

# Network Analysis:

The Hidden Structures behind the Webs We Weave

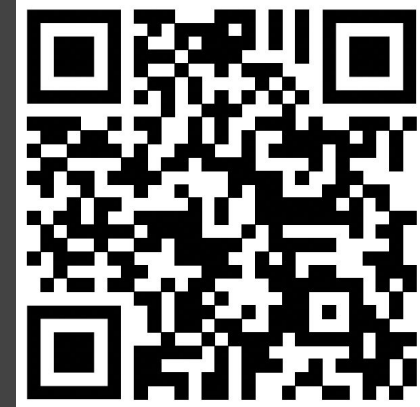
17-213 / 17-668

Connectedness, Random Networks, Triads

Tuesday, September 5, 2023

Patrick Park & Bogdan Vasilescu

2-min Quiz, on Canvas



# Quick Recap – Last Thursday's Lecture

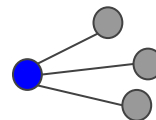
Graph theory as our basic formalism for modeling networks

Basic building blocks: nodes and links



Most basic structure: dyads

Degree and degree distribution



Paths (shortest paths)

The Breadth-first search algorithm to compute distances

Adjacency matrices as an algebraic representation of networks

Network properties as matrix operations!

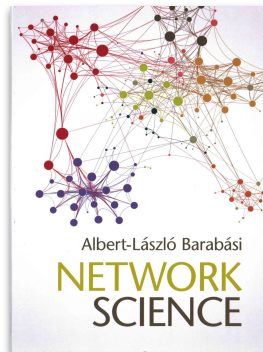
# Plan for Today

More on connectedness and connected components

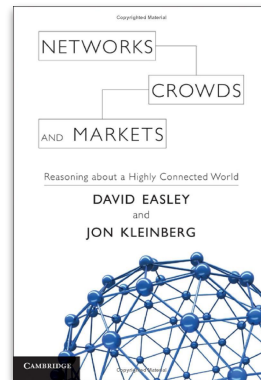
Random graphs, revisiting Six Degrees of Kevin Bacon

Larger building blocks: from dyads to triads

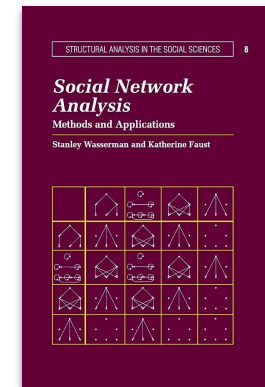
(B Ch. 2.9–2.10, Ch. 3 except 3.9)



(K&E Ch. 4)



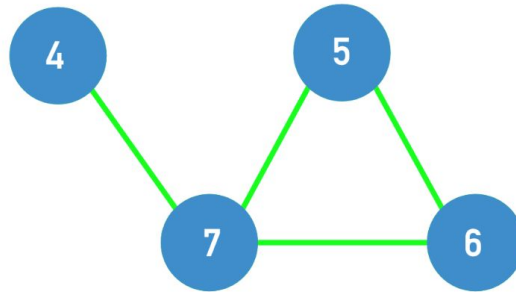
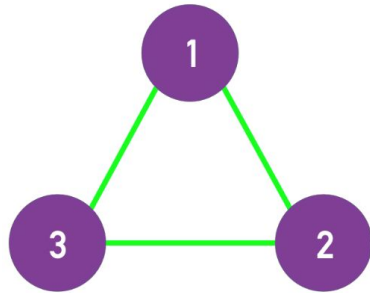
(W&F Ch. 14.1-14.2)



