

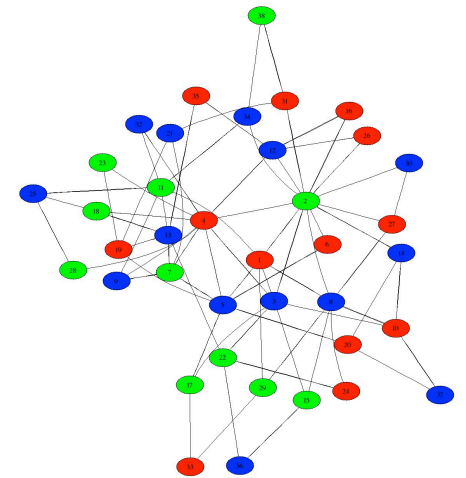
# Models of Network Formation

**Networked Life**

**NETS 112**

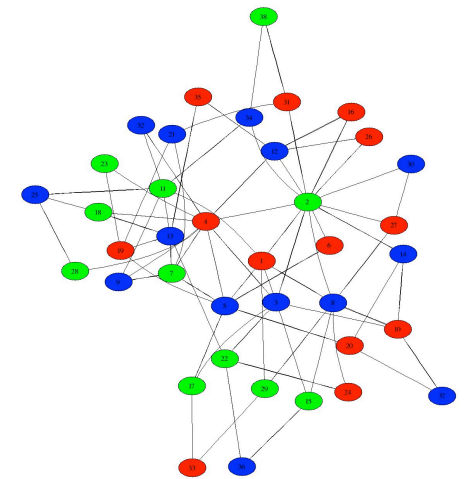
**Fall 2016**

**Prof. Michael Kearns**



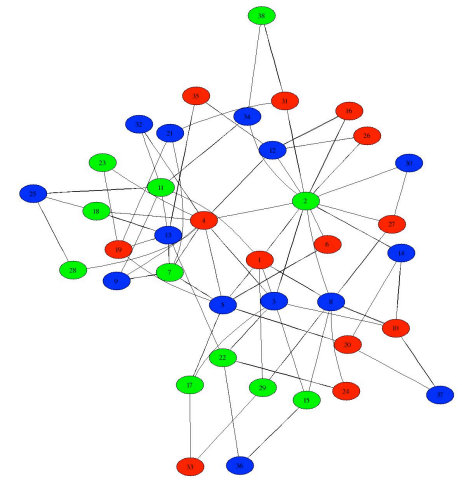
# Roadmap

- Recently: typical large-scale social and other networks exhibit:
  - giant component with small diameter
  - sparsity
  - heavy-tailed degree distributions
  - high clustering coefficient
- These are empirical phenomena
- What could “explain” them?
- One form of explanation: simple models for network formation or growth that give rise to these structural properties
- Next several lectures:
  - Erdős-Renyi (random graph) model
  - “Small Worlds” models
  - Preferential Attachment
- Discussion of structure exhibited (or not) by each



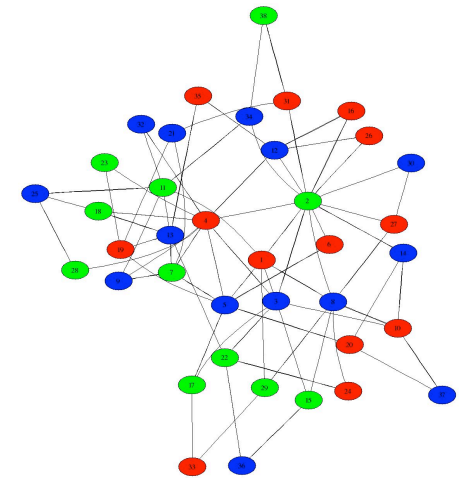
# Models of Network Formation

## I. The Erdős-Renyi (Random Graph) Model



# The Erdős-Renyi (Random Graph) Model

- Really a randomized algorithm for generating networks
- Begin with  $N$  isolated vertices, no edges
- Add edges gradually, one at a time
- Randomly select two vertices not already neighbors, add edge
- So edges are added in a random, unbiased fashion
- About the simplest (dumbest?) formation model possible
- But what can it already explain?

