# INFO I368
# Introduction to Network Science

Filippo Menczer

CNetS Center for Complex Networks and Systems Research

INDIANA UNIVERSITY BLOOMINGTON

SCHOOL OF Informatics and Computing

- Attendance
- HW Questions?
- Readings review / quiz (ch. 2, 3)
  - Degree definitions
  - Path definitions
  - Erdos number
  - Milgram's experiment

# The plan

* In past weeks we introduced the basic concepts and quantities about networks
* You should now be familiar with basic representations and operations on networks using NetworkX
* Next we will go into more depth about some interesting classes of networks and their characteristics, starting with random and social networks

# What's a model?

Reality → Abstraction / conceptualization

Aims

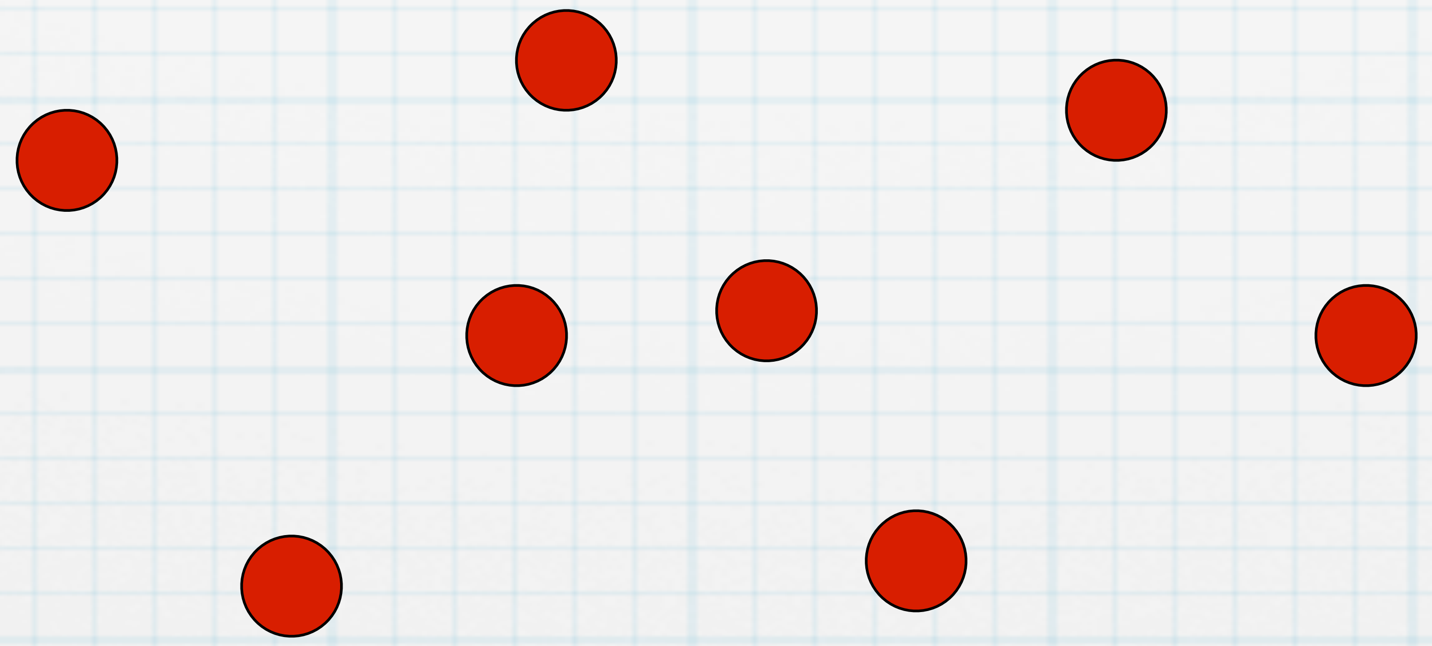
Ingredients

Assumptions

Limitations

Validation

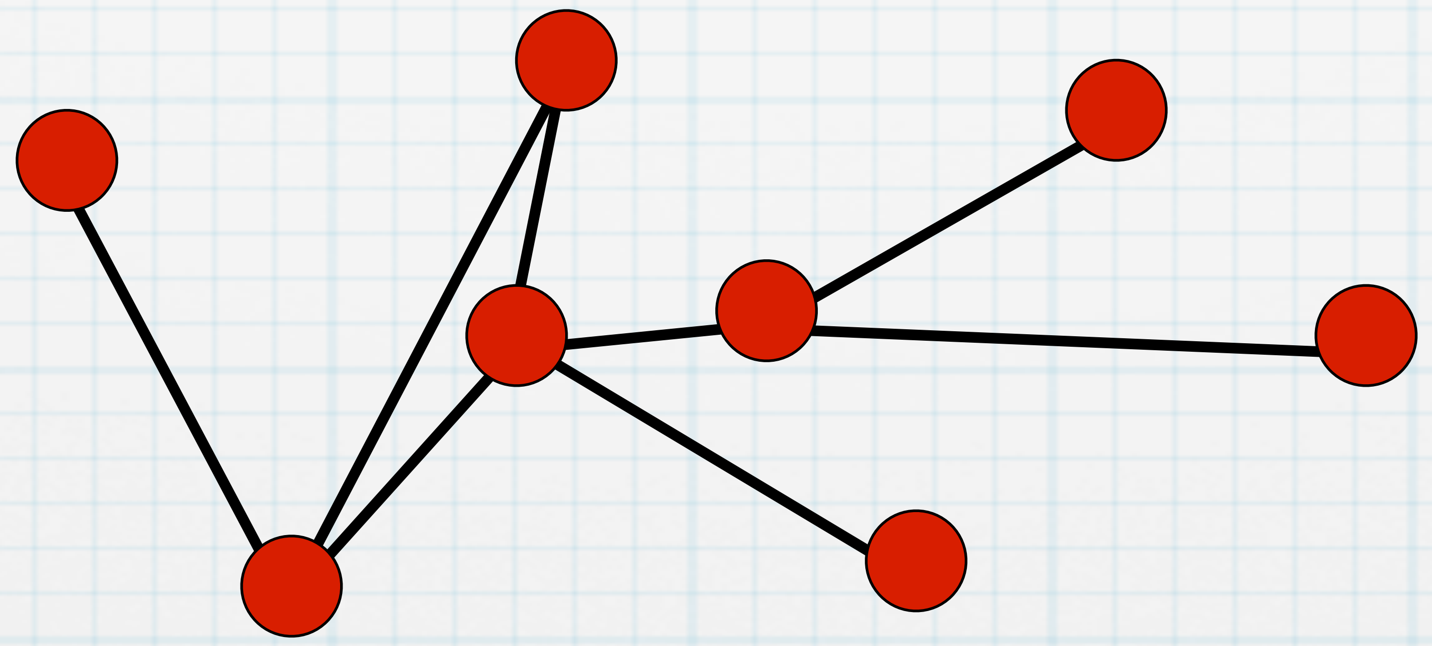
# The simplest graph model: Erdös-Rényi model of random networks





INDIANA UNIVERSITY BLOOMINGTON

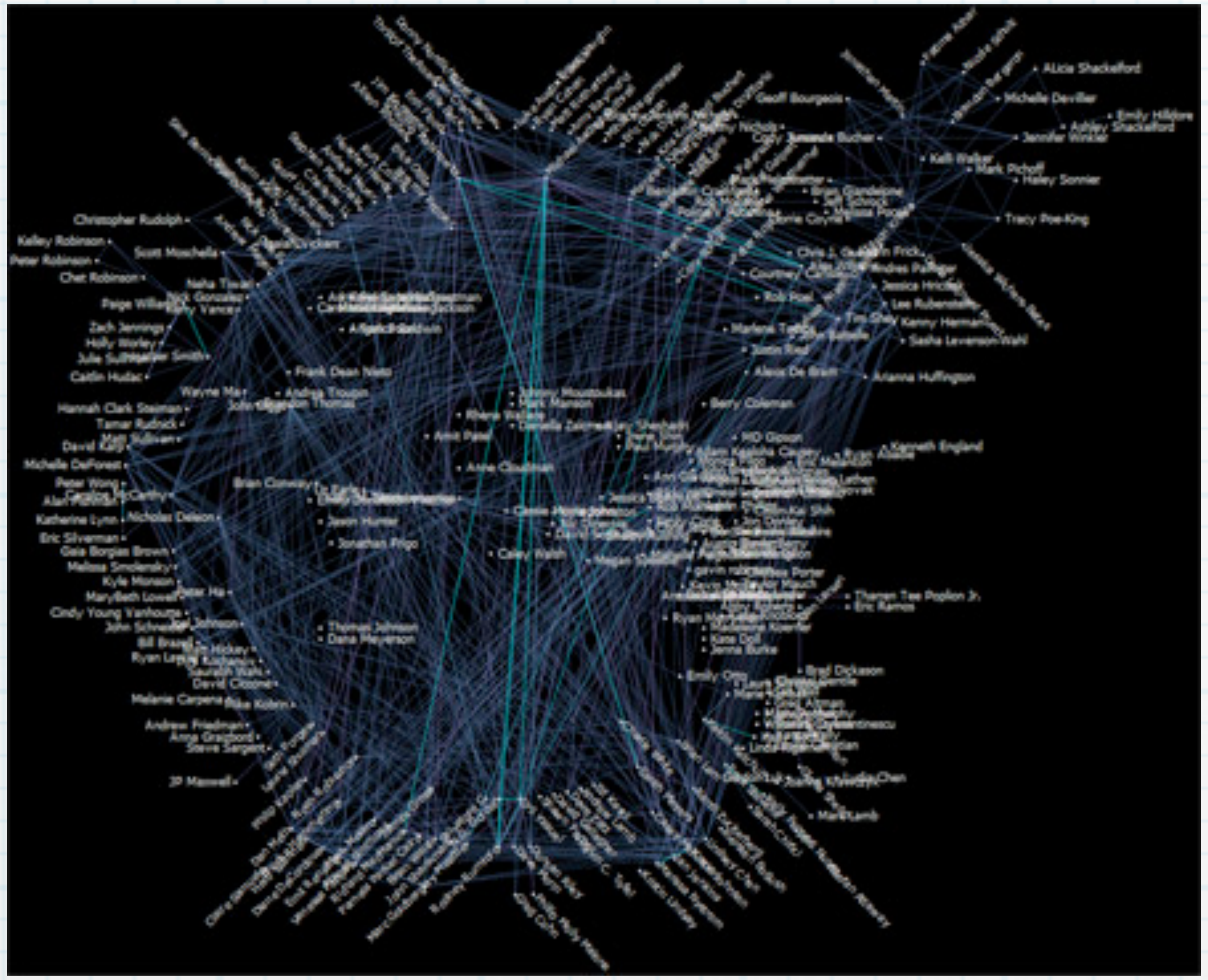
# The simplest graph model: Erdös-Rényi model of random networks





INDIANA UNIVERSITY
BLOOMINGTON

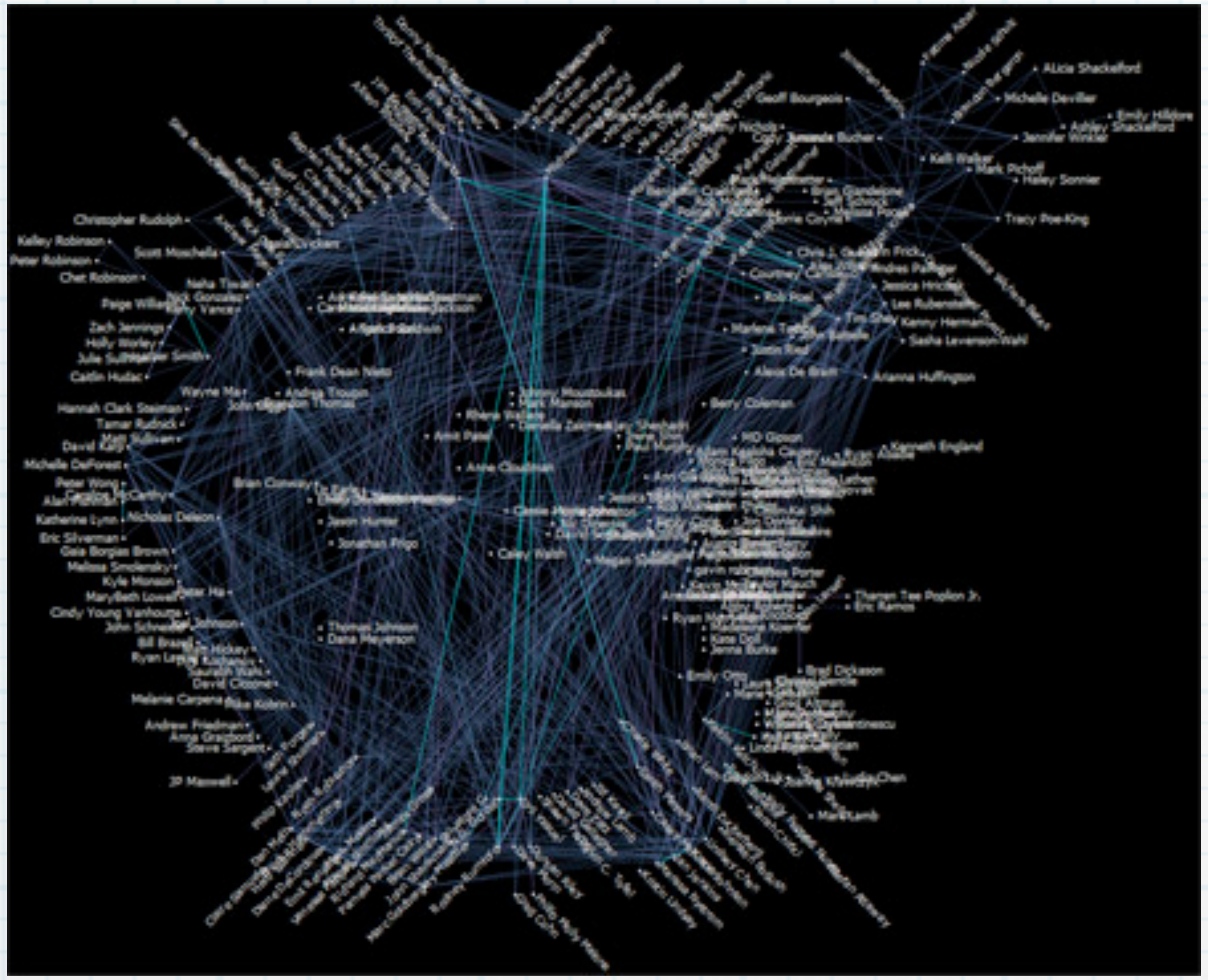
# E-R model

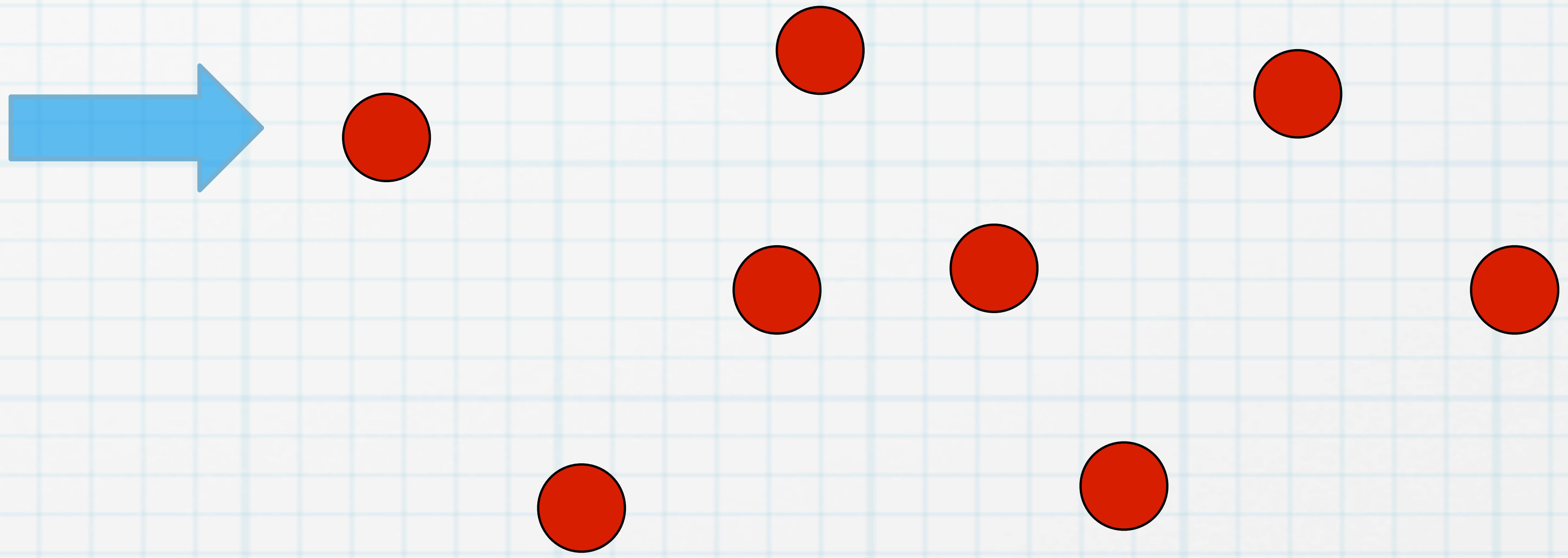


Model ??

N nodes to connect

# E-R model



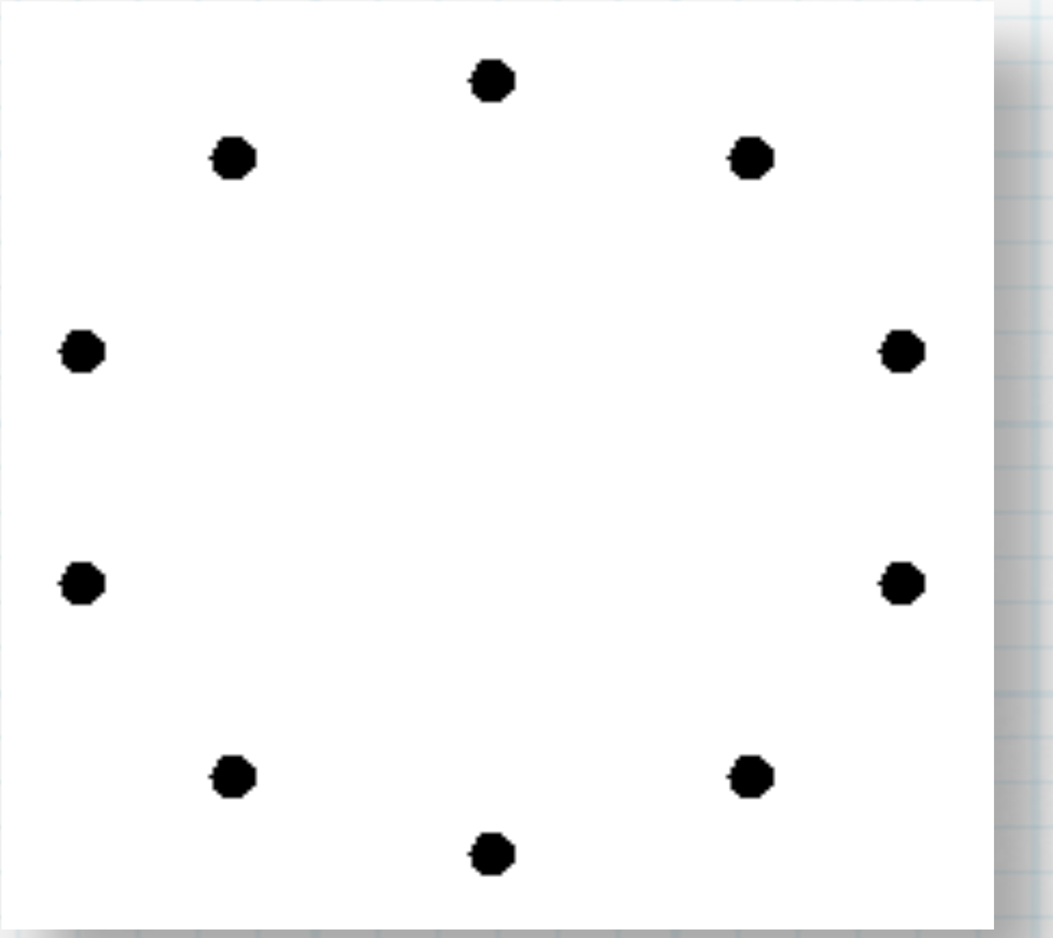


Model ??