# Connectedness

# In a “Connected” Graph, There Is a Path Between Every Pair of Nodes

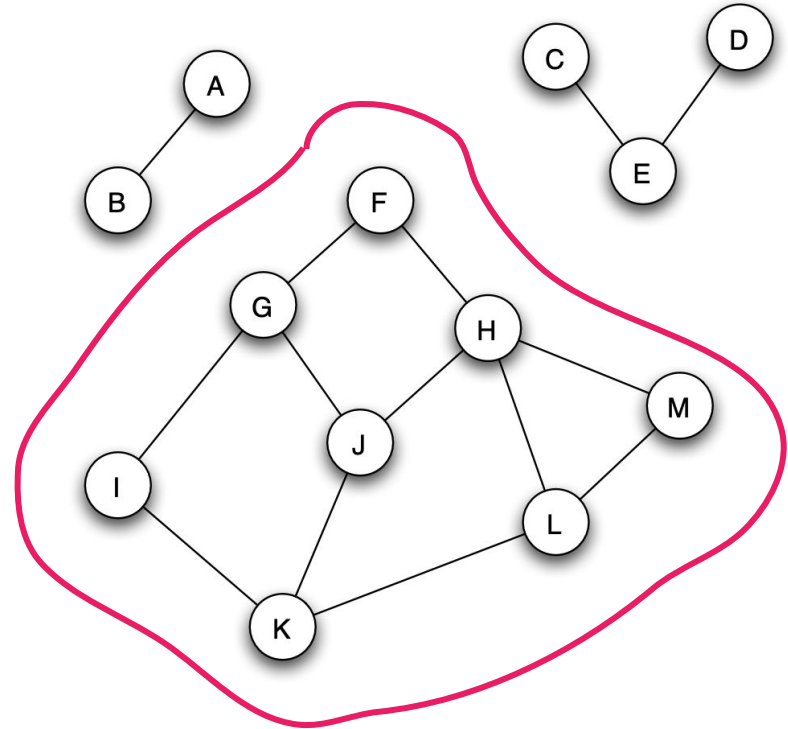
This example shows two disconnected components. If a network has disconnected components, the adjacency matrix (right) can be rearranged into a block diagonal form.



$$\begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix}$$

# When a Network Contains a Giant Component, It Almost Always Contains Only One

Why?



# When a Network Contains a Giant Component, It Almost Always Contains Only One

Imagine there were two giant components in the global friendship network example, each with hundreds of millions of people.

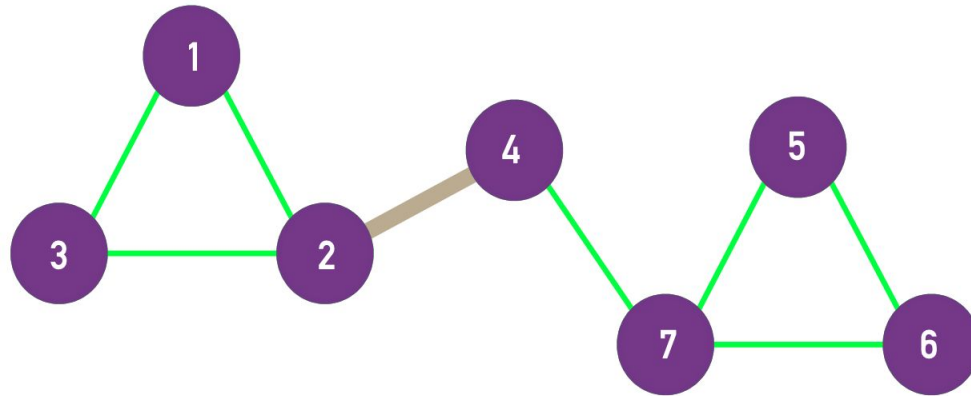
All it would take is a single edge from someone in the first of these components to someone in the second, and the two giant components would merge into a single component!

It's essentially inconceivable that some such edge wouldn't form, and hence two co-existing giant components are almost never seen in real networks.



# A “Bridge” (2–4) Can Turn a Disconnected Network Into a Single Connected Component.

Note: The adjacency matrix cannot be written in a block diagonal form.



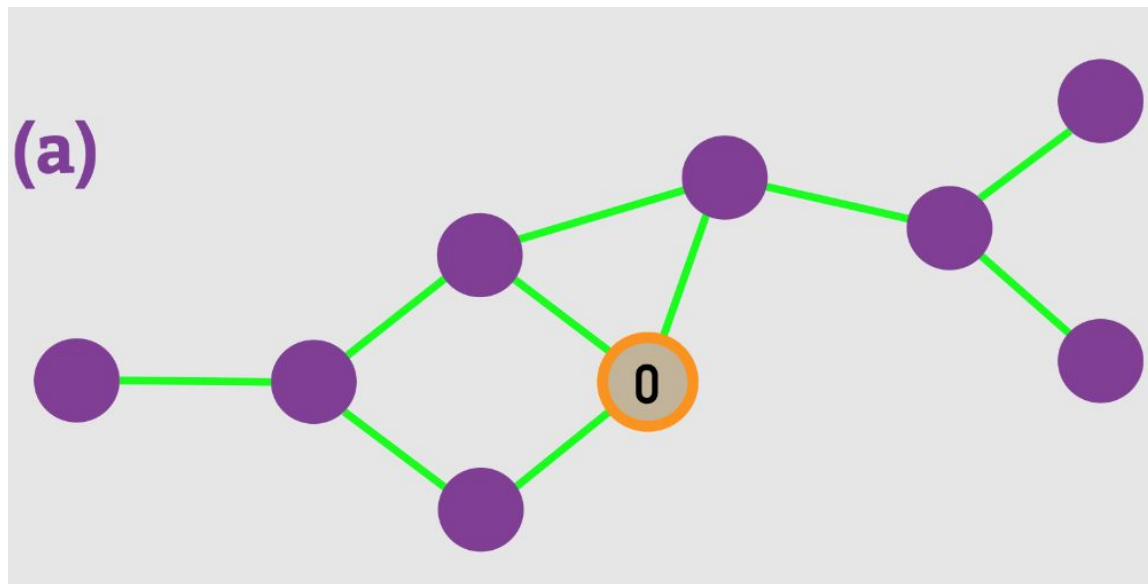
$$\begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix}$$