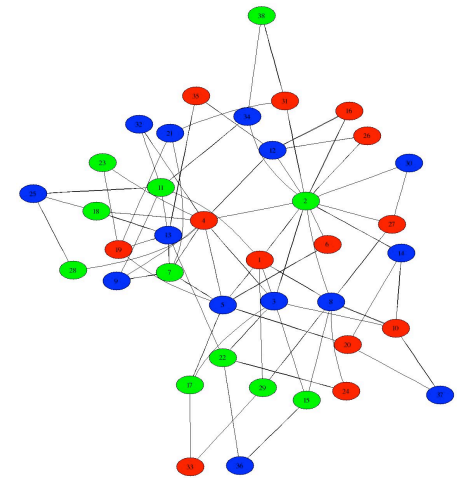


# The Erdős-Renyi (Random Graph) Model

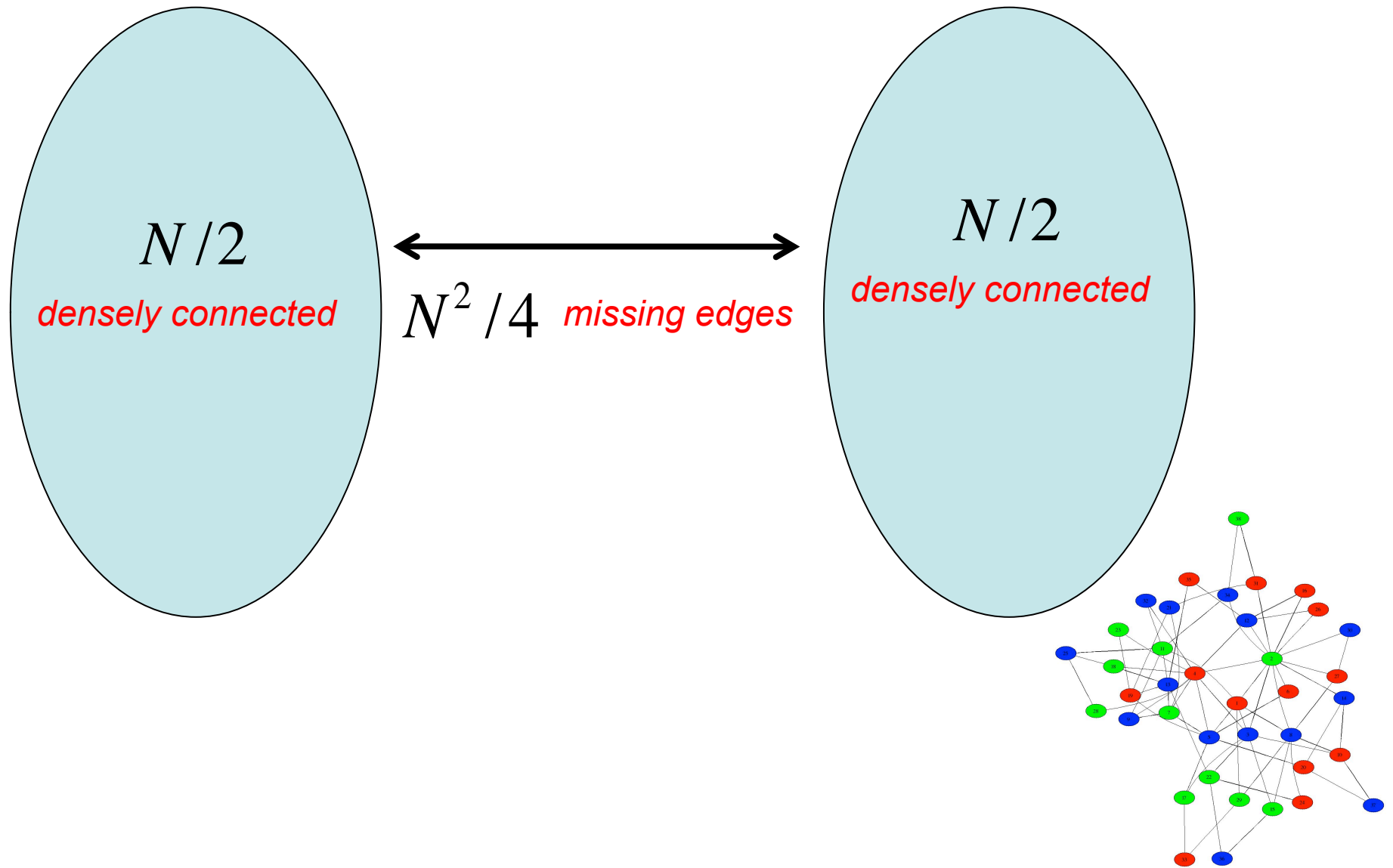
- After adding  $E$  edges, edge density is

