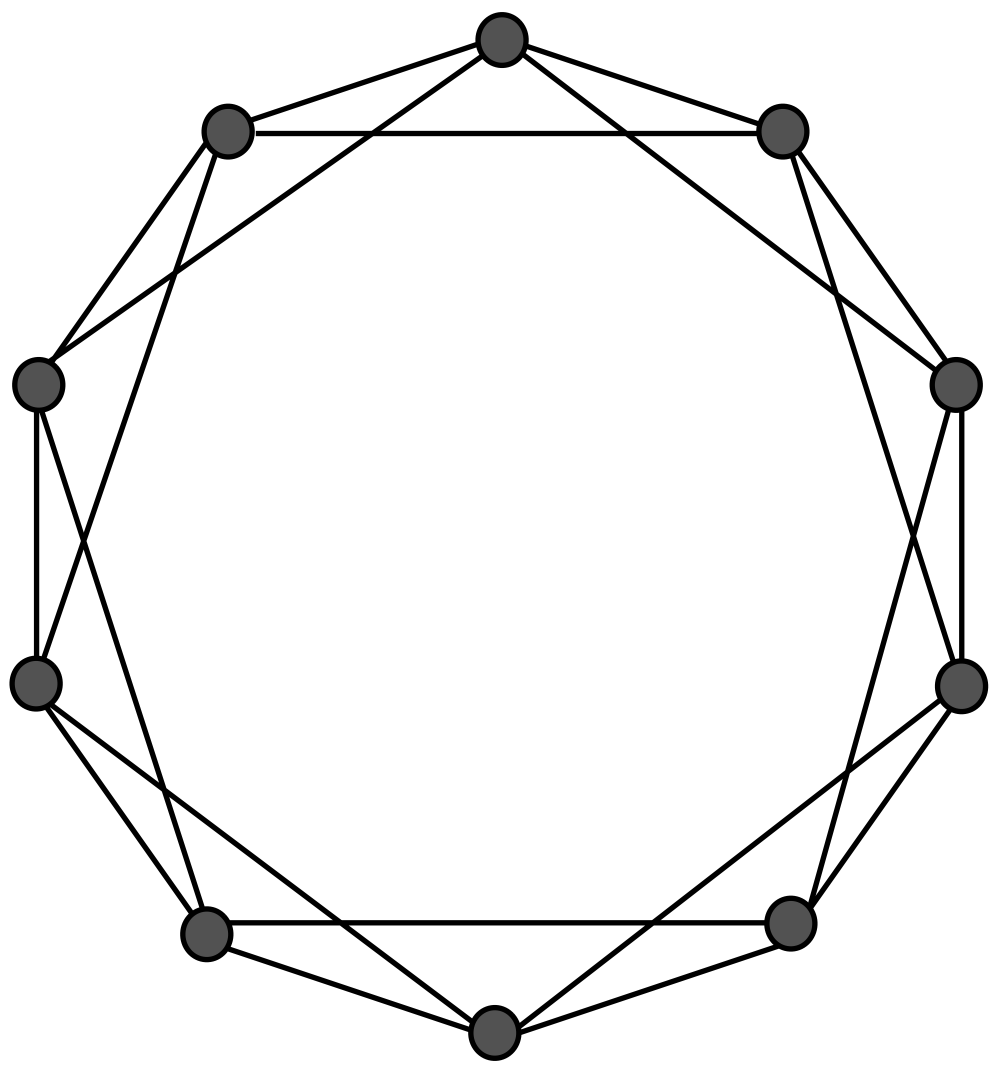
Empirically, social networks have a low average path length (like a random network) and high clustering (like a regular network)





So intuitively, we might want something between a random network and a regular network…

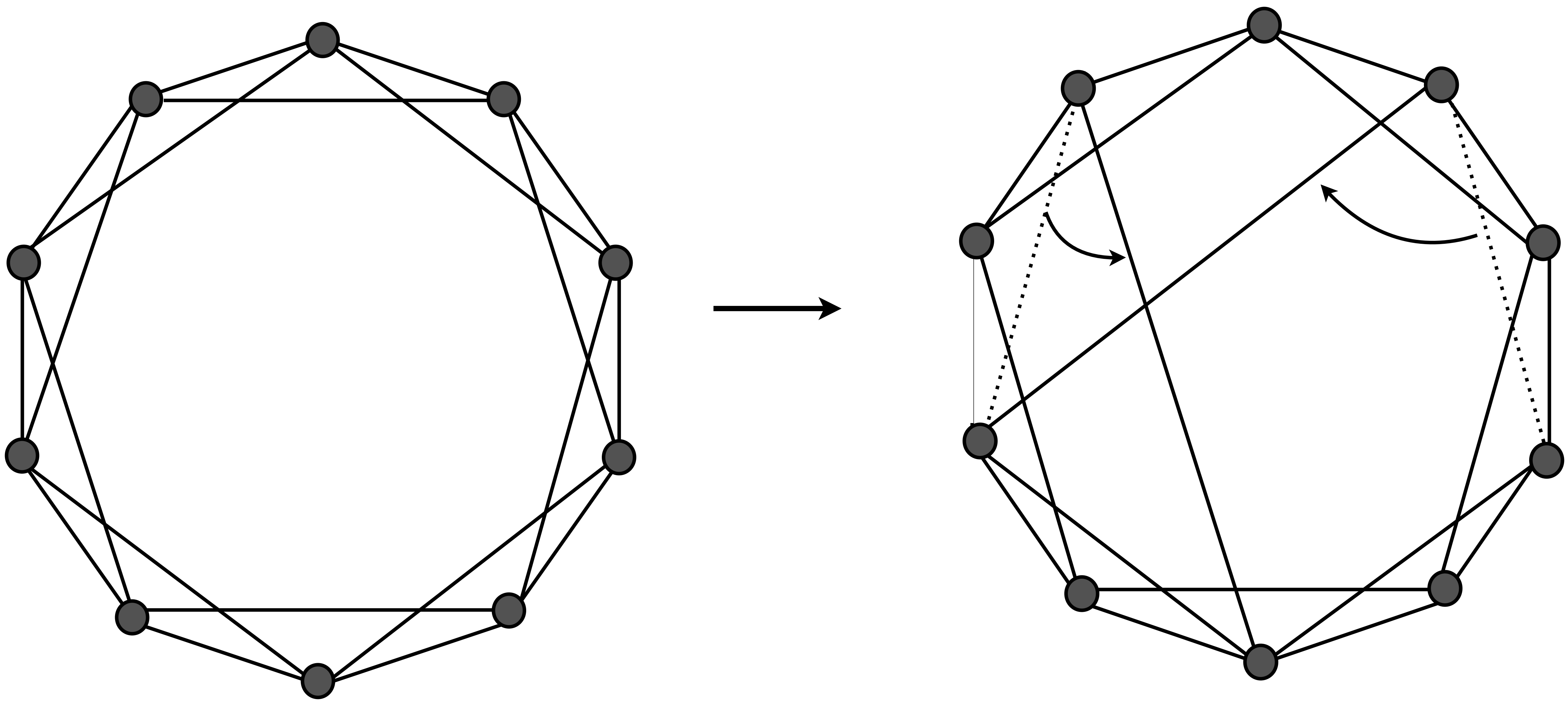
Models of Social Networks: Watts-Strogatz Small World



The Watts-Strogatz network:

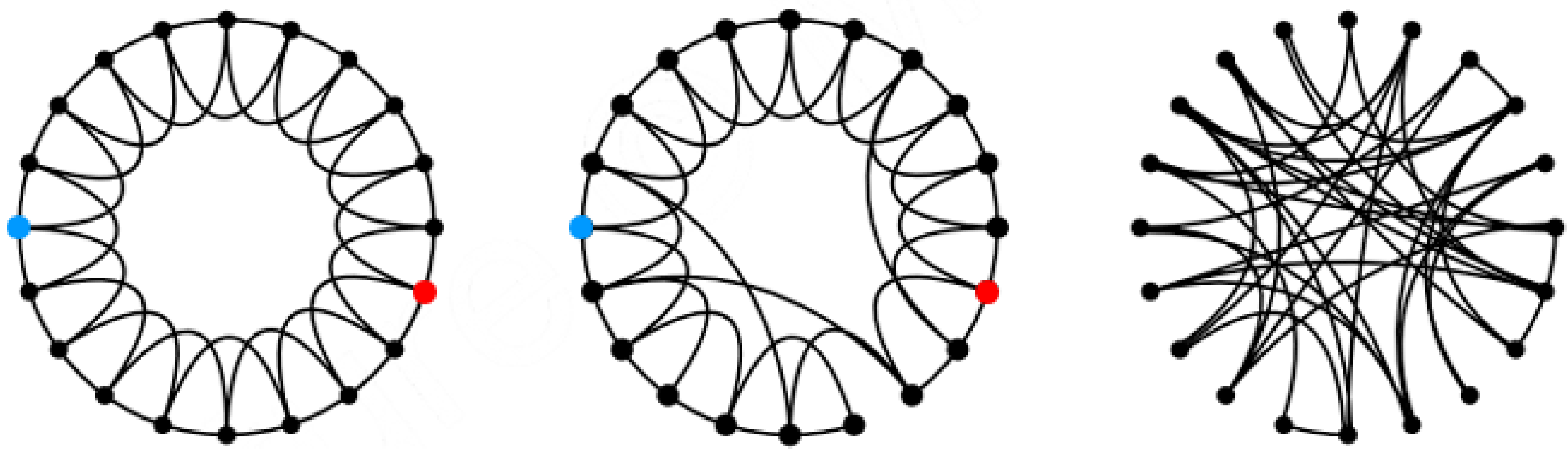
- Start with a regular network, degree d
- Take each pair of nodes, and with probability p, add a link between them