(Barabasi, 2016)

# Recall the BFS Algorithm

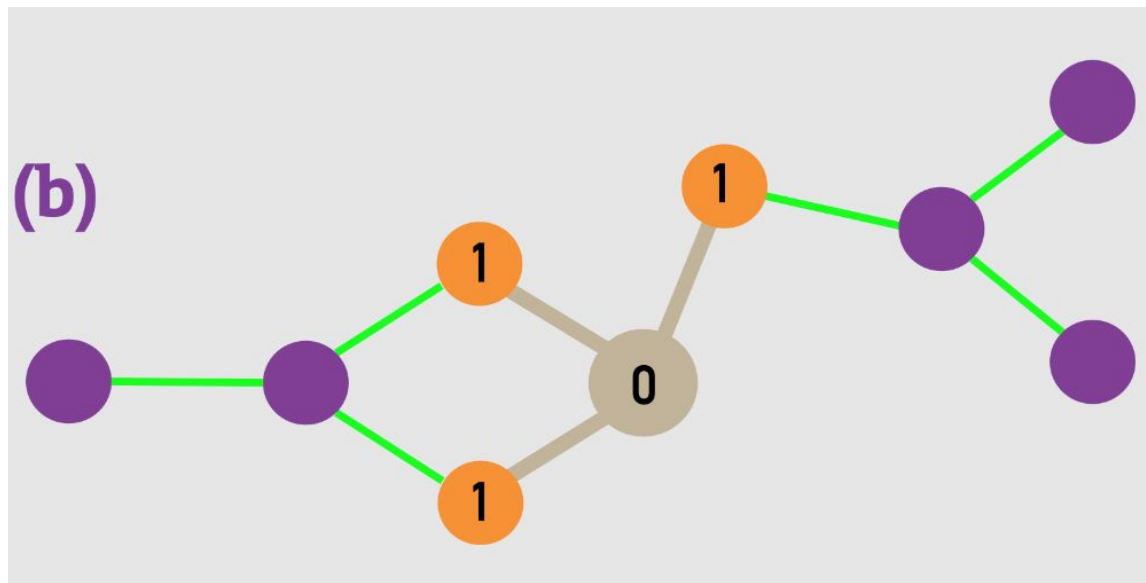
Assume we're starting from the orange node, labeled "0."

First, we identify all its neighbors, labeling them "1".