$$p = E / (N(N - 1) / 2)$$

- As  $E$  increases,  $p$  goes from 0 to 1
- Q: What are the likely structural properties at density  $p$ ?
  - e.g. as  $p = 0 \rightarrow 1$ , small diameter occurs; single connected component
- At what values of  $p$  do “natural” structures emerge?
- We will see:
  - many natural and interesting properties arise at rather “small”  $p$
  - furthermore, they arise very suddenly (tipping/threshold)
- Let’s examine the Erdős-Renyi [simulator](#)

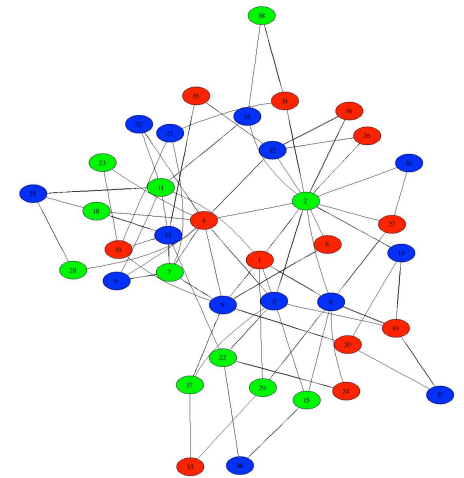


# Why Can't There Be Two Large Components?



# Threshold Phenomena in Erdős-Renyi

- Theorem: In Erdős-Renyi, as  $N$  becomes large:
  - If  $p < 1/N$ , probability of a giant component (e.g. 50% of vertices) goes to 0
  - If  $p > 1/N$ , probability of a giant component goes to 1, and all other components will have size at most  $\log(N)$
- Note: at edge density  $p$ , expected/average degree is  $p(N-1) \sim pN$
- So at  $p \sim 1/N$ , average degree is  $\sim 1$ : incredibly sparse
- So model “explains” giant components in real networks
- General “tipping point” at edge density  $q$  (may depend on  $N$ ):
  - If  $p < q$ , probability of property goes to 0 as  $N$  becomes large
  - If  $p > q$ , probability of property goes to 1 as  $N$  becomes large
- For example, could examine property “diameter 6 or less”

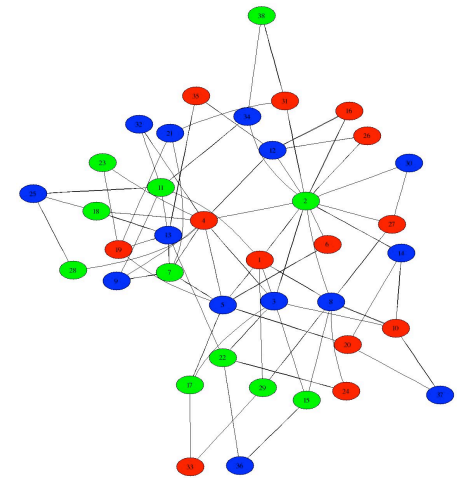


# Threshold Phenomena in Erdős-Renyi

- Theorem: In Erdős-Renyi, as  $N$  becomes large:
  - Threshold at

$$p \sim \log(N) / N^{5/6}$$

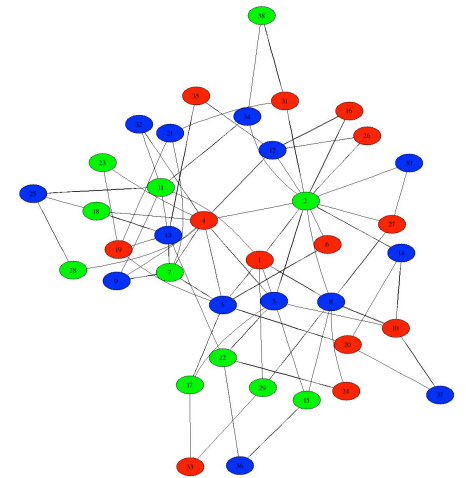
- for diameter 6.
- Note: degrees growing (slightly) with  $N$
- If  $N = 300M$  (U.S. population) then average degree  $pN \sim 500$
- If  $N = 7BN$  (world population) then average degree  $pN \sim 1000$
- Not unreasonable figures...
- At  $p$  not too far from  $1/N$ , get strong connectivity
- Very efficient use of edges