Models of Social Networks: Watts-Strogatz Small World



The result is somewhere between a regular network and a random network

Models of Social Networks: Watts-Strogatz Small World



Regular Network

Random Network

↑

p = 0 ⟶ 0 < p < 1 ⟶ p = 1

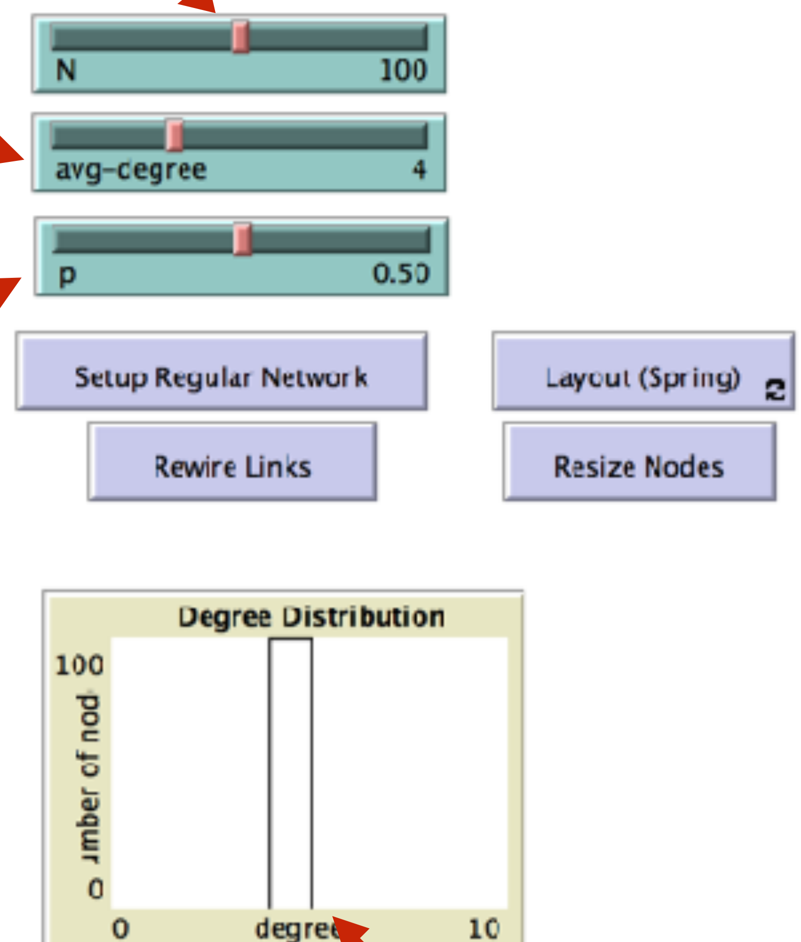
Figure from Watts and Strogatz (1998)

A simulation of the Watts-Strogatz Small World Network

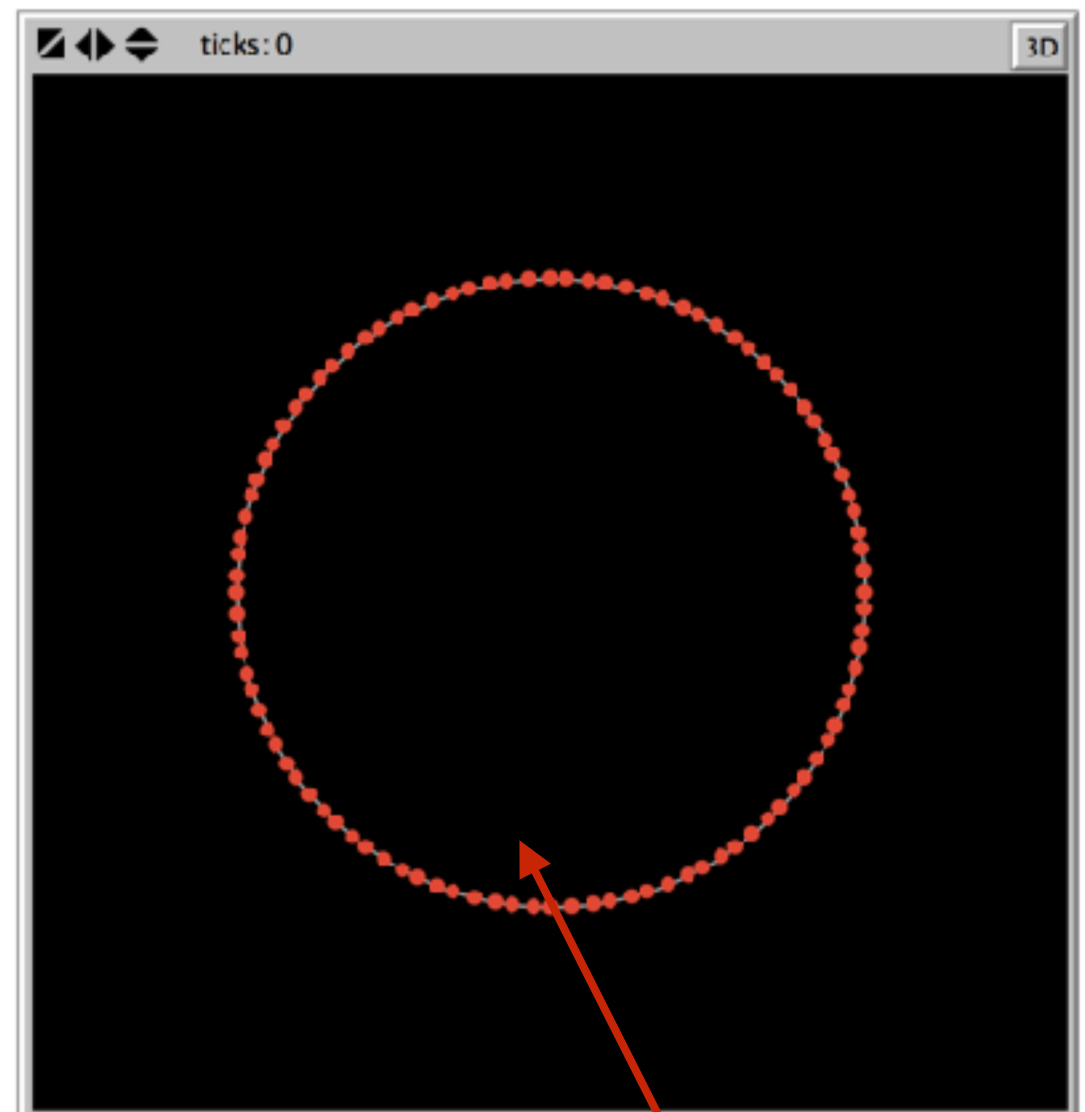
number of nodes

number of connections the nodes start with

probability of rewiring each link



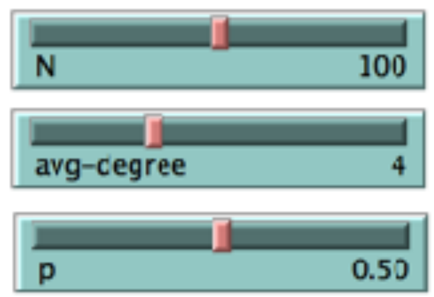
degree distribution (updated each time step)

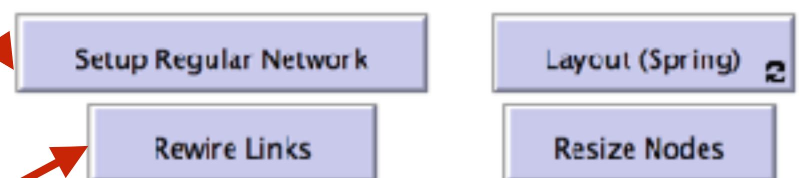


visualization

A simulation of the Watts-Strogatz Small World Network

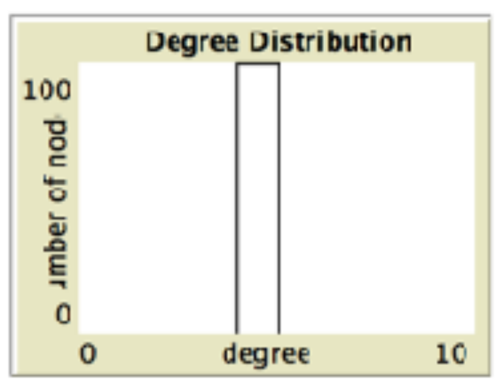
this sets up the initial, random network—press this every time you want to rerun it