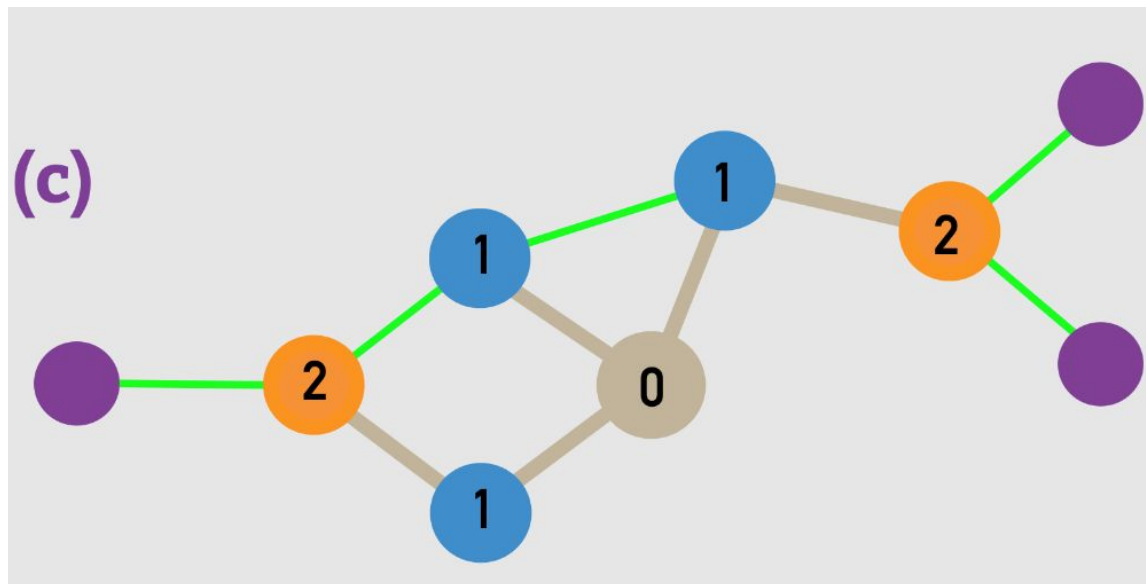
# Recall the BFS Algorithm

Next we label “2” the unlabeled neighbors of all nodes labeled “1”, and so on, in each iteration increasing the label number, until no node is left unlabeled.



# Recall the BFS Algorithm

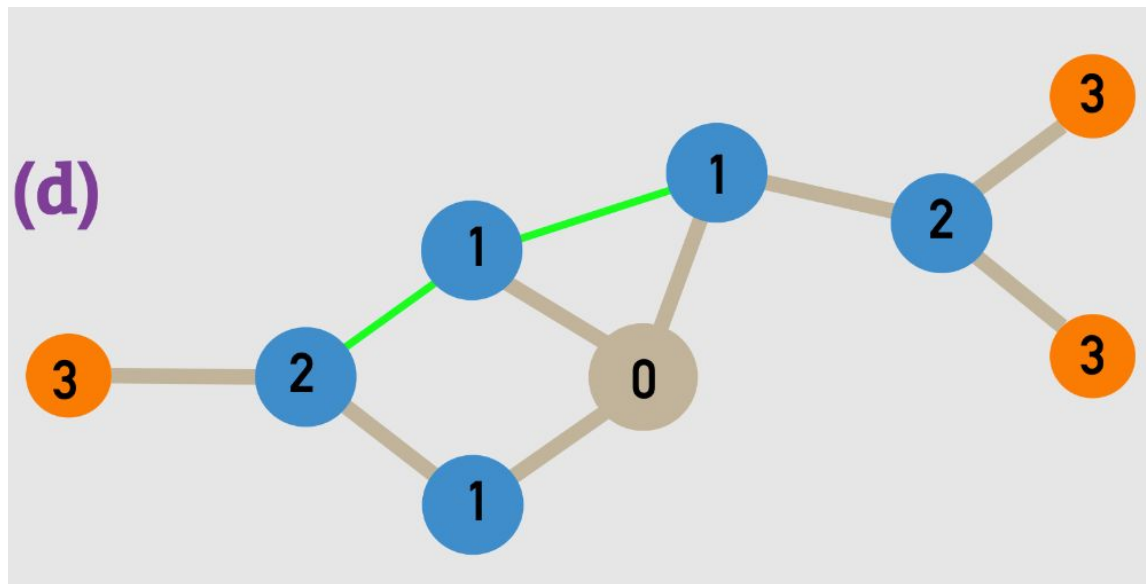
Next we label “2” the unlabeled neighbors of all nodes labeled “1”, and so on, in each iteration increasing the label number, until no node is left unlabeled.



# Recall the BFS Algorithm

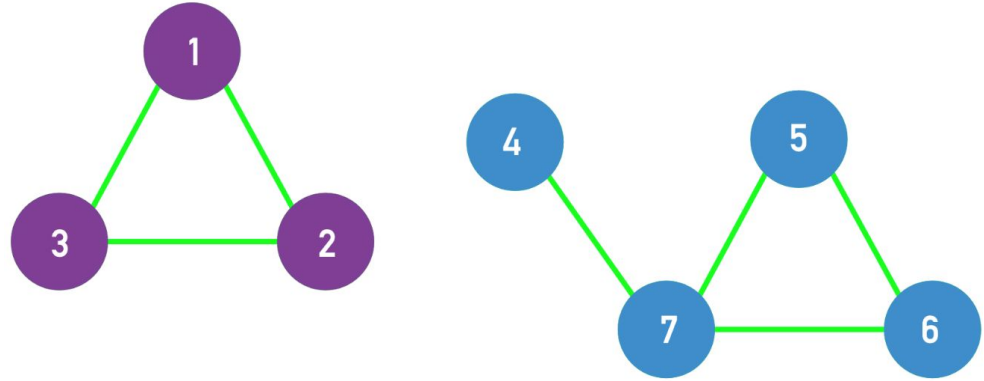
Ultimately, the length of the shortest path (or the distance  $d_{0i}$  between node 0 and any other node  $i$  in the network is given by the label of node  $i$ .