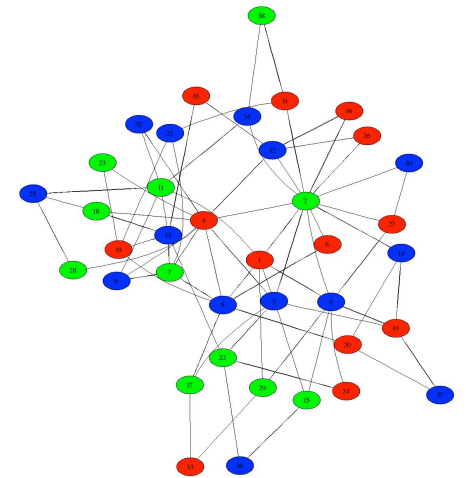
# Threshold Phenomena in Erdős-Renyi

- In fact: Any *monotone property* of networks exhibits a threshold phenomenon in Erdős-Renyi
  - monotone: property continues to hold if you add edges to the networks
  - e.g. network has a group of  $K$  vertices with at least 71% neighbors
  - e.g. network has a cycle of at least  $K$  vertices
- Tipping is the rule, not the exception



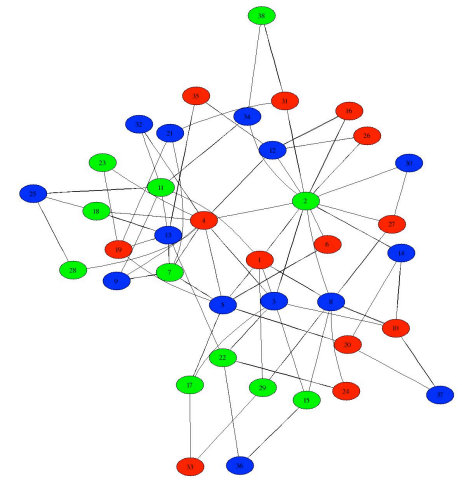
# What Doesn't the Model Explain?

- Erdős-Renyi explains giant component and small diameter
- But:
  - degree distribution not heavy-tailed; exponential decay from mean (Poisson)
  - clustering coefficient is *exactly*  $p$
- To explain these, we'll need richer models with greater realism



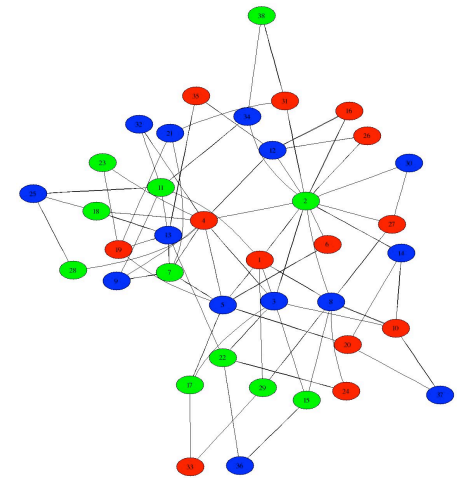
# Models of Network Formation

## II. Clustering Models



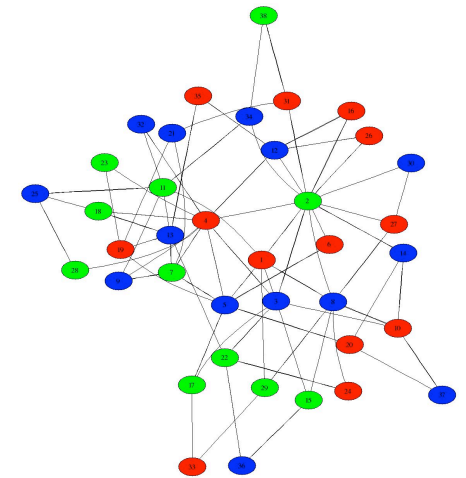
# Roadmap

- So far:
  - Erdős-Renyi exhibits small diameter, giant connected component
  - Does not exhibit high edge clustering or heavy-tailed degree distributions
- Next: network formation models yielding high clustering
  - Will also get small diameter “for free”
- Two different approaches:
  - “program” or “bake” high clustering into the model
  - balance “local” or “geographic” connectivity with long-distance edges