N nodes to connect

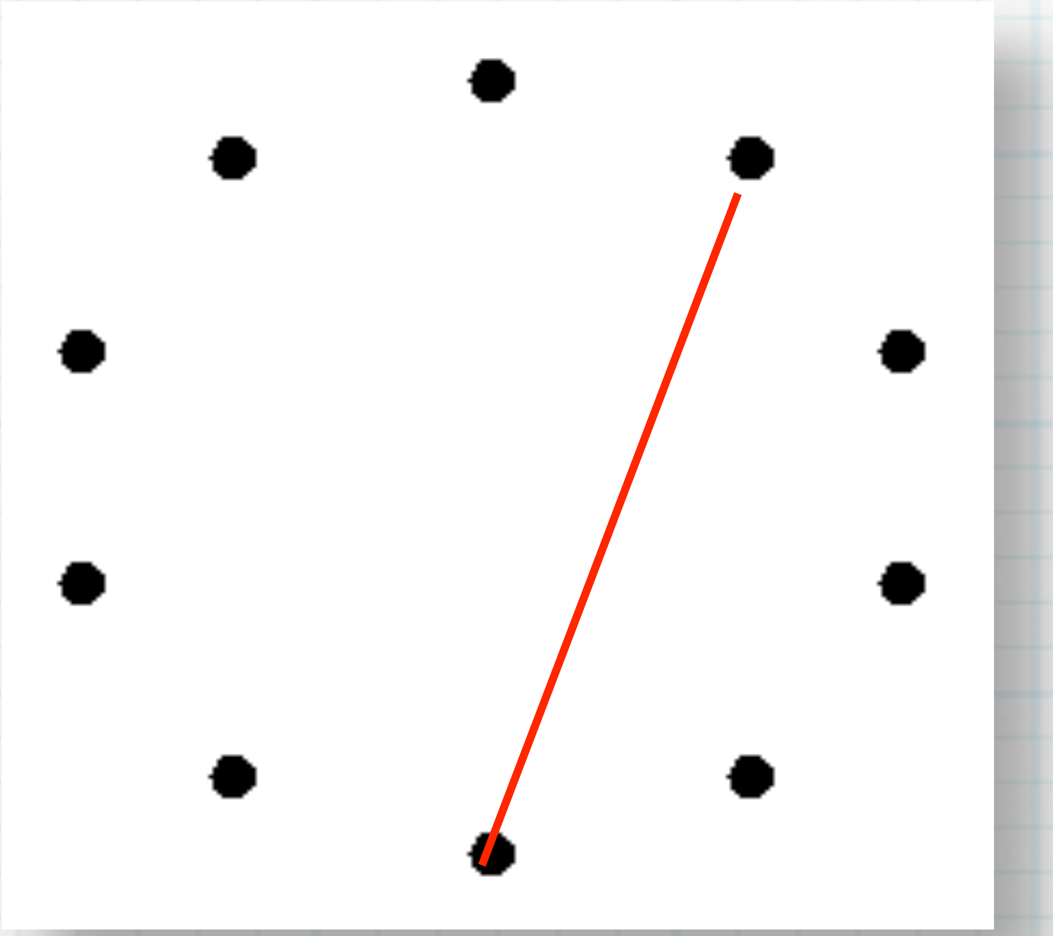
# Erdös-Rényi model: Random Graph

With probability p an edge is established between each couple of vertices



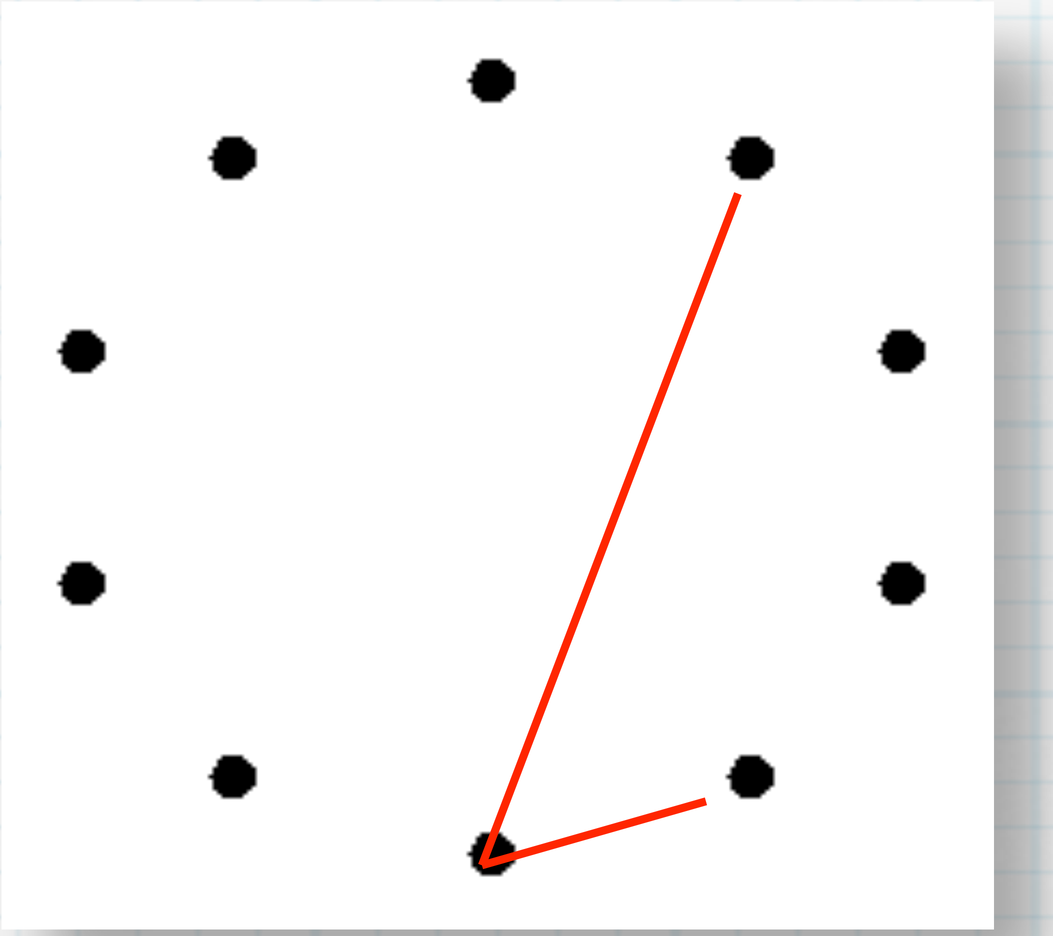
# Erdös-Rényi model: Random Graph

With probability p an edge is established between each couple of vertices



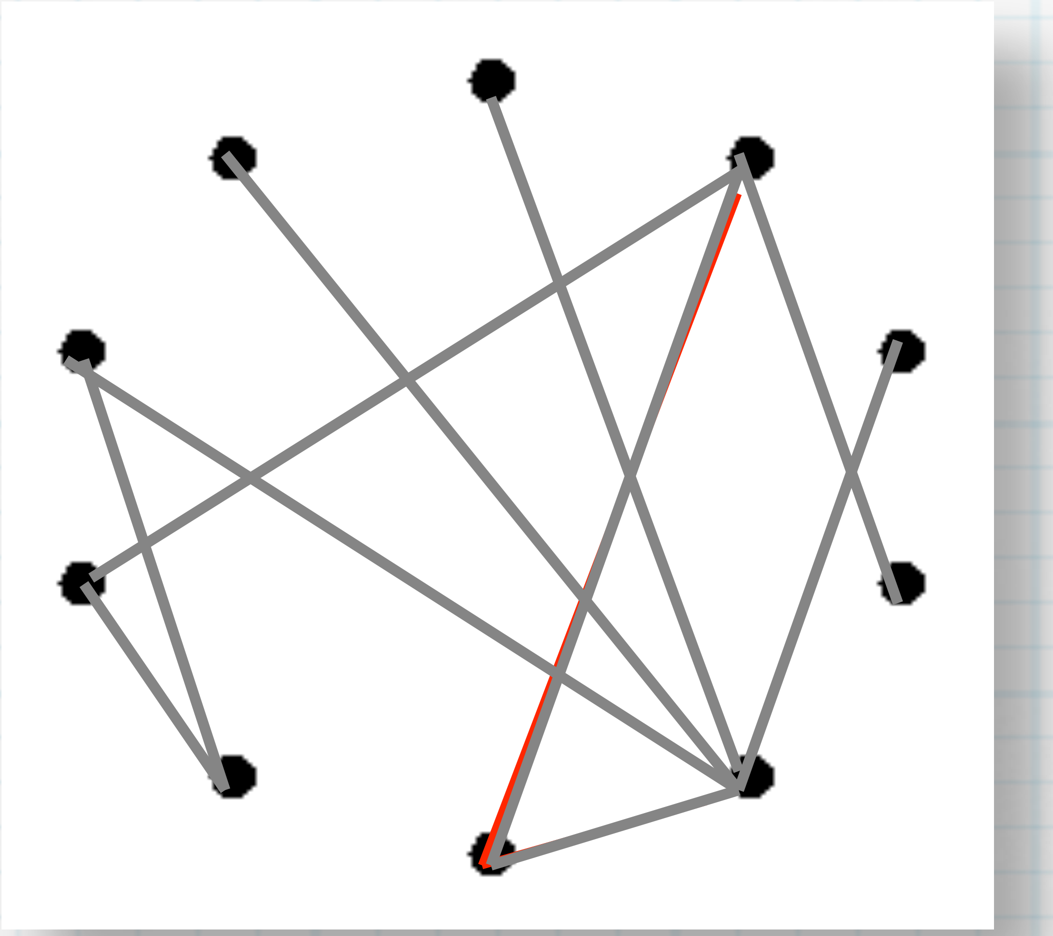
# Erdös-Rényi model: Random Graph

With probability p an edge is established between each couple of vertices



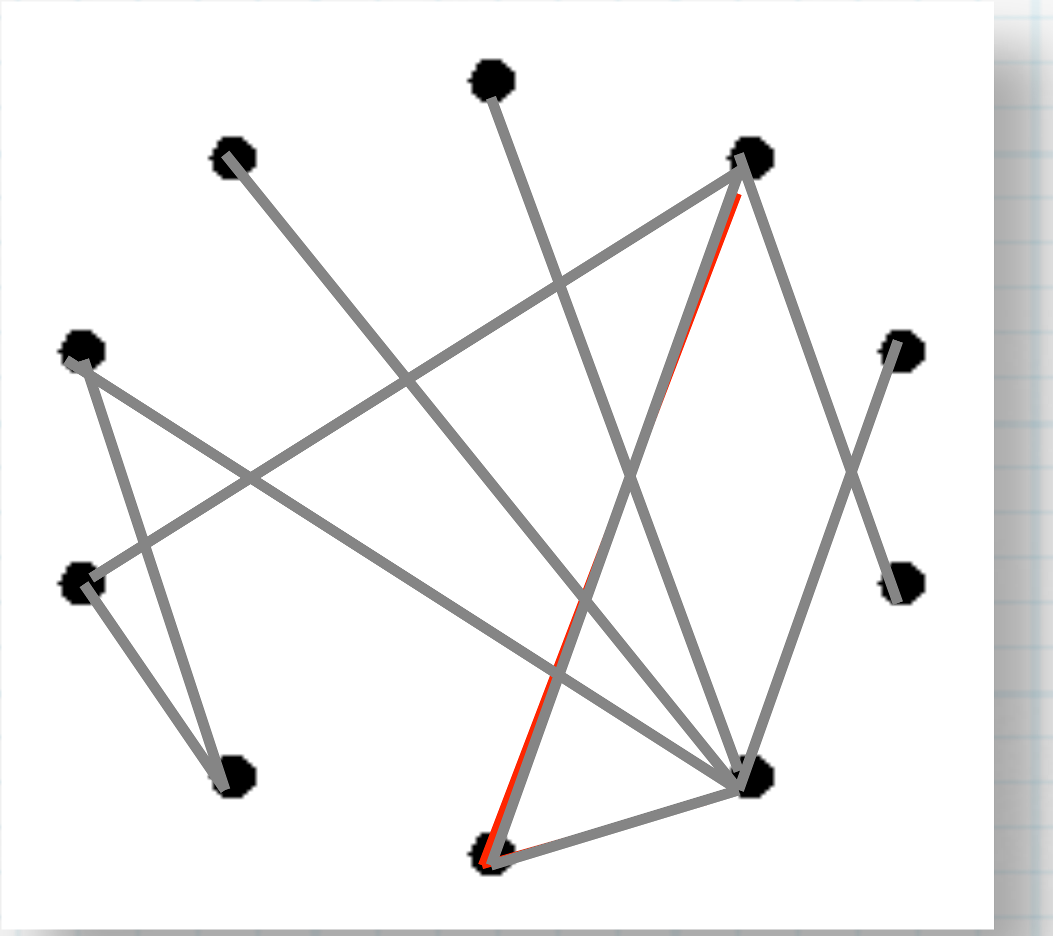
# Erdös-Rényi model: Random Graph

With probability p an edge is established between each couple of vertices



# Erdös-Rényi model: Random Graph

With probability p an edge is established between each couple of vertices

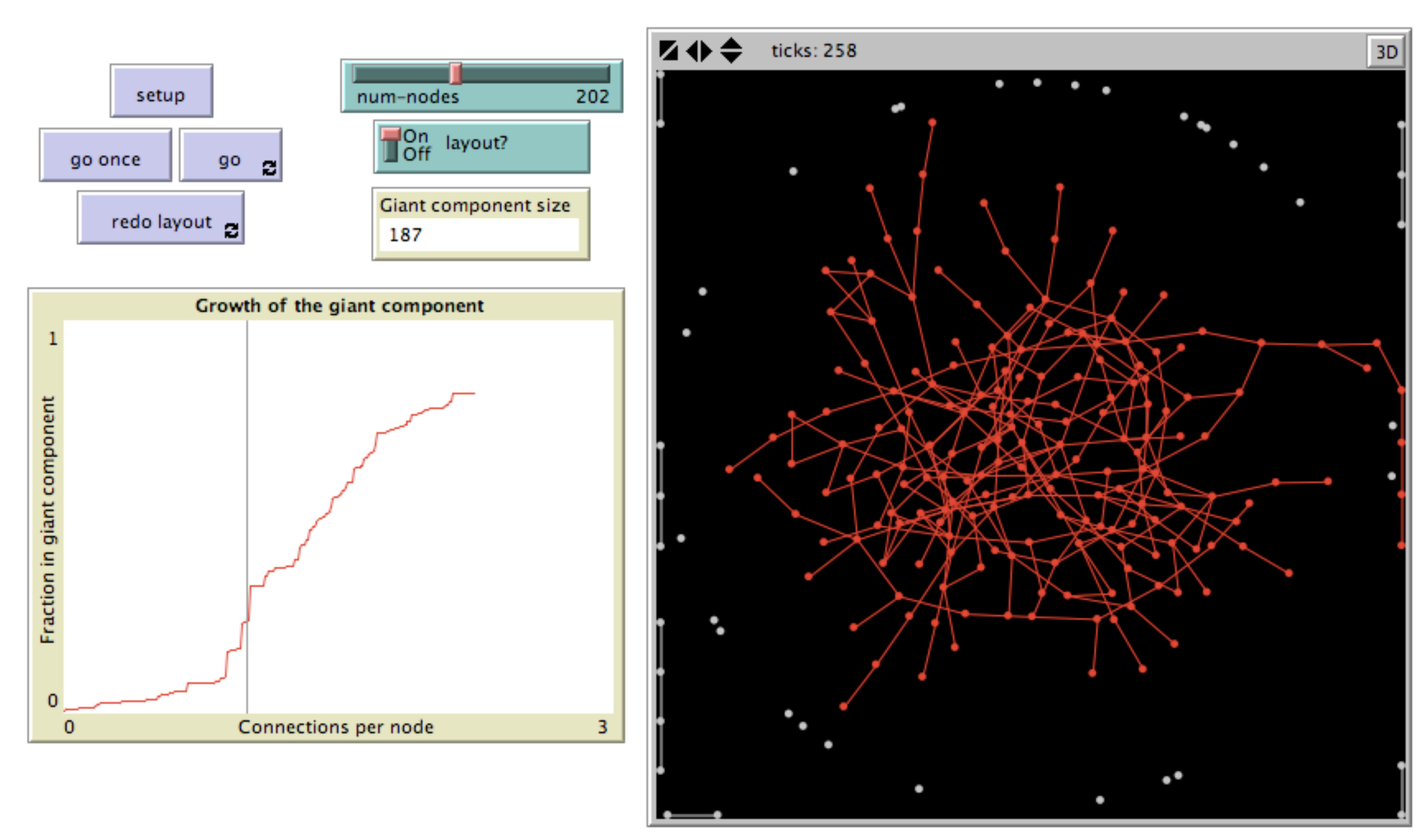


$\langle k \rangle = 2E/N = p\,(N-1)$

If you know N and the desired $\langle k \rangle$:

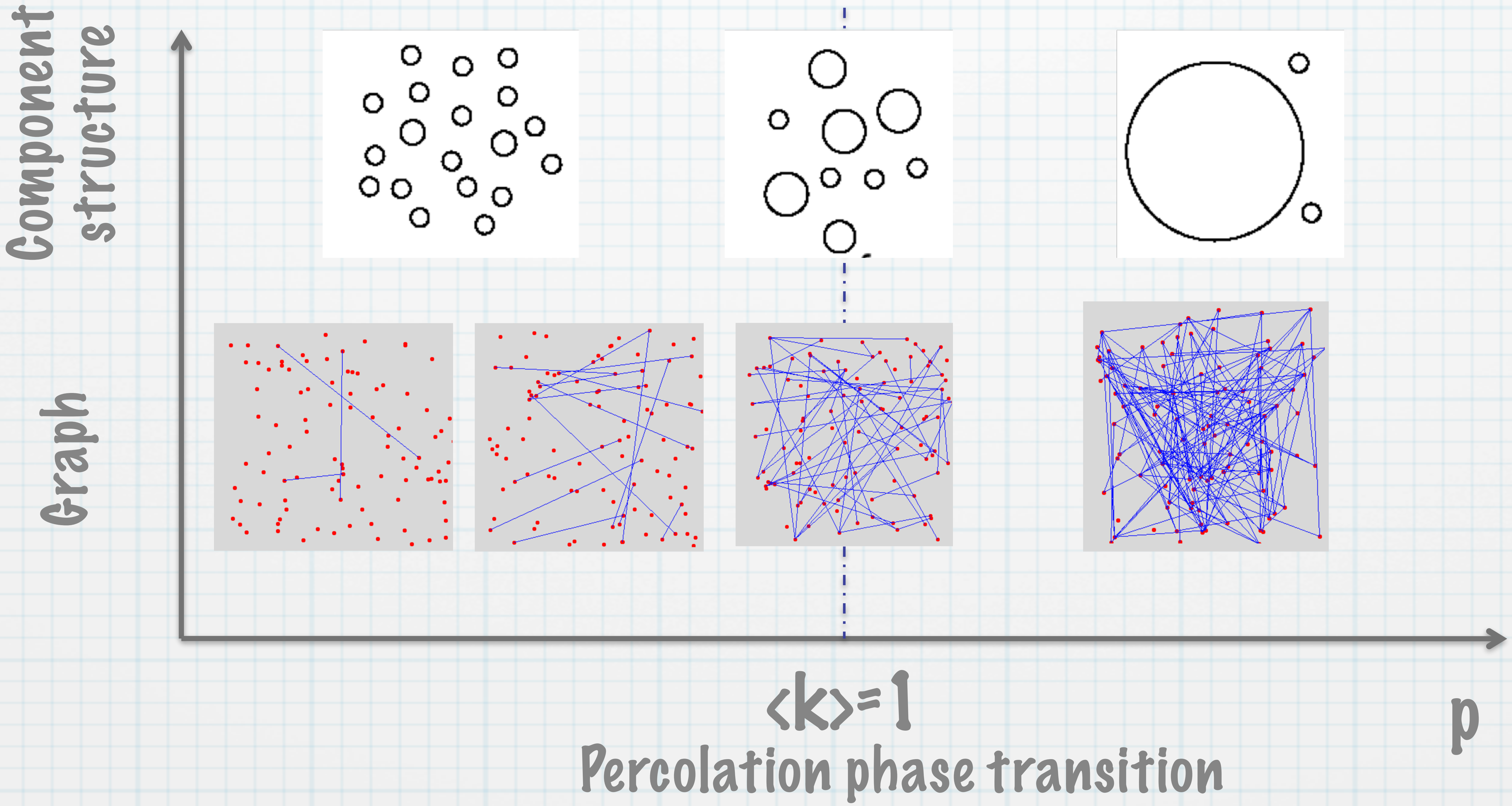
$p = \langle k \rangle/(N-1)$

# NetLogo: Giant Component

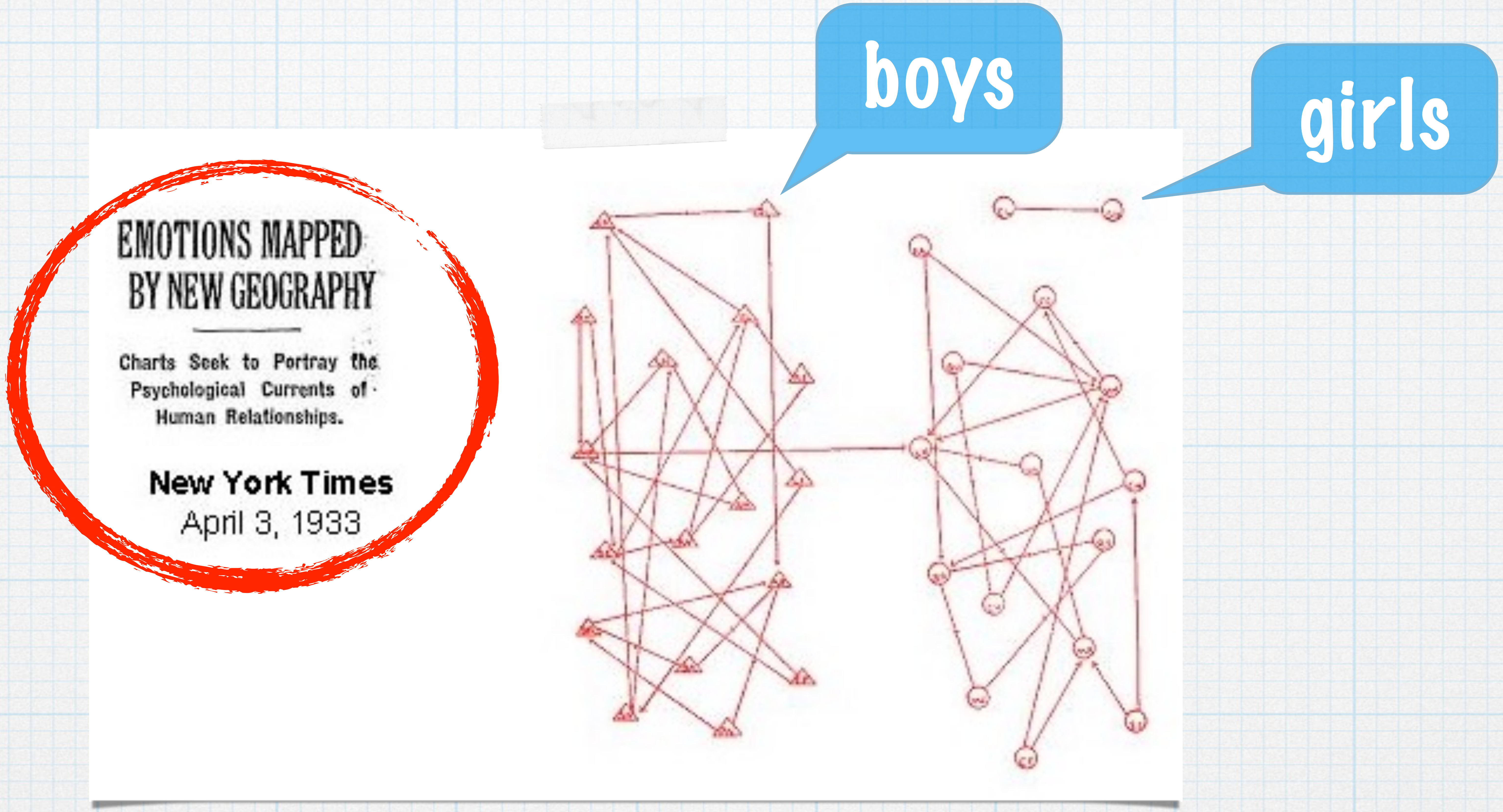


# Erdös-Rényi model: Random Graph

Graph structure changes as a function of p

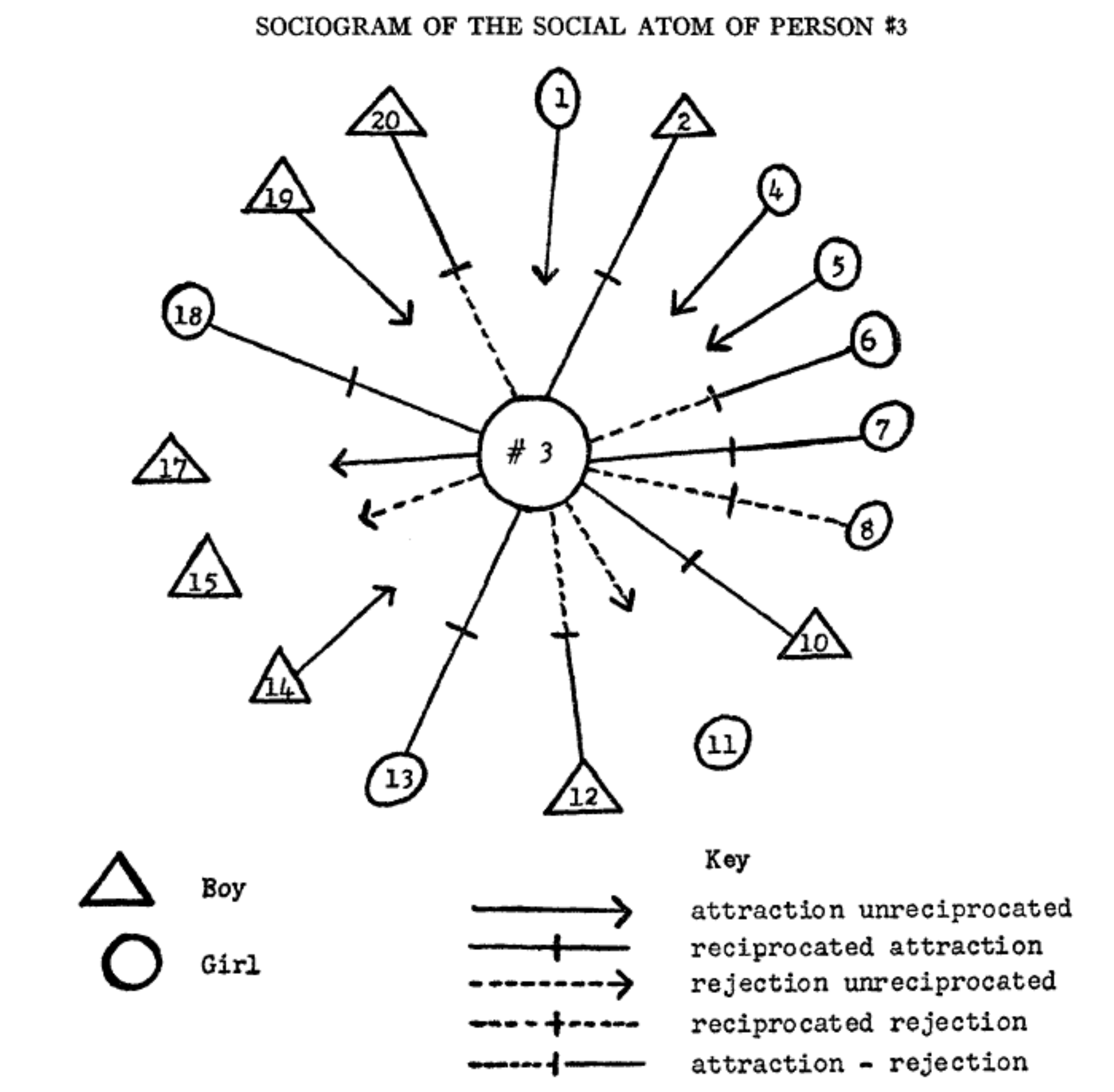


# Social networks

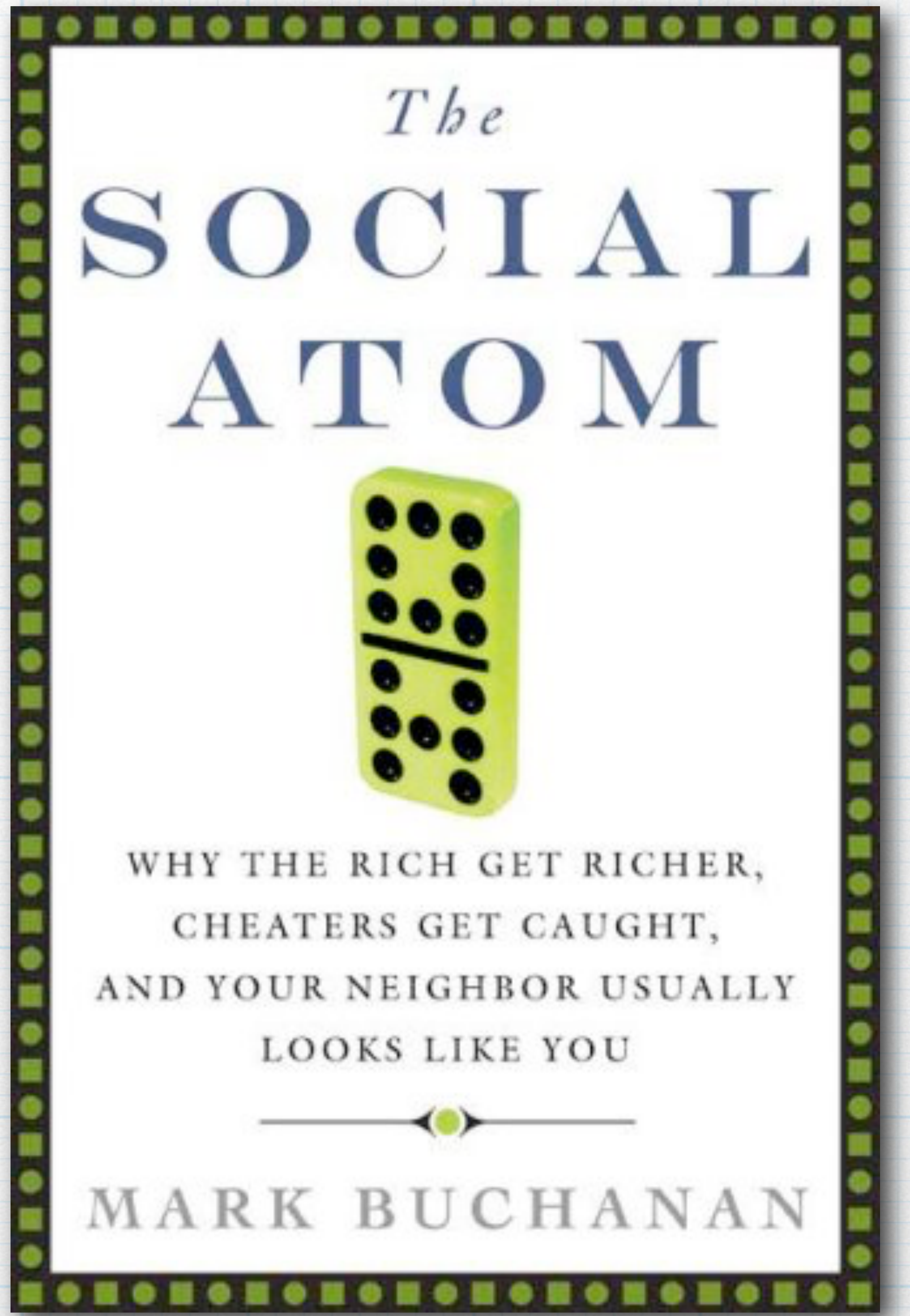



# Jacob Moreno's "sociograms"

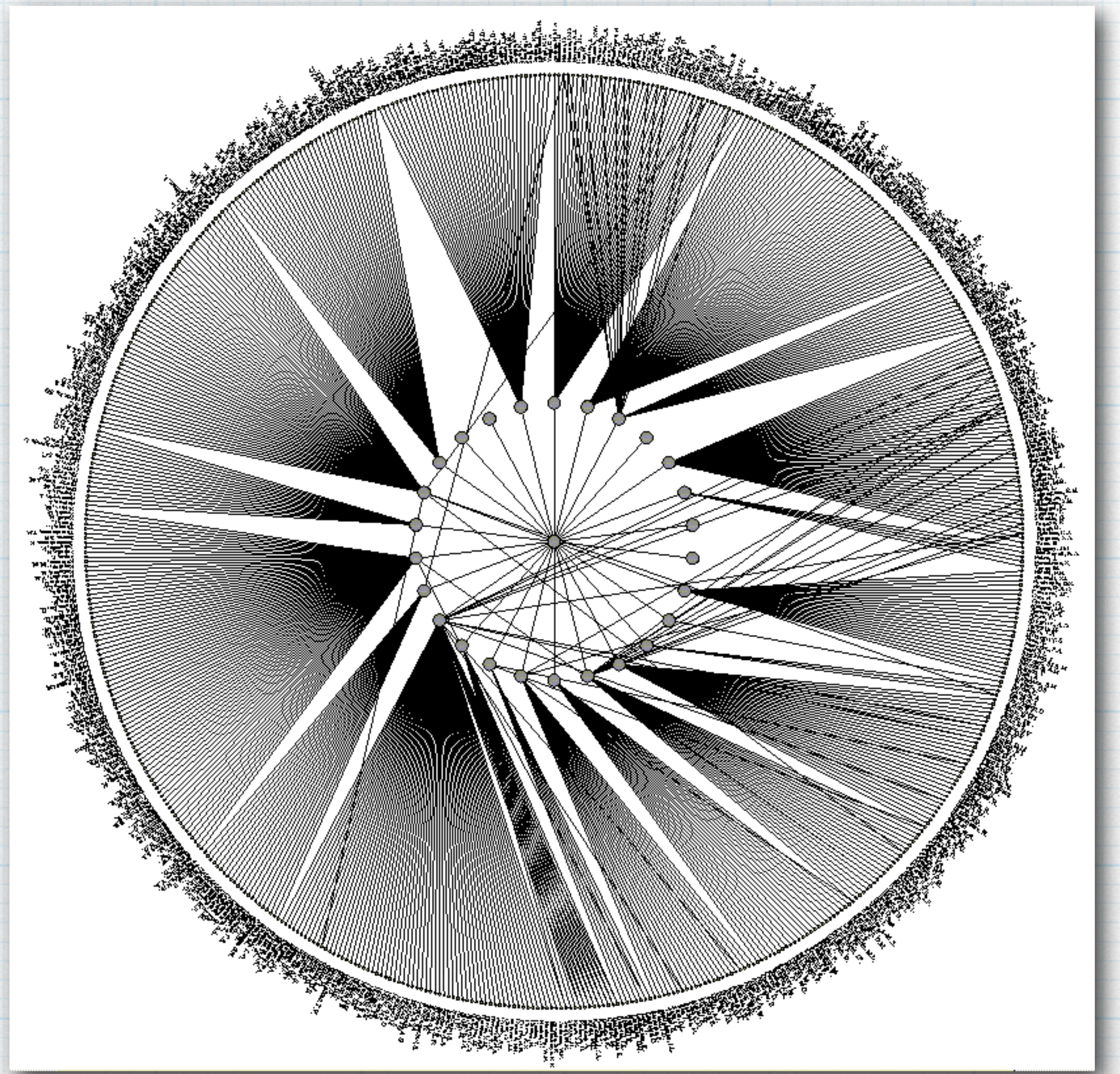
# Moreno's "social atom"




A study of the social atom of #3 shows #3 to be much desired as an associate by her classmates; but that she is very "choosy" about those with whom she will "pal." Her close boy and girl friends are shown, too.