For example, the distance between node 0 and the leftmost node is  $d = 3$ .



(Barabasi, 2016)

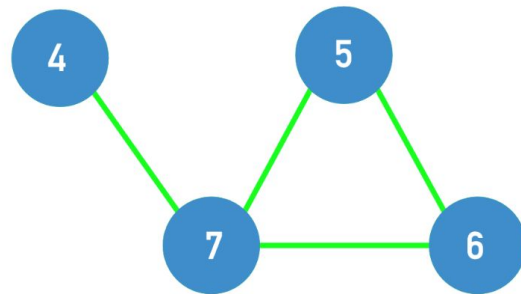
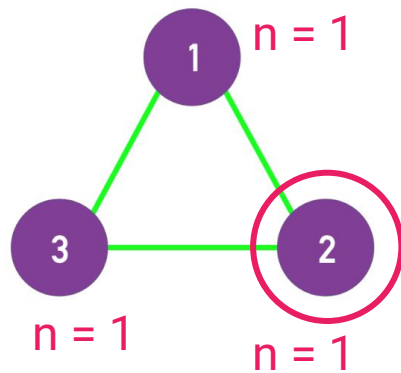
# Can We Identify Connected Components Using BFS?



(Barabasi, 2016)

# We Can Identify Connected Components Using BFS!

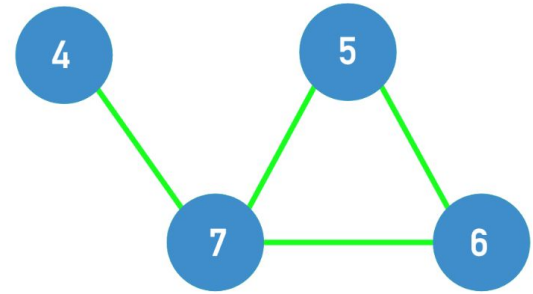
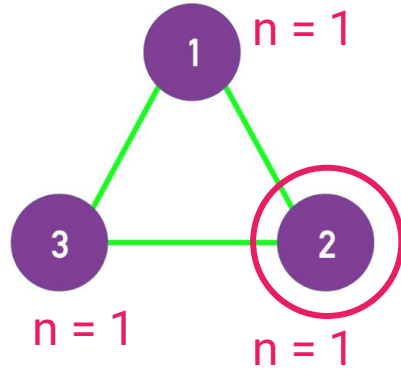
(1) Start from a randomly chosen node  $i$  and perform a BFS. Label all nodes reached this way with  $n = 1$ .



# We Can Identify Connected Components Using BFS!

(2) If the total number of labeled nodes equals  $N$ , then the network is connected.

If the number of labeled nodes is smaller than  $N$ , the network consists of several components.





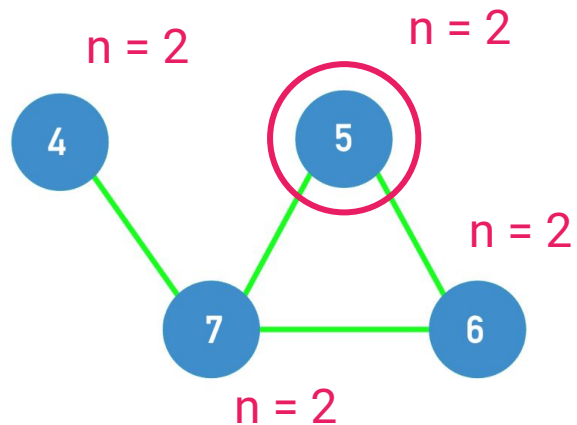
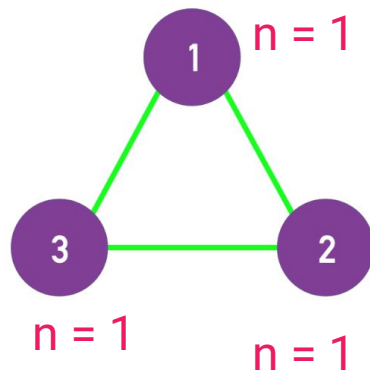
# We Can Identify Connected Components Using BFS!

(3) Increase the label  $n \rightarrow n + 1$ .

Choose an unmarked node  $j$ ,  
label it with  $n$ .

Use BFS to find all nodes  
reachable from  $j$ , label them all  
with  $n$ .

Return to step 2.