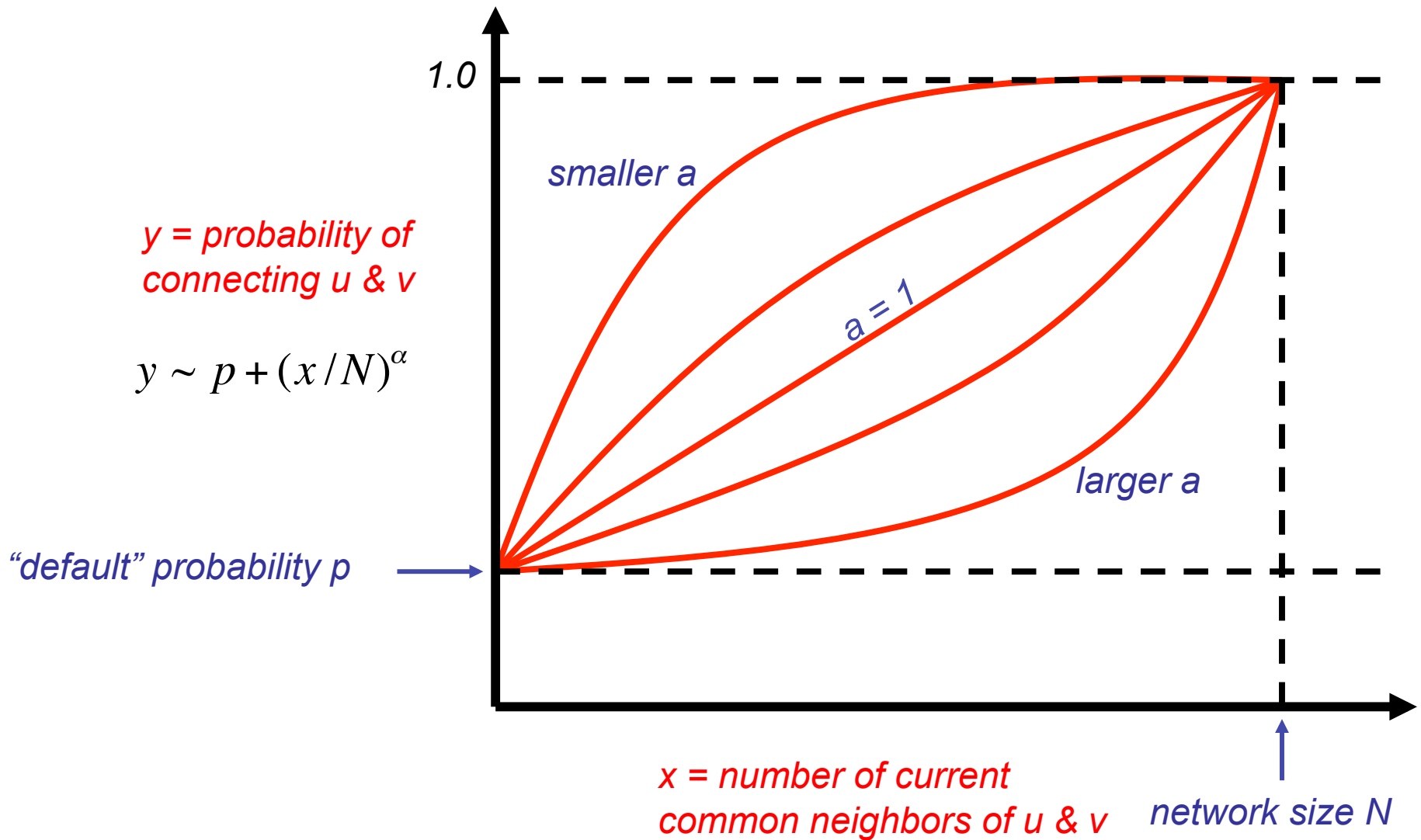


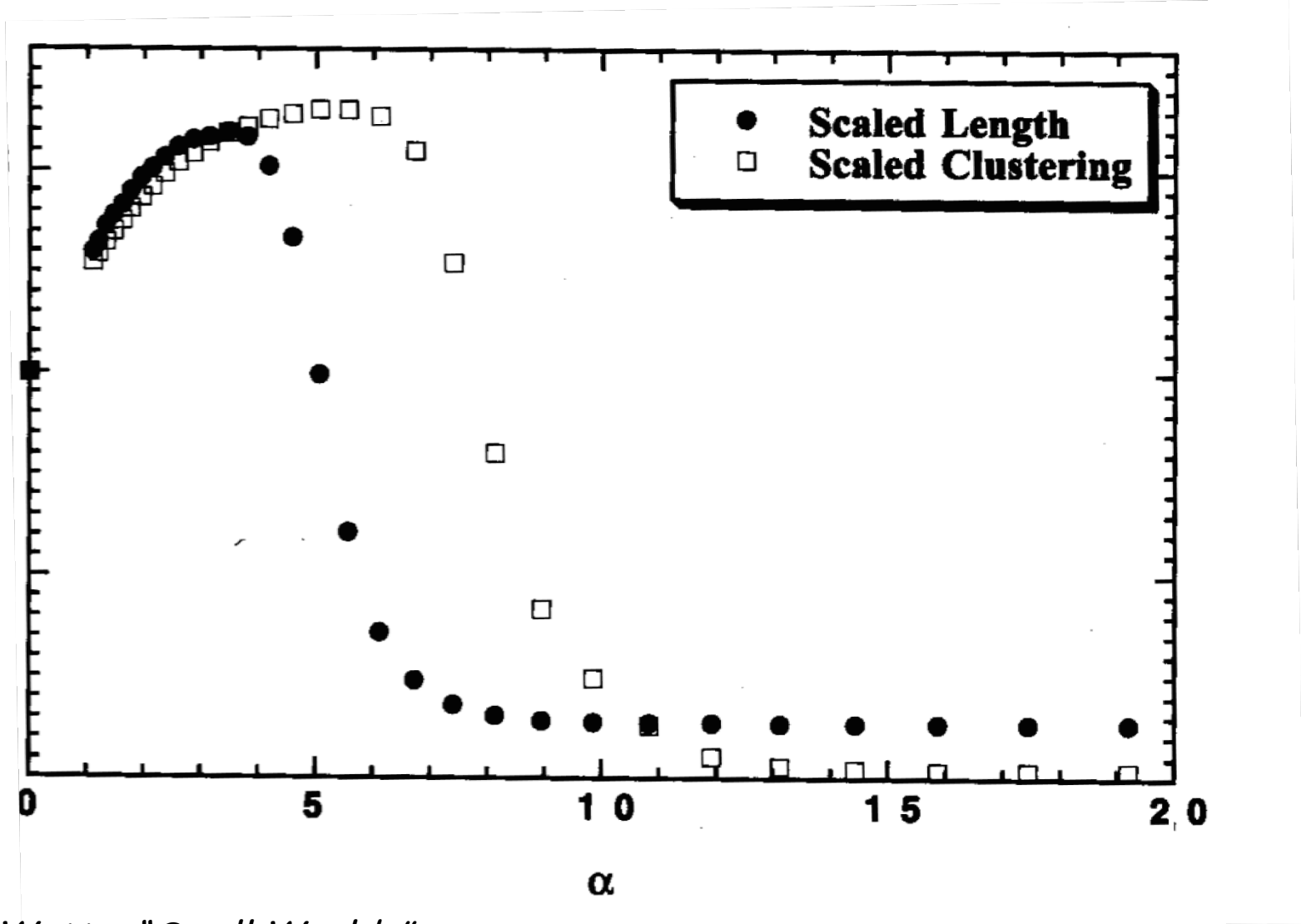
# “Programming” Clustering

- Erdős-Renyi:
  - global/background edge density  $p$
  - all edges appear independently with probability  $p$
  - no bias towards connecting friends of friends (distance 2)  $\rightarrow$  no high clustering
- But in real networks, such biases often exist:
  - people introduce their friends to each other
  - people with common friends may share interests (homophily)
- So natural to consider a model in which:
  - the more common neighbors two vertices share, the more likely they are to connect
  - still some “background” probability of connecting
  - still selecting edges randomly, but now with a bias towards friends of friends