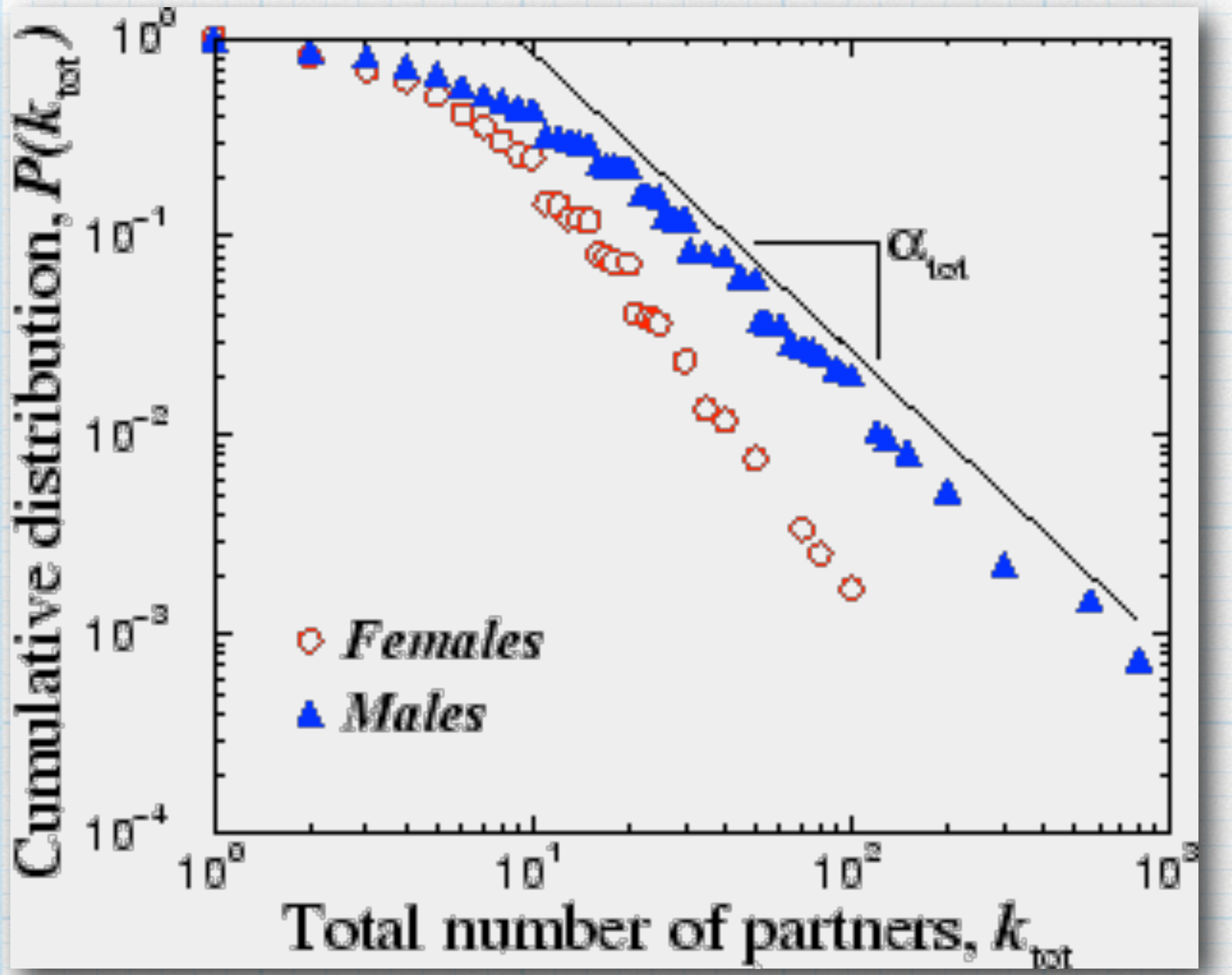
# Science Coauthorship



**Nodes**: scientist (authors)
**Links**: write paper together

# Sex

**Nodes:** people (females, males)
**Links:** sexual relationships

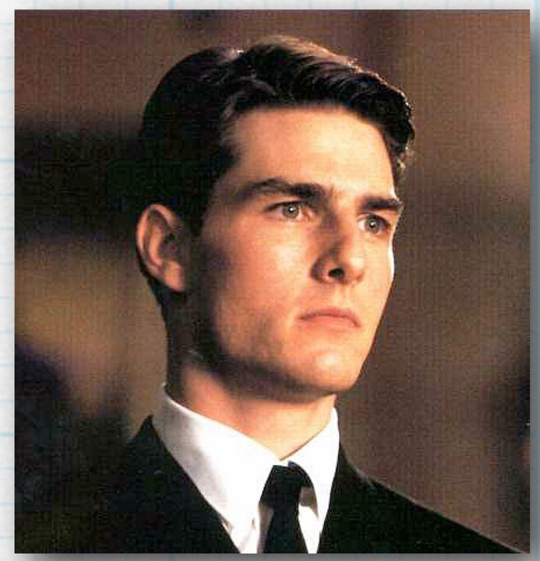



4781 Swedes age 18-74
59% response rate

Liljeros et al. Nature 2001

# Actors

**Nodes**: actors
**Links**: cast jointly



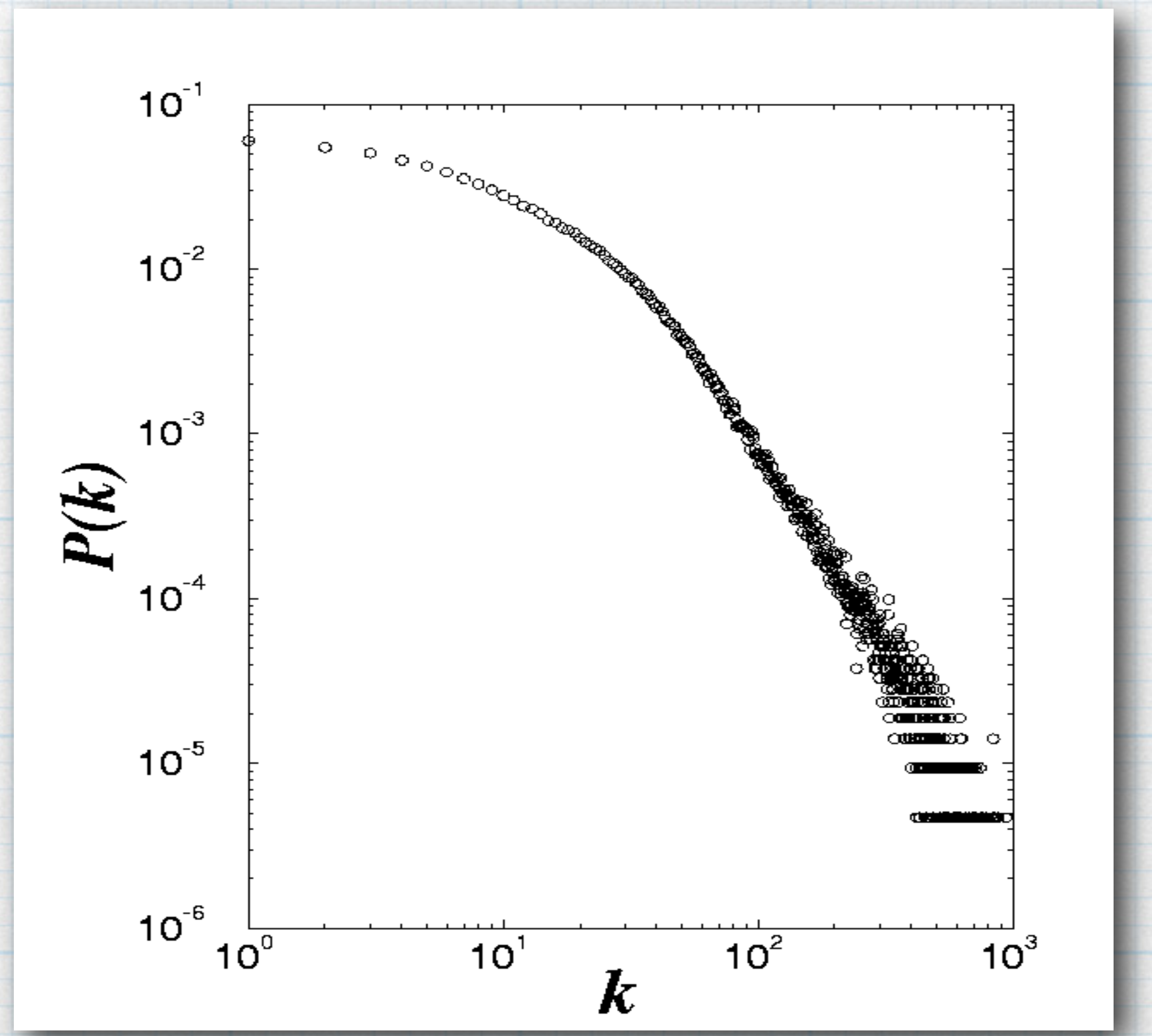
Days of Thunder (1990)
Far and Away (1992)
Eyes Wide Shut (1999)