# Case Study Scenario

(Barabasi, 2016)

Imagine organizing a party for a hundred guests who initially do not know each other.

You bought a couple of exquisite cakes from La Gourmandine, plus many fillers from the grocery store.

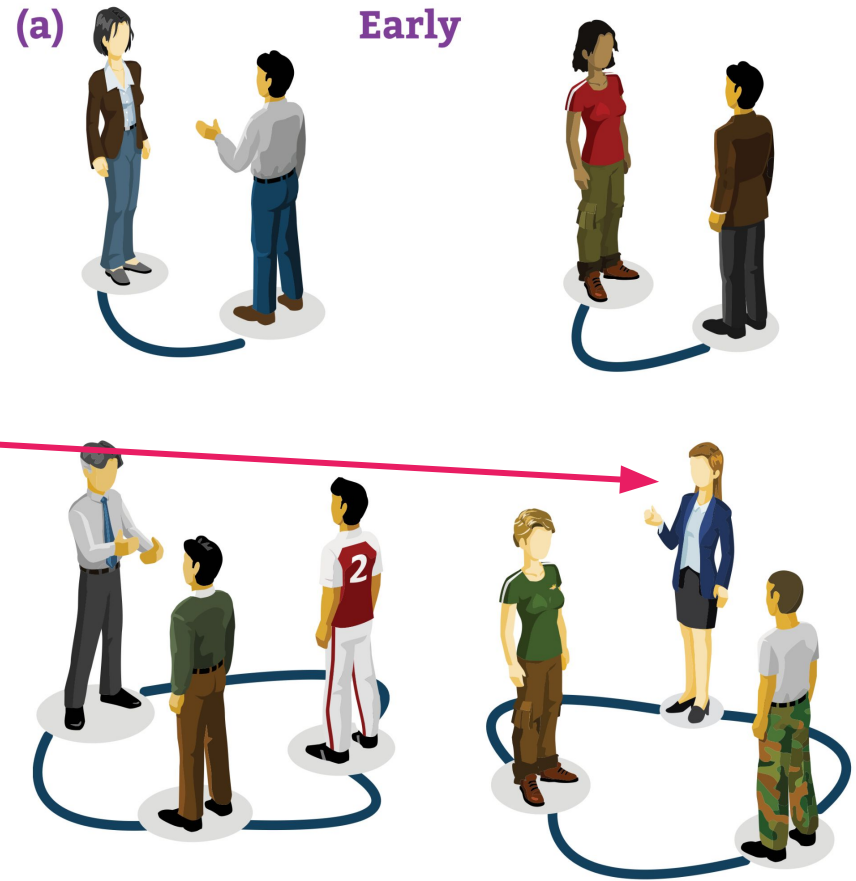
Your guests don't know about the La Gourmandine gems.



People come and start chatting, in small groups.

Now mention to Mary, one of your guests, which cakes came from La Gourmandine.

If she shares this info only with her acquaintances, your expensive cake appears to be safe; she only had time to meet a few others so far.



Early on the guests form isolated groups.

As time goes on, the guests will mingle, becoming increasingly interwoven by subtle paths between them.

How long before you run out of premium cake?



As people mingle, changing groups, an invisible network emerges that connects everyone into a single network.<sup>21</sup>

Clearly, after all guests get to know each other, everyone would be eating the superior cake.

But if each encounter took only ten minutes, meeting all ninety-nine others would take ~16h.

Thus, you could reasonably hope that a few pieces of your premium cake would be left for you to enjoy once the guests are gone.



As people mingle, changing groups, an invisible network emerges that connects everyone into a single network.<sup>22</sup>

What if I told you we don't have to wait until *all* individuals get to know each other for our expensive cake to be in danger?

Rather, soon after each person meets at least one other guest, an invisible network will emerge that will allow the information to reach all of them.

Hence in no time everyone will be enjoying the better cake!



(b)

