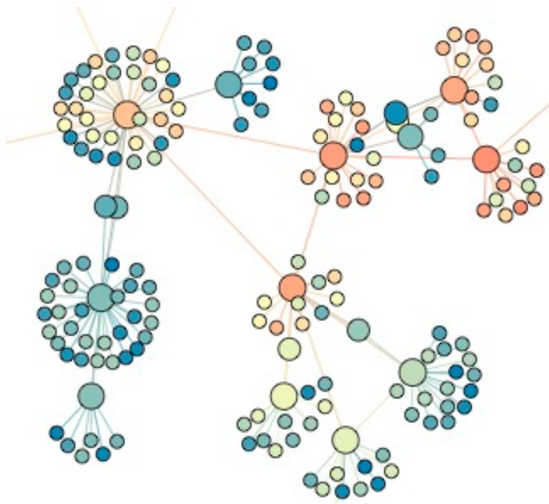


Random Graphs

CS224W

Network models

- Why model?
  - simple representation of complex network
  - can derive properties mathematically
  - predict properties and outcomes
- Also: to have a strawman
  - In what ways is your real-world network different from hypothesized model?
  - What insights can be gleaned from this?

Downloading NetLogo

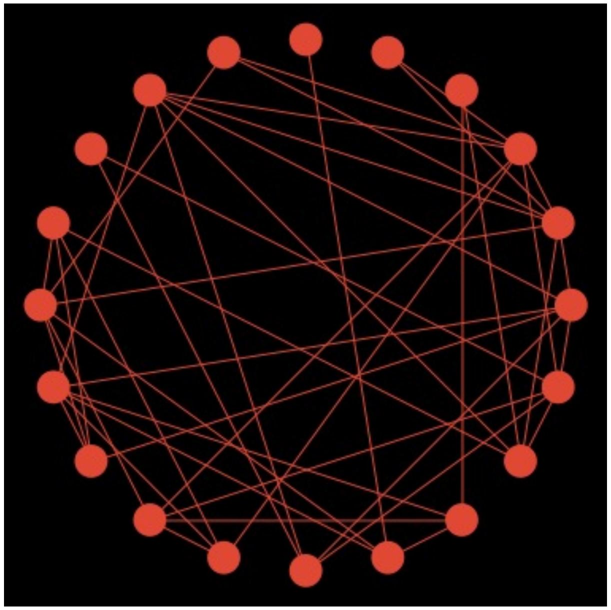
- https://ccl.northwestern.edu/netlogo/
- Models specific to this class: http://web.stanford.edu/class/cs224w/NetLogo/

Erdös and Rényi

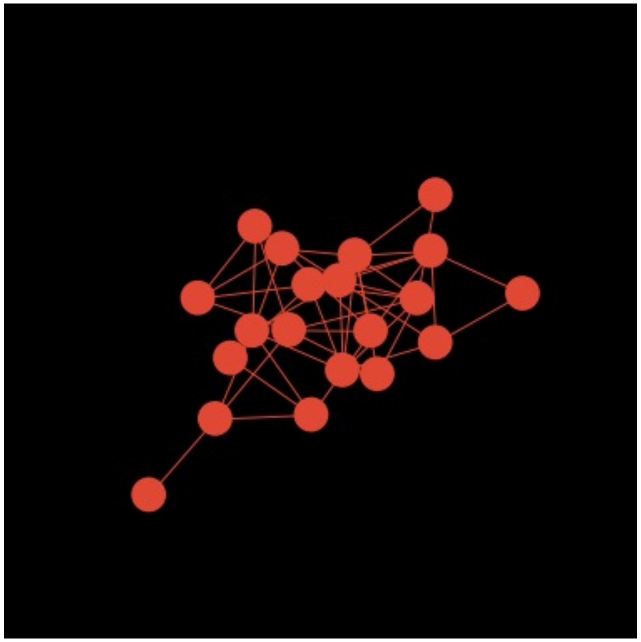
Erdös-Renyi: simplest network model

- Assumptions
  - nodes connect at random
  - network is undirected
- Key parameter (besides number of nodes N) : p or M
  - p = probability that any two nodes share and edge
  - M = total number of edges in the graph

what they look like



after spring layout



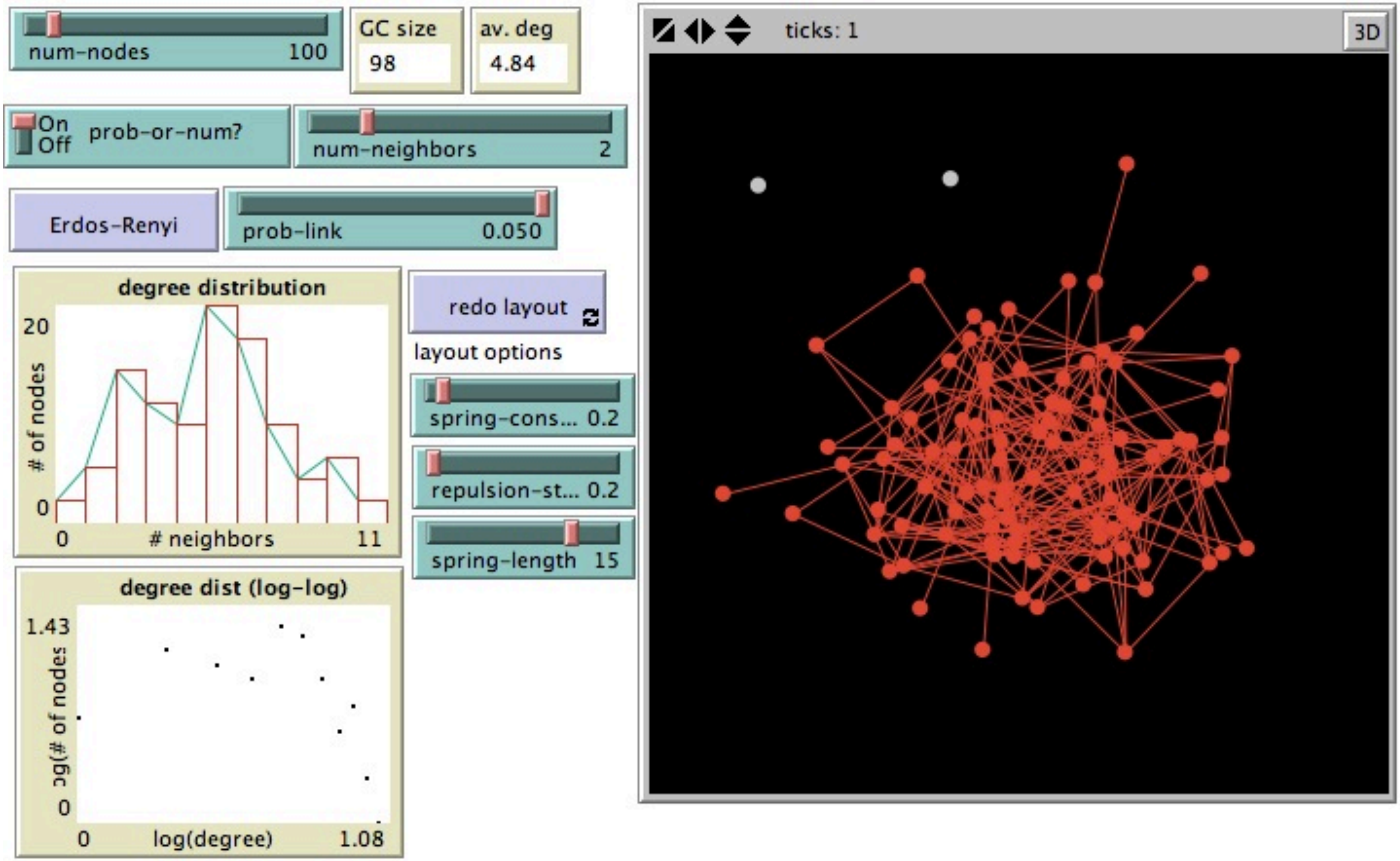
Degree distribution

- (N,p)-model: For each potential edge we flip a biased coin
  - with probability p we add the edge
  - with probability (1-p) we don't
- Alternate notation: G_{np}

Quiz Q:

- As the size of the network increases, if you keep ***p***, the probability of any two nodes being connected, the same, what happens to the average degree
  - a) stays the same
  - b) increases
  - c) decreases

http://web.stanford.edu/class/cs224w/NetLogo/ErdosRenyiDegDist.nlogo



http://web.stanford.edu/class/cs224w/NetLogo/ErdosRenyiDegDist.nlogo

Degree distribution

- What is the probability that a node has 0,1,2,3… edges?
- Probabilities sum to 1

How many edges per node?

- Each node has (N – 1) tries to get edges
- Each try is a success with probability p
- The binomial distribution gives us the probability that a node has degree k:

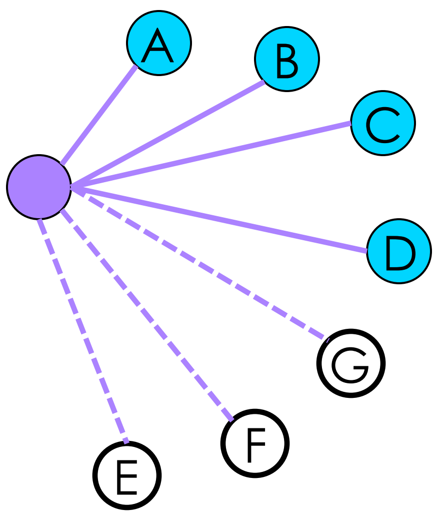
$$B(N-1;k;p) = \binom{N-1}{k} p^k (1-p)^{N-1-k}$$

Quiz Q:

- The maximum degree of a node in a simple (no multiple edges between the same two nodes) N node graph is
  - a) N
  - b) N - 1
  - c) N / 2

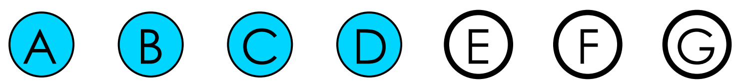
Explaining the binomial distribution

- 8 node graph, probability ***p*** of any two nodes sharing an edge
- What is the probability that a given node has degree 4?

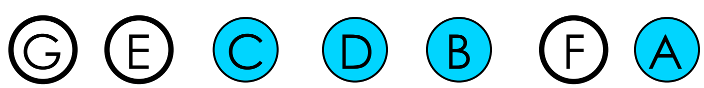


Binomial coefficient: choosing 4 out of 7

Suppose I have 7 blue and white nodes, each of them uniquely marked so that I can distinguish them. The blue nodes are ones I share an edge with, the white ones I don't.