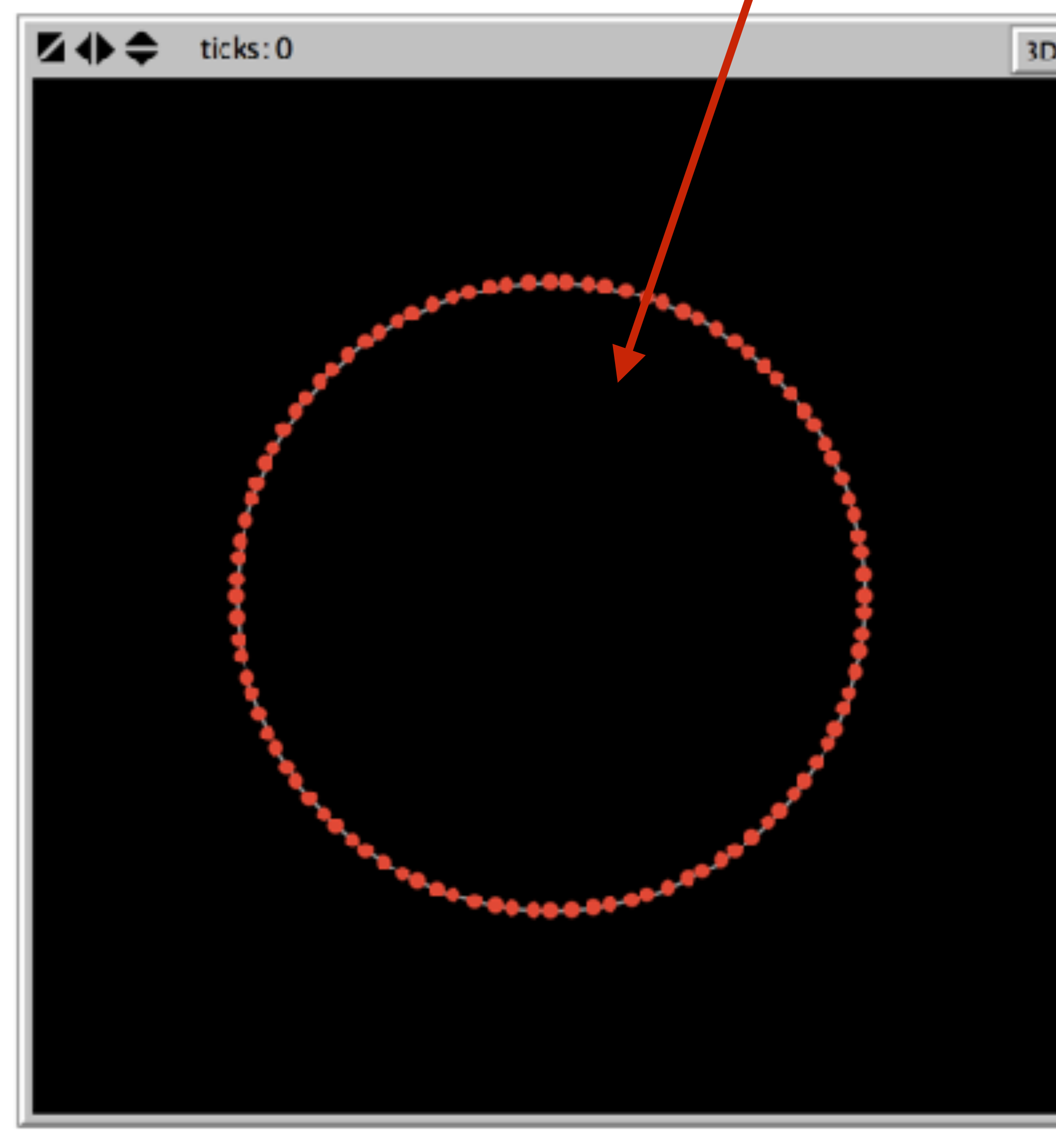
layout to look more natural

rewire each link with probability p



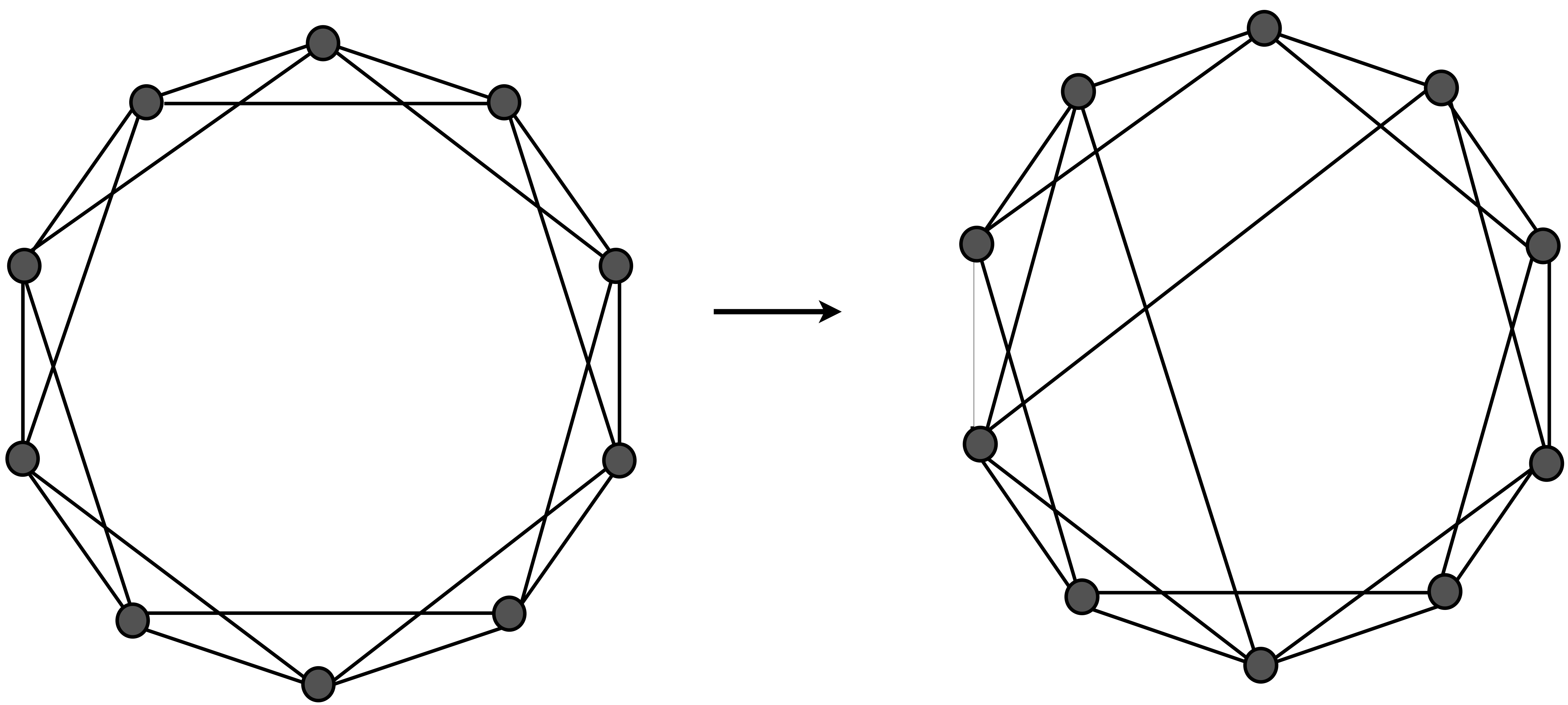
size and color by degree

You start with a regular network



Models of Social Networks: Watts-Strogatz Small World

What would you *expect* to be the effect of a small number of rewired links? What would happen to clustering? What would happen to average distance?



Models of Social Networks: Watts-Strogatz Small World

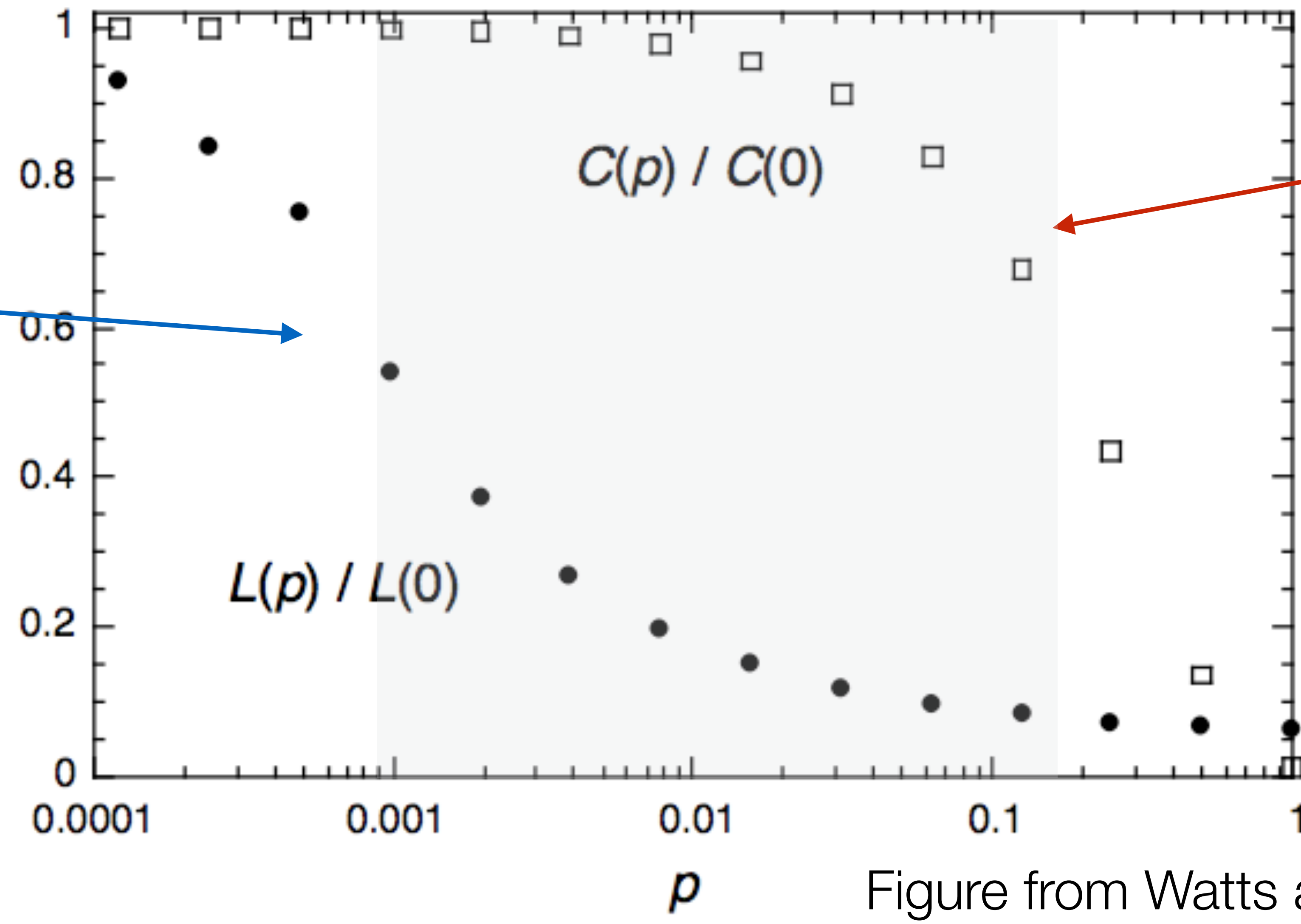
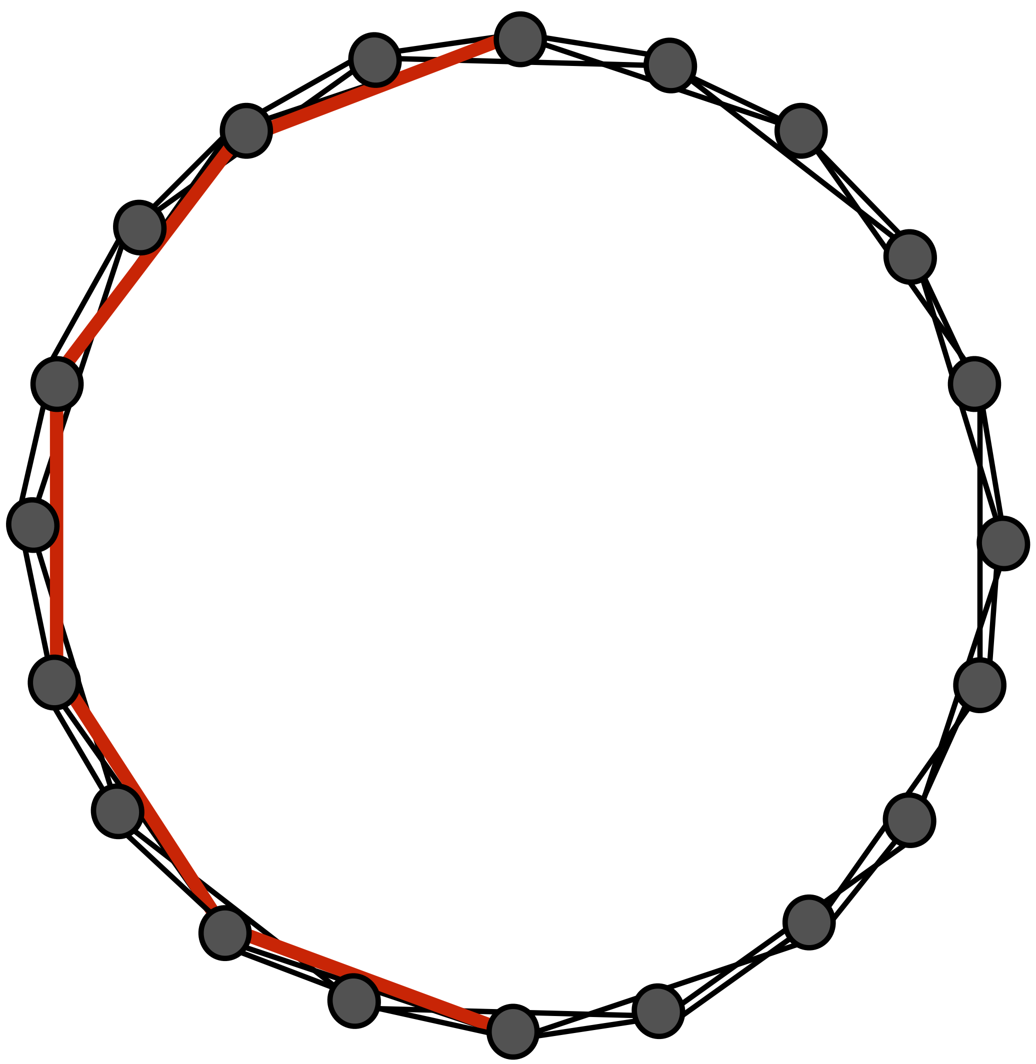