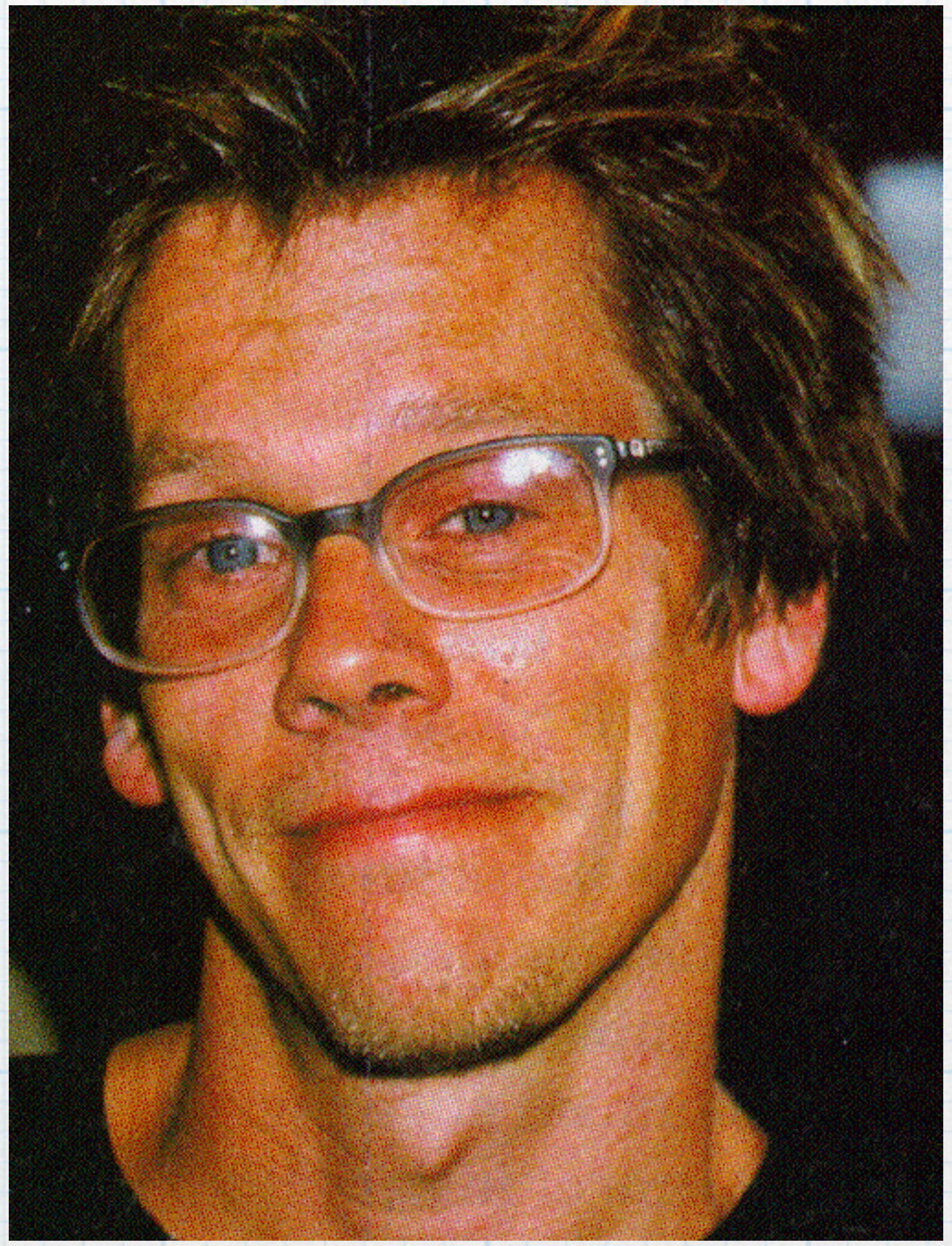
N = 212,250 actors

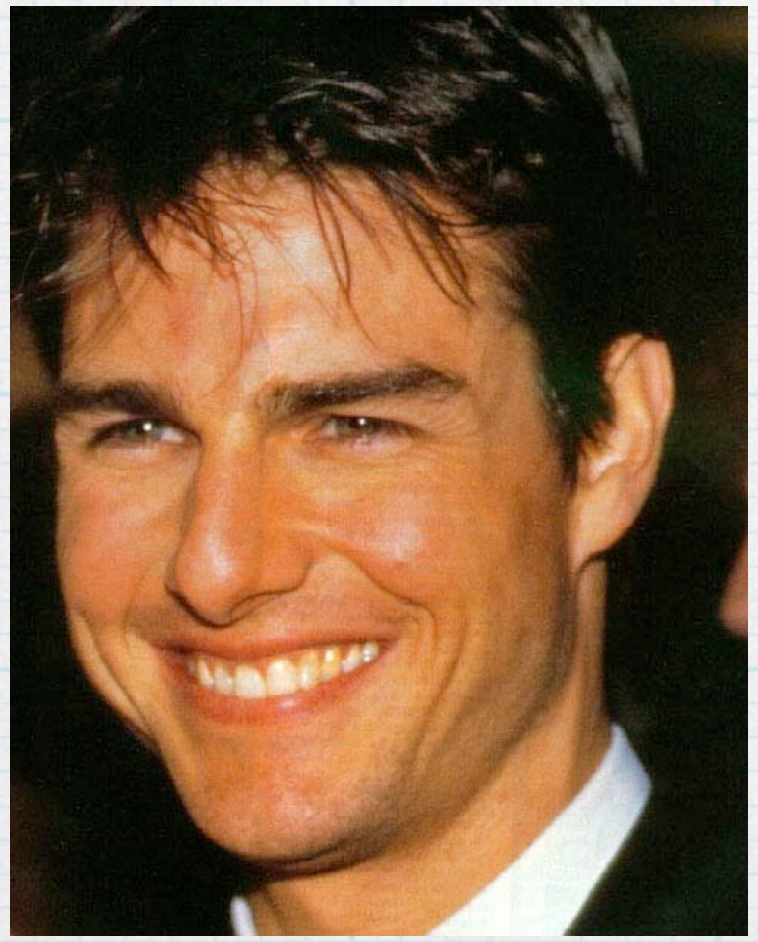
$\langle k \rangle = 28.78$

$P(k) \sim k^{-\gamma}$

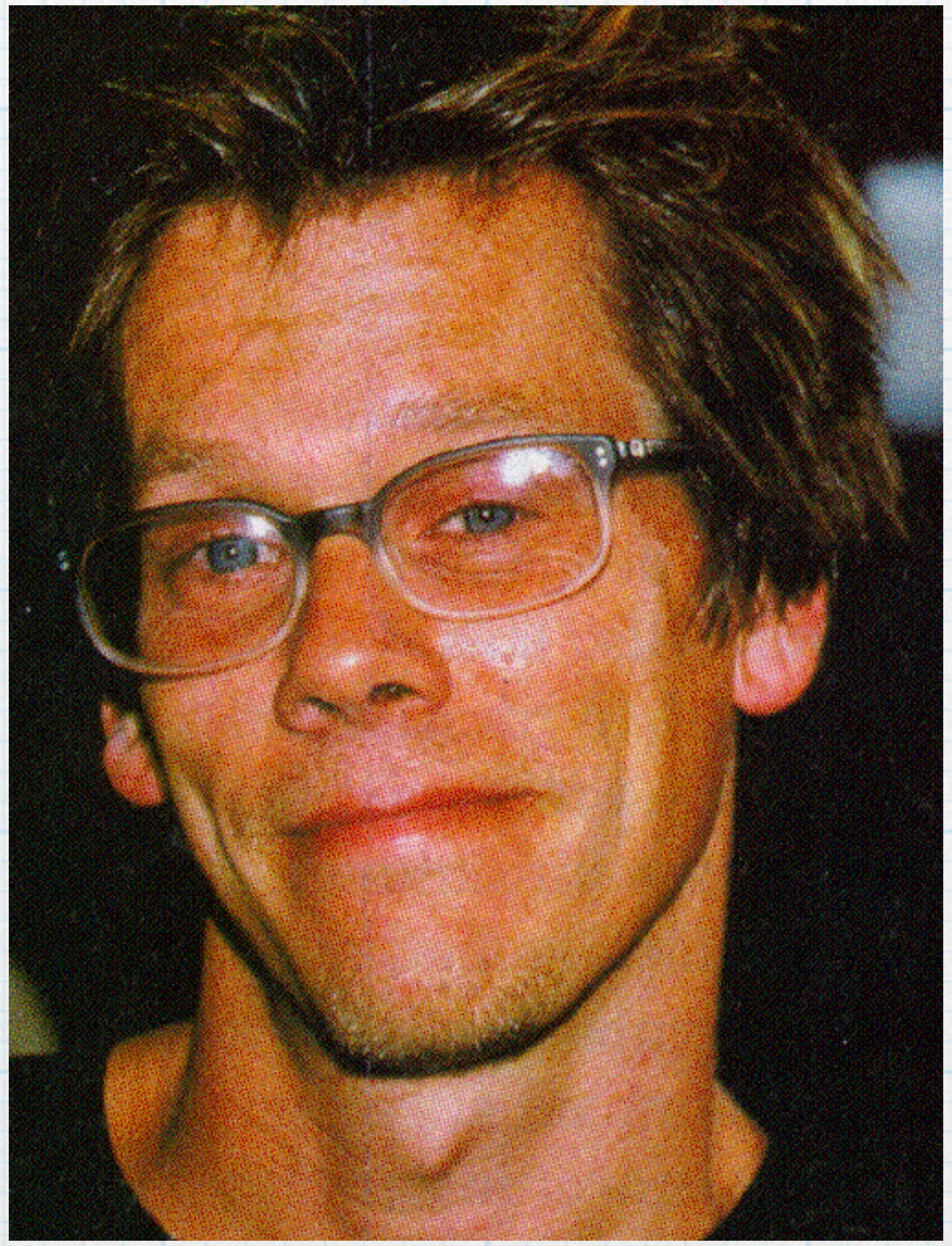
$\gamma = 2.3$







A Few Good Man

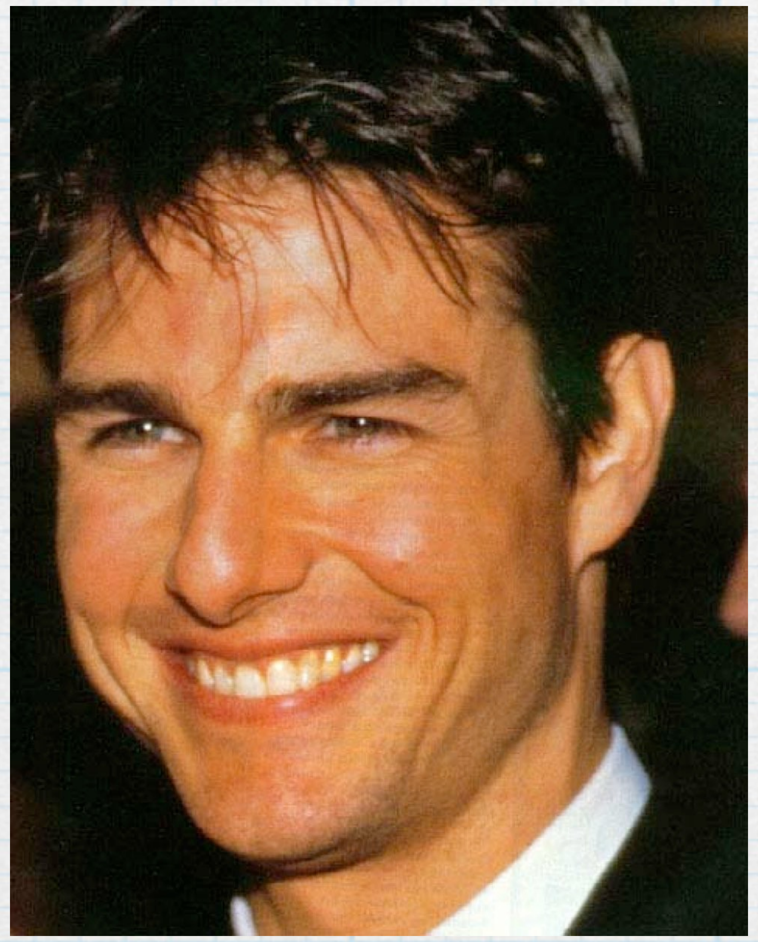


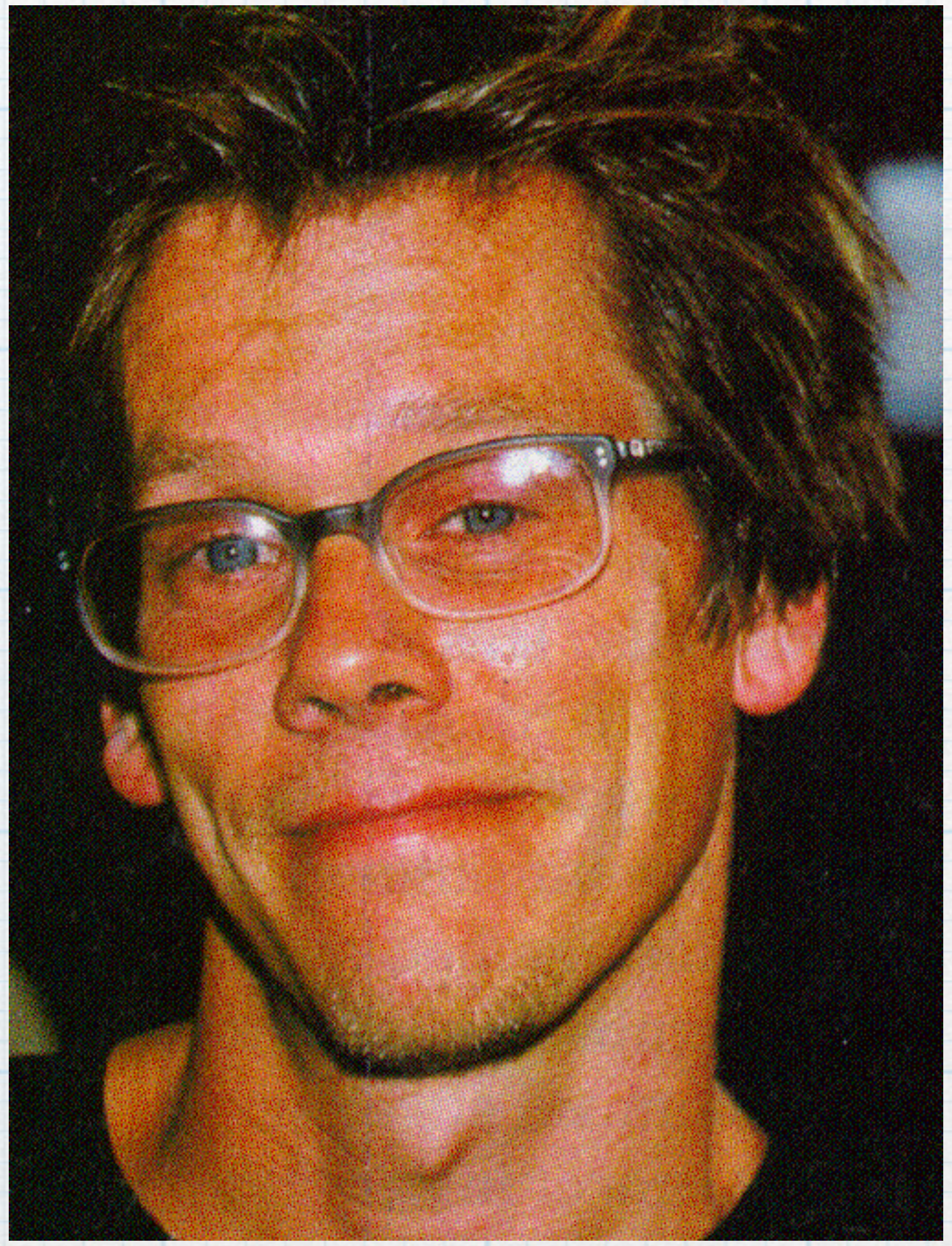
Austin Powers: The spy who shagged me

Robert Wagner

Wild Things

A Few Good Man





Austin Powers: The spy who shagged me

Let's make it legal

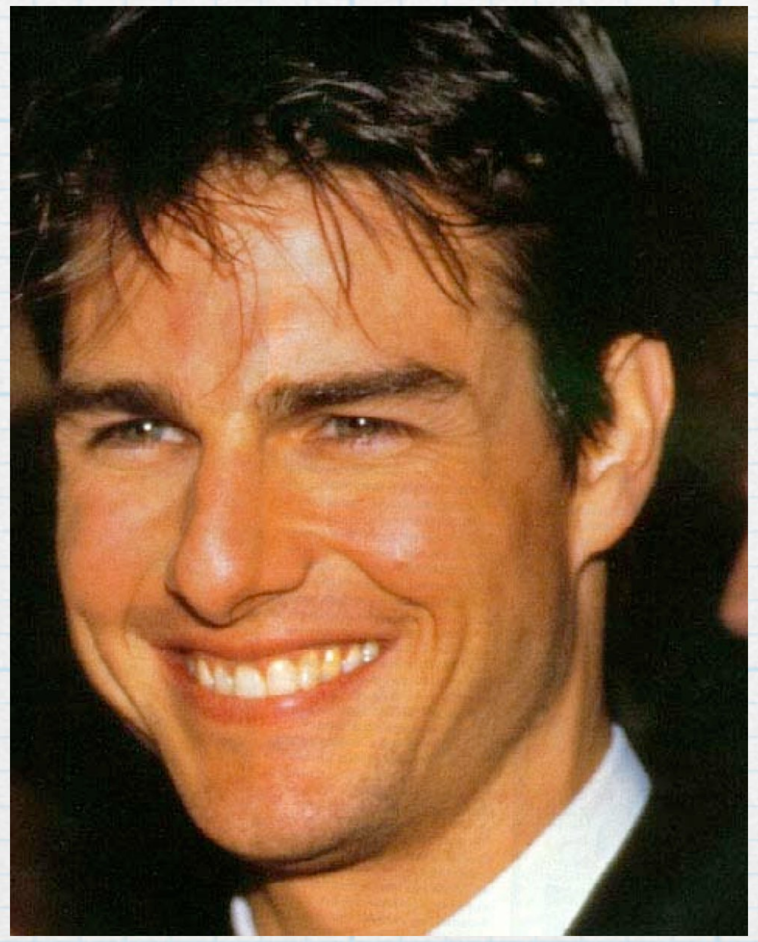
Robert Wagner

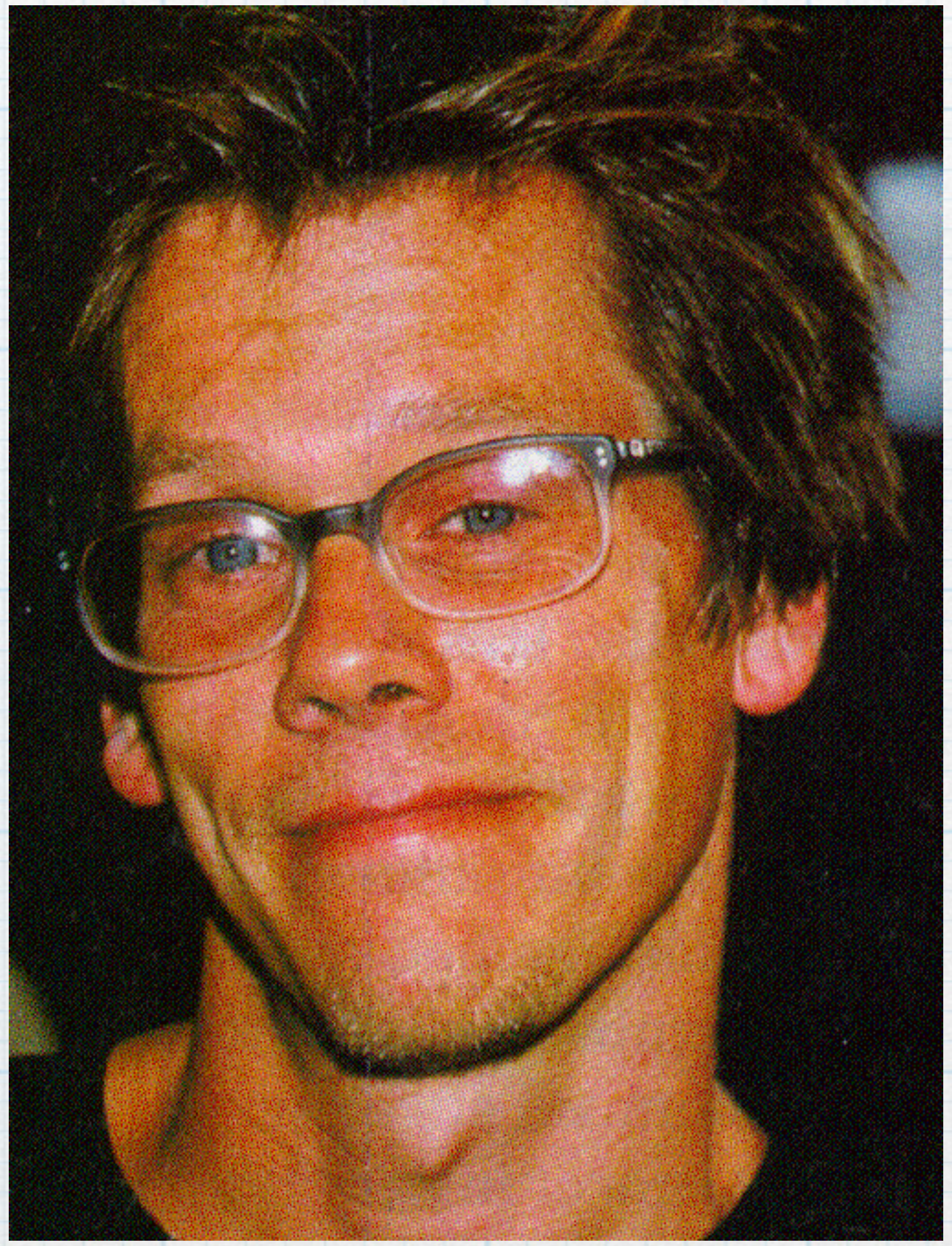
Wild Things

What Price Glory

Barry Norton

A Few Good Man



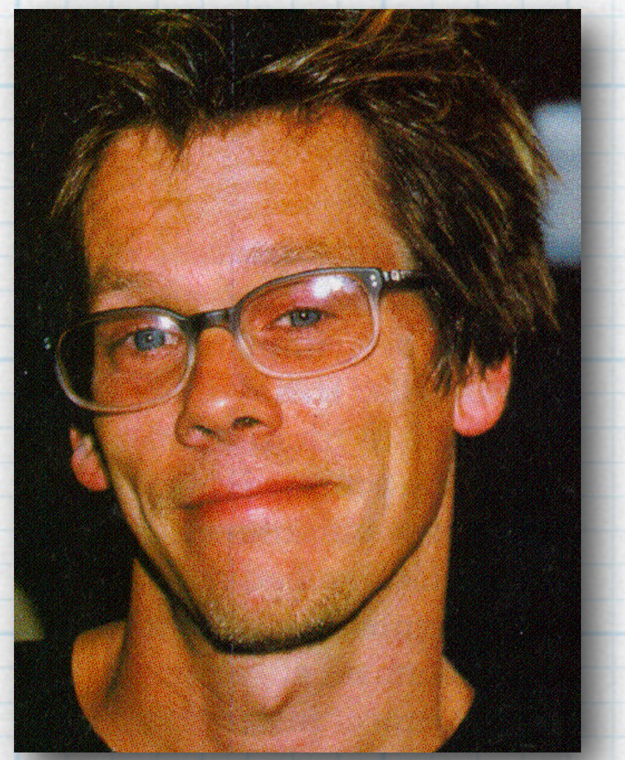


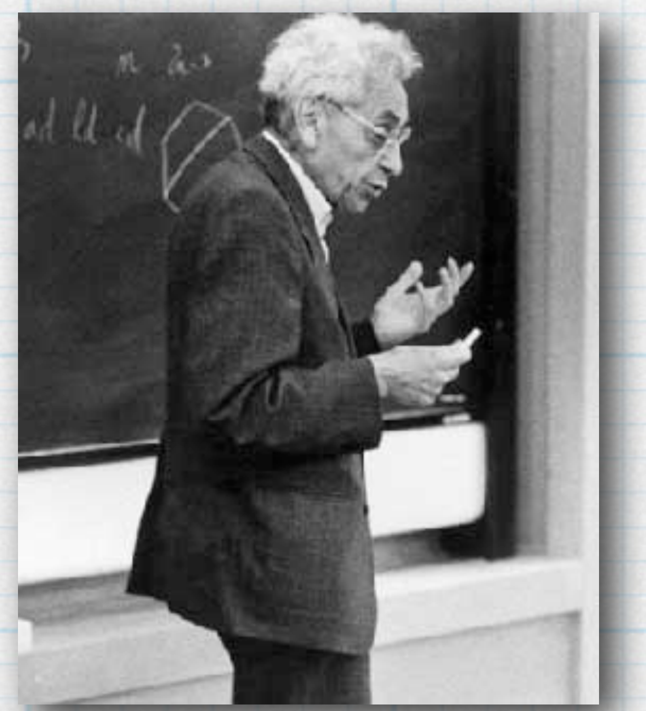
Monsieur Verdoux

# Structural properties

* When we talk about social networks, three structural properties are important:
  * Broad degree distribution — will get back to this later
  * Short paths, or small APL and diameter — defined in past weeks
  * Clustering , or cohesiveness of the neighborhood — we need to define this property formally

# Social networks





- Let's play a game:
  - Six degrees of Kevin Bacon
    - http://oracleofbacon.org/
  - Alternative game you can play: Erdös number
    - http://academic.research.microsoft.com/VisualExplorer
    - http://www.ams.org/mathscinet/collaborationDistance.html

# What have we learned?

# Milgram's experiment

* Stanley Milgram, psychologist at Harvard (famous for another exp.)
* 1967 experiment to measure "social distance" between any 2 people in the US
    * First idea was in the short story "Chains" by Hungarian writer Frigyes Karinthy in 1929
    * John Guare's 1991 play coined the term "six degrees of separation" (movie, too)

2 targets in Mass: the wife of a student in Sharon and a stockbroker in Boston

160 letters to people in Omaha, NE and Wichita, KS

Instructions: send to personal acquaintance who is more likely to know target

2 targets in Mass: the wife of a student in Sharon and a stockbroker in Boston

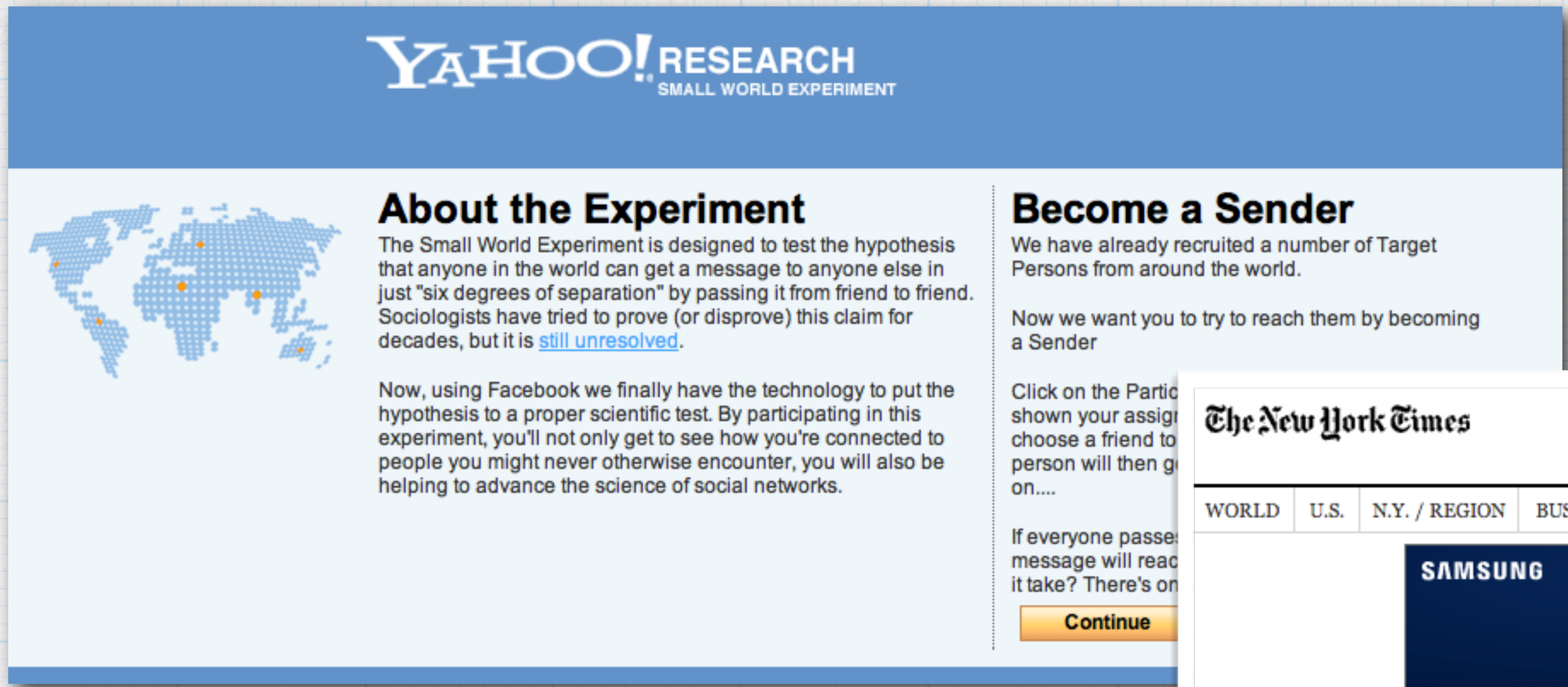
160 letters to people in Omaha, NE and Wichita, KS


Figure 1

**Instructions:** send to personal acquaintance who is more likely to know target

# Milgram's experiment

* 42 letters made it back (only 26%)
* Range: 3-12 steps
* Average: 5.5 intermediates (6.5 steps)
* Much lower than most people expected!
* "Small world" effect is still surprising
* Half of letters arrived via 3 friends of target: "gatekeepers"?


YAHOO! RESEARCH
SMALL WORLD EXPERIMENT

## About the Experiment

The Small World Experiment is designed to test the hypothesis that anyone in the world can get a message to anyone else in just "six degrees of separation" by passing it from friend to friend. Sociologists have tried to prove (or disprove) this claim for decades, but it is still unresolved.

Now, using Facebook we finally have the technology to put the hypothesis to a proper scientific test. By participating in this experiment, you'll not only get to see how you're connected to people you might never otherwise encounter, you will also be helping to advance the science of social networks.

## Become a Sender

We have already recruited a number of Target Persons from around the world.

Now we want you to try to reach them by becoming a Sender

Click on the Partic
shown your assig
choose a friend to
person will then g
on....

If everyone passe
message will reac
it take? There's on

Continue


The New York Times

Business Day
Technology

WORLD | U.S. | N.Y. / REGION | BUSINESS | TECHNOLOGY | SCIENCE | HEALTH | SPORTS | OPI



# Separating You and Me? 4.74 Degrees

By JOHN MARKOFF and SOMINI SENGUPTA
Published: November 21, 2011

The world is even smaller than you thought.

Enlarge This Image

Adding a new chapter to the research that cemented the phrase "six degrees of separation" into the language, scientists at Facebook and the University of Milan reported on Monday that the average number of acquaintances separating any two people in the world was not six but 4.74.

The original "six degrees" finding, published in 1967 by the psychologist Stanley Milgram,

RECOMMEND
TWITTER
LINKEDIN
SIGN IN TO E
MAIL
PRINT
REPRINTS
SHARE



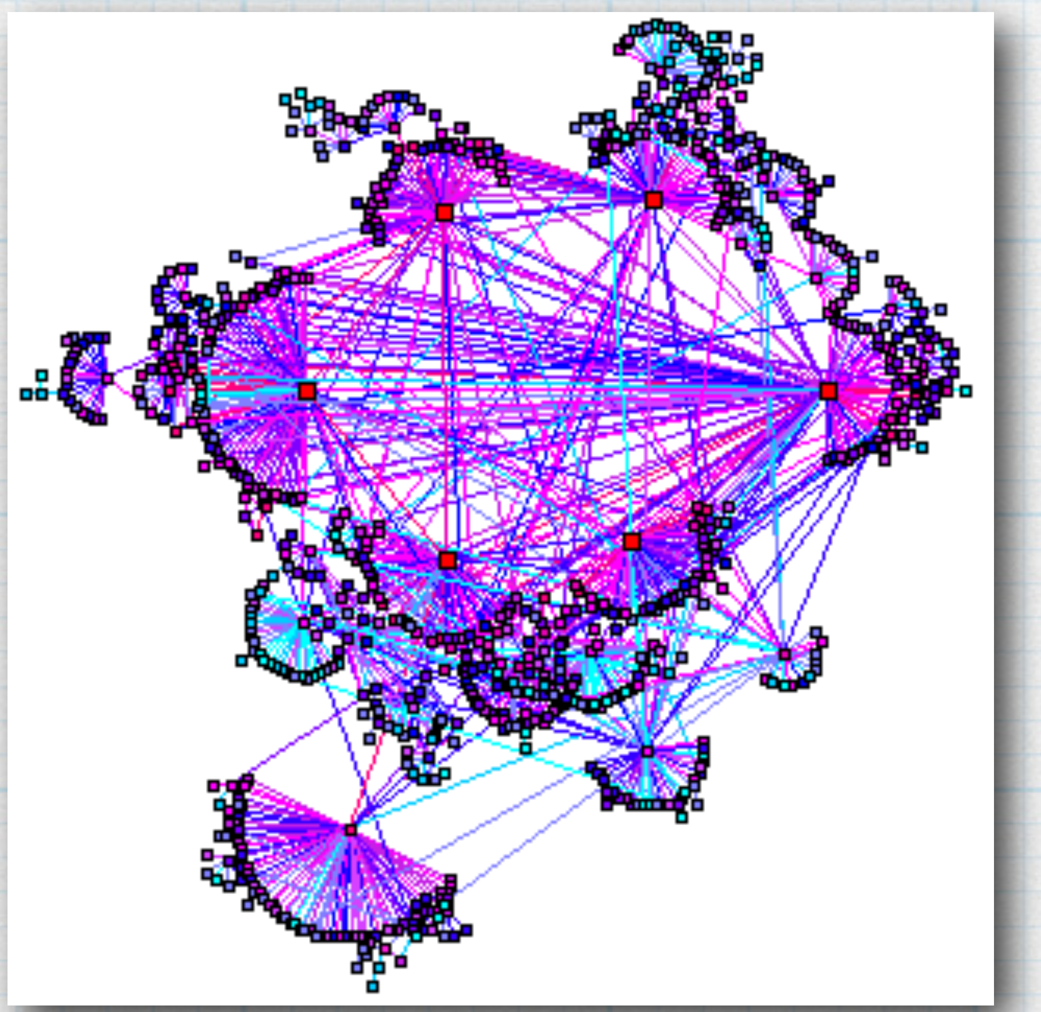
- Repeated in 2003 by Watts at Yahoo Labs using email
- APL=4 (6 when accounting for broken chains)

Even shorter online!

# The Web also has a small diameter

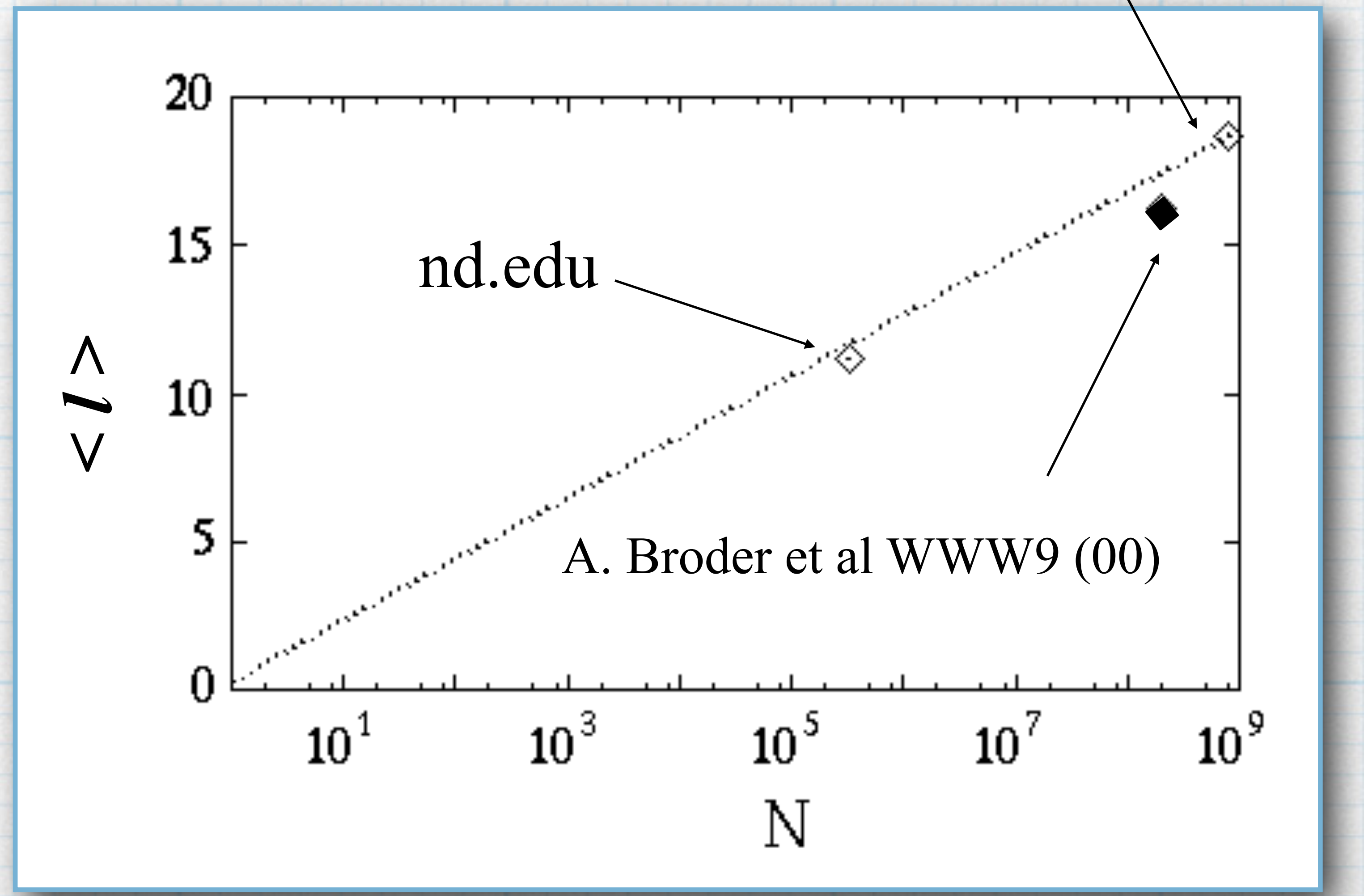
**Nodes**: WWW documents
**Links**: URL links

800 million documents
(Lawrence & Giles, 1999)



**CRAWLER:** collects all URL's found in a document and follows them recursively

19 degrees of separation
R. Albert et al Nature (99)

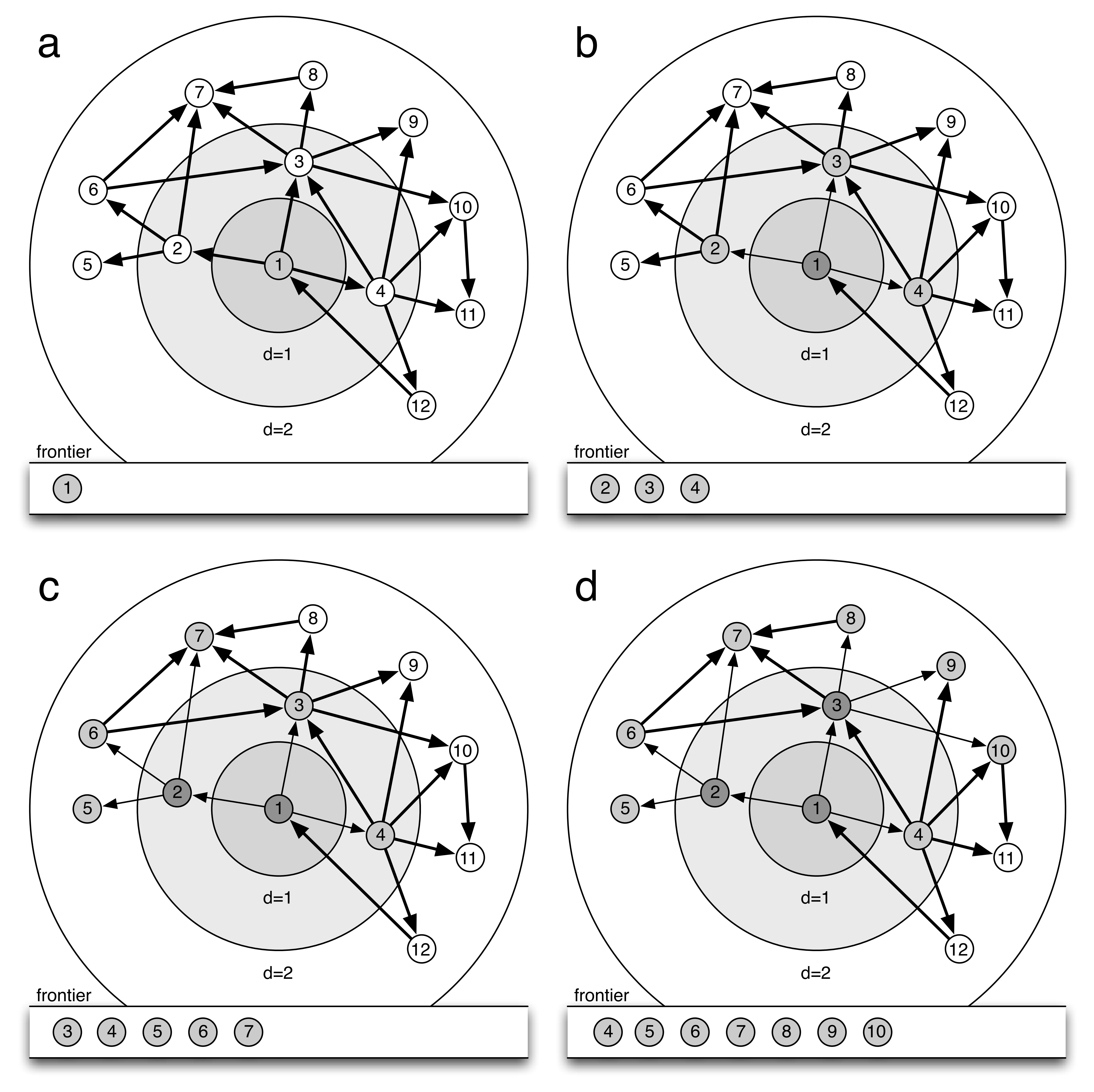



APL $\propto$ 2.06 log(N)



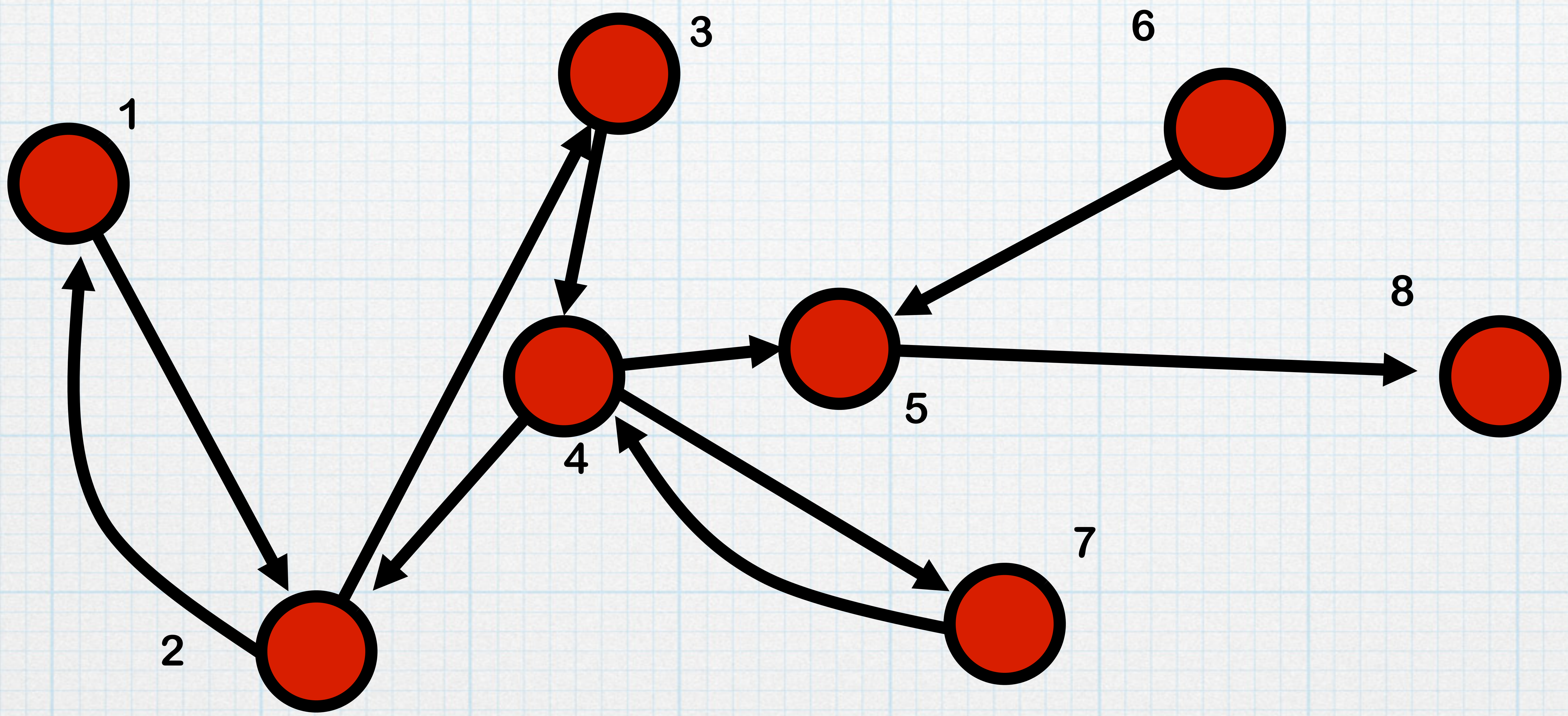
INDIANA UNIVERSITY BLOOMINGTON

# Finding Shortest Paths




- Breadth-first search algorithm (BFS)
- O(E+V) runtime
  - is that good or bad?