## Making it More Precise: the a-model

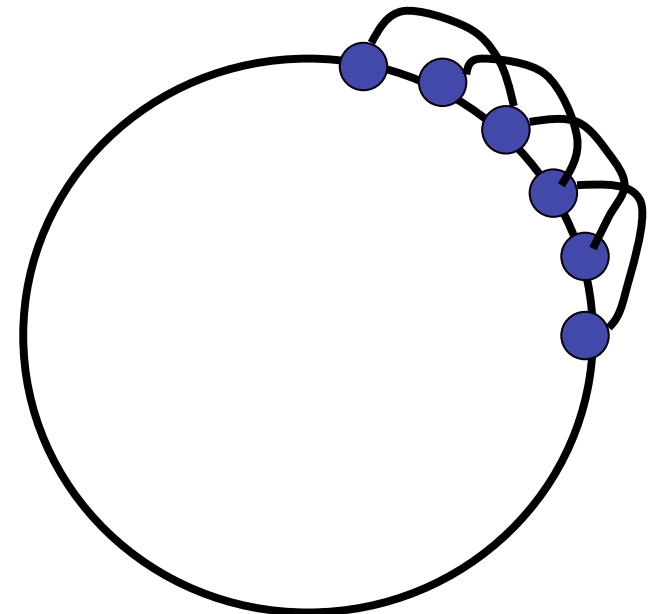




*From D. Watts, "Small Worlds"*

# Clustering Coefficient Example 2

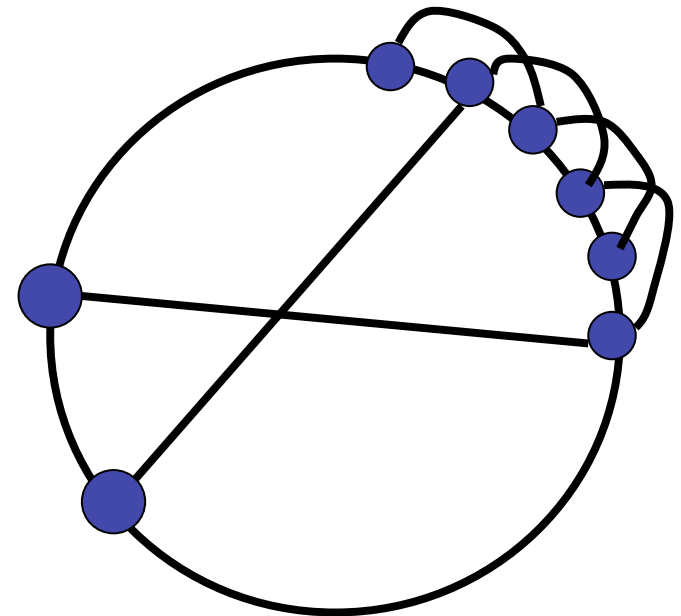
- Network: simple cycle + edges to vertices 2 hops away on cycle
- By symmetry, all vertices have the same clustering coefficient
- Clustering coefficient of a vertex  $v$ :
  - Degree of  $v$  is 4, so the number of *possible* edges between pairs of neighbors of  $v$  is  $4 \times 3/2 = 6$
  - How many pairs of  $v$ 's neighbors actually *are* connected? 3 --- the two clockwise neighbors, the two counterclockwise, and the immediate cycle neighbors
  - So the c.c. of  $v$  is  $3/6 = 1/2$
- Compare to overall edge density:
  - Total number of edges =  $2N$
  - Edge density  $p = 2N/(N(N-1)/2) \sim 4/N$
  - As  $N$  becomes large,  $1/2 \gg 4/N$
  - So this cyclical network is highly clustered