Figure from Watts and Strogatz (1998)

The result: adding a few random links dramatically decreases average distance in the regular network without affecting average clustering

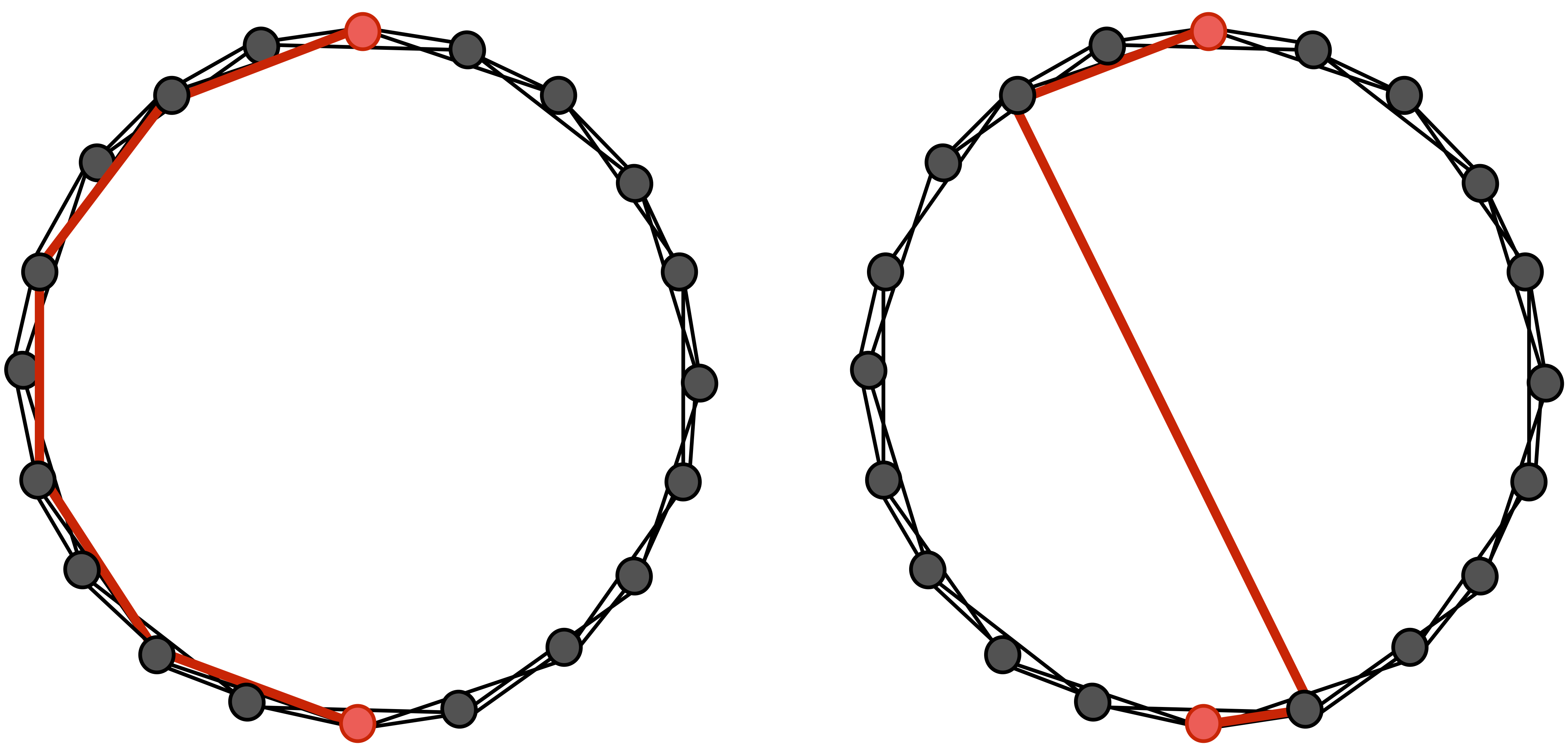
Models of Social Networks: Watts-Strogatz Small World

The reason that the regular network has a long average distance is that you have to traverse the entire ring to get from one side to the other…



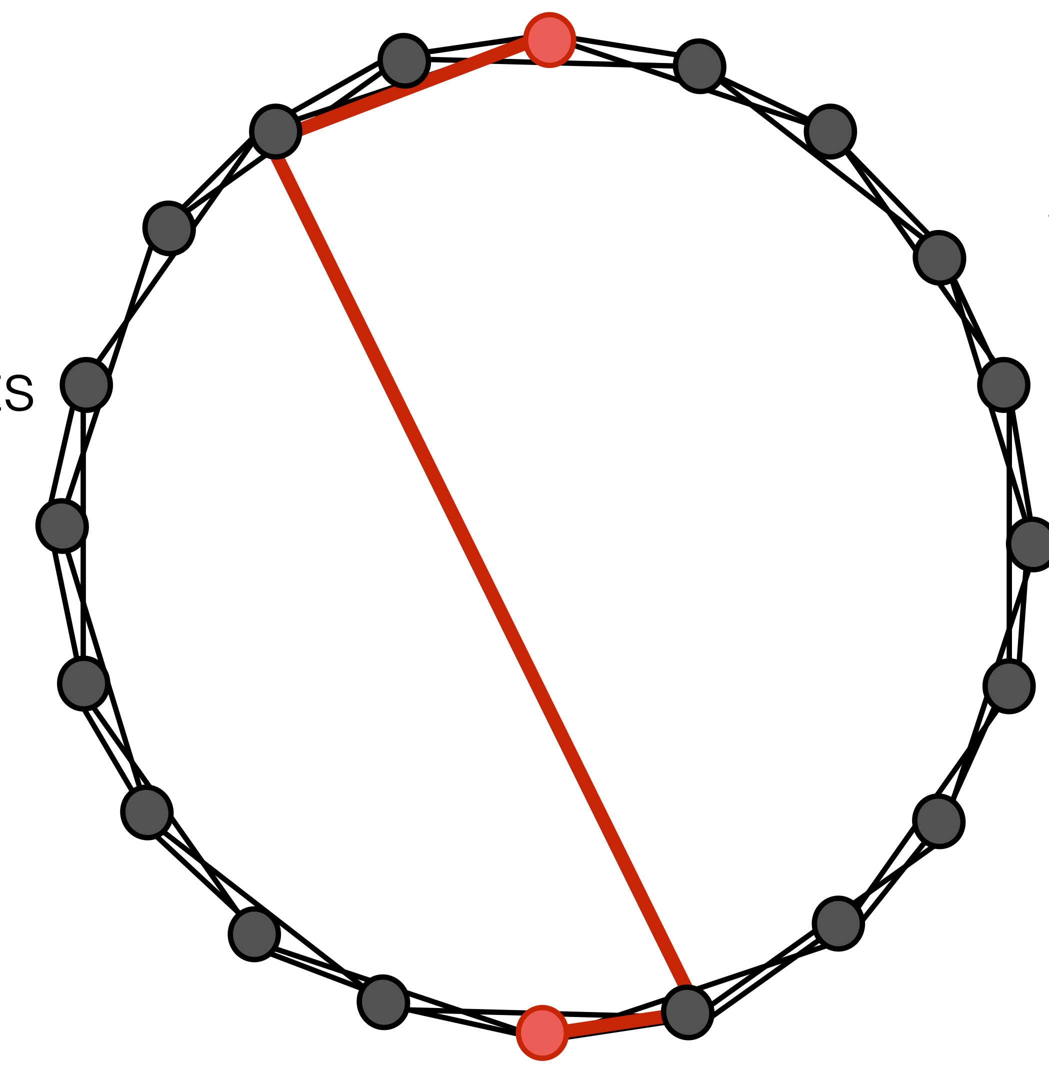
Models of Social Networks: Watts-Strogatz Small World

Even a single rewired link can drastically shorten the distance between many pairs of nodes.



Models of Social Networks: Watts-Strogatz Small World

It only takes a small number of long-distance links to connect distant parts of the network.