How many different samples can I draw containing the same nodes but in a different order (the order could be e.g. the order in which the edges are added (or not)? e.g.



binomial coefficient explained

G E C D B F A

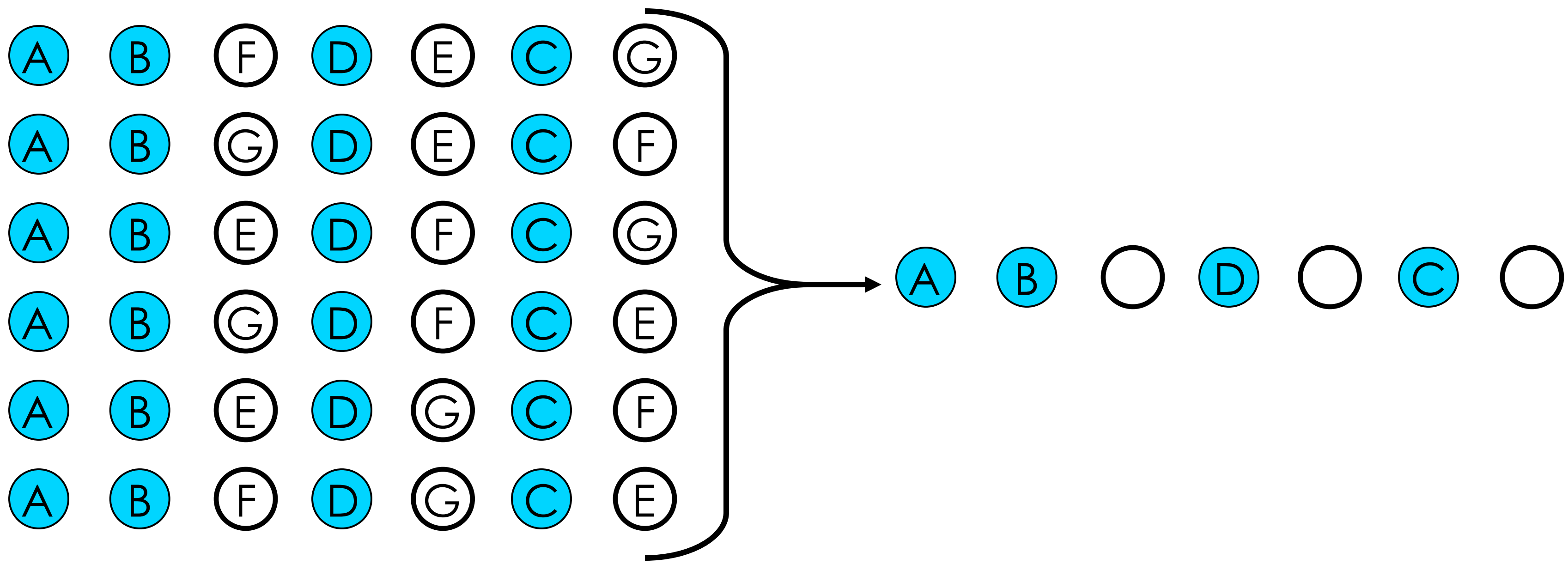
If order matters, there are **7!** different orderings:

I have 7 choices for the first spot, 6 choices for the second (since I've picked 1 and now have only 6 to choose from),
5 choices for the third, etc.

7! = 7 * 6 * 5 * 4 * 3 * 2 * 1

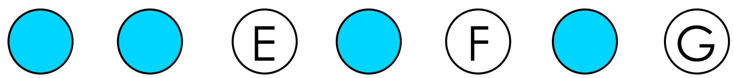
binomial coefficient

Suppose the order of the nodes I don't connect to (white) doesn't matter.
All possible arrangements (3!) of white nodes look the same to me.



Instead of 7! combinations, we have 7!/3! combinations

binomial coefficient explained



The same goes for the blue nodes, if we can’t tell them apart, we lose a factor of 4!

binomial coefficient explained

number of ways of choosing k items out of (n-1)

$$= \frac{\text{number of ways of arranging } \boldsymbol{n-1} \text{ items}}{(\text{\# of ways to arrange } \boldsymbol{k} \text{ things})*(\text{\# ways to arrange } \boldsymbol{n}\text{-1-}\boldsymbol{k} \text{ things})}$$

$$= \frac{\boldsymbol{n-1}!}{\boldsymbol{k}!\,(\boldsymbol{n}-1-\boldsymbol{k})!}$$

Note that the binomial coefficient is symmetric – there are the same number of ways of choosing ***k*** or ***n-1-k*** things out of ***n-1***

Quiz Q:

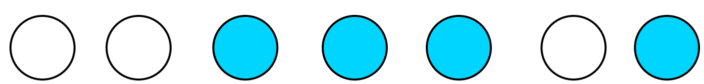
- What is the number of ways of choosing 2 items out of 5?
  - 10
  - 120
  - 6
  - 5

Now the distribution

- p = probability of having edge to node (blue)
- (1-p) = probability of not having edge (white)
- The probability that you connect to 4 of the 7 nodes in some particular order (two white followed by 3 blues, followed by a white followed by a blue) is

P(white)*P(white)*P(blue)*P(blue)*P(blue)*P(white)*P(blue)

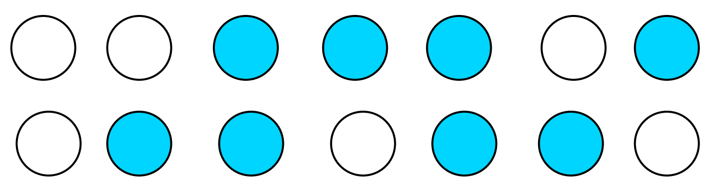
$= p^4 * (1-p)^3$



Binomial distribution

- If order doesn't matter, need to multiply probability of any given arrangement by number of such arrangements:

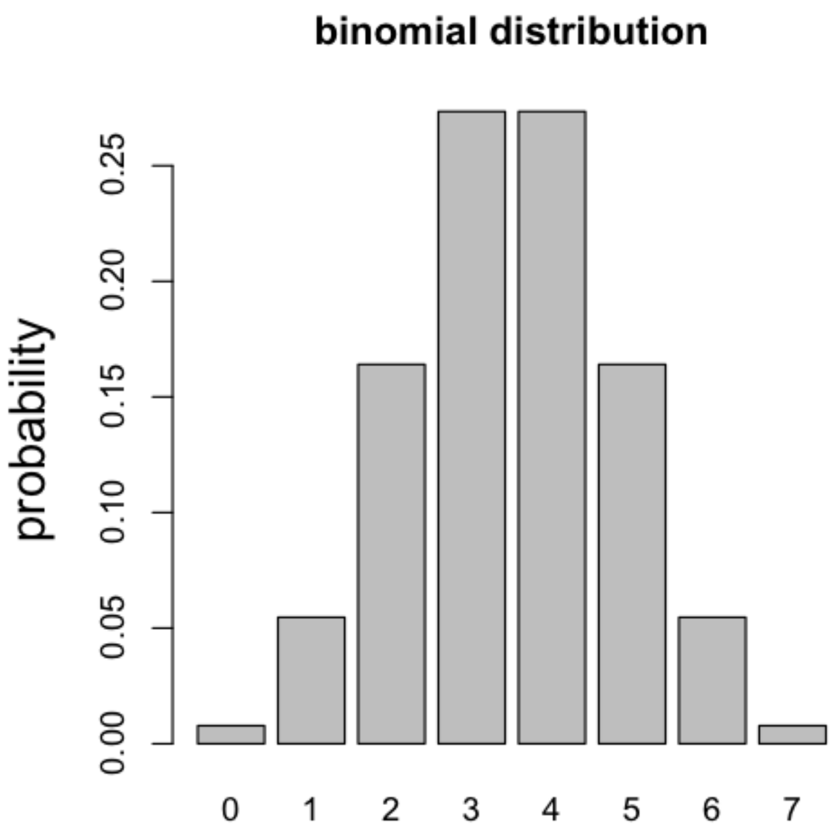
$$B(7;4;p) = \binom{7}{4} p^4 (1-p)^3$$



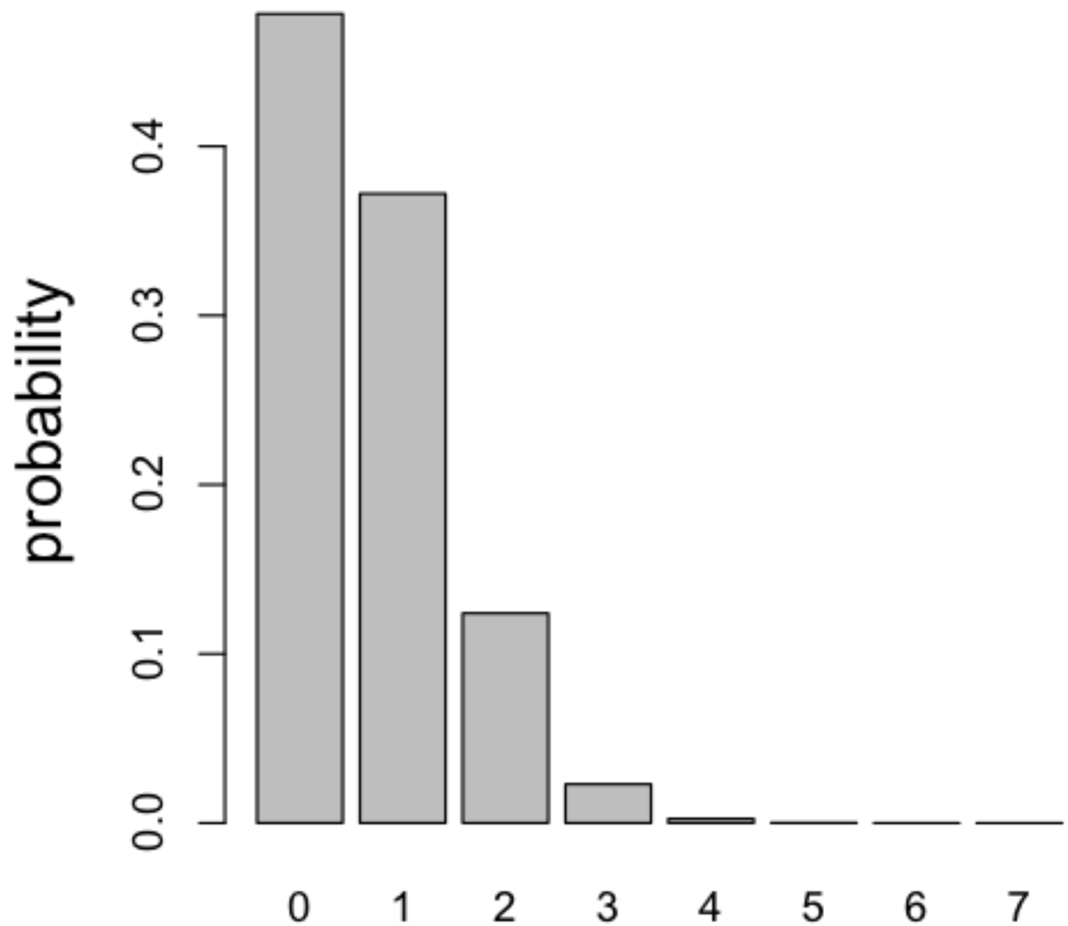
+

....