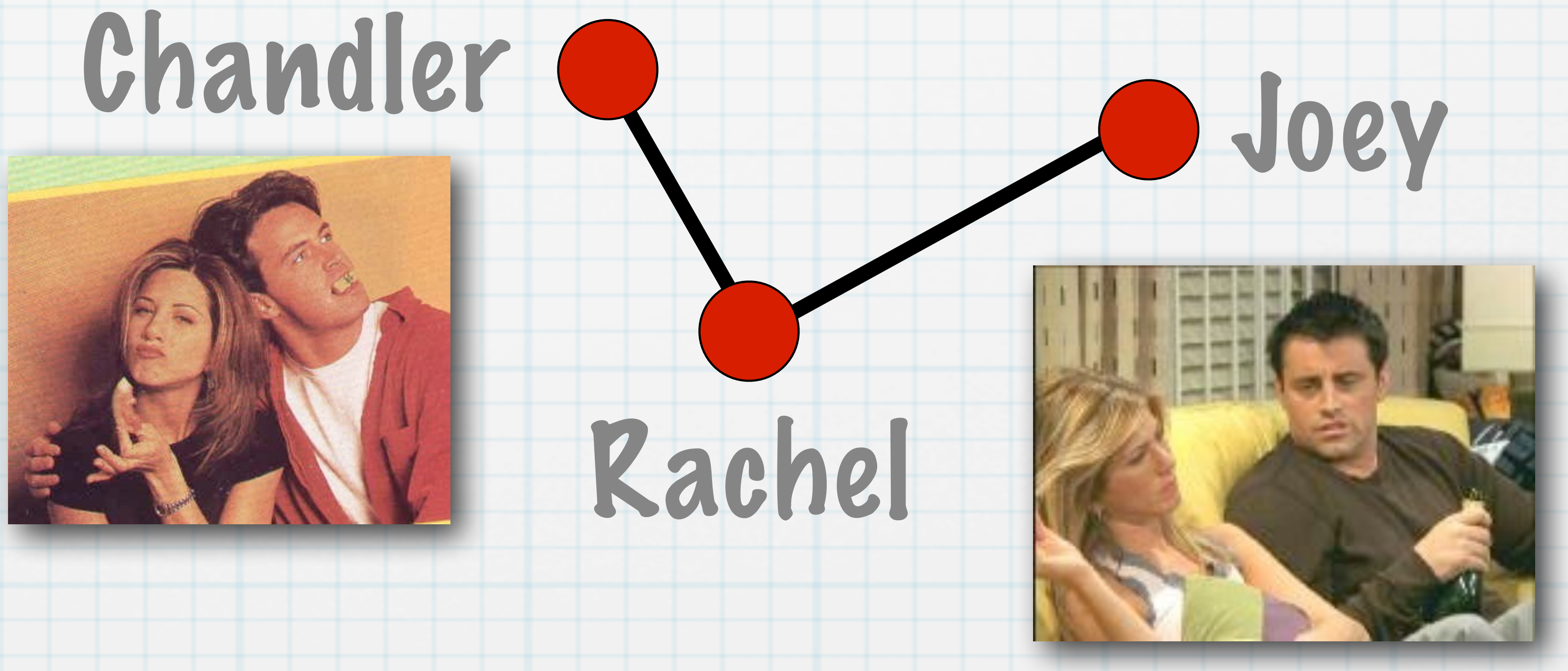
# BFS: let's practice



# Cohesiveness

- **Clustering coefficient:** measure of node interconnectivity

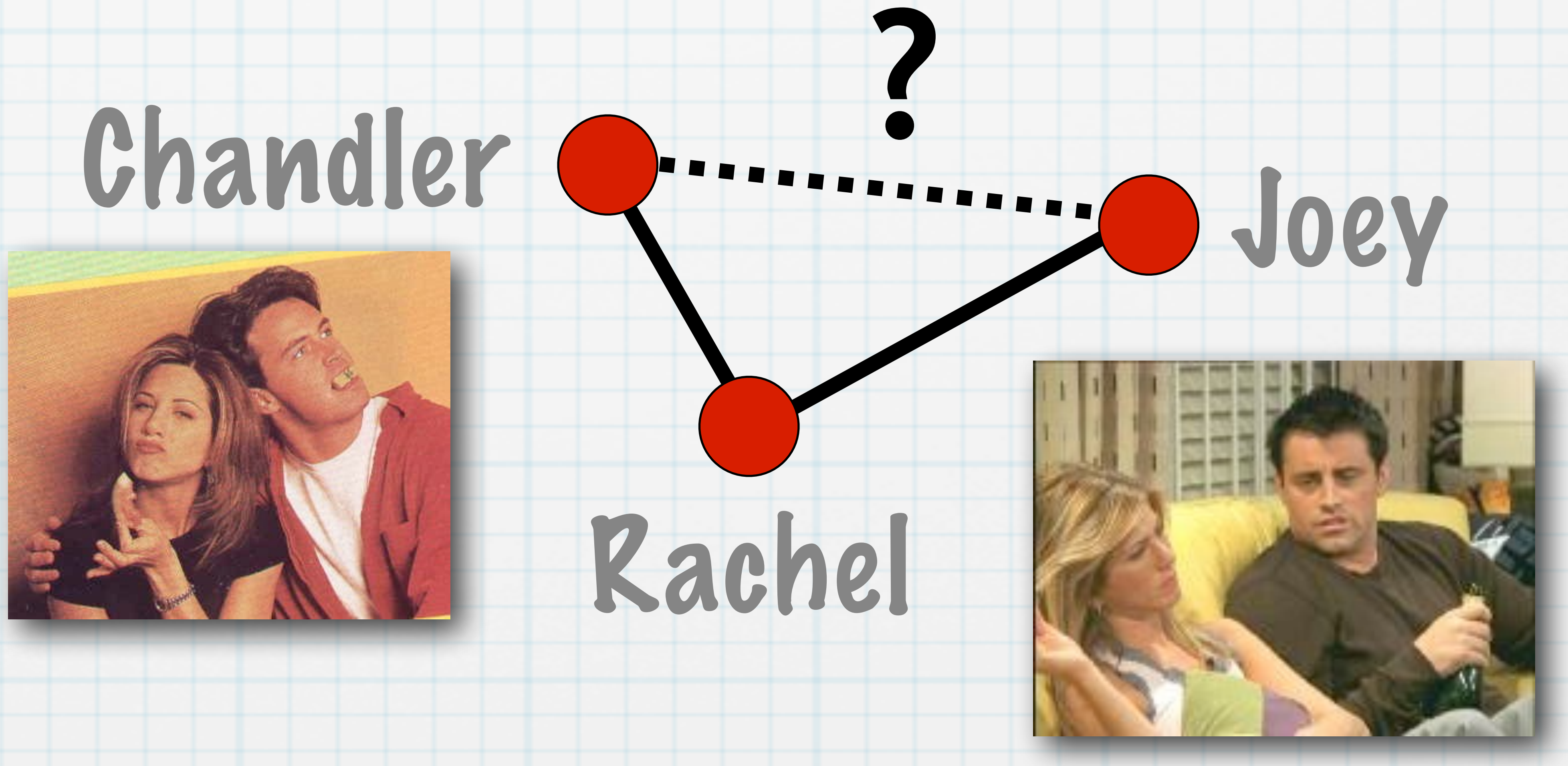
*Are friends of friends friends of each other ?*



# Cohesiveness

* **Clustering coefficient:** measure of node interconnectivity

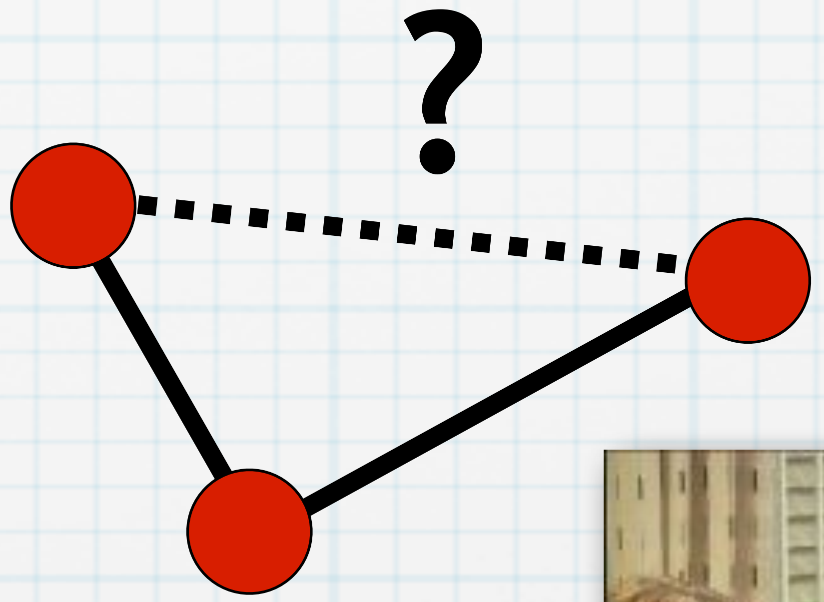
*Are friends of friends friends of each other ?*



# Cohesiveness

* **Clustering coefficient:** measure of node interconnectivity

*Are friends of friends friends of each other ?*

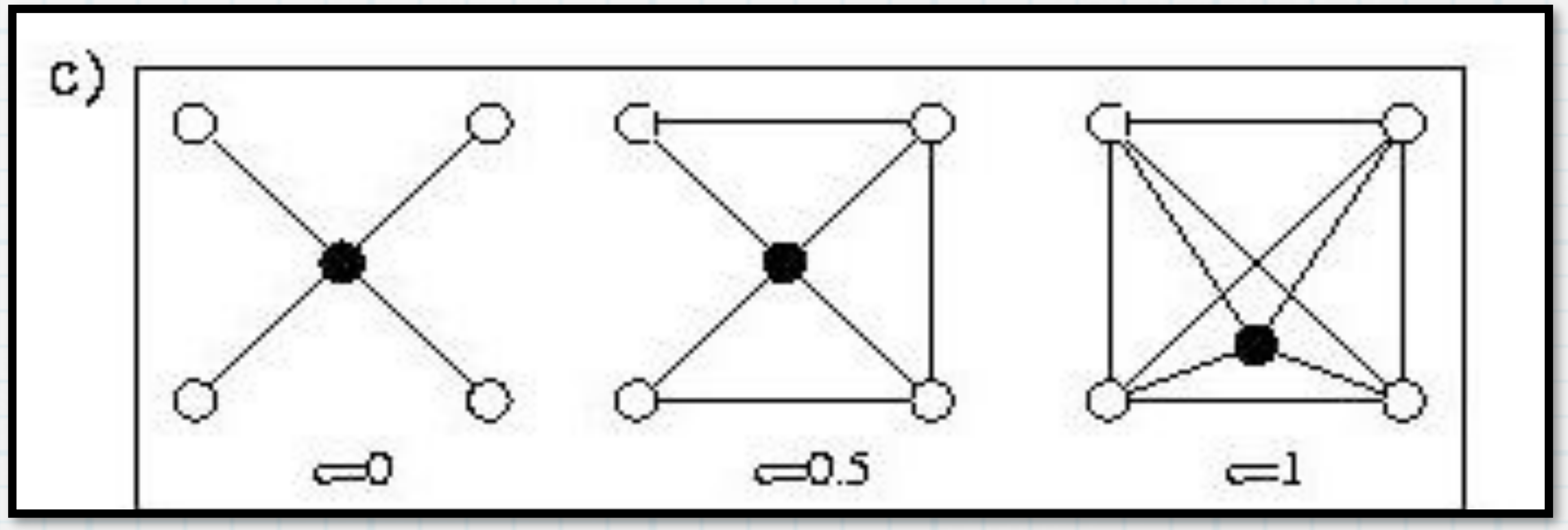



# Clustering coefficient

* What portion of your neighbors are connected?
    * Node i with degree $k_i$

$e_i$ : actual number of triangles in which i participates

$$C_i = \frac{2e_i}{k_i(k_i - 1)}$$

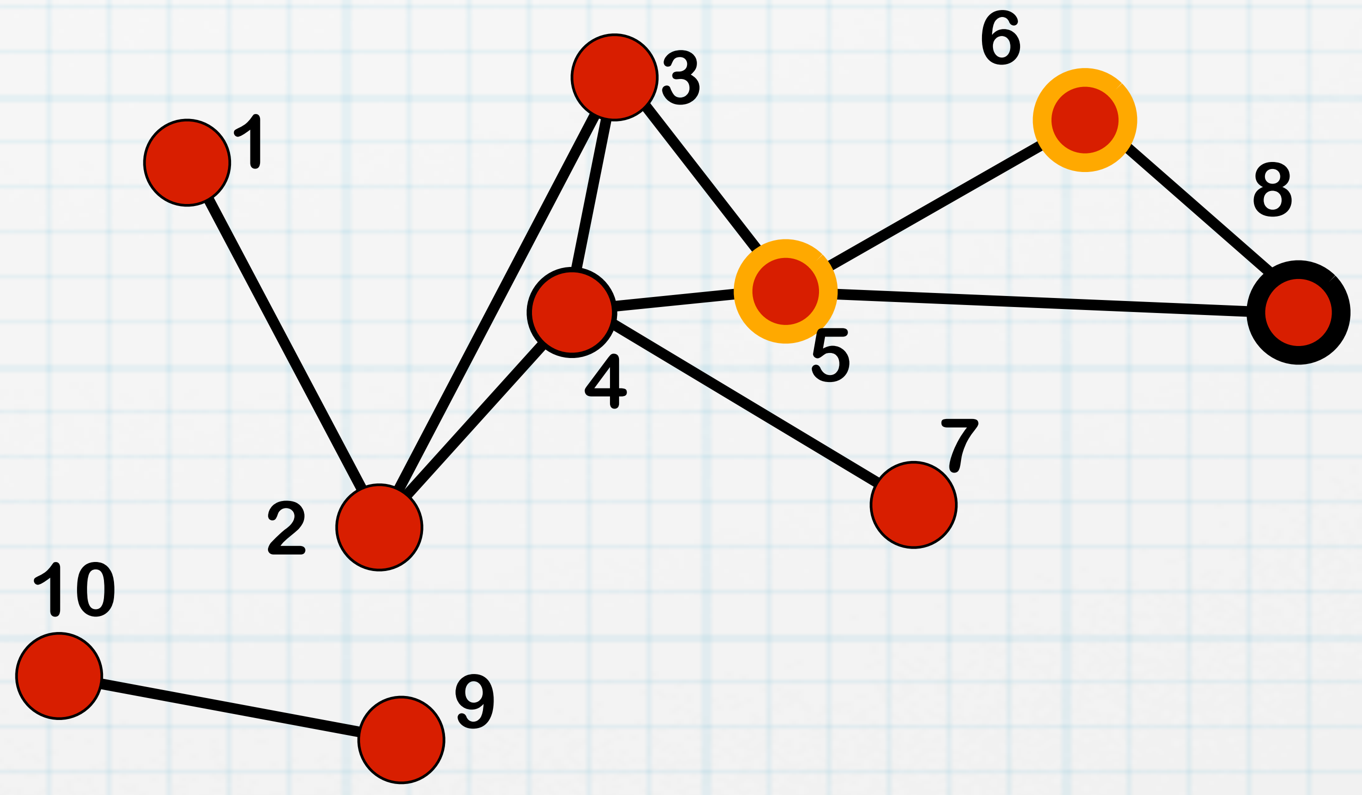



* $C_i$ in [0,1]

# Clustering coefficient

* What portion of your neighbors are connected?
* Node i with degree $k_i$

$$C_i = \frac{2e_i}{k_i(k_i - 1)}$$

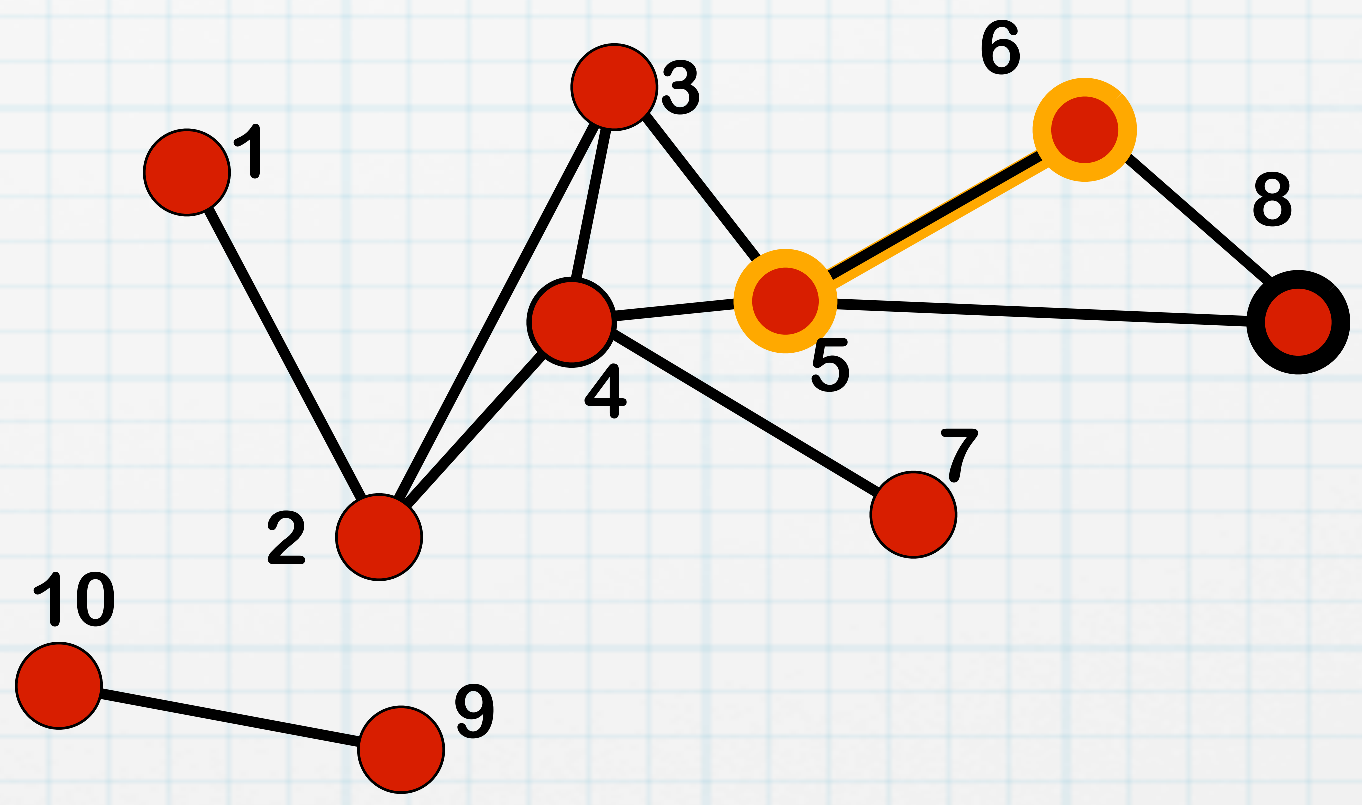



$i=8$: $k_8=2$, $e_8=1$, $MAX=2*1/2=1$ → $C_8=1/1=1$

# Clustering coefficient

- What portion of your neighbors are connected?
- Node i with degree $k_i$

$$C_i = \frac{2e_i}{k_i(k_i - 1)}$$

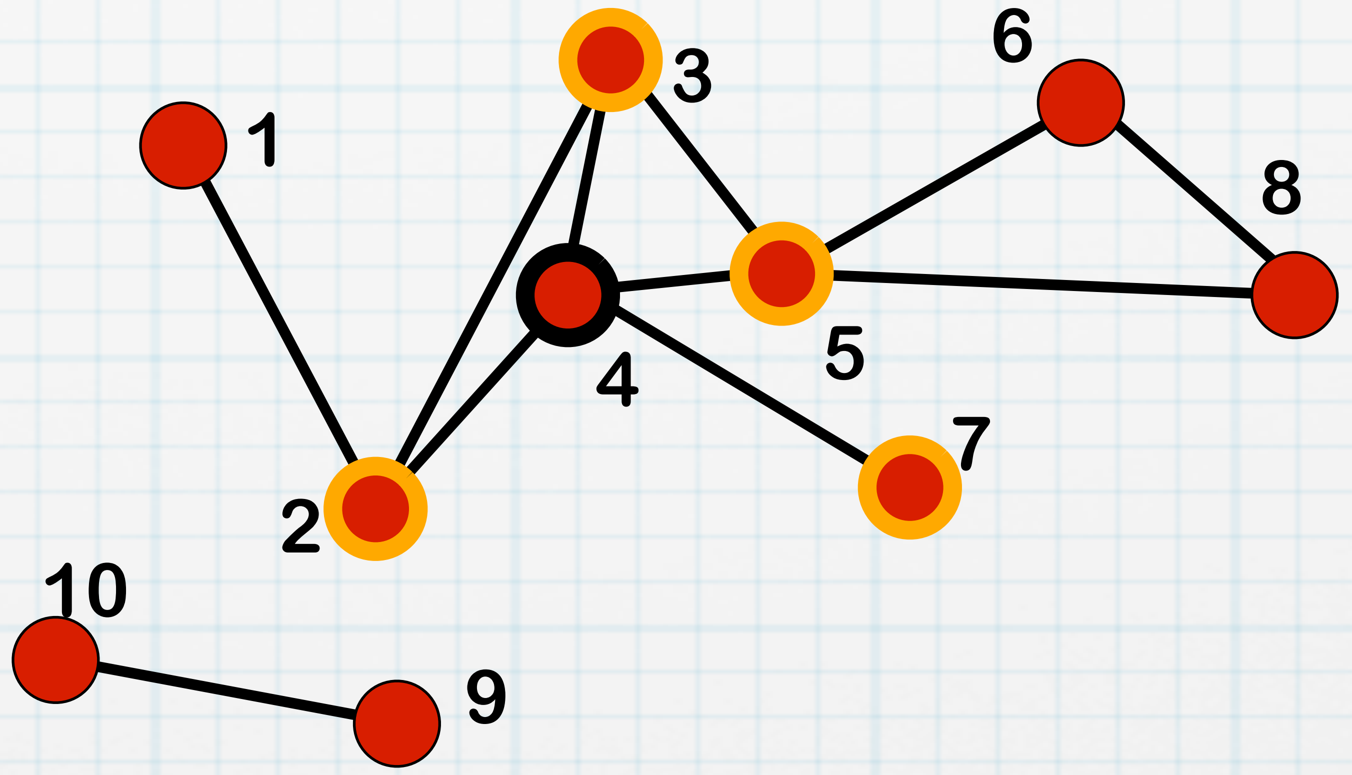



$i=8$: $k_8=2$, $e_8=1$, $MAX=2*1/2=1$ → $C_8=1/1=1$

# Exercise

- Clustering coefficient: what portion of your neighbors are connected?
  - Node i with degree $k_i$