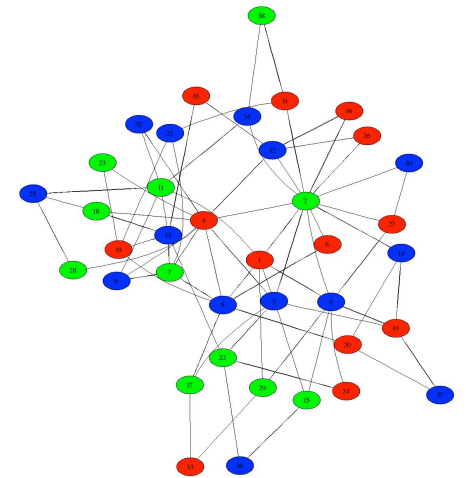
# An Alternative Model

- A different model:
  - start with all vertices arranged on a ring or cycle (or a grid)
  - connect each vertex to all others that are within  $k$  cycle steps
  - with probability  $q$ , **rewire** each local connection to a **random** vertex
- Initial cyclical structure models “local” or “geographic” connectivity
- Long-distance rewiring models “long-distance” connectivity
- $q=0$ : high clustering, high diameter
- $q=1$ : low clustering, low diameter ( $\sim$  Erdős-Renyi)
- Again is a “magic range” of  $q$  where we get both high clustering and low diameter
- Let’s look at this [demo](#)



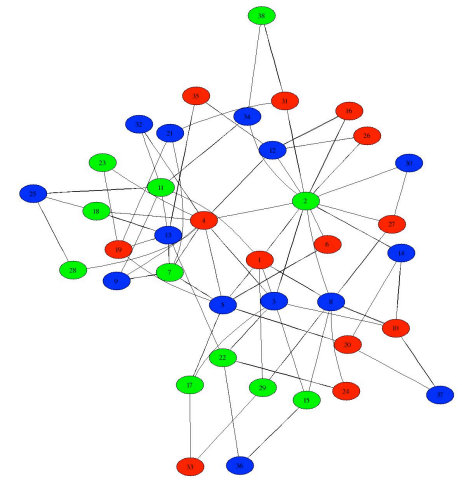
# Summary

- Two rather different ways of getting high clustering, low diameter:
  - bias connectivity towards shared friendships
  - mix local and long-distance connectivity
- Both models require proper “tuning” to achieve simultaneously
- Both a bit more realistic than Erdős-Renyi
- Neither model exhibits heavy-tailed degree distributions



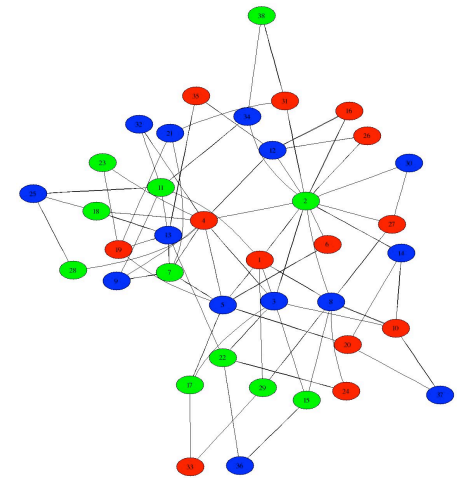
# Models of Network Formation

## III. Preferential Attachment



# Rich-Get-Richer Processes

- Processes in which the more someone has of something, the more likely they are to get more of it
- Examples:
  - the more friends you have, the easier it is to make more
  - the more business a firm has, the easier it is to win more
  - the more people there are at a nightclub, the more who want to go
- Such processes will amplify inequality
- One simple and general model: if you have amount  $x$  of something, the probability you get more is proportional to  $x$ 
  - so if you have twice as much as me, you're twice as likely to get more
- Generally leads to heavy-tailed distributions (power laws)
- Let's look at a simple “nightclub” demo...



# Preferential Attachment

- Start with two vertices connected by an edge
- At each step, add one *new* vertex  $v$  with one edge back to *previous* vertices
- Probability a previously added vertex  $u$  receives the new edge from  $v$  is *proportional to the (current) degree of  $u$* 
  - more precisely, probability  $u$  gets the edge = (current degree of  $u$ )/(sum of all current degrees)
- Vertices with high degree are likely to get *even more* links!
  - ...just like the crowded nightclub
- *Generates a power law distribution of degrees*
- Variation: each new vertex initially gets  $k$  edges
- Here's another [demo](#)

# Summary

- Now have provided network formation models exhibiting each of the universal structure arising in real-world networks
- Often got more than one property at a time:
  - Erdős-Renyi: giant component, small diameter
  - $\alpha$  model, local+long-distance: high clustering, small diameter
  - Preferential Attachment: heavy-tailed degree distribution, small diameter
- Can we achieve all of them simultaneously?
- Probably: mix together aspects of all the models
- Won't be as simple and appealing, though