if p = 0.5

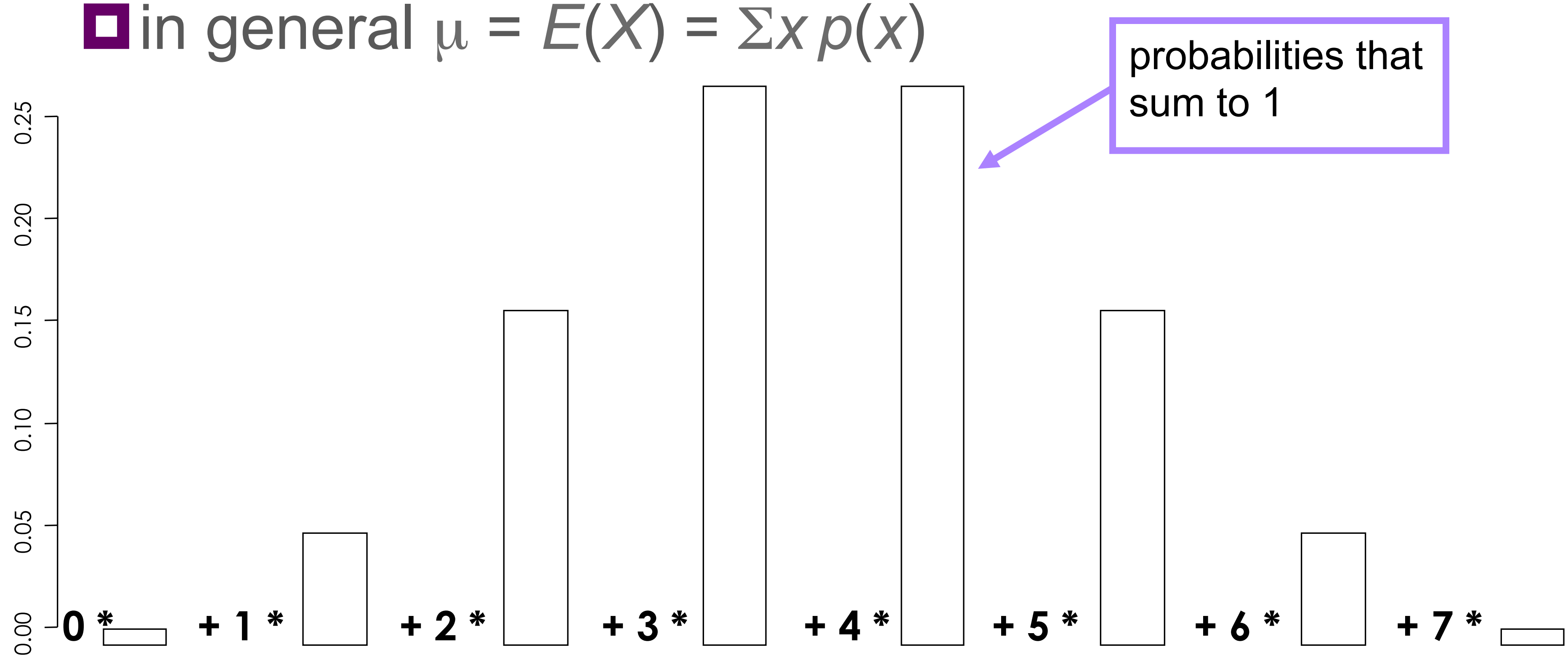

binomial distribution
probability
0.00
0.05
0.10
0.15
0.20
0.25
0
1
2
3
4
5
6
7

p = 0.1



What is the mean?

- Average degree $\langle k \rangle = z = (n-1)*p$
- in general $\mu = E(X) = \Sigma x\, p(x)$



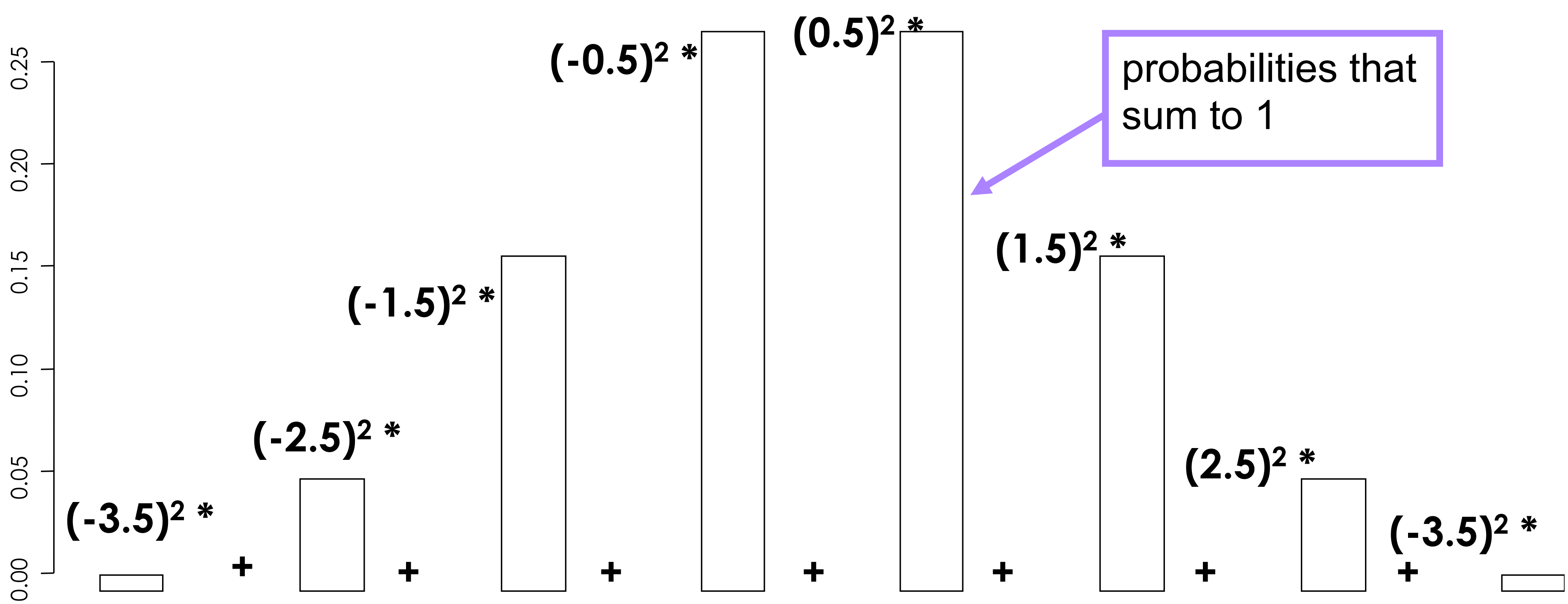
$\mu = 3.5$

Quiz Q:

- What is the average degree of a graph with 10 nodes and probability $p = 1/3$ of an edge existing between any two nodes?
  - 1
  - 2
  - 3
  - 4

What is the variance?

- variance in degree σ^2=**(n-1)*p*(1-p)**
- in general $\sigma^2 = E[(X-\mu)^2] = \Sigma\ (x-\mu)^2\ p(x)$



Approximations

$$p_k = \binom{n-1}{k} p^k (1-p)^{n-1-k}$$

Binomial

↓ limit *p* small

$$p_k = \frac{z^k e^{-z}}{k!}$$

Poisson

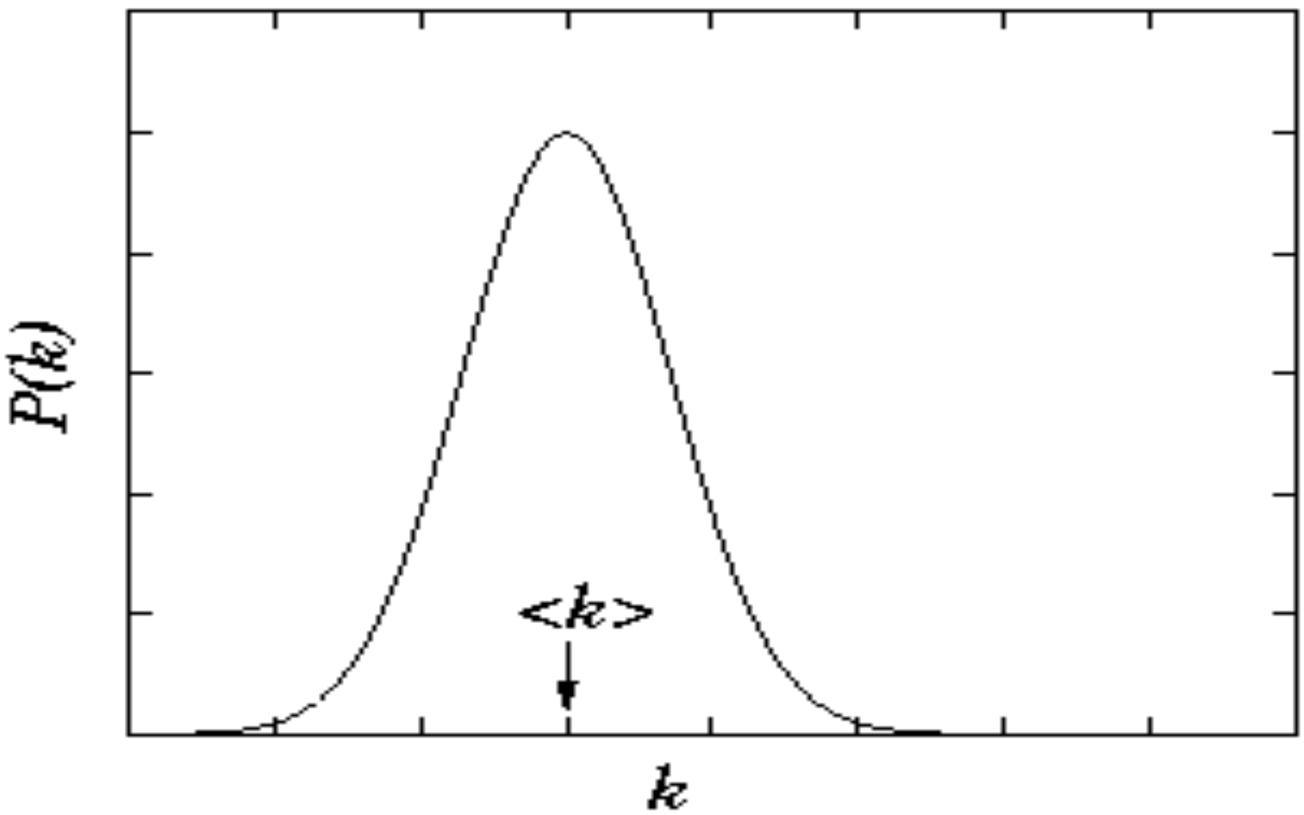
↓ limit large *n*

$$p_k = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(k-z)^2}{2\sigma^2}}$$