$$C_i = \frac{2e_i}{k_i(k_i - 1)}$$




*i=4: $k_4$=4, → $C_4$=?*

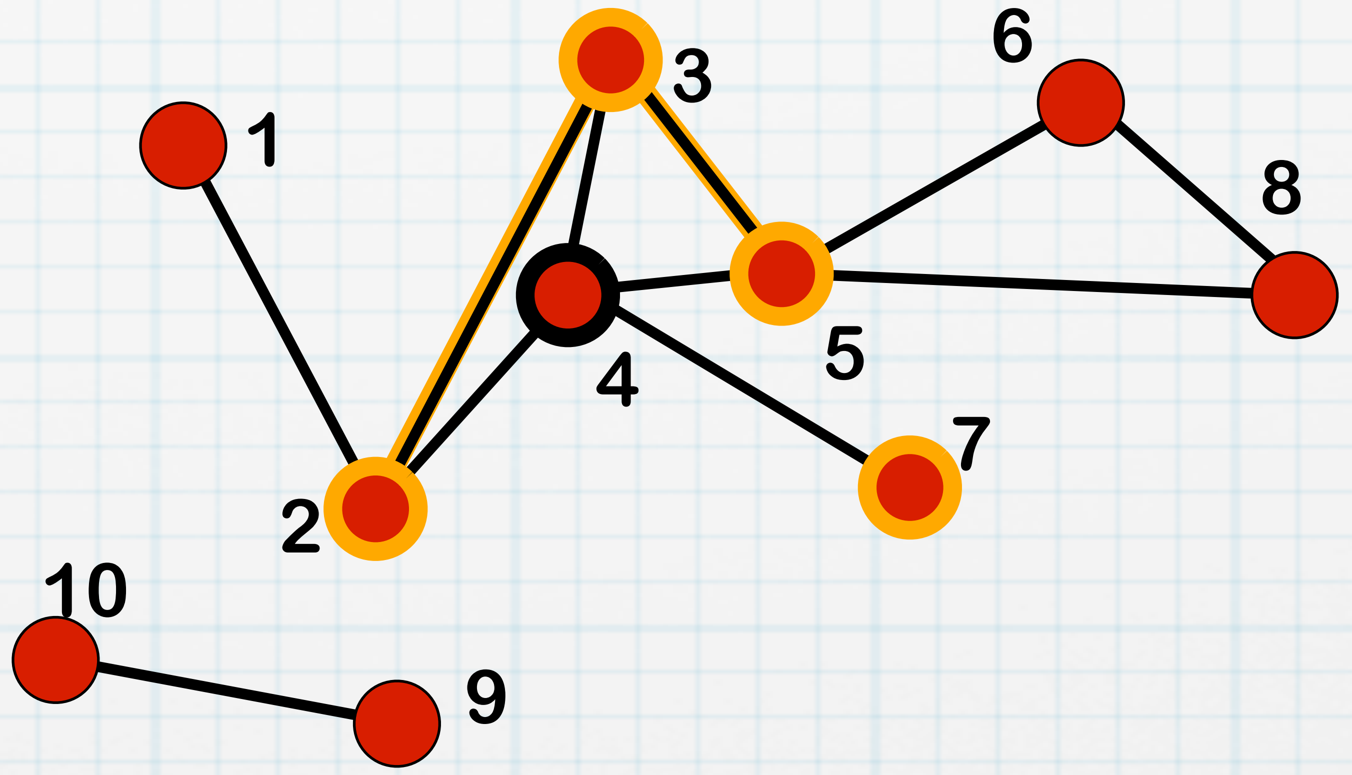


INDIANA UNIVERSITY BLOOMINGTON

# Exercise

* Clustering coefficient: what portion of your neighbors are connected?
  * Node i with degree $k_i$

$$C_i = \frac{2e_i}{k_i(k_i - 1)}$$

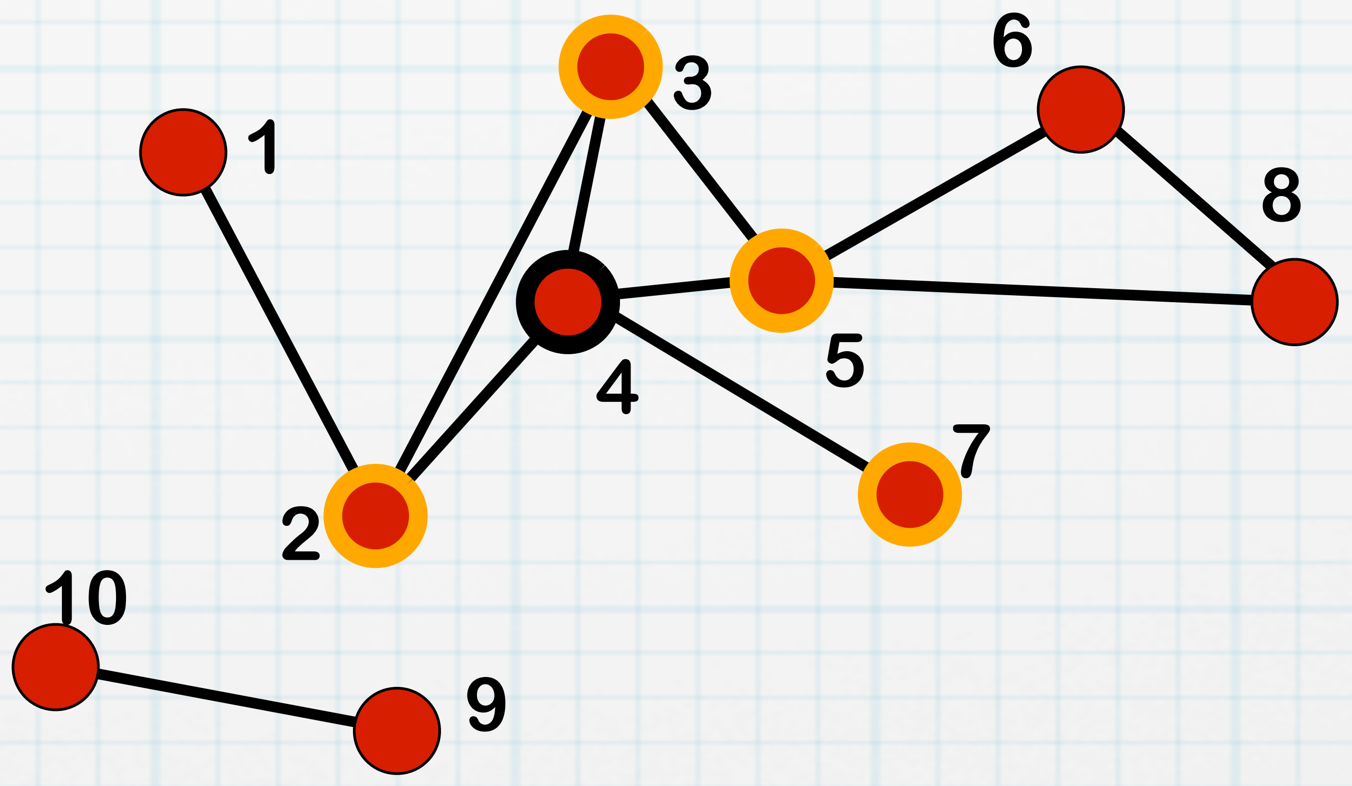



$i=4$: $k_4=4$, → $C_4=?$

# Exercise

* Clustering coefficient: what portion of your neighbors are connected?
  * Node i with degree $k_i$

$$C_i = \frac{2e_i}{k_i(k_i - 1)}$$

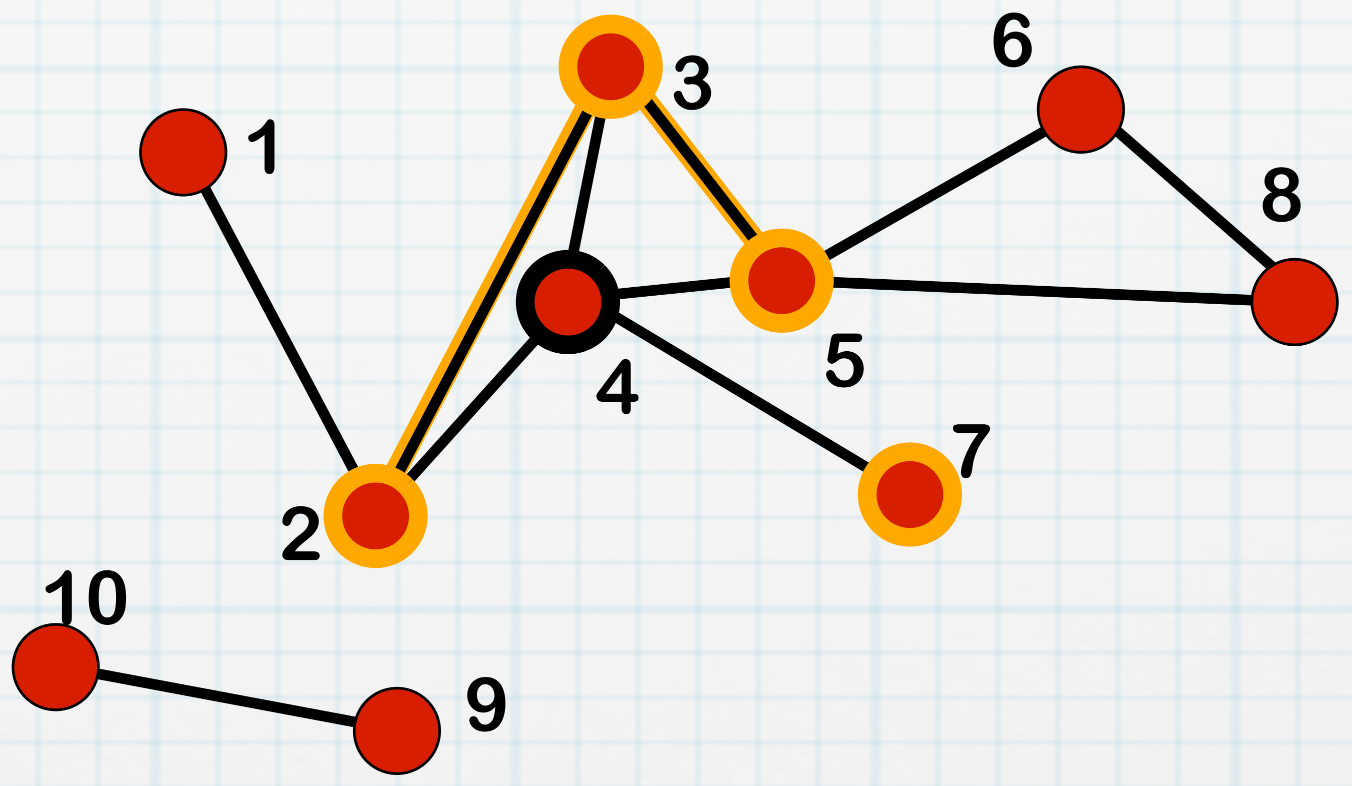



$i=4$: $k_4=4$, $e_4=2$, $MAX=4*3/2=6$ → $C_4=2/6=1/3$

# Exercise

- Clustering coefficient: what portion of your neighbors are connected?
  - Node i with degree $k_i$

$$C_i = \frac{2e_i}{k_i(k_i - 1)}$$




$i=4$: $k_4=4$, $e_4=2$, $MAX=4*3/2=6$ → $C_4=2/6=1/3$

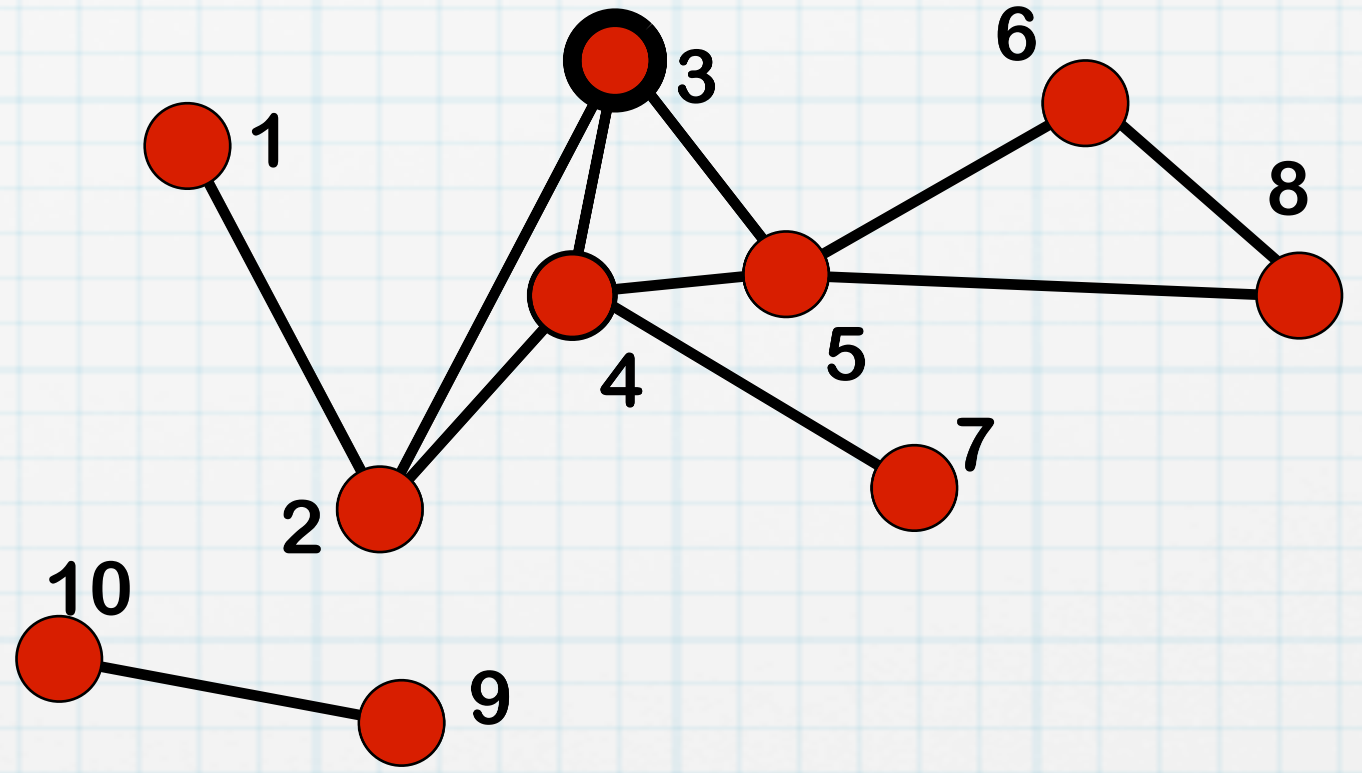


INDIANA UNIVERSITY BLOOMINGTON

# Exercise

* Clustering coefficient: what portion of your neighbors are connected?
  * Node i with degree $k_i$

$$C_i = \frac{2e_i}{k_i(k_i - 1)}$$

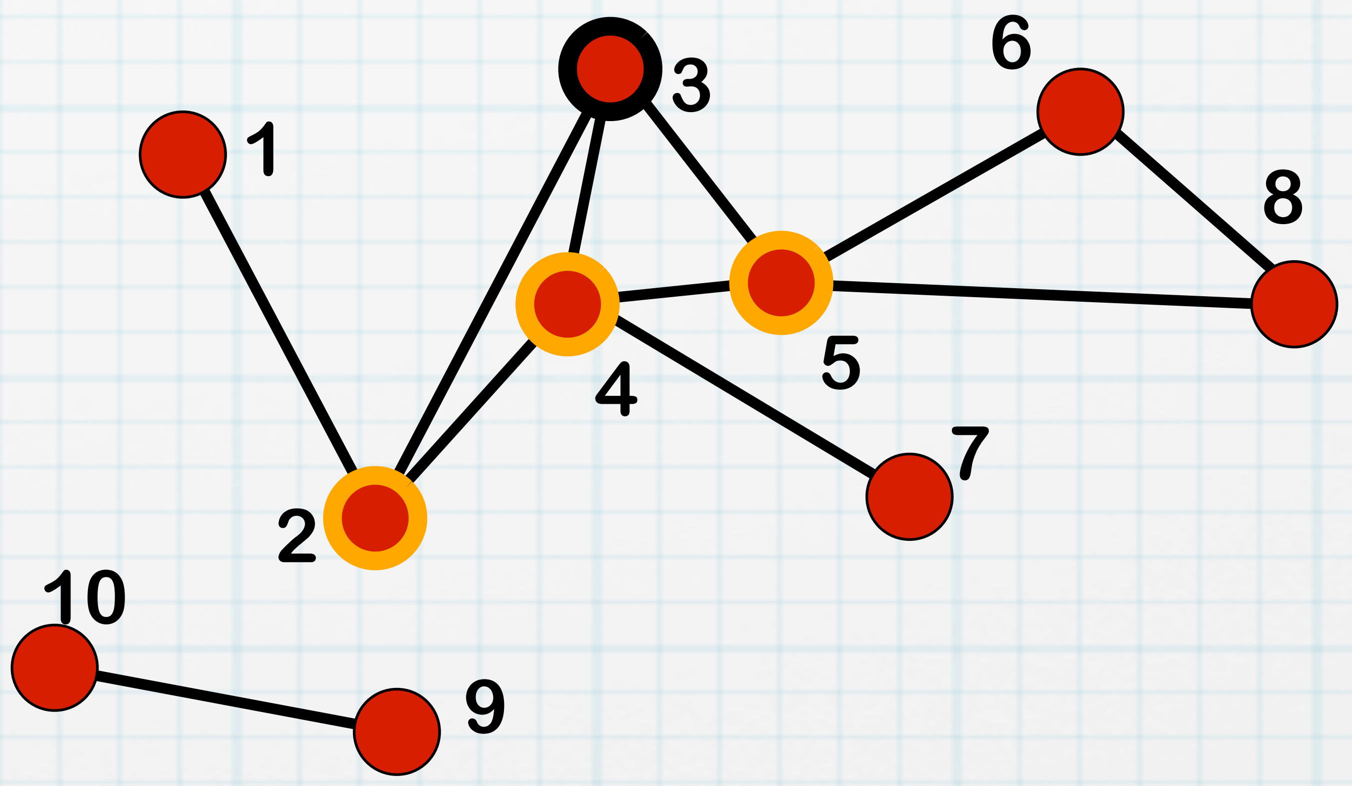



$i=3$: $k_3=3$ → $C_3$= ???

# Exercise

* Clustering coefficient: what portion of your neighbors are connected?
   * Node i with degree $k_i$

$$C_i = \frac{2e_i}{k_i(k_i - 1)}$$

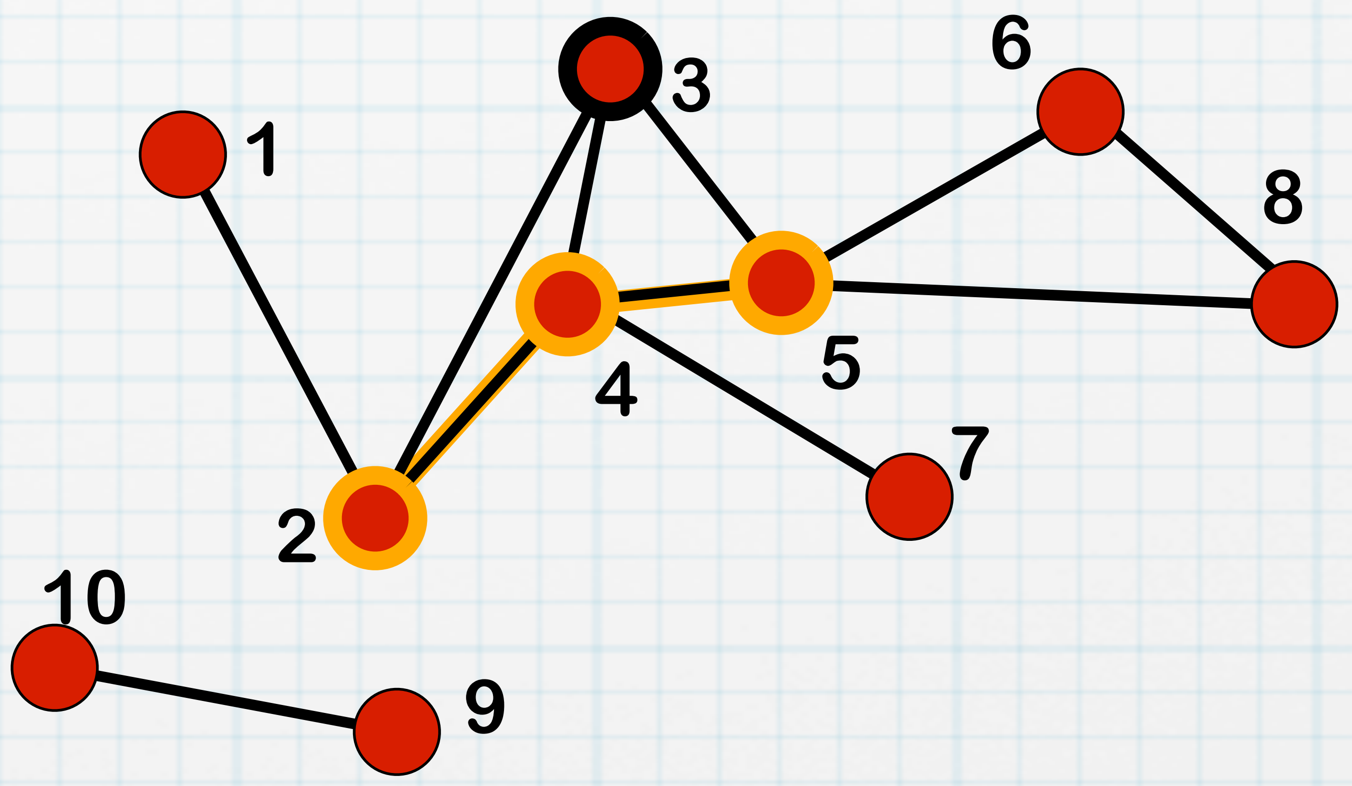



*i=3: $k_3$=3, $e_3$=2, MAX=3*2/2=3 → $C_3$= 2/3*

# Exercise

* Clustering coefficient: what portion of your neighbors are connected?
  * Node i with degree $k_i$

$$C_i = \frac{2e_i}{k_i(k_i - 1)}$$

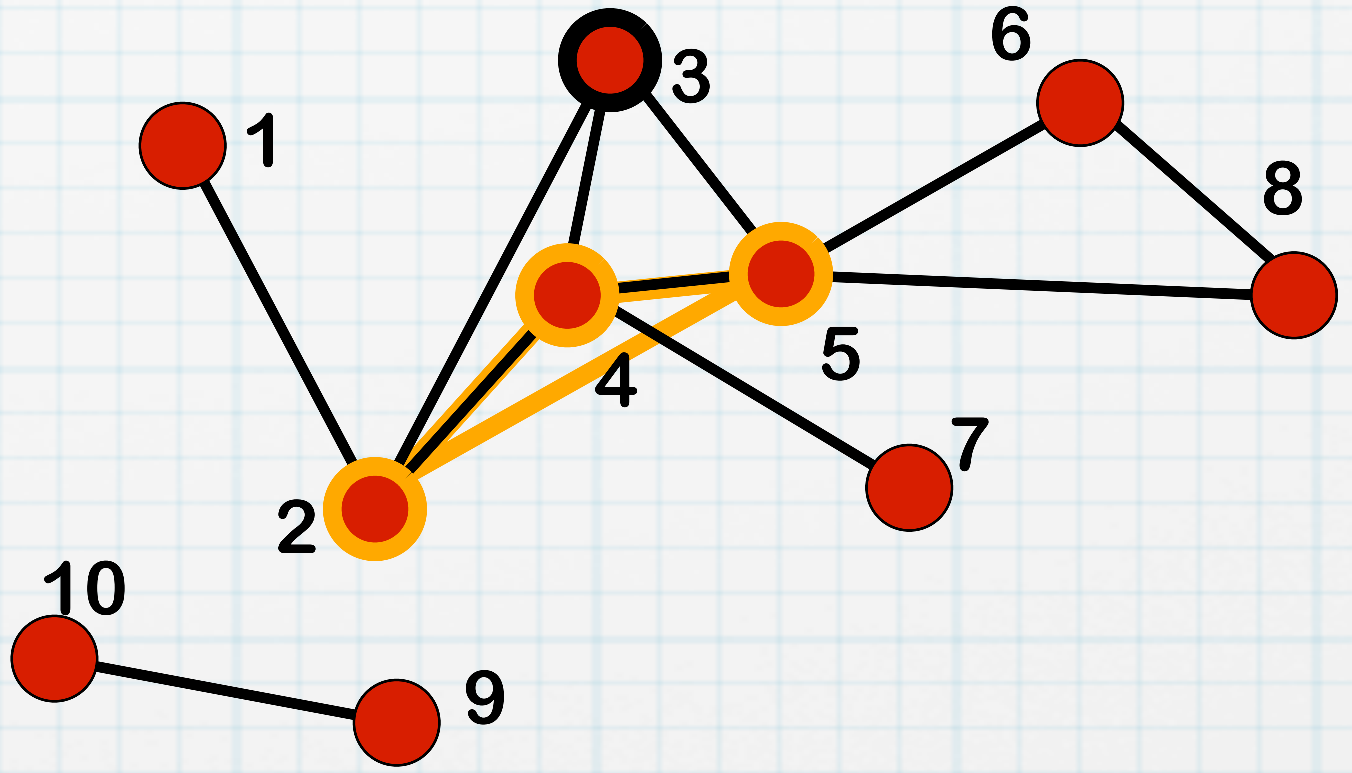



*i=3: $k_3$=3, $e_3$=2, MAX=3*2/2=3 → $C_3$= 2/3*

# Exercise

* Clustering coefficient: what portion of your neighbors are connected?
  * Node i with degree $k_i$

$$C_i = \frac{2e_i}{k_i(k_i - 1)}$$



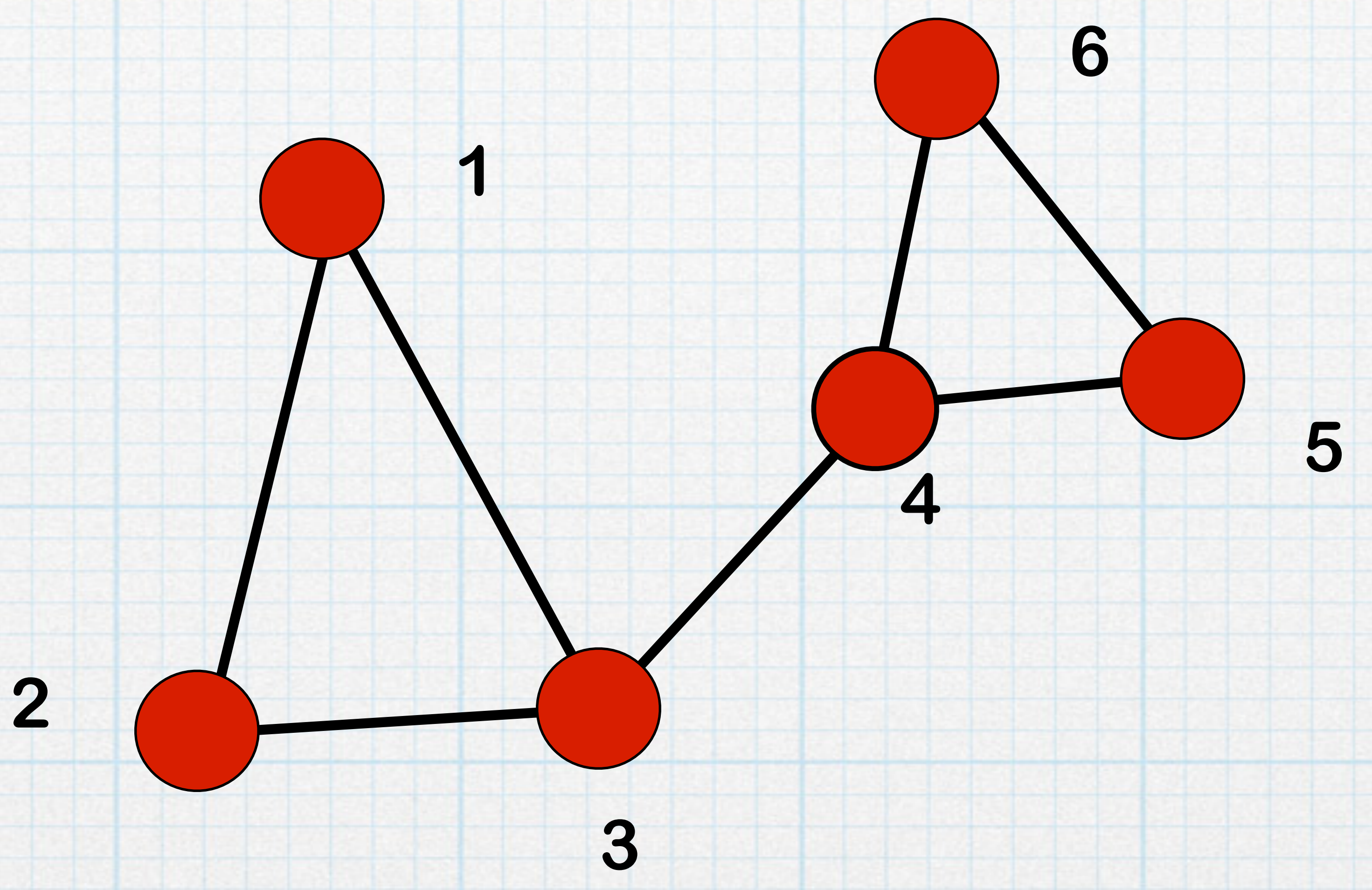

*i=3: $k_3$=3, $e_3$=2, MAX=3*2/2=3 → $C_3$= 2/3*

# Network stats review

* The network clustering coefficient is simply the average of the clustering coefficients of the individual nodes
* The average path length (apl) is the average of the (shortest) path lengths across the individual node pairs
    * How is this different from the diameter?