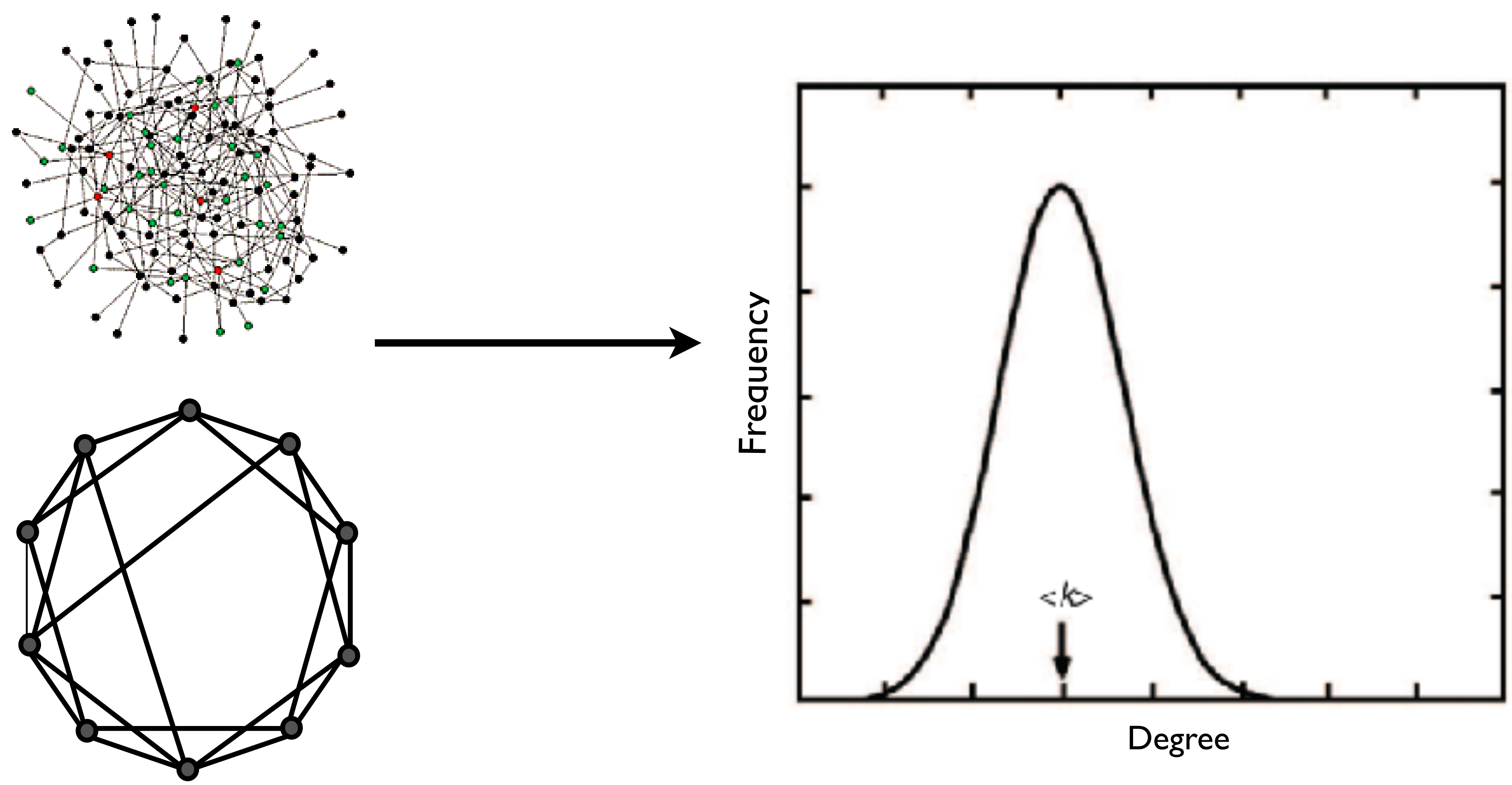


Most of the links in the network are still local, so clustering remains high.

The result is a “small world”: a social network that is both highly local and highly global!

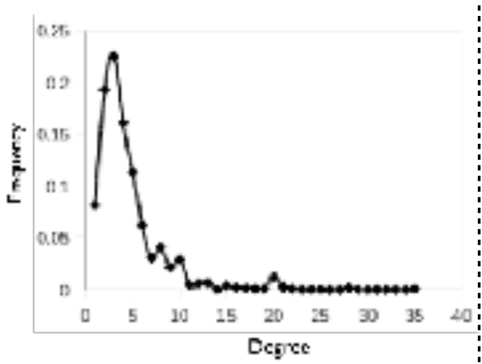
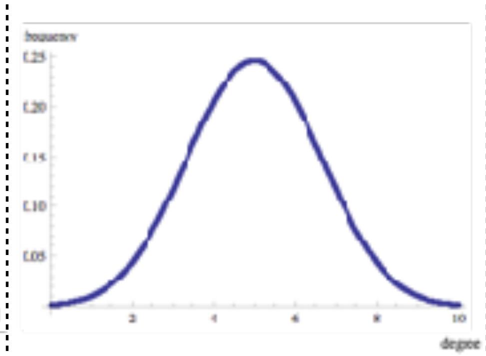
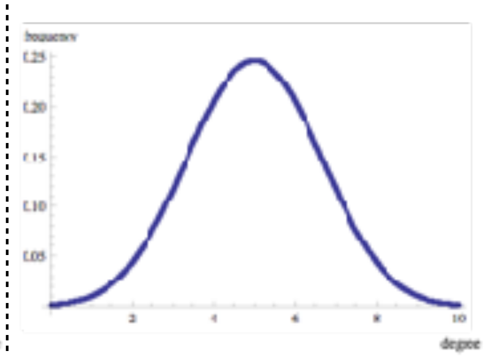
Models of Social Networks

But note that the degree distribution doesn't look like a real social network: because every link is rewired with equal probability, the small world network will have a symmetric degree distribution



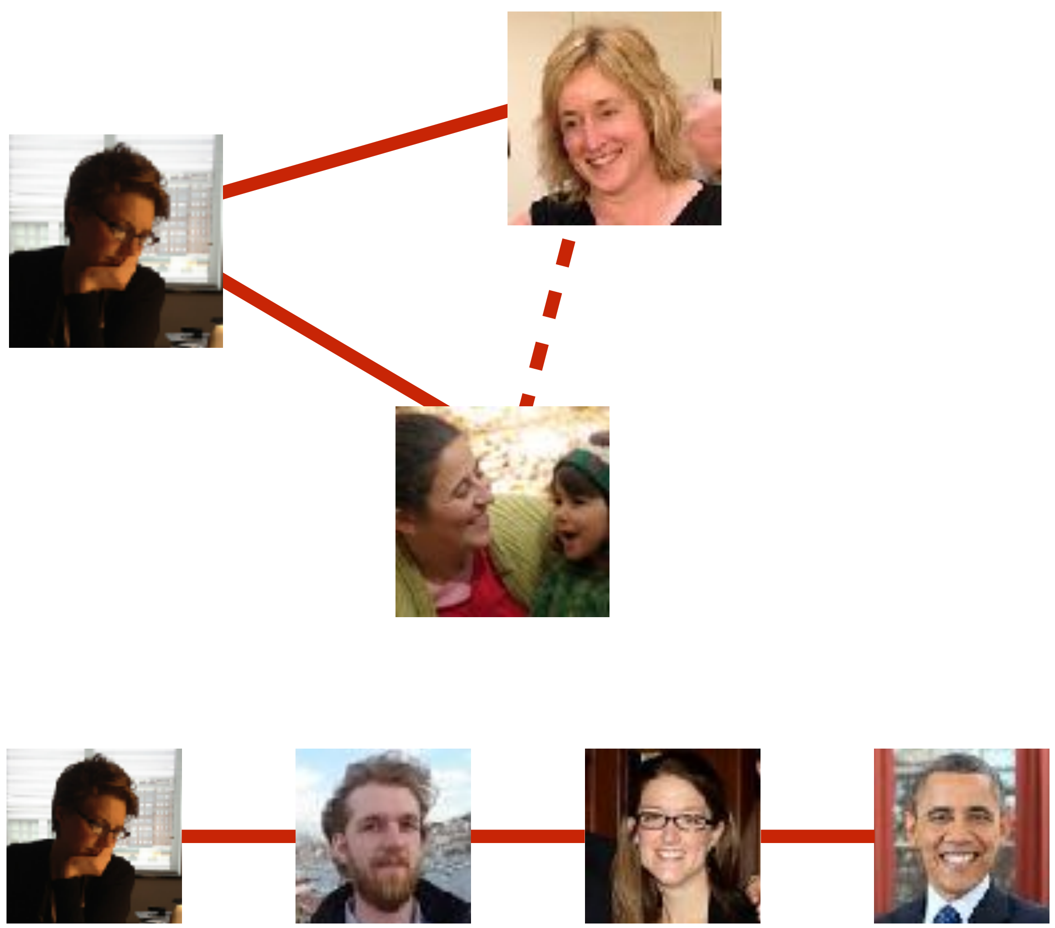
Fun fact: When it's random (*p = 1*), the distribution is the result of a set of Bernoulli random trials, so the degree distribution is binomial

Network Models Summary

	Empirical	Erdös-Renyi	Watts-Strogatz	Preferential Attachment
Average Distance	Low	Low	Low	
Clustering	High	Low	High	
Degree Distribution				

Models of Social Networks

Again, the small world network doesn't match all of the characteristics of a real-world social network. But it still help us understand something useful!

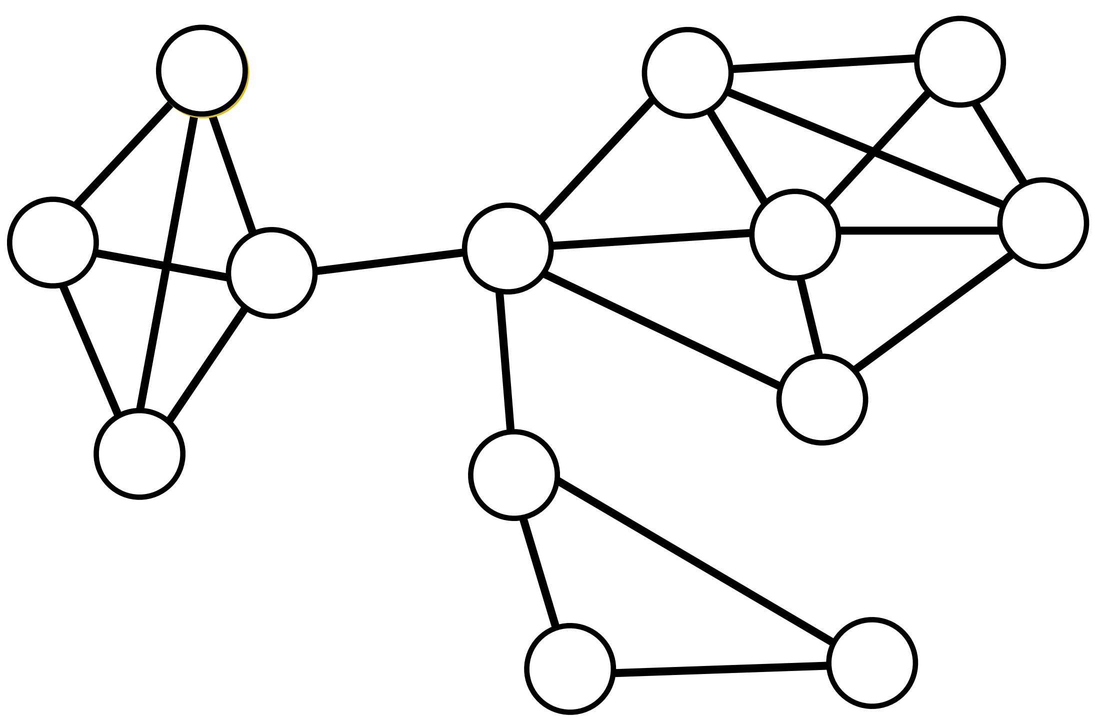


Social networks seem to be full of contradictions:

- We have close-knit groups of friends, who tend to be friends with each other
- But we are also tied to people in far-flung corners of the social world