Normal

Poisson distribution

Poisson distribution



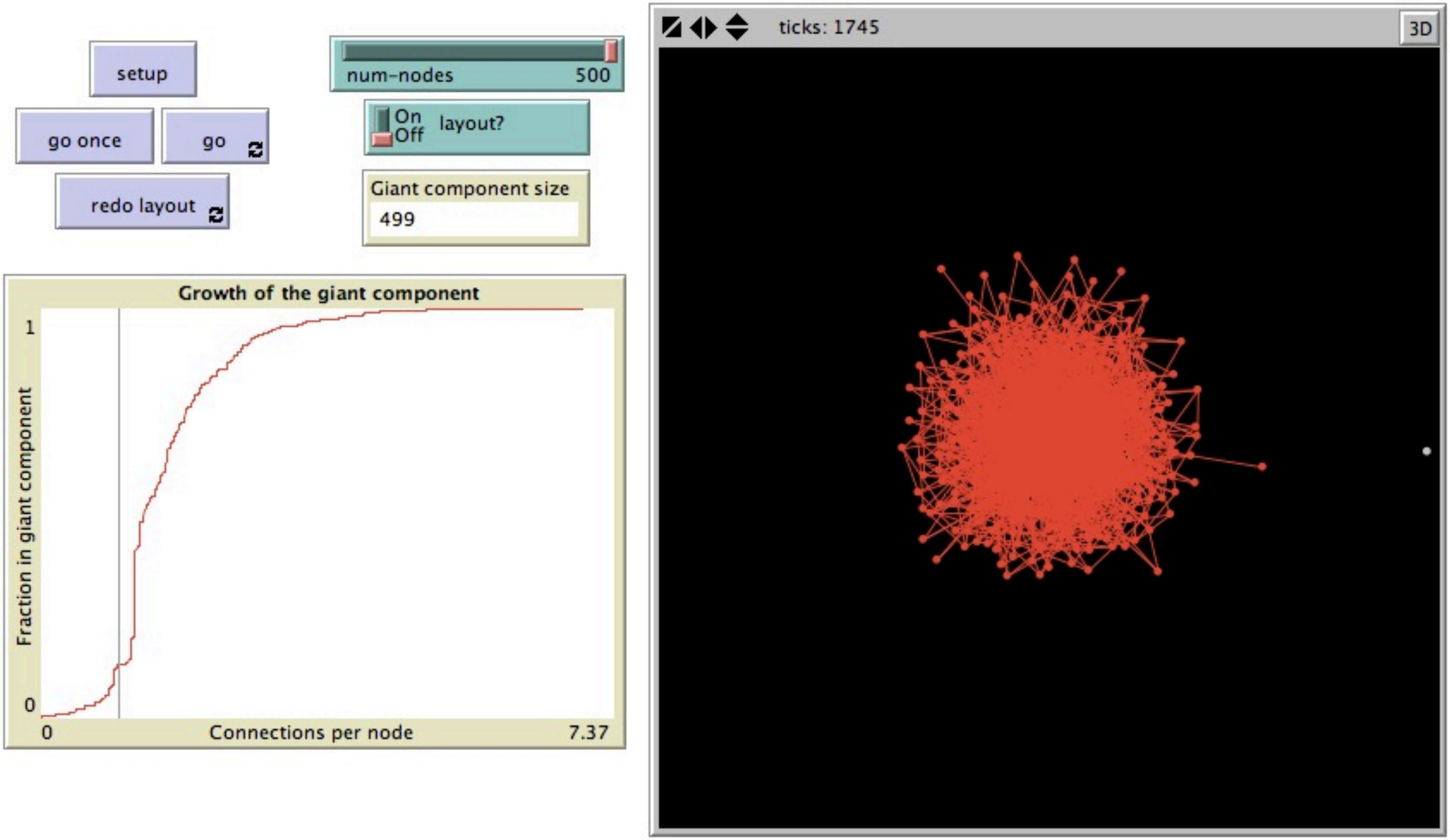
What insights does this yield? No hubs

- You don’t expect large hubs in the network

Insights

- Previously: degree distribution / absence of hubs
- Emergence of giant component
- Average shortest path

Emergence of the giant component

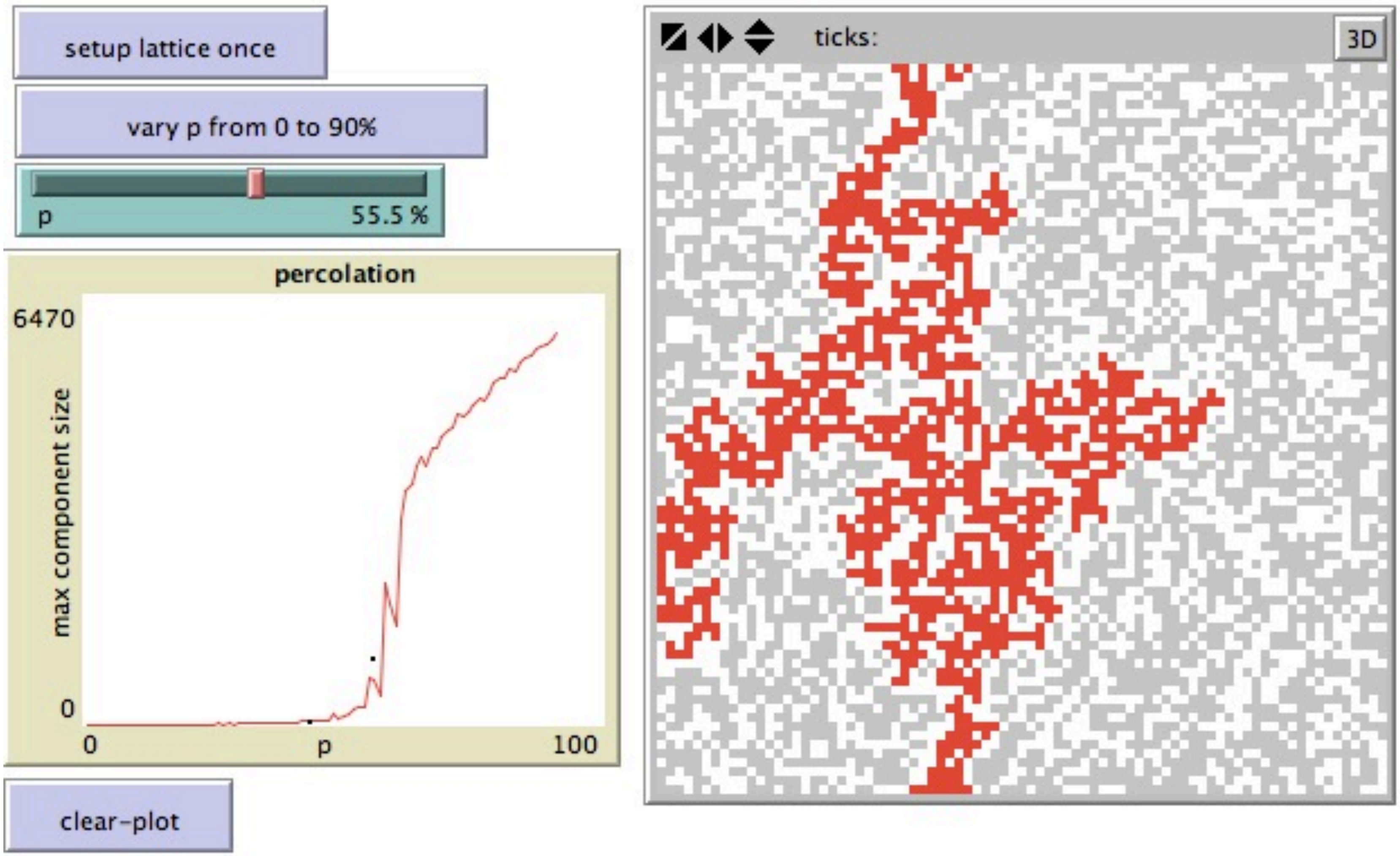


[(standard model in NetLogo library) http://ccl.northwestern.edu/netlogo/models/GiantComponent](http://ccl.northwestern.edu/netlogo/models/GiantComponent)

Quiz Q:

- What is the average degree z at which the giant component starts to emerge?
  - 0
  - 1
  - 3/2
  - 3

Percolation on a 2D lattice

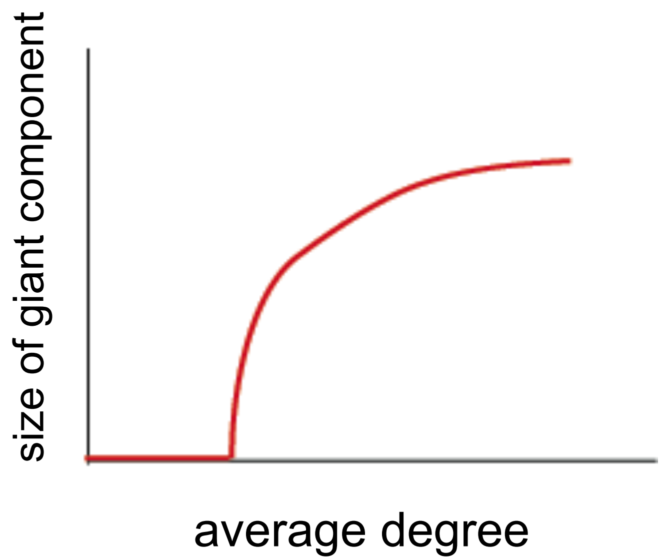


http://web.stanford.edu/class/cs224w/NetLogo/LatticePercolation.nlogo

Quiz Q:

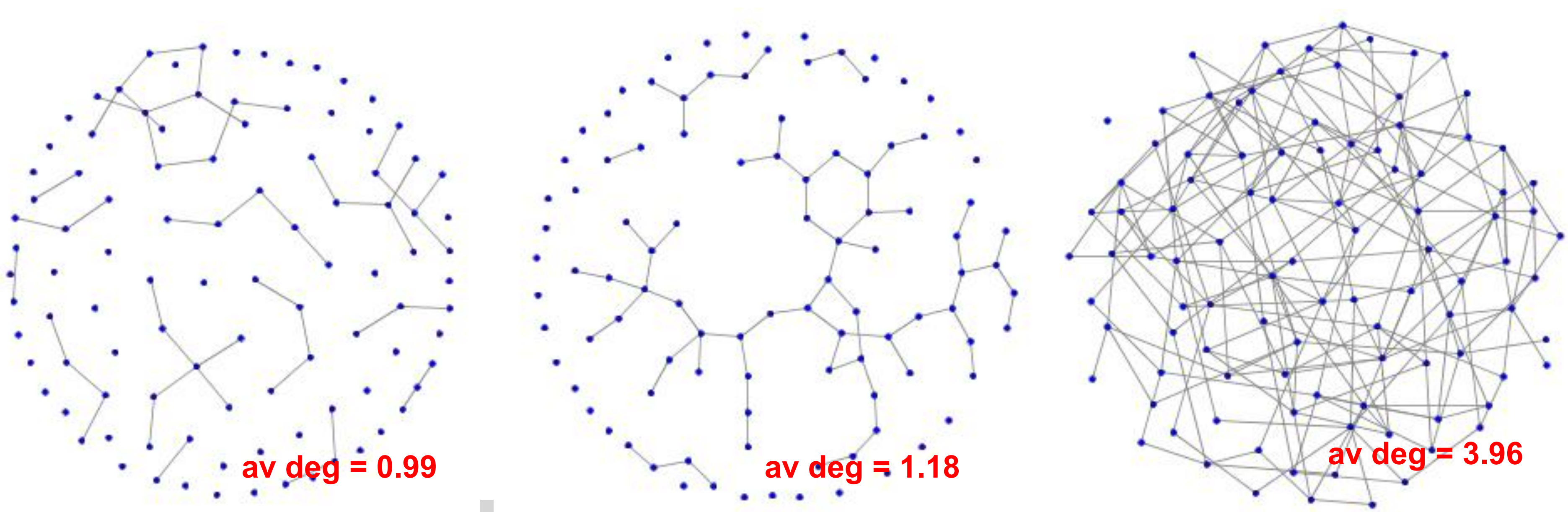
- What is the percolation threshold of a 2D lattice: fraction of sites that need to be occupied in order for a giant connected component to emerge?
  - 0
  - ¼
  - 1/3
  - 1/2

Percolation threshold



Percolation threshold: how many edges need to be added before the giant component appears?

As the average degree increases to z = 1, a giant component suddenly appears



"Evolution" of the G_{np}

What happens to G_{np} when we vary p?

Back to Node Degrees of G_{np}

- **Remember, expected degree** $E[X_v] = (n-1)p$
- **If want $E[X_v]$ be independent of n**

let: $p=c/(n-1)$

Probability of a node being isolated

- **Observation:** If we build random graph G_{np} with $p=c/(n-1)$ we have many isolated nodes
- **Why?**

$$P[v \text{ has degree } 0] = (1-p)^{n-1} = \left(1 - \frac{c}{n-1}\right)^{n-1} \xrightarrow[n\to\infty]{} e^{-c}$$

$$\lim_{n\to\infty}\left(1-\frac{c}{n-1}\right)^{n-1} = \left(1-\frac{1}{x}\right)^{-x\cdot c} = \left[\lim_{x\to\infty}\left(1-\frac{1}{x}\right)^{-x}\right]^{-c} = e^{-c}$$

Use substitution $\frac{1}{x} = \frac{c}{n-1}$

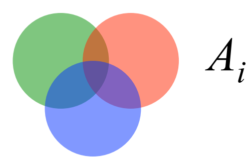
e (by definition)

No Isolated Nodes

- **How big do we have to make p before we are likely to have no isolated nodes?**
- We know: $P[v \text{ has degree } 0] = e^{-c}$
- Event we are asking about is:
  - I = some node is isolated
  - $I = \bigcup_{v \in N} I_v$ where I_v is the event that v is isolated
- **We have:**

$$P(I) = P\left(\bigcup_{v \in N} I_v\right) \leq \sum_{v \in N} P(I_v) = ne^{-c}$$

Union bound



$$\left|\bigcup_i A_i\right| \leq \sum_i |A_i|$$

No Isolated Nodes

- **We just learned:** $P(I) = n\, e^{-c}$
- Let's try:
  - $c = \ln n$ then: $n\, e^{-c} = n\, e^{-\ln n} = n \cdot 1/n = 1$
  - $c = 2 \ln n$ then: $n\, e^{-2 \ln n} = n \cdot 1/n^2 = 1/n$
- **So if:**
  - $p = \ln n$ then: $P(I) = 1$
  - $p = 2 \ln n$ then: $P(I) = 1/n \rightarrow 0$ as $n \rightarrow \infty$

"Evolution" of a Random Graph

- **Graph structure of G_{np} as p changes:**

- **Emergence of a Giant Component:**

avg. degree $k=2E/n$ or $p=k/(n-1)$

- $k=1-\varepsilon$: all components are of size $\Omega(\log n)$
- $k=1+\varepsilon$: 1 component of size $\Omega(n)$, others have size $\Omega(\log n)$

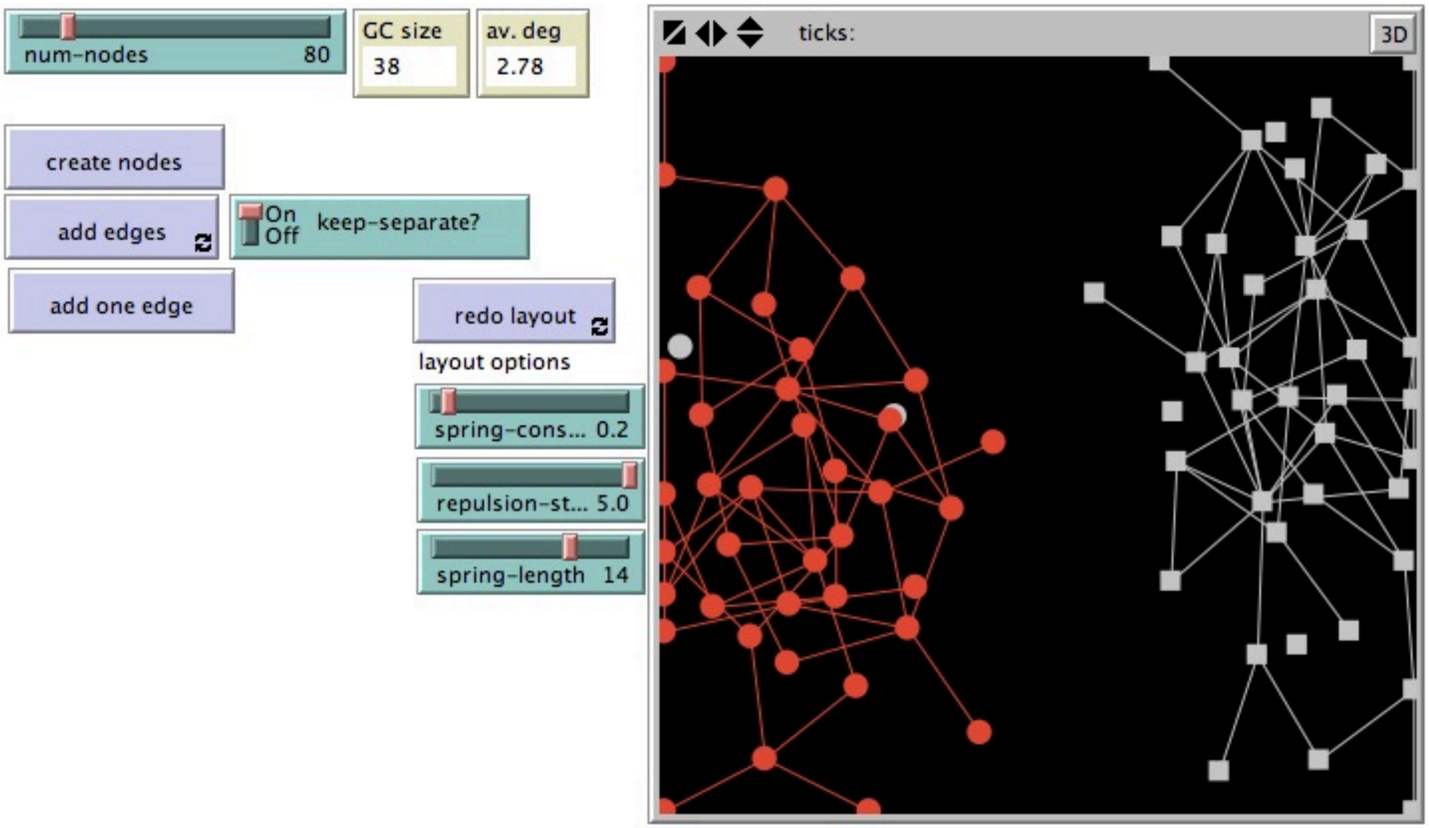
Giant component – another angle

- How many other friends besides you does each of your friends have?
- By property of degree distribution
  - the average degree of your friends, you excluded, is z
  - so at z = 1, each of your friends is expected to have another friend, who in turn have another friend, etc.
  - the giant component emerges

Giant component illustrated

Why just one giant component?