# Network statistics

* C=?
* L=?



# NetworkX

```
nx.has_path(G,source,target)

nx.shortest_path(G,source=s,target=t)
nx.shortest_path(G,source=s) # dict with all targets
nx.shortest_path(G) # dict for all sources and targets

nx.shortest_path_length(G,source=s,target=t)
nx.shortest_path_length(W,source=s,target=t,weight='weight')

nx.average_shortest_path_length(G)
nx.diameter(G)
```

```
nx.triangles(G) # dict w/no. triangles for each node

nx.clustering(G,node)
nx.clustering(G) # dict w/cc of each node

nx.average_clustering(G)
```

# Do social networks have high clustering?

* Facebook data:

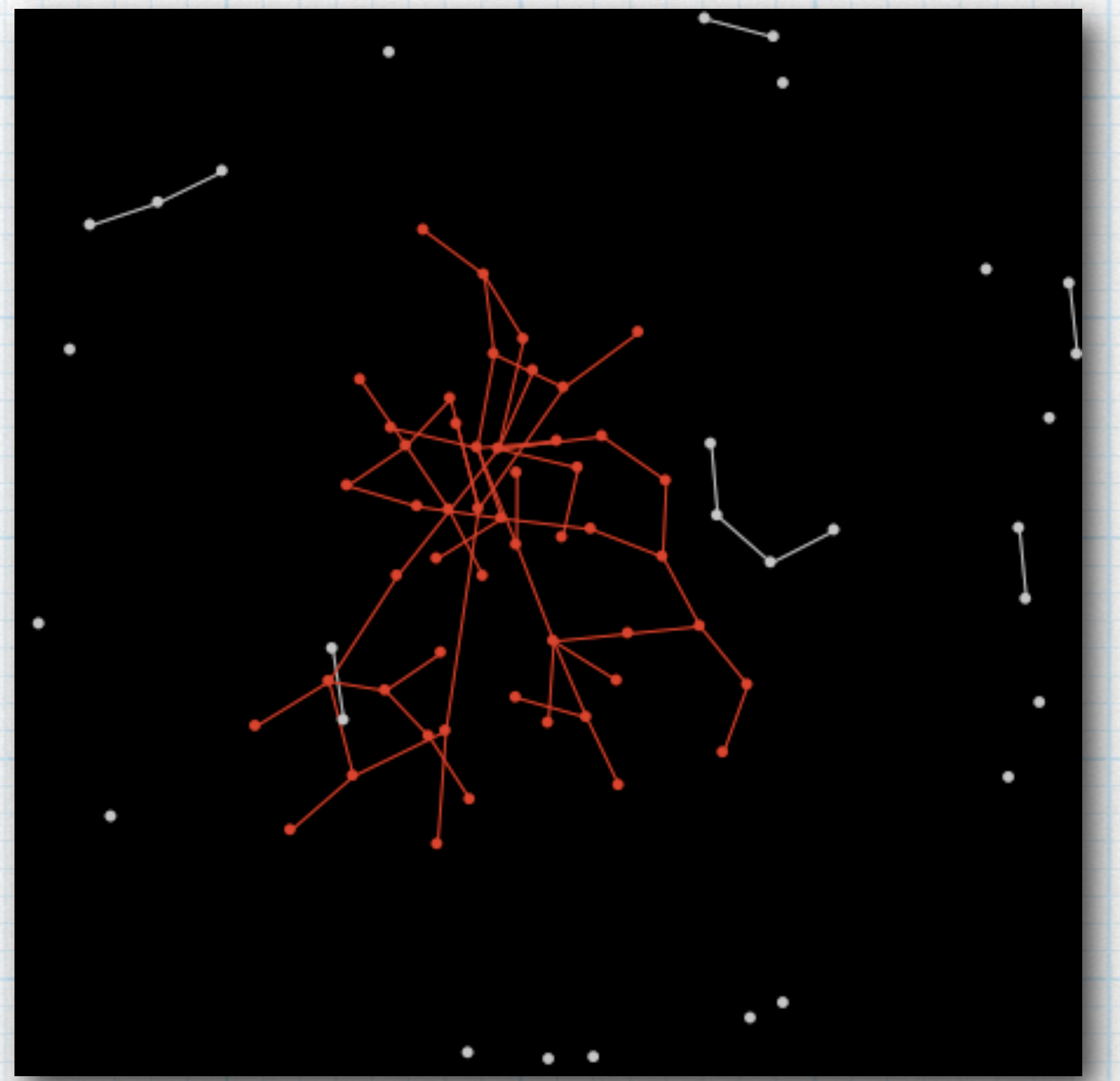
```
import networkx as nx
G = nx.read_edgelist('/Users/fil/I400/Data/facebook.txt')
print nx.average_clustering(G)
>>> 0.6055
```

* Coauthorship networks: $C > 0.5$

# Social networks are "small worlds"

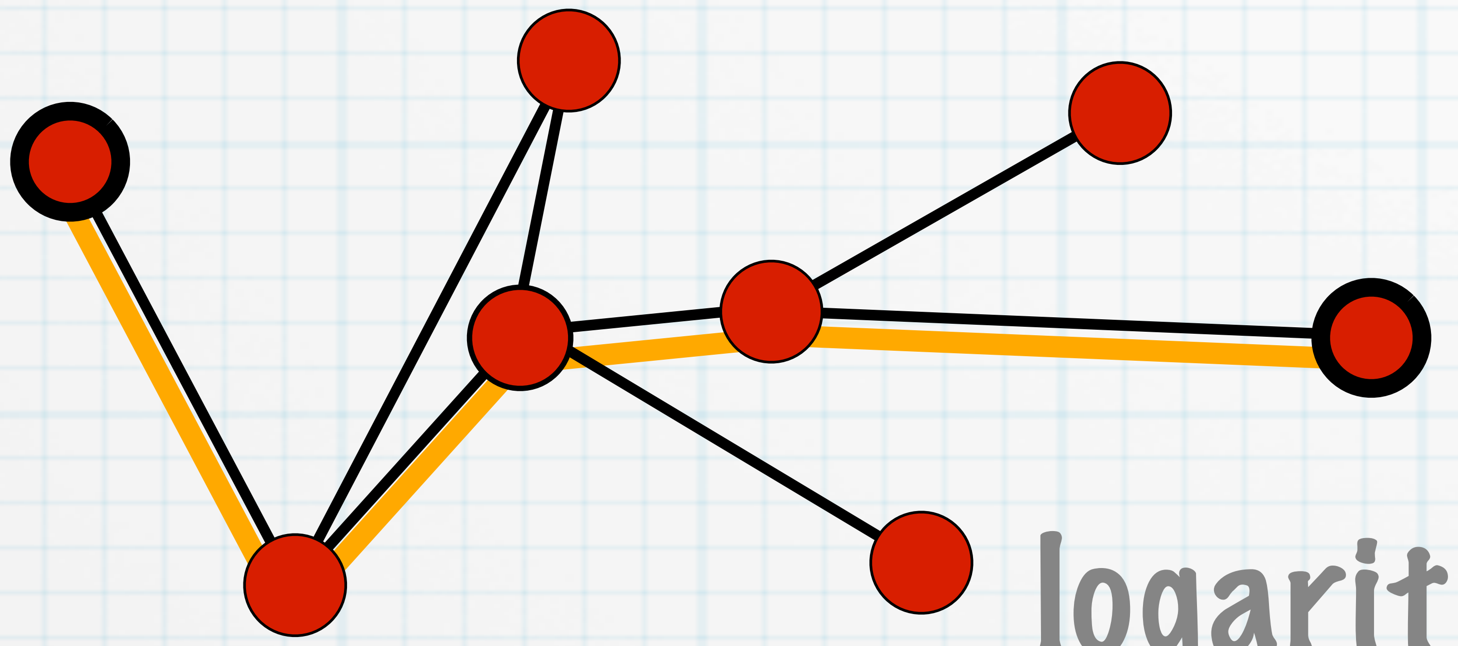
* Two properties:
  1. "small" diameter or apl
  2. "high" clustering coefficient

* Based on this definition, are random graphs "small worlds"?



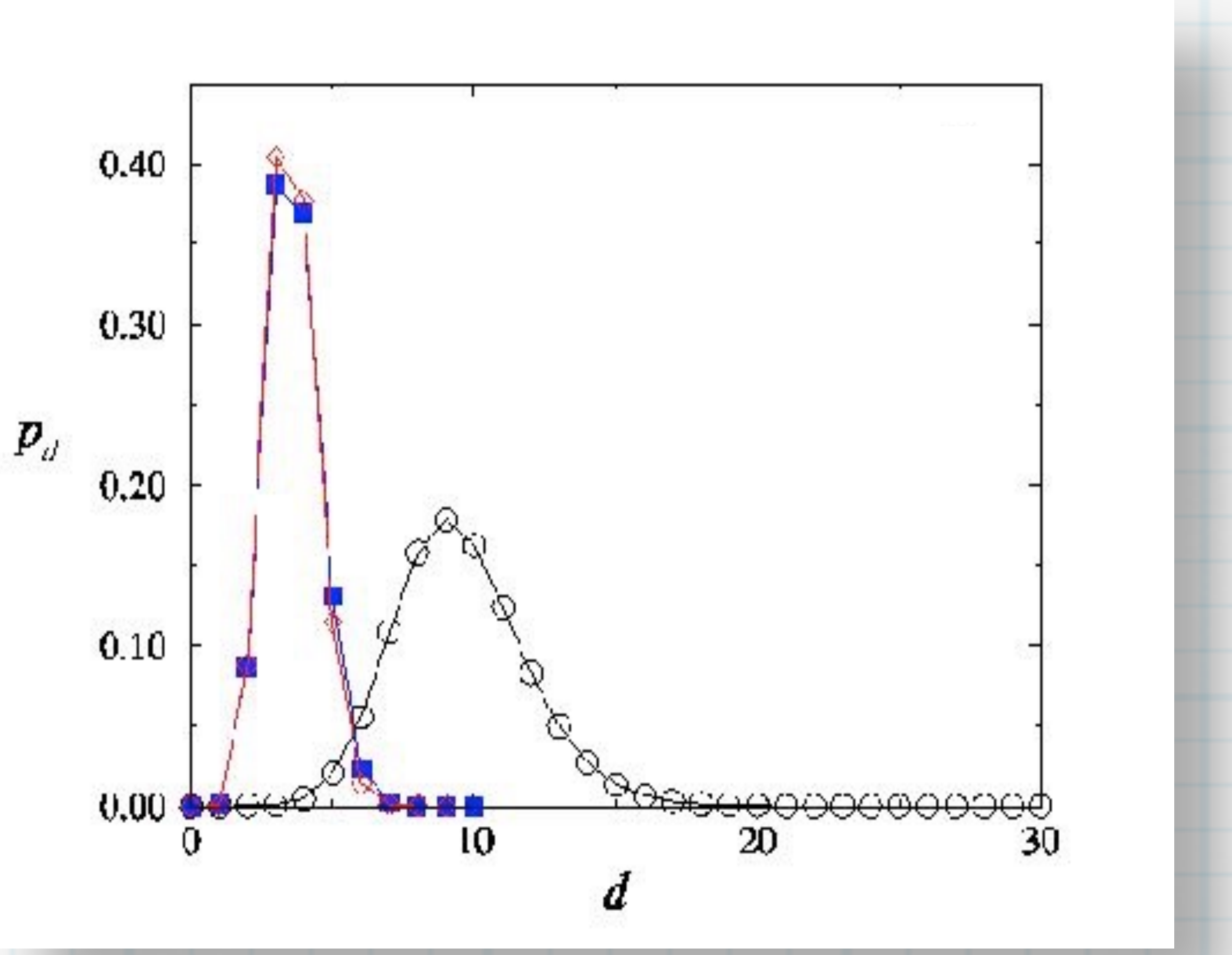
# Random Graphs (Erdös-Rényi model)

☑ Paths and average distance



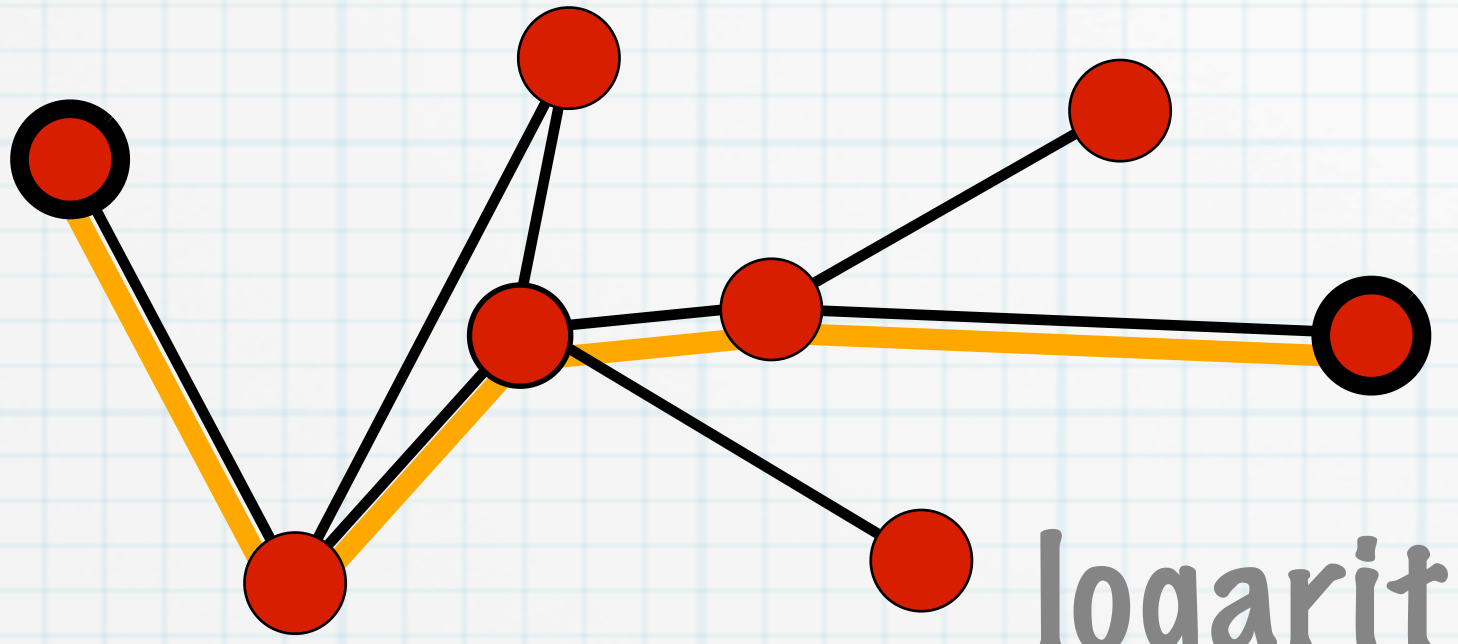
$$\langle \ell \rangle \simeq \frac{\log N}{\log\langle k \rangle}$$

logarithmically small

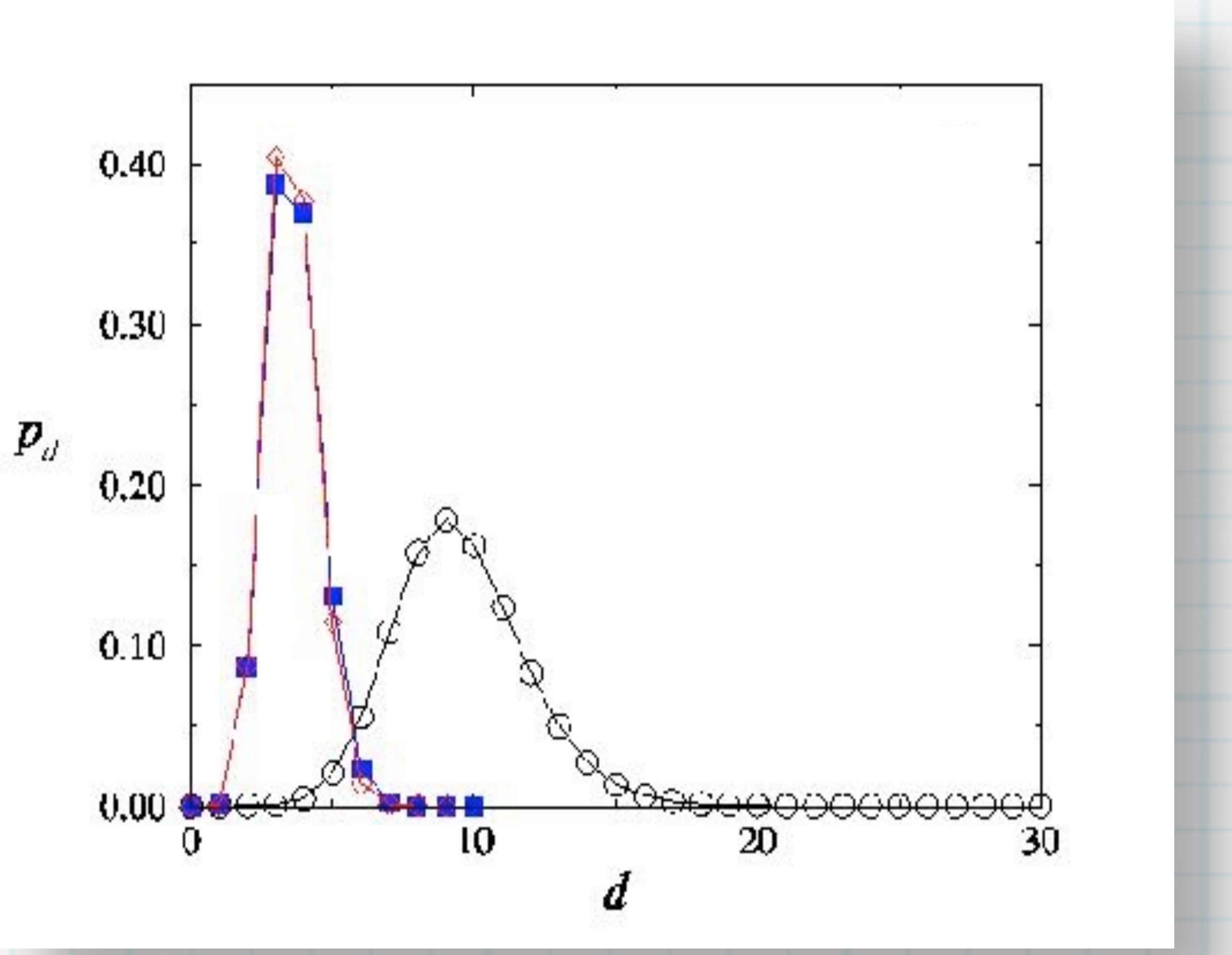


# Random Graphs (Erdös-Rényi model)

☑ Paths and average distance



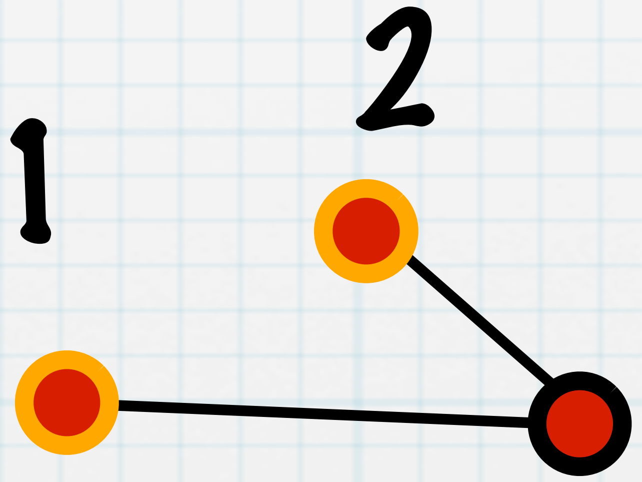
$$\langle \ell \rangle \simeq \frac{\log N}{\log \langle k \rangle}$$




logarithmically small

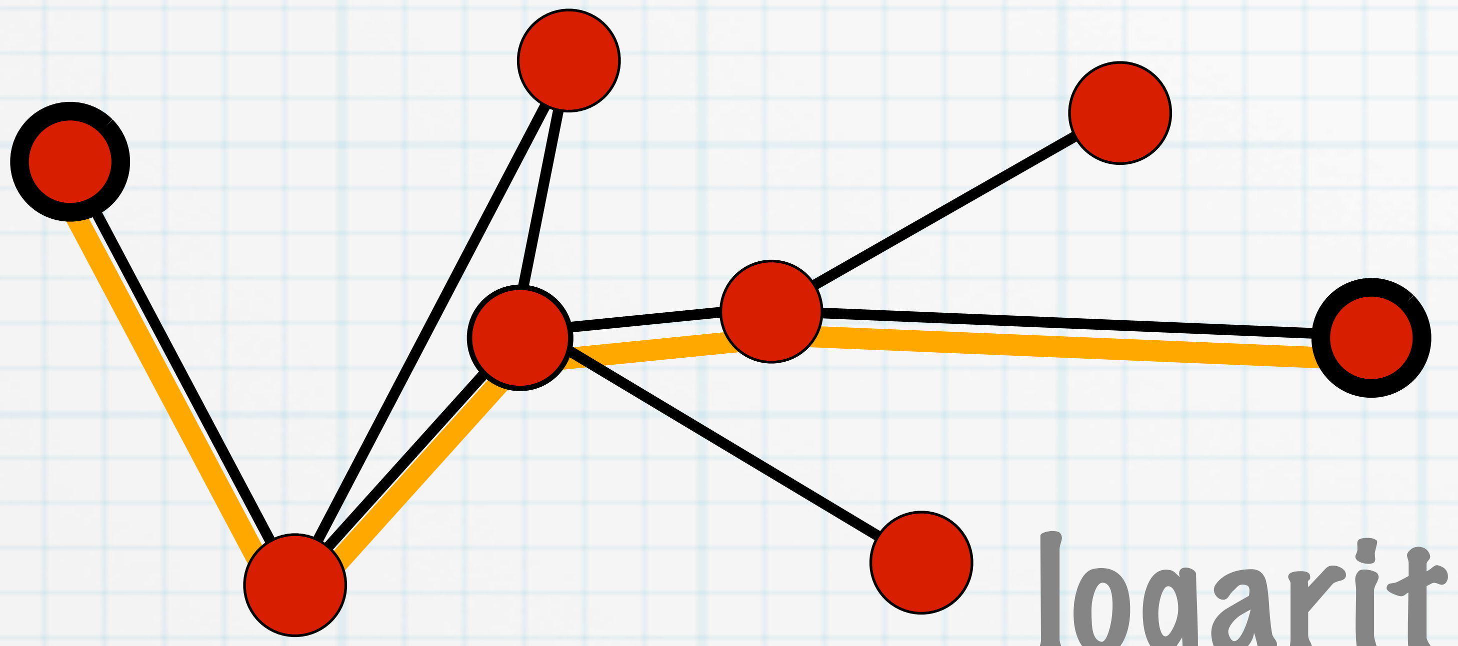
☐ Clustering coefficient

What is the probability that nodes 1 and 2 are linked?