Models of Social Networks

But the Watts-Strogatz model gives us a way to resolve those contradictions

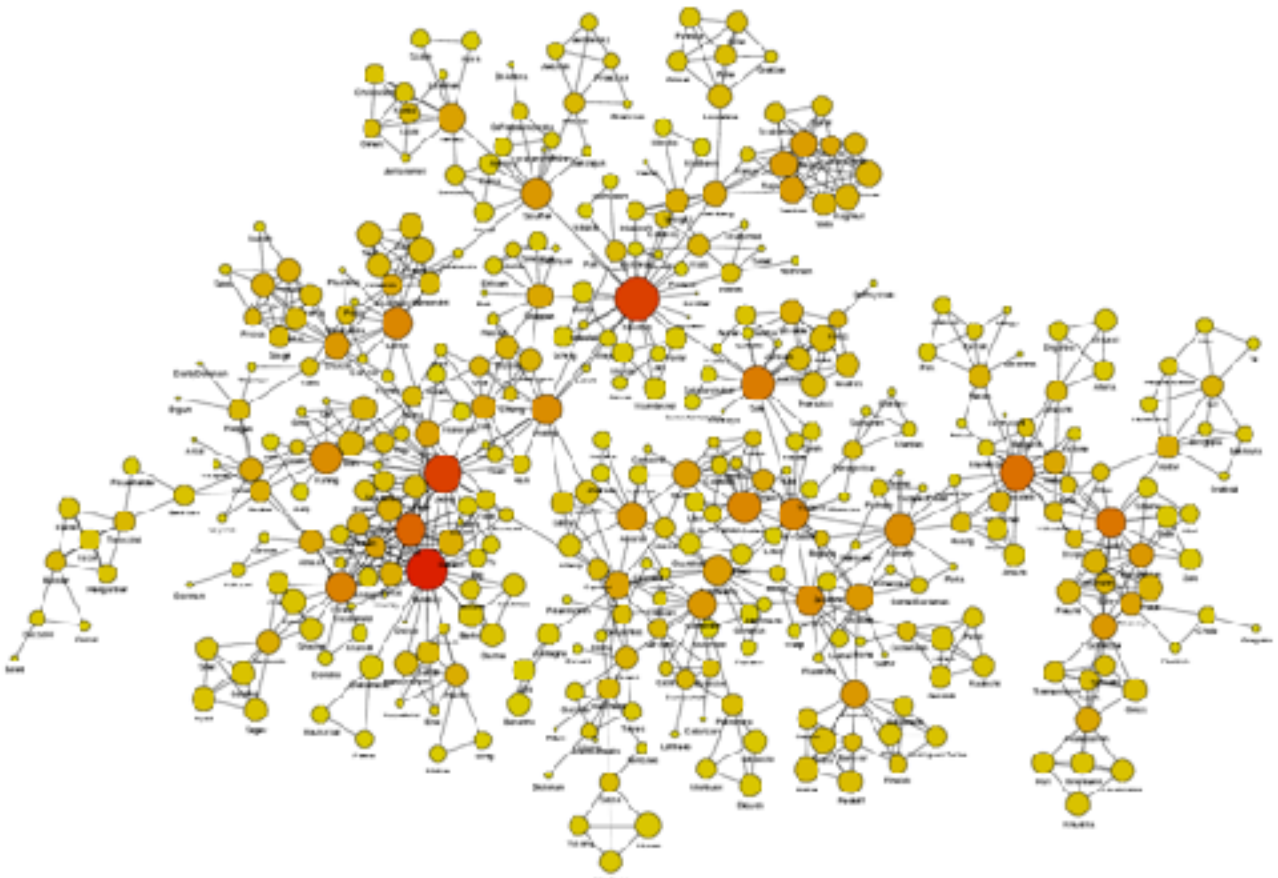


Our tight social groups are our “short distance” links

And it doesn’t take many “long distance” links to shorten the average distance between nodes in the network

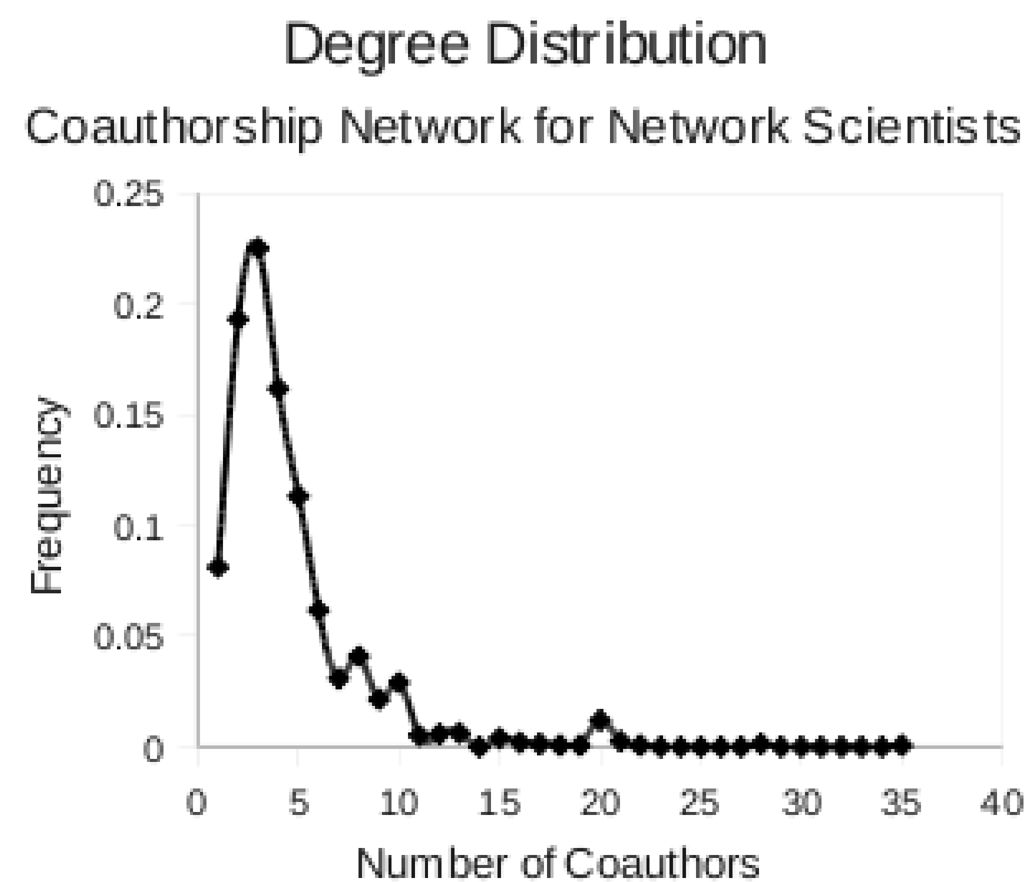
Models of Social Networks

But now, what about that degree distribution?



Network Science Coauthorship Network

Ref: Mark Newman (2006)



Models of Social Networks: Preferential Attachment

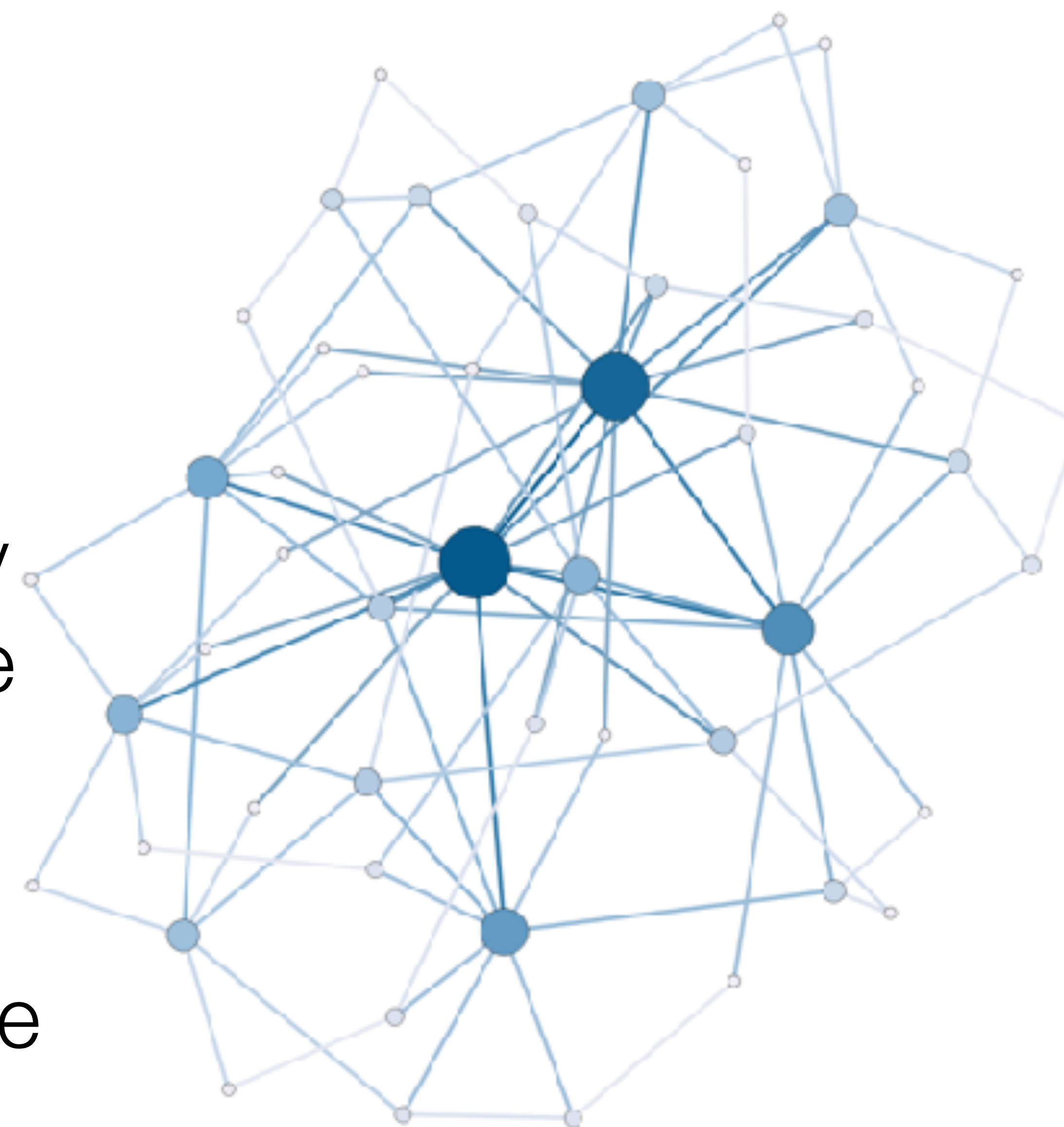
Preferential attachment is a model of network *growth*

Procedure:

- Start with a handful of nodes
- Each time step, add a new node, and link it to *m* existing nodes
- When entering, you are more likely to link to nodes with higher degree