Later



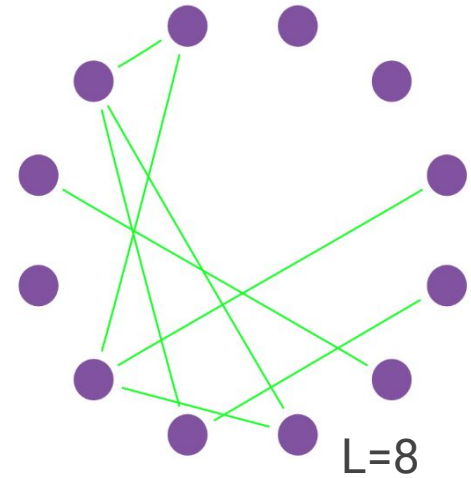
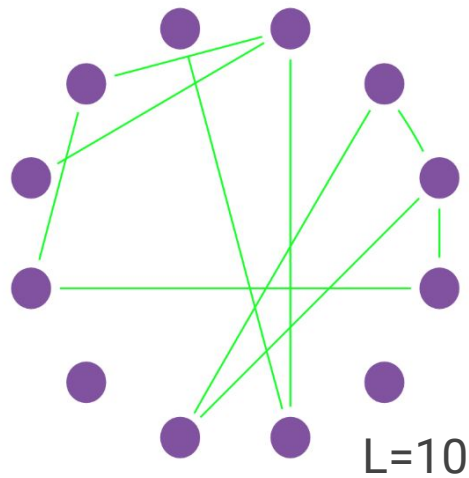
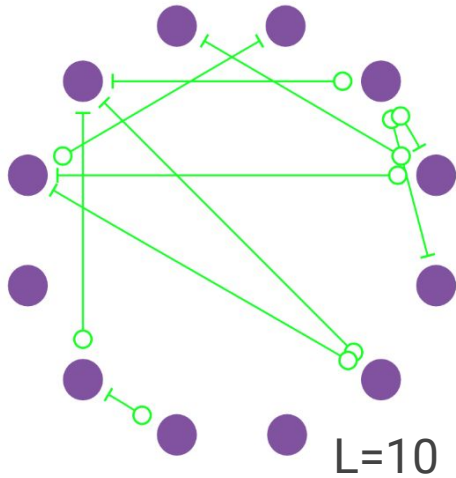
As people mingle, changing groups, an invisible network emerges that connects everyone into a single network.<sup>23</sup>

# The Random Network Model



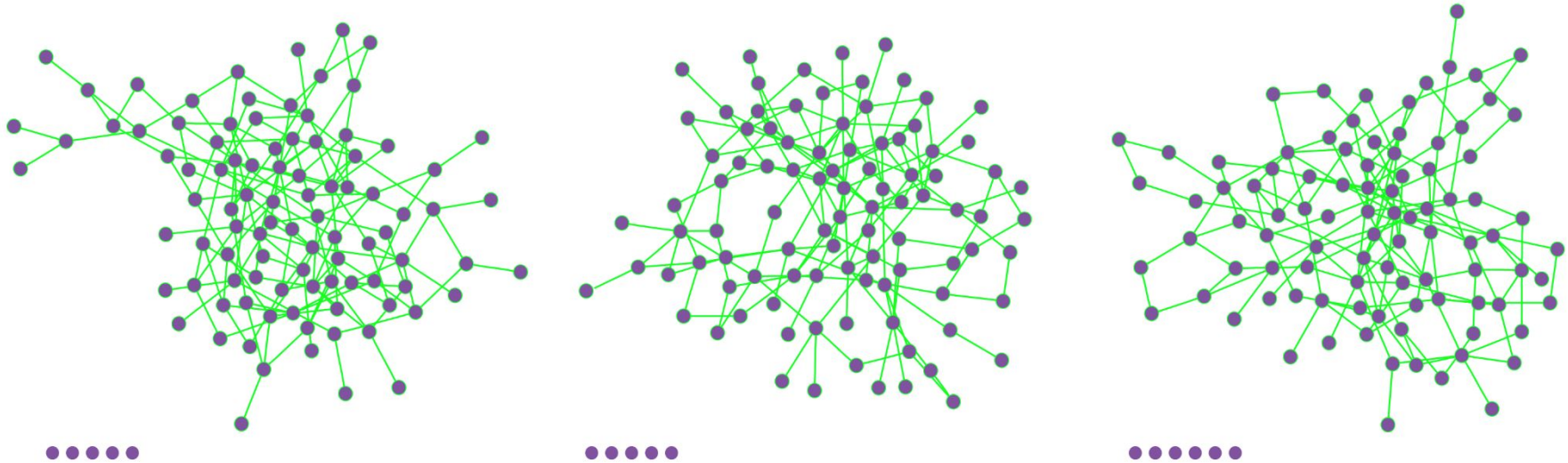
# A random network consists of $N$ nodes where each node pair is connected with probability $p$

Aka “Erdős-Rényi network” – from random graph theory (1959–1968)



Three realizations of a random network generated with the same parameters  $p=1/6$  and  $N=12$ .

A random network consists of  $N$  nodes where each node pair is connected with probability  $p$



Three realizations of a random network with  $p=0.03$  and  $N=100$ . Several nodes have degree  $k=0$ , shown as isolated nodes at the bottom.