- What if you had 2, how long could they be sustained as the network densifies?



http://web.stanford.edu/class/cs224w/NetLogo/ErdosRenyiTwoComponents.nlogo

Quiz Q:

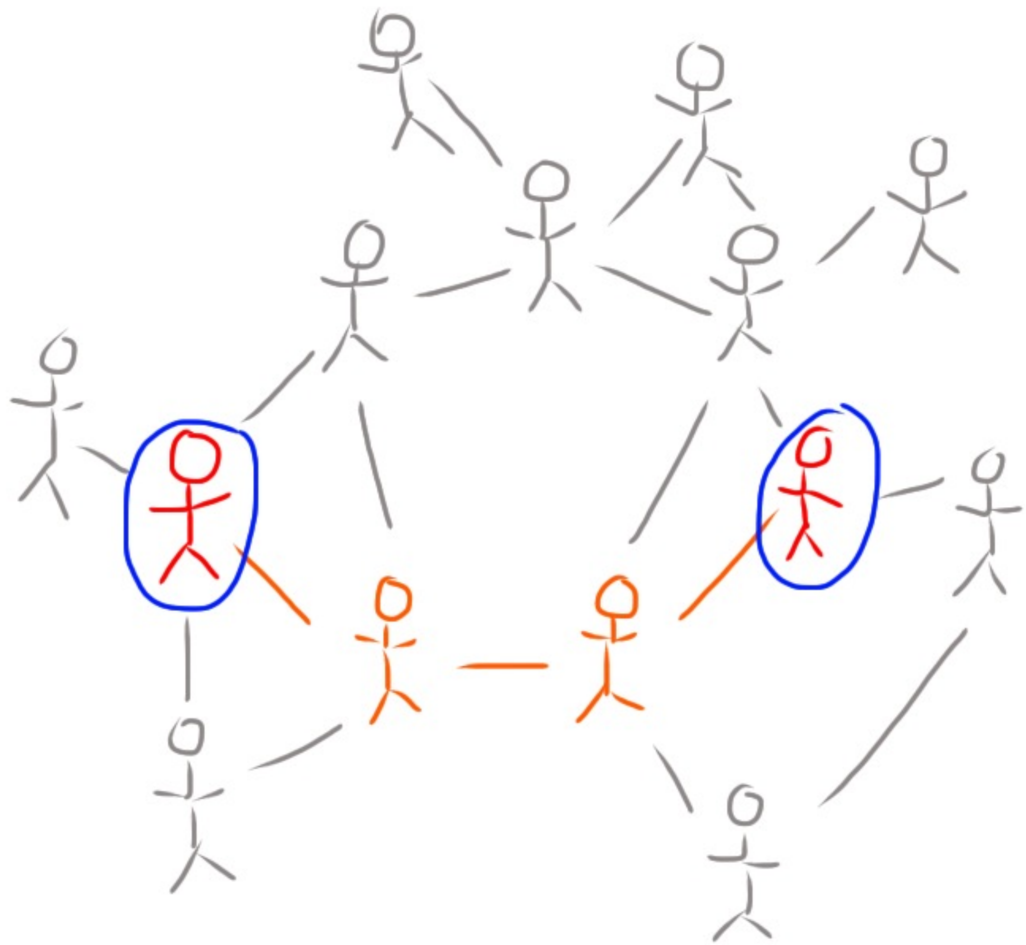
- If you have 2 large-components each occupying roughly 1/2 of the graph, how long does it typically take for the addition of random edges to join them into one giant component
  - 1-4 edge additions
  - 5-20 edge additions
  - over 20 edge additions

Average shortest path

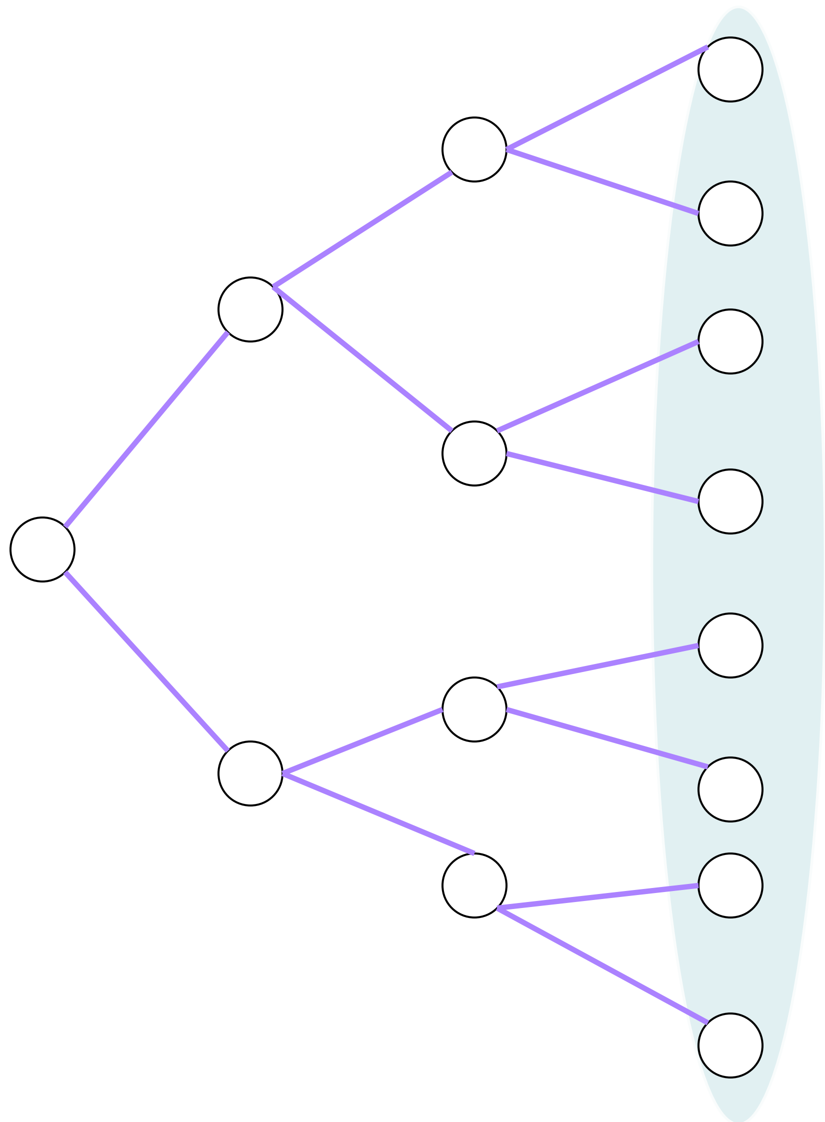
- How many hops on average between each pair of nodes?
- again, each of your friends has ***z* = *avg. degree*** friends besides you
- ignoring loops, the number of people you have at distance l is

$$z^l$$

Average shortest path



friends at distance l



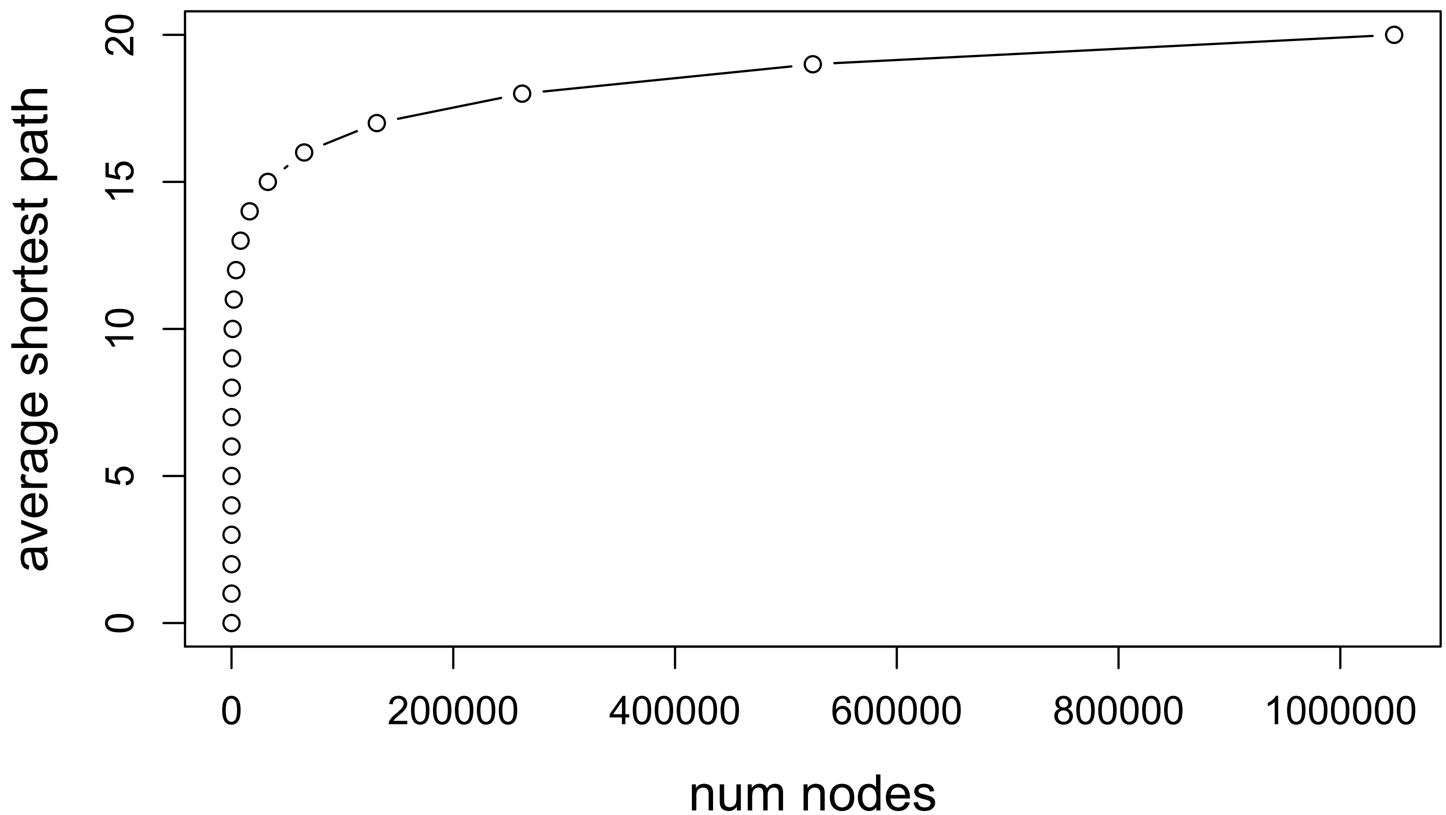
$$N_l = z^l$$

scaling:
average shortest path l_{av}

$$l_{av} \sim \frac{\log N}{\log z}$$

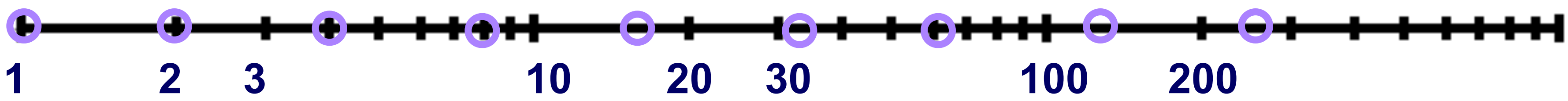
What this means in practice

- Erdös-Renyi networks can grow to be very large but nodes will be just a few hops apart