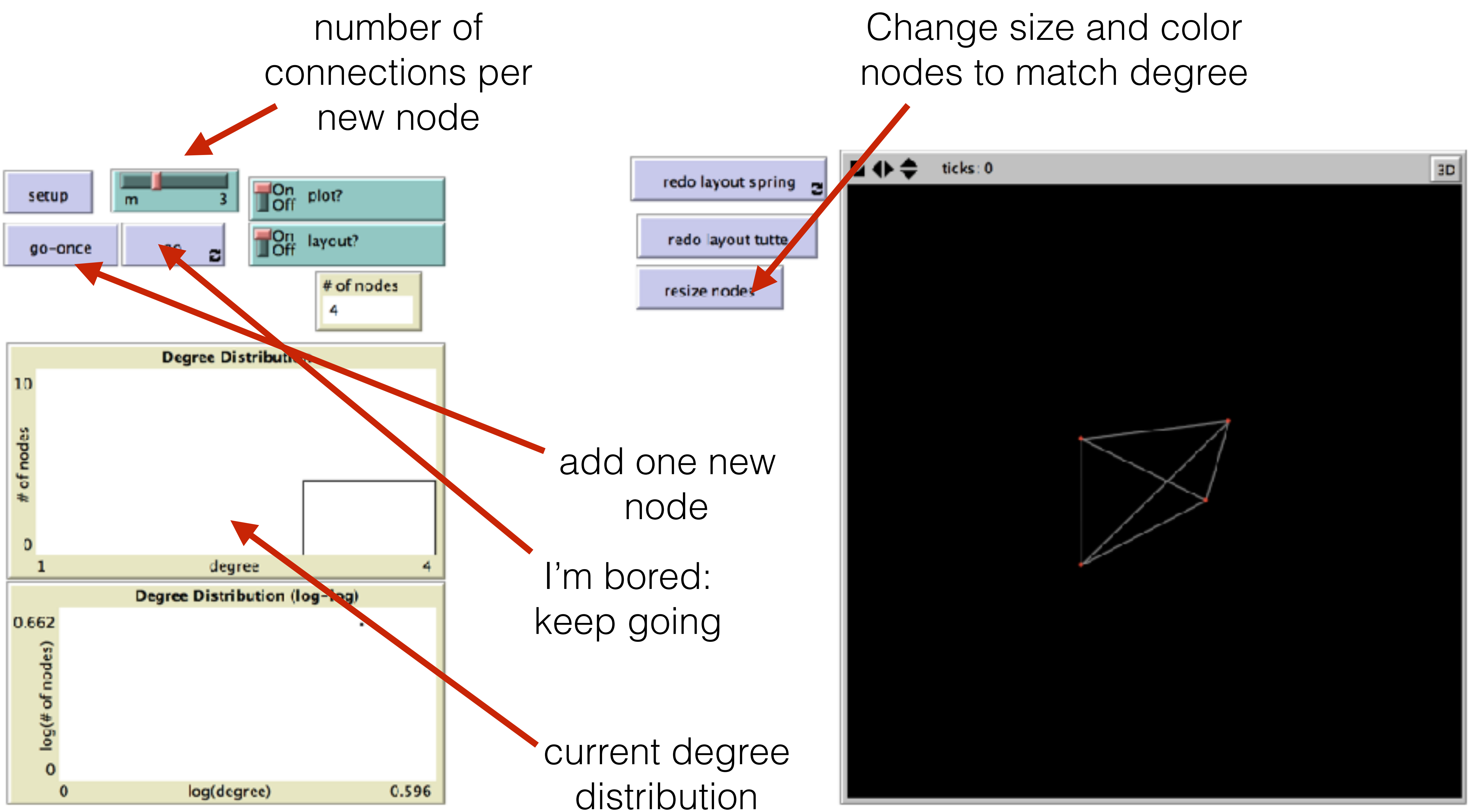
  → more precisely:

  $\text{Prob(link to node i)} \propto d_i$
- Repeat, adding one node at a time


m = 2

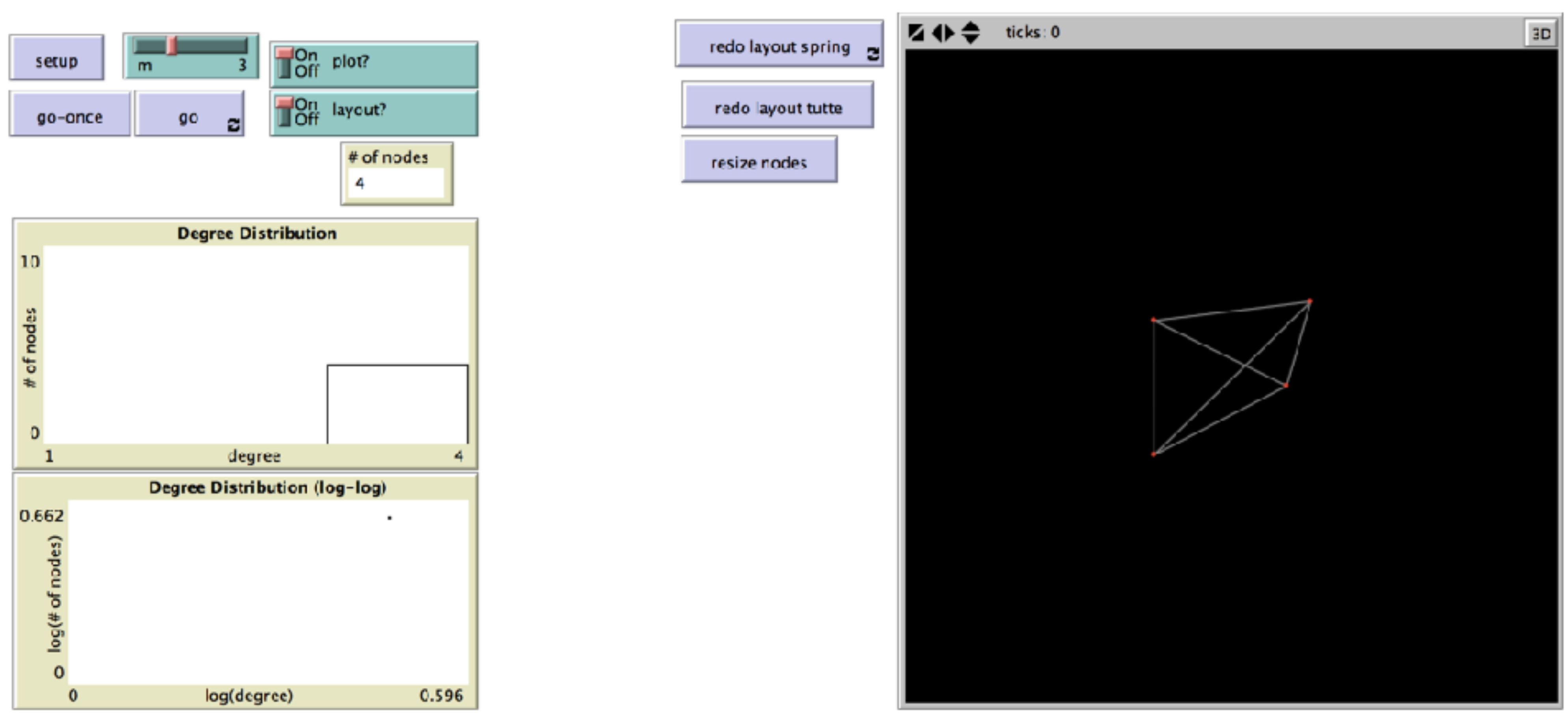
Now open: Preferential Attachment.nlogo

A simulation of the preferential attachment model



Now take a look at what happens:

1) Set up the nodes (m = 3 to start)
2) Hit “go-once” to add one node at a time. Every once in a while, resize and color the nodes by degree. What is happening to the network? To the degree distribution?
3) When you get bored, hit “go”. What happens as more and more nodes are added?
4) Try it again. Do you get the same thing? Now try it will a different value for “m” (the number of connections each new node makes). Do you have any idea what is going on?

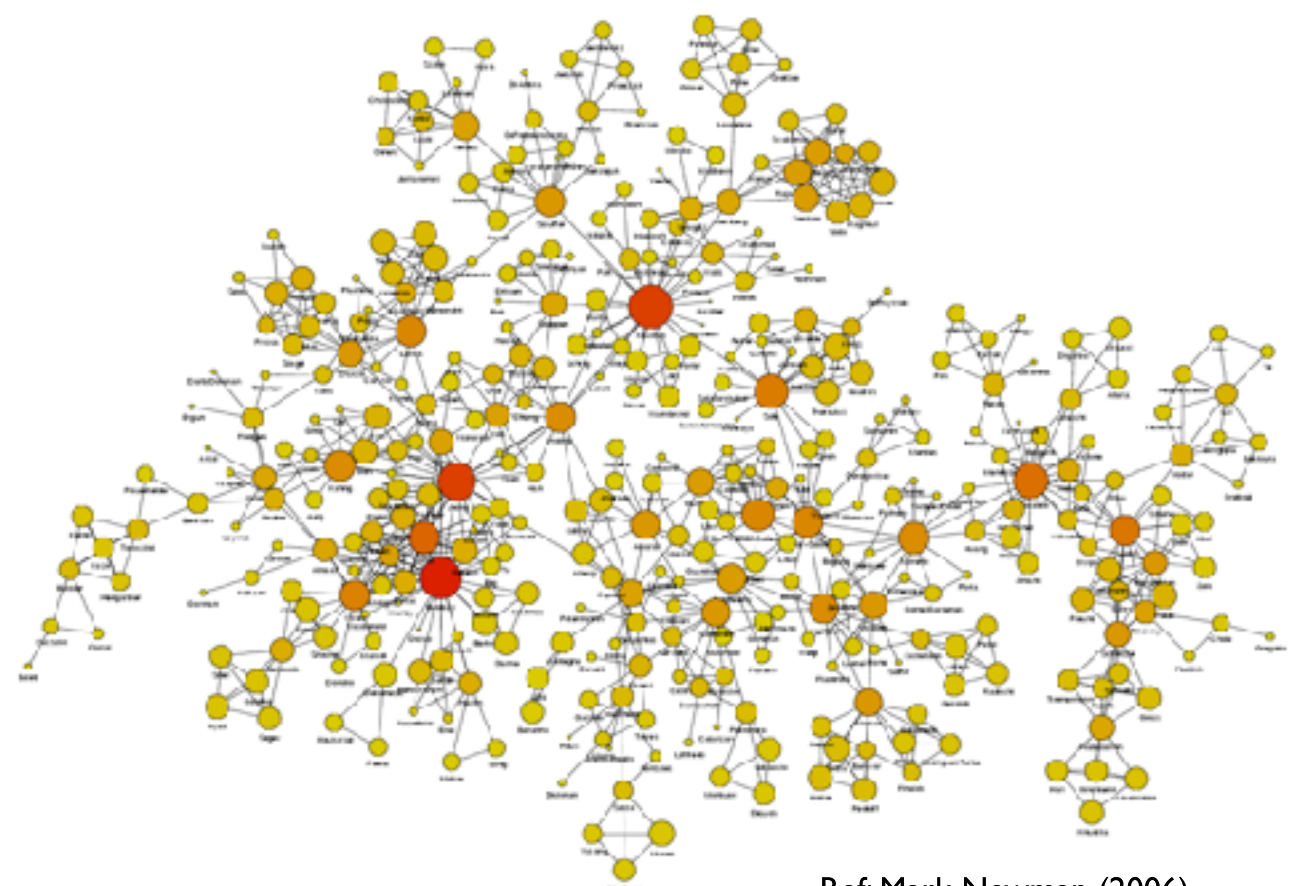


Models of Social Networks: Preferential Attachment

So what does the preferential attachment model tell us about social networks?