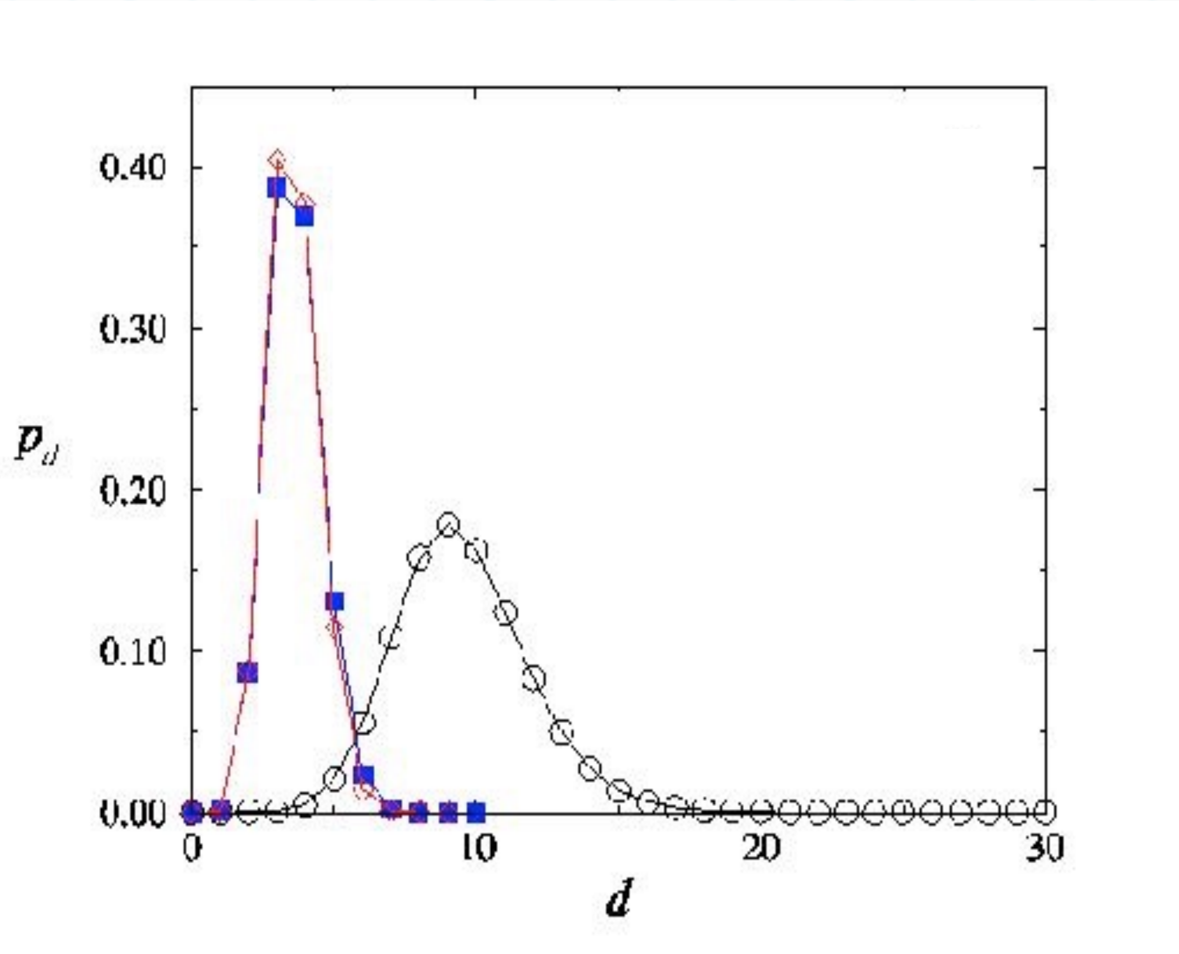


# Random Graphs (Erdös-Rényi model)

☑ Paths and average distance



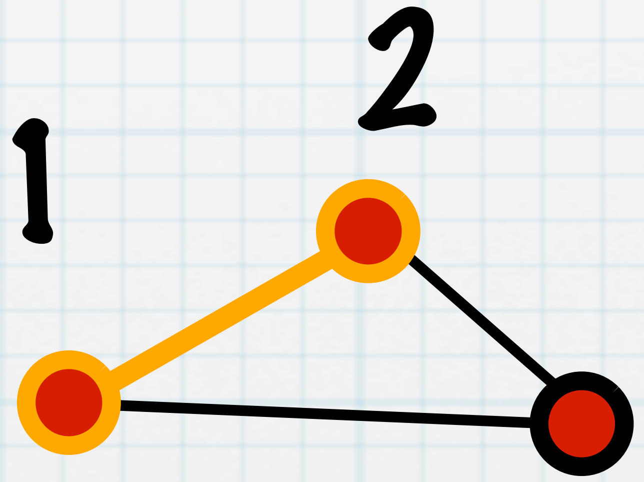
$$\langle \ell \rangle \simeq \frac{\log N}{\log \langle k \rangle}$$




logarithmically small

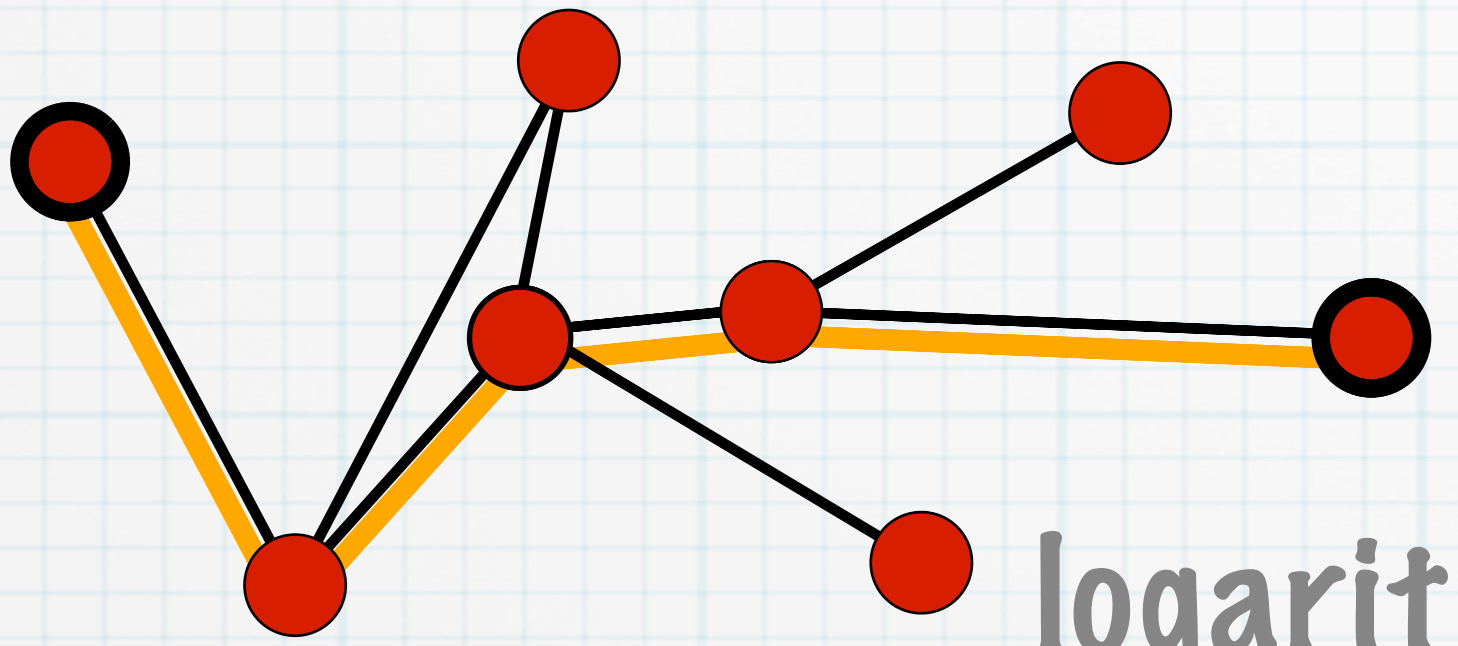
☐ Clustering coefficient

What is the probability that nodes 1 and 2 are linked?

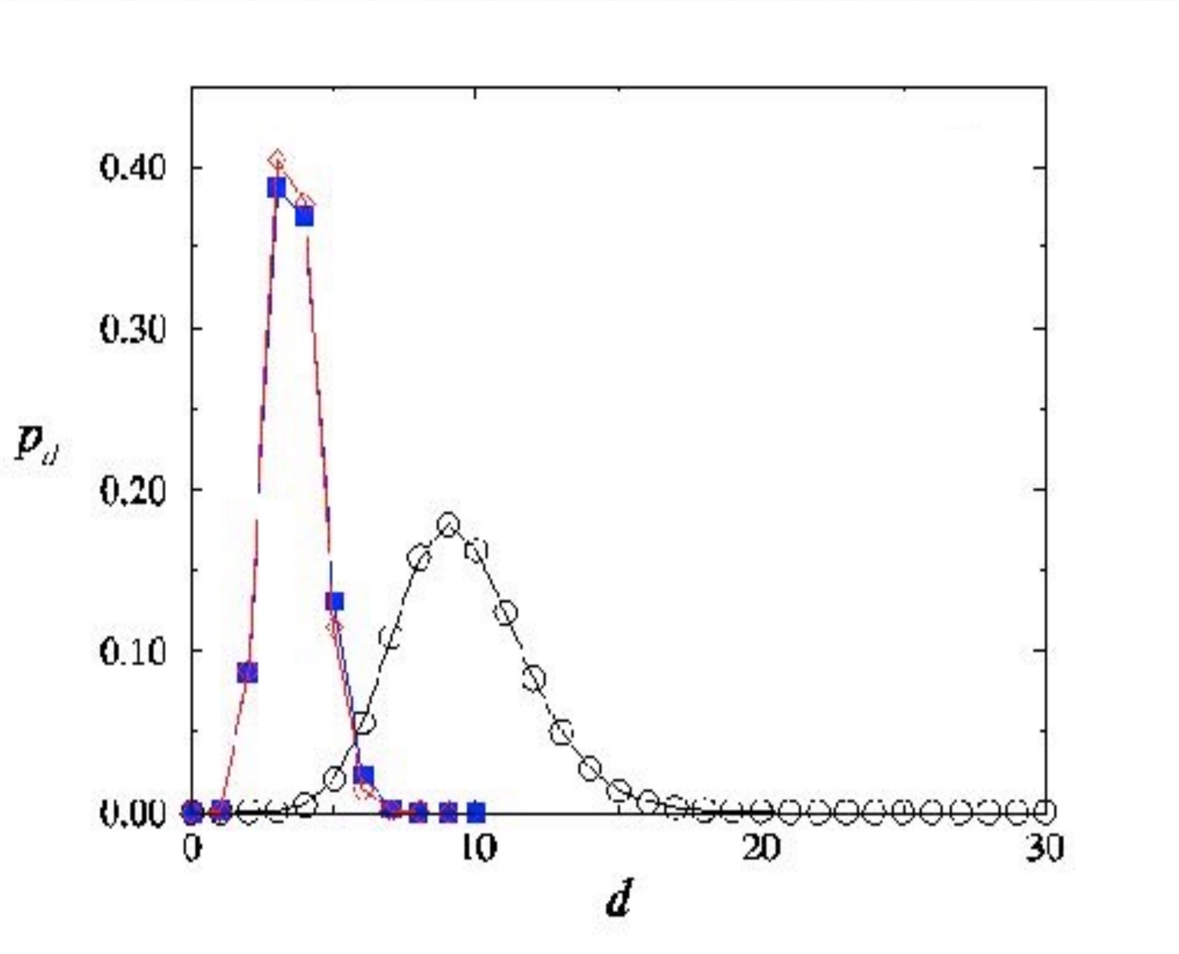


# Random Graphs (Erdös-Rényi model)

## ☑ Paths and average distance



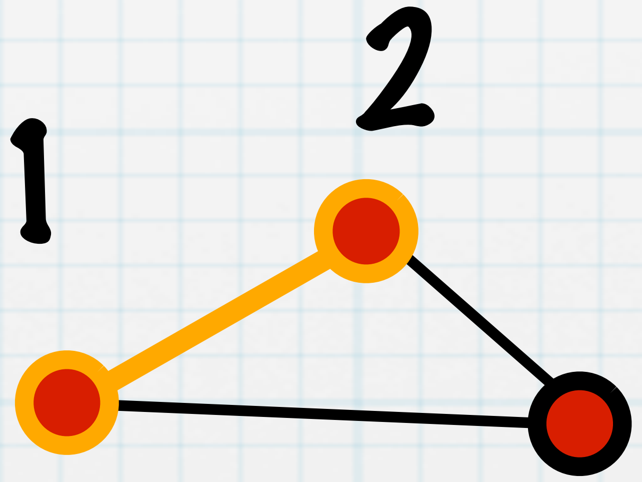
$$\langle \ell \rangle \simeq \frac{\log N}{\log \langle k \rangle}$$




logarithmically small

## ☐ Clustering coefficient

What is the probability that nodes 1 and 2 are linked?



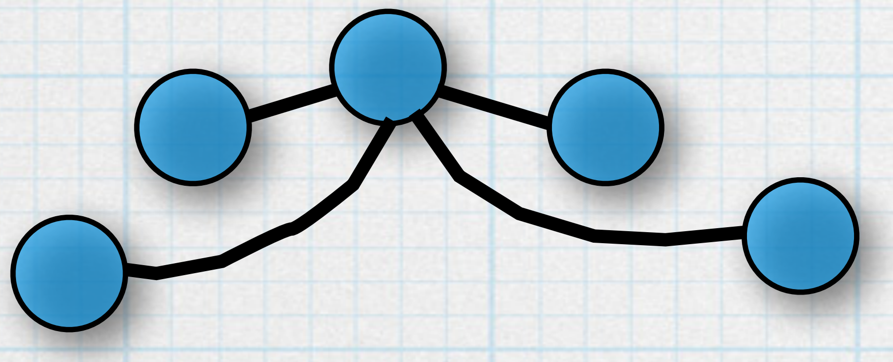

C = p = <k>/(N-1)

vanishingly small for large network sizes

# What have we learned?

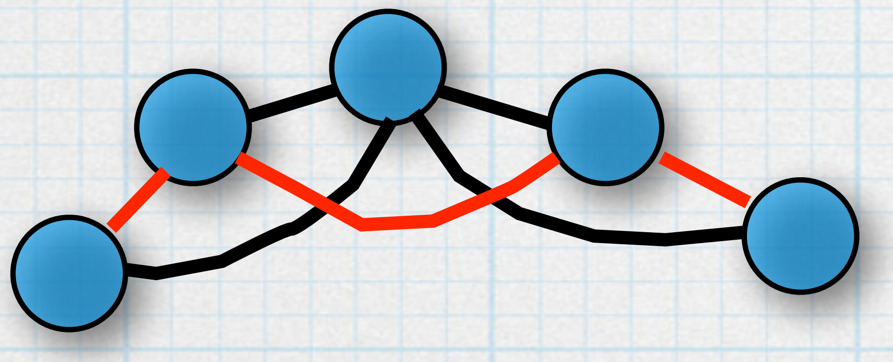
- Social (and many other) networks have both short paths and high clustering
- Random networks have short paths but not high clustering
  - If coauthorship networks were random graphs, they would have:
    $C = p = \langle k \rangle/(N-1) \approx 10^{-5} \ll 0.5$
- So, how can we reproduce both?

# Watts-Strogatz "Small-World" model



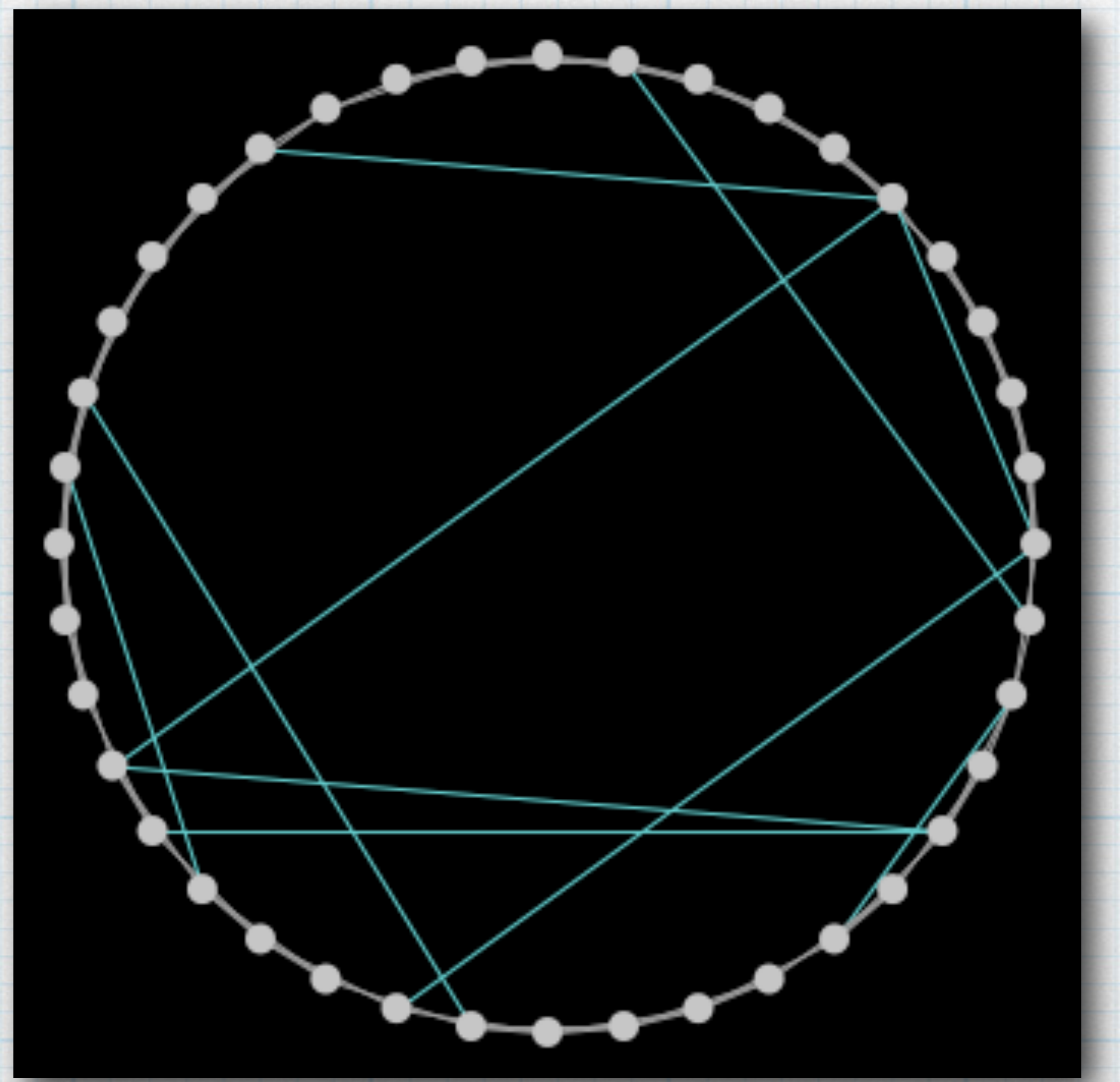
* Start with a "regular network" (lattice)
  * Each node linked to, say, 4 neighbors on a line
  * High clustering coefficient: C=?
* Pick a random link and rewire it at random with probability p; repeat...

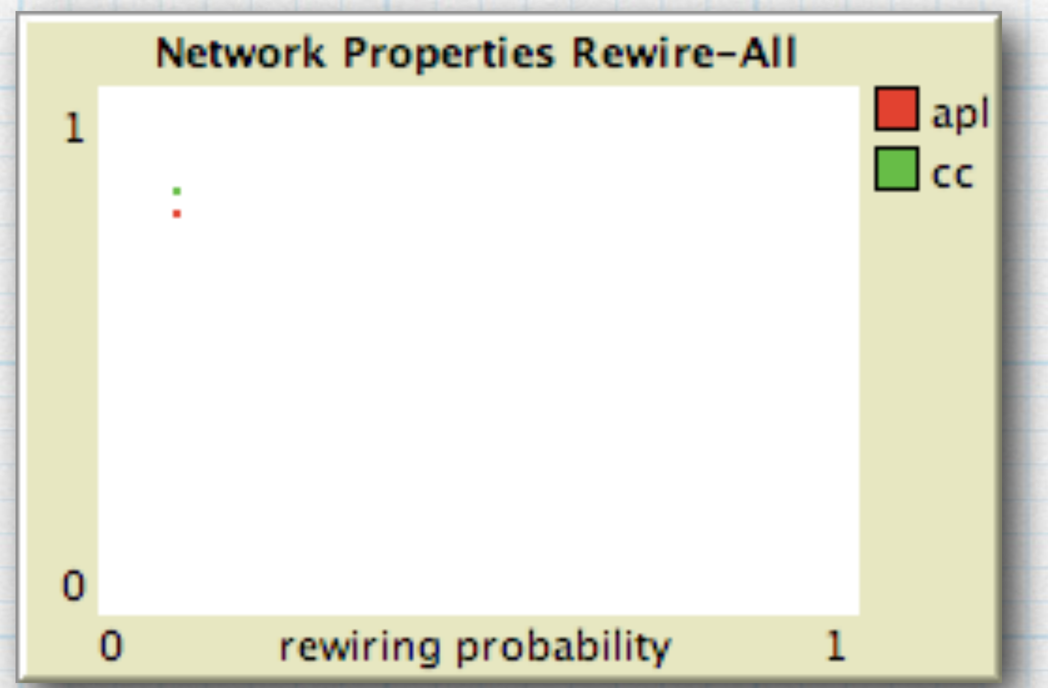
# Watts-Strogatz "Small-World" model



* Start with a “regular network” (lattice)
  * Each node linked to, say, 4 neighbors on a line
  * High clustering coefficient: C=?
* Pick a random link and rewire it at random with probability p; repeat...

* What happens to the diameter and apl as we rewire edges?
* What happens to the clustering coefficient?

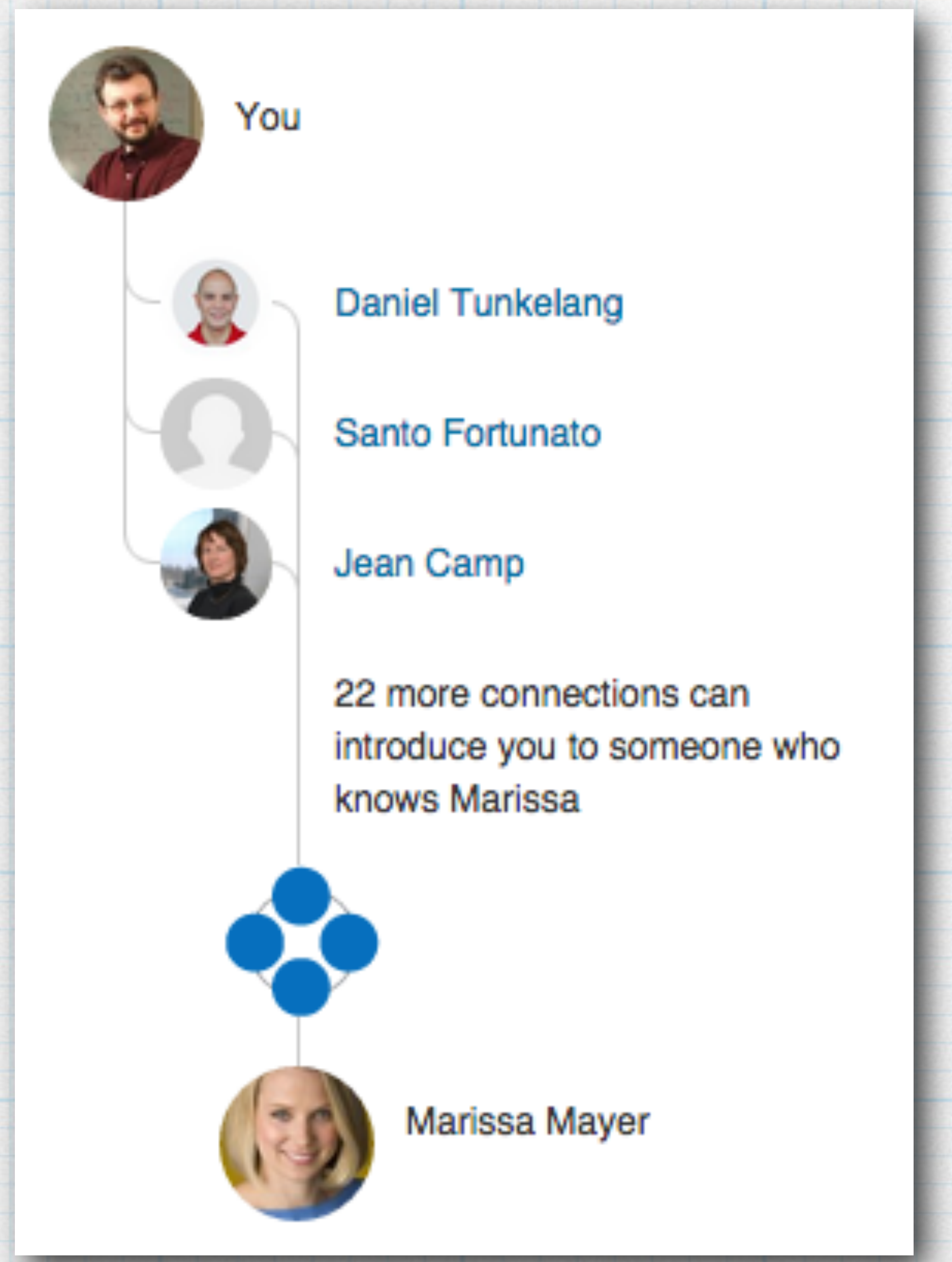





# Let's play NetLogo!

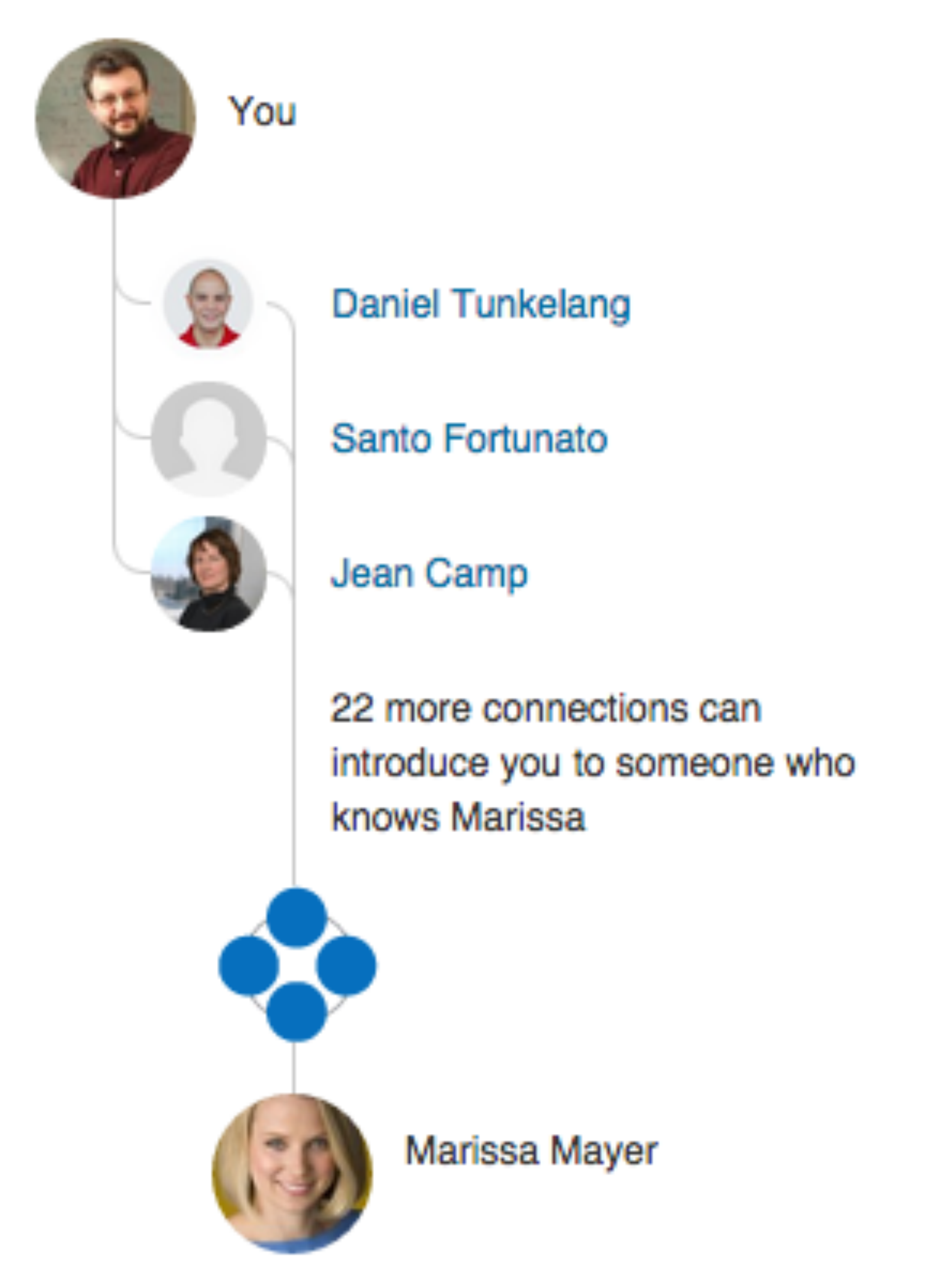
Models Library : Networks: Small Worlds

# Search & navigation in (social) networks




* How do we find a short path between two actors to play six degrees of Kevin Bacon?
* How do we determine a scholar's Erdös Number?
* How did the subjects in the Migram and Watts experiments forward the messages?
* How do Web crawlers find every page?
* How do we find a file in a peer network?

# Path-finding in networks




* Many different algorithms based on:
  * application
  * network structure and size
  * information about nodes, e.g., geography, homophily...
* We will discuss the simplest algorithm (best-first search) later in the course