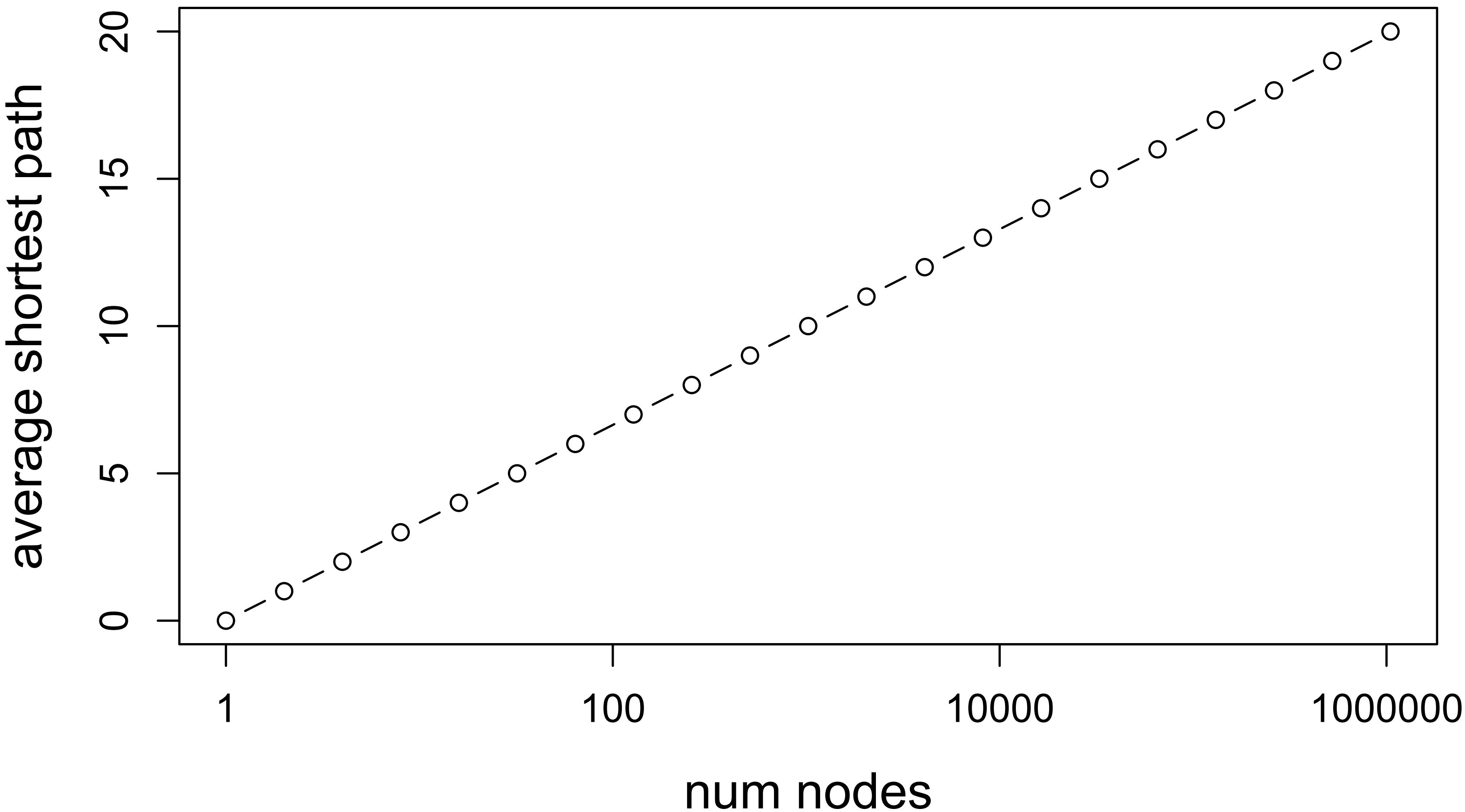
Logarithmic axes

- powers of a number will be uniformly spaced



- $2^0=1,\ 2^1=2,\ 2^2=4,\ 2^3=8,\ 2^4=16,\ 2^5=32,\ 2^6=64, \ldots.$

Erdös-Renyi avg. shortest path



Quiz Q:

- If the size of an Erdös-Renyi network increases 100 fold (e.g. from 100 to 10,000 nodes), how will the average shortest path change
  - it will be 100 times as long
  - it will be 10 times as long
  - it will be twice as long
  - it will be the same
  - it will be 1/2 as long

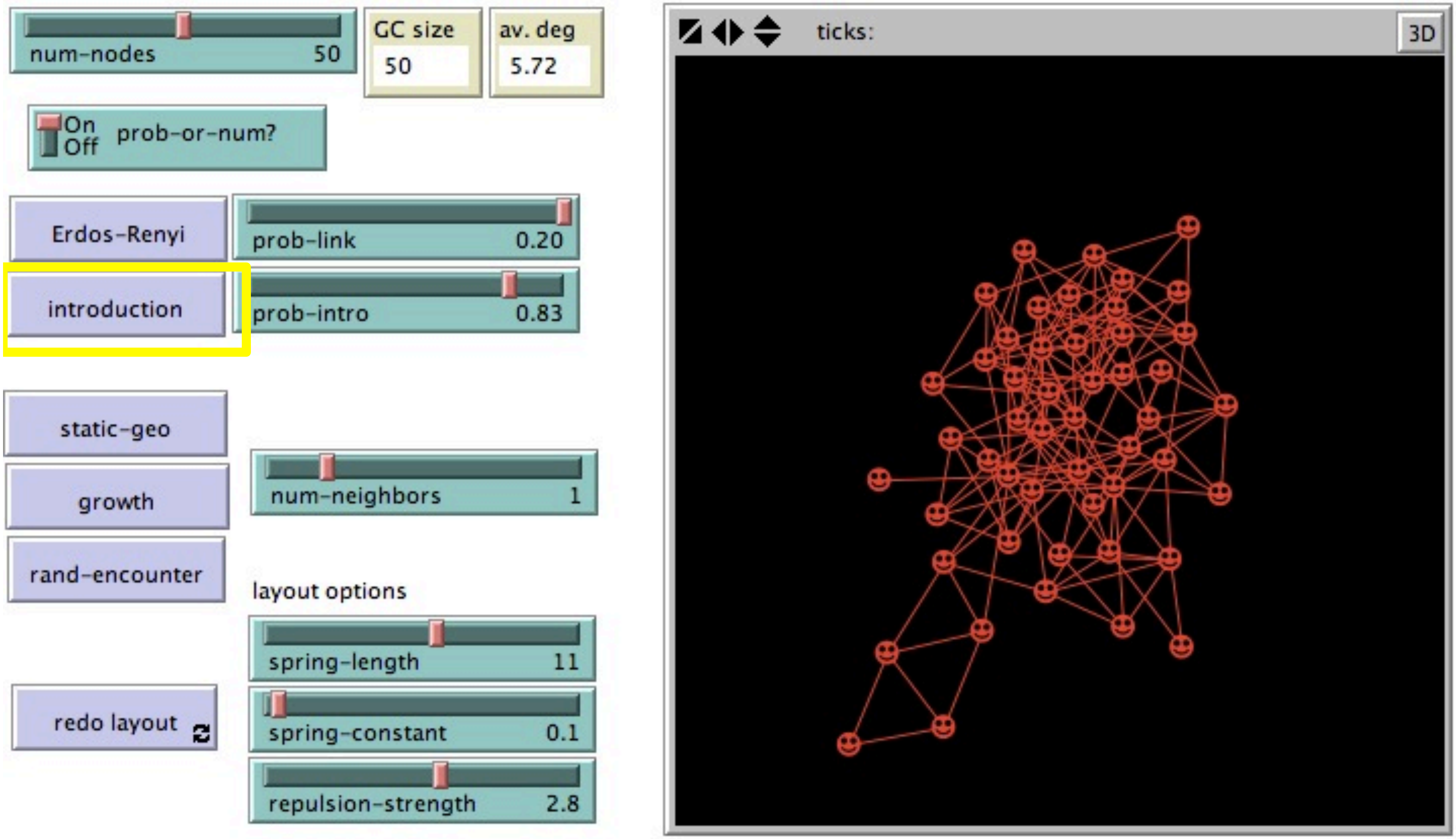
Realism

- Consider alternative mechanisms of constructing a network that are also fairly "random".
- How do they stack up against Erdös-Renyi?
- http://web.stanford.edu/class/cs224w/NetLogo/RandomGraphs.nlogo

Introduction model

- Prob-link is the p (probability of any two nodes sharing an edge) that we are used to
- But, with probability prob-intro the other node is selected among one of our friends' friends and not completely at random

Introduction model



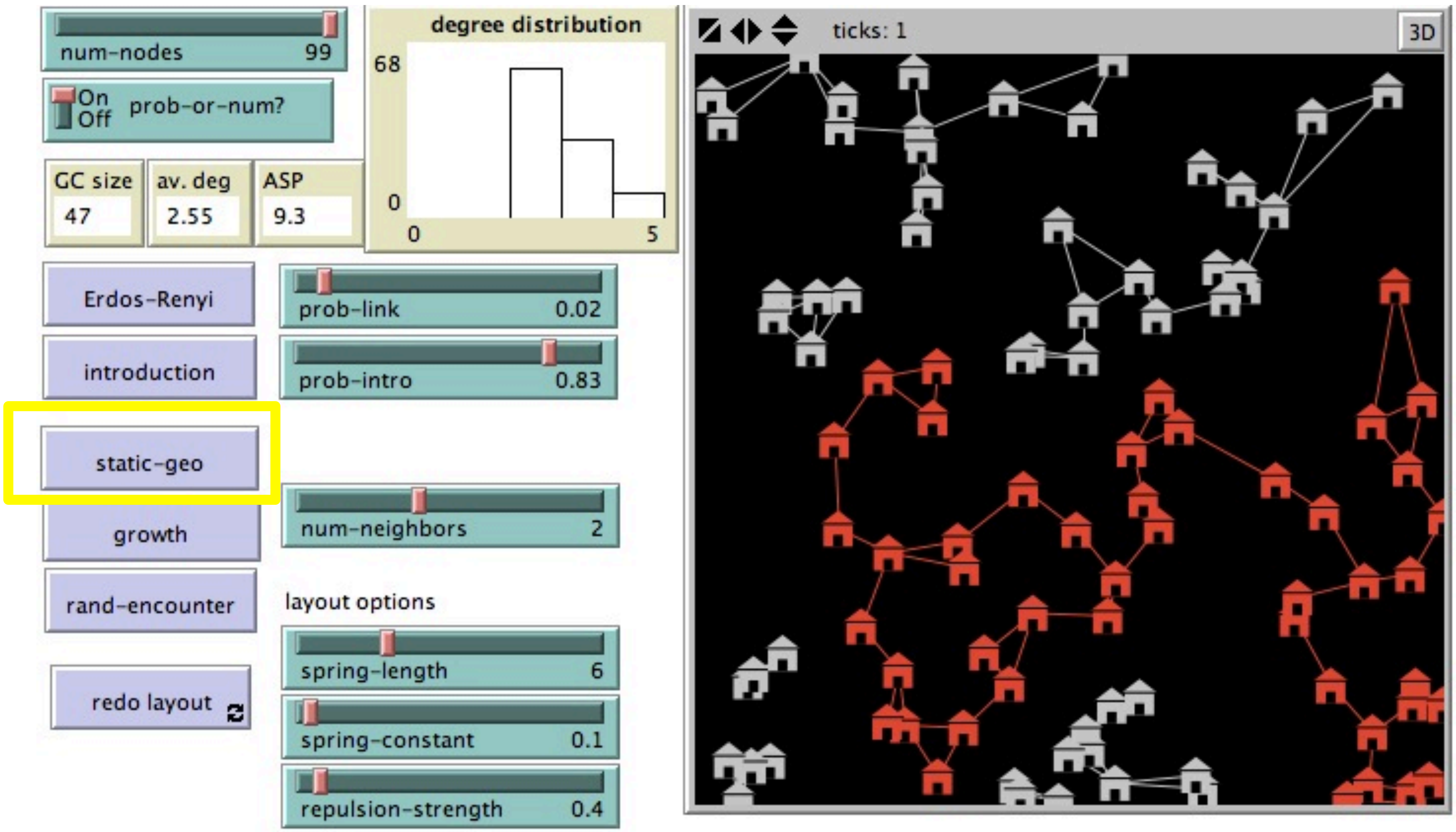
Quiz Q:

- Relative to ER, the introduction model has:
  - more edges
  - more closed triads
  - longer average shortest path
  - more uneven degree
  - smaller giant component at low p

Static Geographical model

- Each node connects to num-neighbors of its closest neighbors
- use the num-neighbors slider, and for comparison, switch PROB-OR-NUM to ‘off’ to have the ER model aim for num-neighbors as well
- turn off the layout algorithm while this is running, you can apply it at the end

static geo


num-nodes 99
On
Off
prob-or-num?
GC size
47
av. deg
2.55
ASP
9.3
degree distribution
68
0
0
5
Erdos-Renyi
introduction
static-geo
growth
rand-encounter
redo layout
prob-link 0.02
prob-intro 0.83
num-neighbors 2
layout options
spring-length 6
spring-constant 0.1
repulsion-strength 0.4
ticks: 1
3D

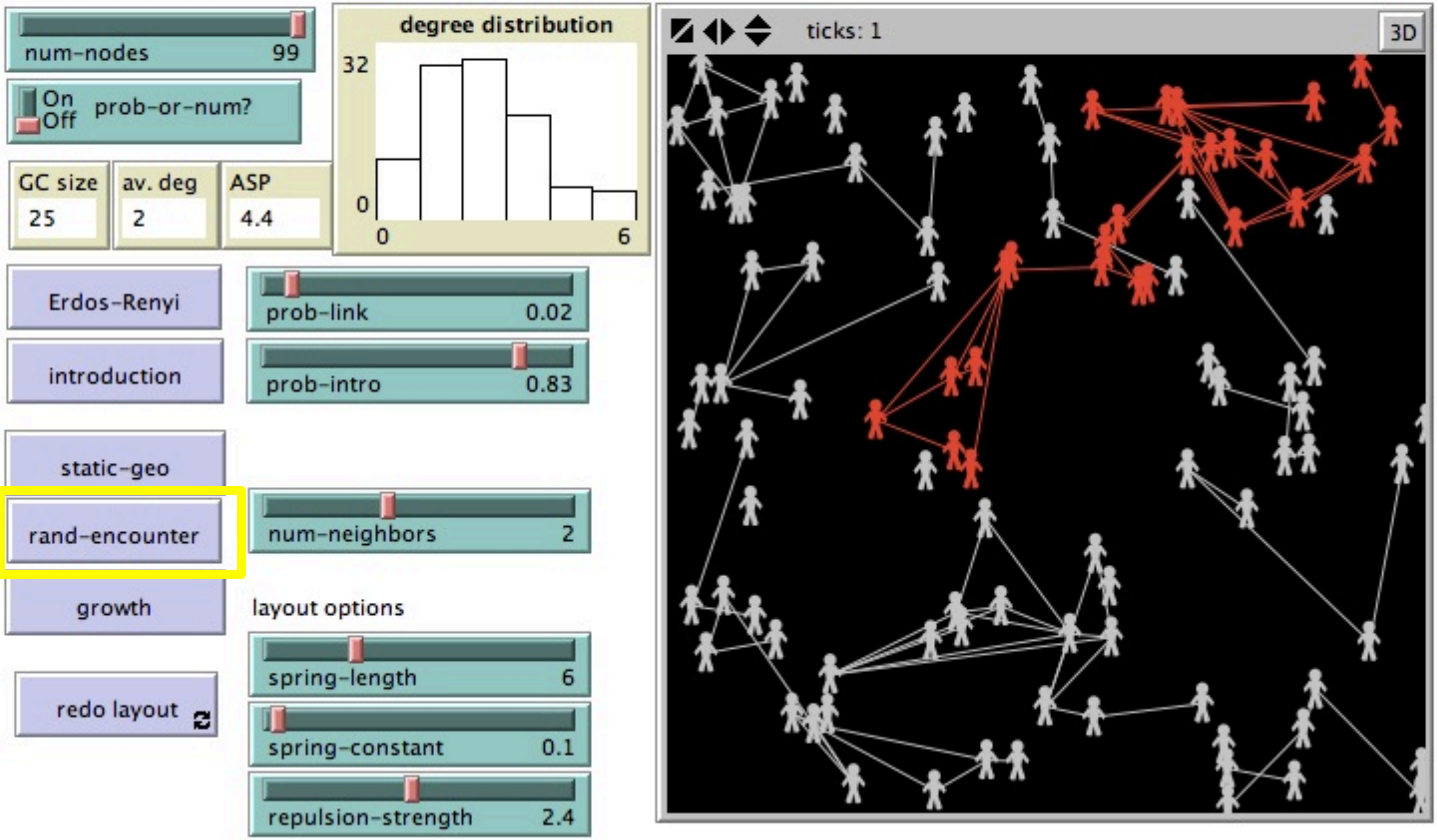
Quiz Q:

- Relative to ER, the static geographical model has :
  - longer average shortest path
  - shorter average shortest path
  - narrower degree distribution
  - broader degree distribution
  - smaller giant component at a low number of neighbors
  - larger giant component at a low number of neighbors

Random encounter

- People move around randomly and connect to people they bump into
- use the num-neighbors slider, and for comparison, switch PROB-OR-NUM to ‘off’ to have the ER model aim for num-neighbors as well
- turn off the layout algorithm while this is running (you can apply it at the end)

random encounters



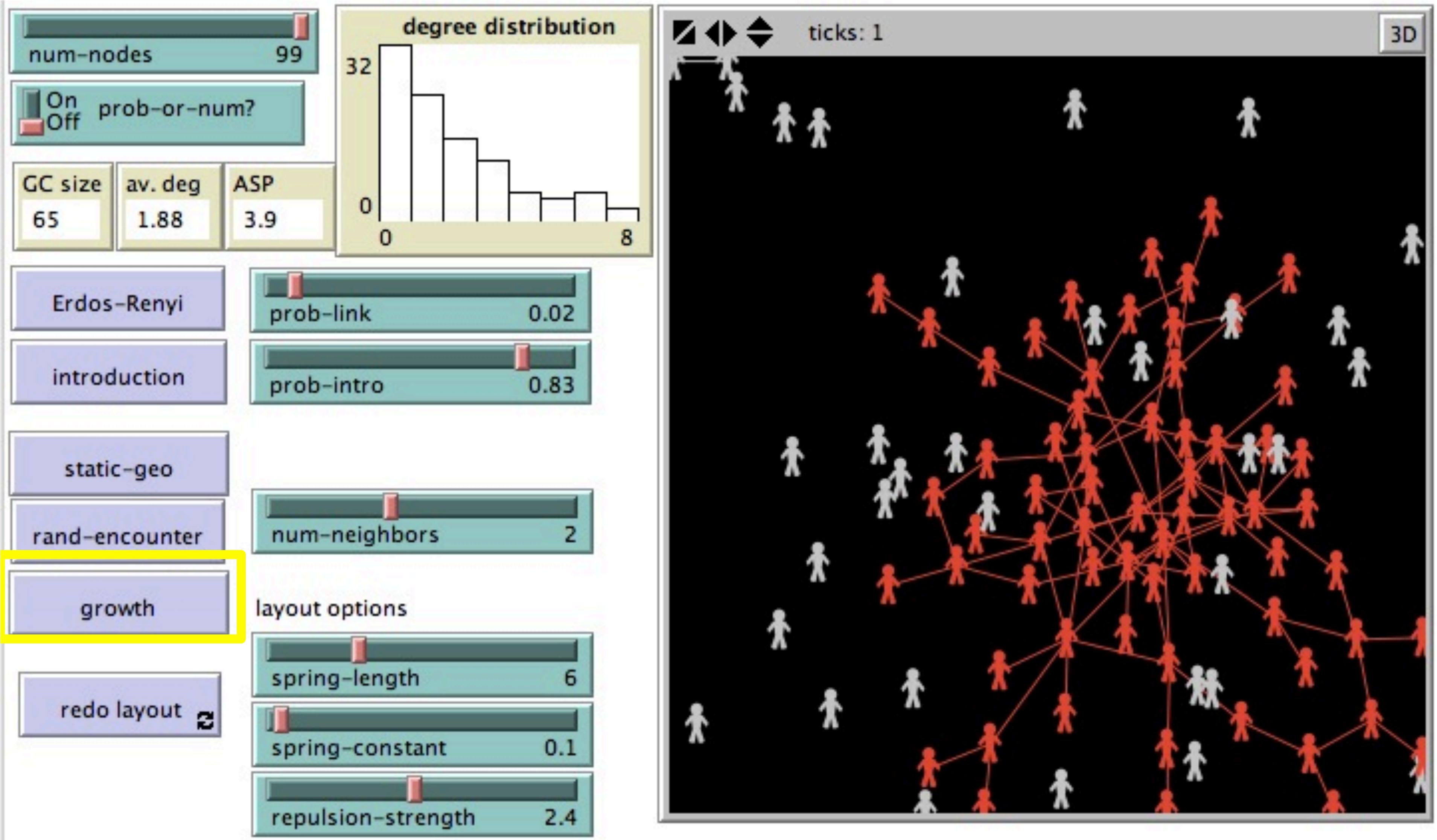
Quiz Q:

- Relative to ER, the random encounters model has :
  - more closed triads
  - fewer closed triads
  - smaller giant component at a low number of neighbors
  - larger giant component at a low number of neighbors

Growth model

- Instead of starting out with a fixed number of nodes, nodes are added over time
- use the num-neighbors slider, and for comparison, switch PROB-OR-NUM to ‘off’ to have the ER model aim for num-neighbors as well

growth model



Quiz Q:

- Relative to ER, the growth model has :
  - more hubs
  - fewer hubs
  - smaller giant component at a low number of neighbors
  - larger giant component at a low number of neighbors

other models

- in some instances the ER model is plausible
- if dynamics are different, ER model may be a poor fit