It gives us one possible explanation for high degree nodes: links beget links!

Nodes that are already high degree attract even more links over time



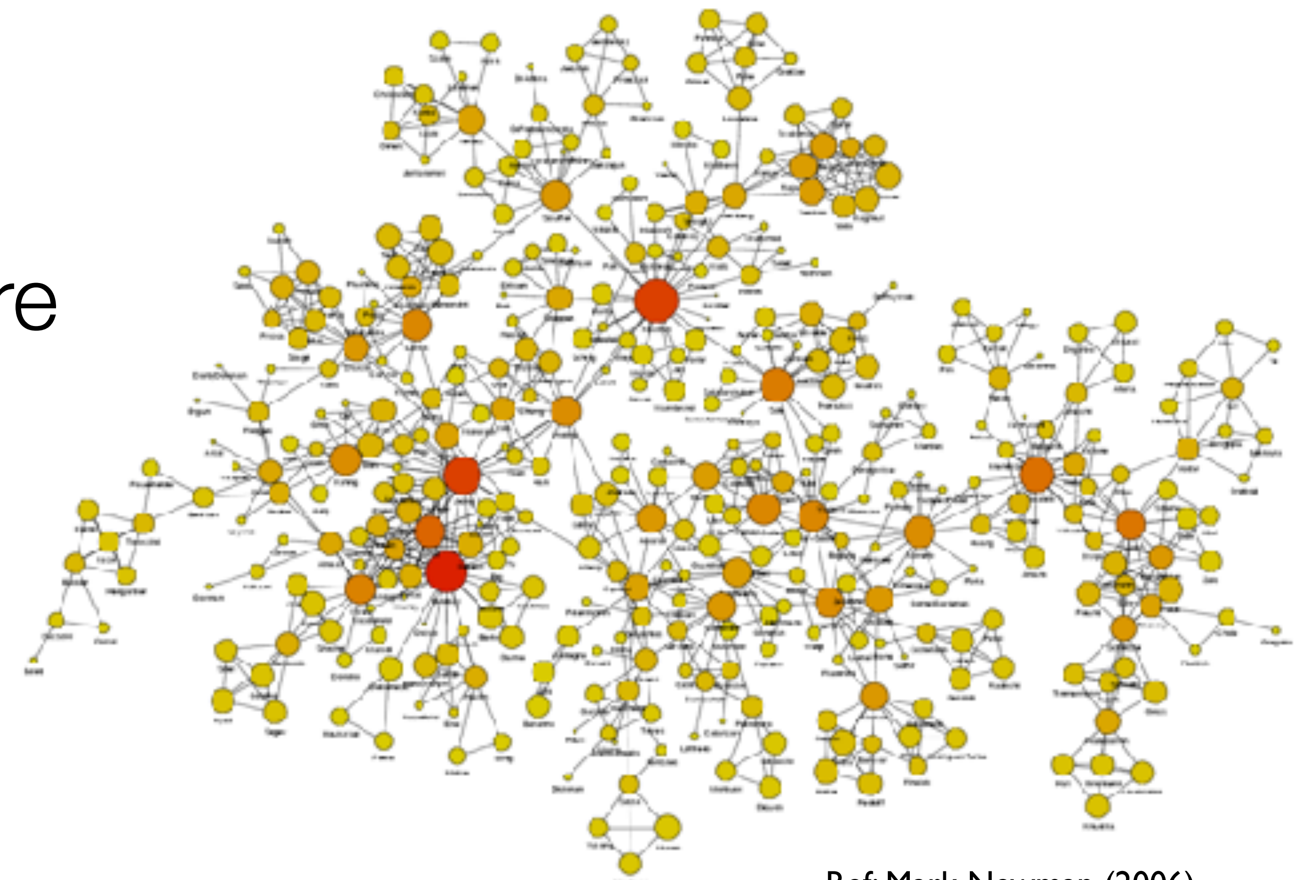
Ref: Mark Newman (2006)

This kind of positive feedback loop is called the “rich-get-richer” effect

Models of Social Networks: Preferential Attachment

This rich-get-richer effect is found in all kinds of contexts:

Twitter accounts with lots of followers tend to attract more followers

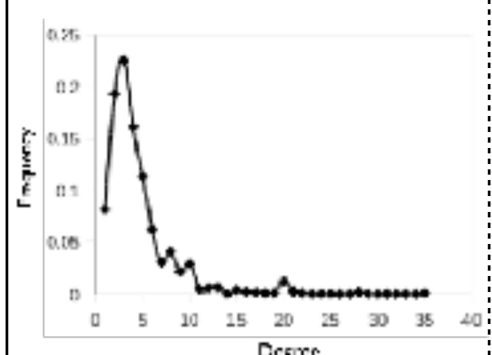
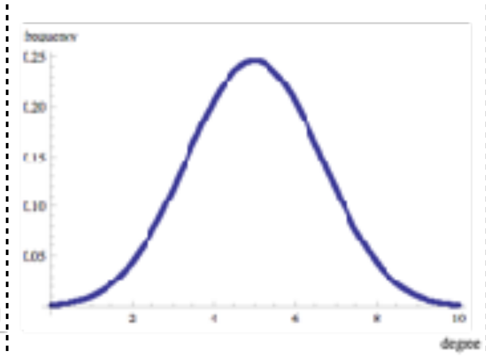
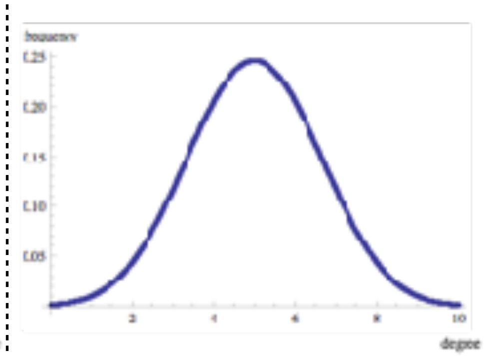
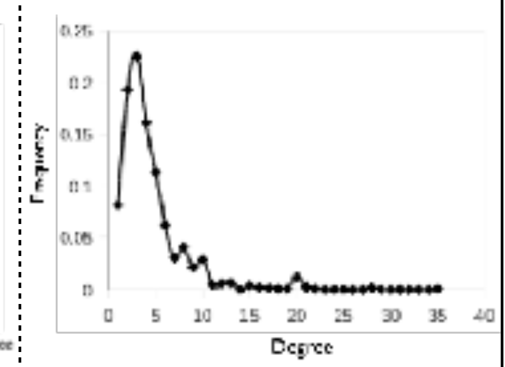


Ref: Mark Newman (2006)

Videos that are popular tend to become even more popular

People who have lots of money tend to make even more money

Network Models Summary

	Empirical	Erdös-Renyi	Watts-Strogatz	Preferential Attachment
Average Distance	Low	Low	Low	Low
Clustering	High	Low	High	~
Degree Distribution				

Lessons for today

Models give us insight into the general by abstracting away from the specific. We've looked at three models that all give us some insight into why networks look the way they do:

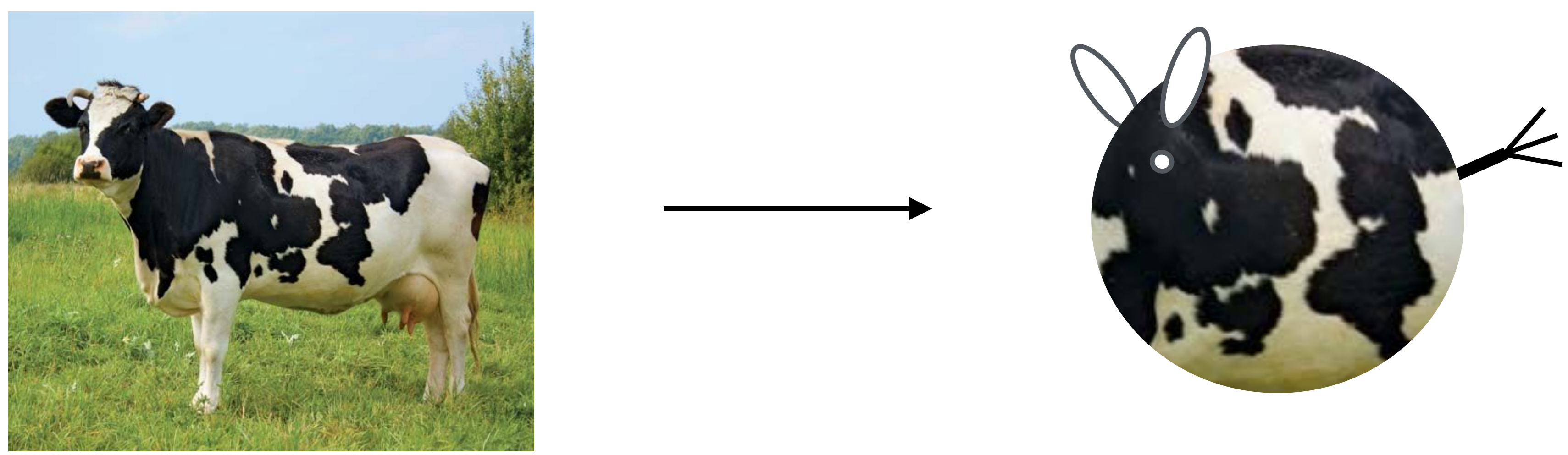
Erdös-Renyi (random) networks: network growth is a non-linear process

Watts-Strogatz: a few long-distance connections make the world both locally clustered and globally small

Preferential Attachment: when links beget links, the result is a network with a few individuals who have a large number of links

Lessons for today

Also, different models of cows help us understand different things about cows



When someone tells you something about cows, make sure you understand the assumptions they are making about cows…or I suppose other things as well.