# Common question: How many links can we expect for a particular realization of a random network with fixed $N$ and $p$ ?

The probability that a random network has exactly  $L$  links is:

$$\langle L \rangle = p \frac{N(N-1)}{2}$$

(recall, the second term is the max possible number of pairs)

(Barabasi Ch. 3.3)

The average degree of a random network is:

$$\langle k \rangle = \frac{2\langle L \rangle}{N} = p(N-1)$$

(recall, the second term is the max possible node degree)

# Common question: How many links can we expect for a particular realization of a random network with fixed $N$ and $p$ ?

The number of links in a random network varies between realizations.

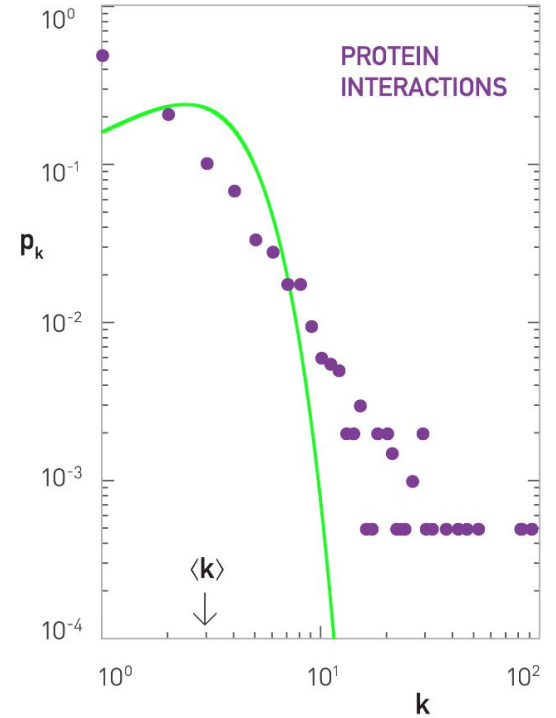
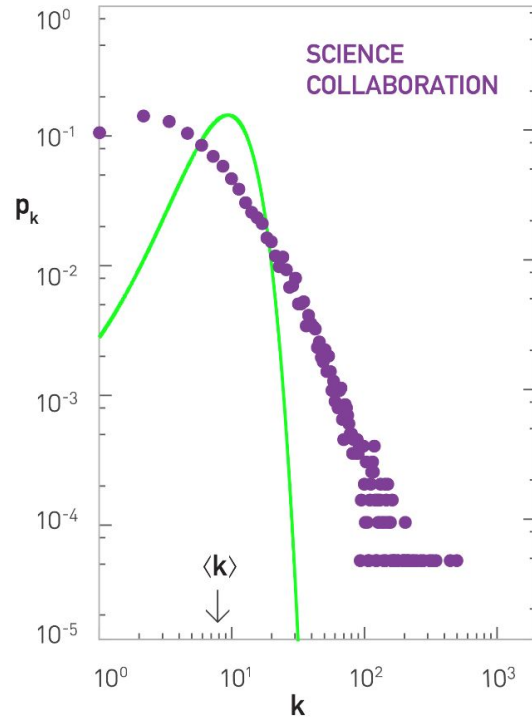
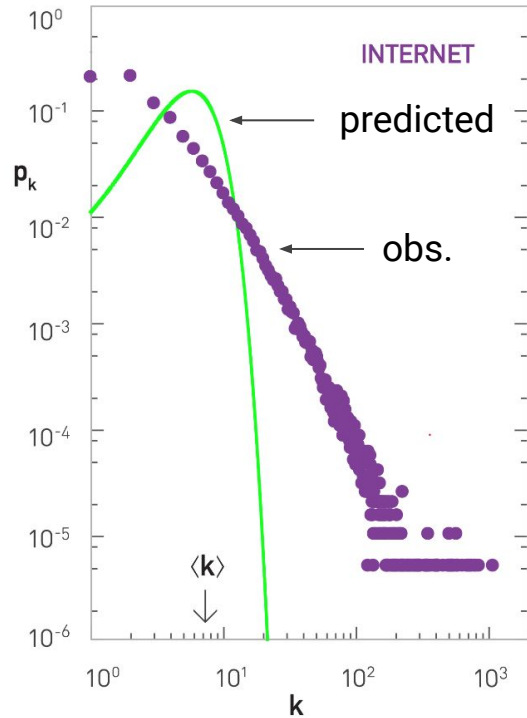
Its expected value is determined by  $N$  and  $p$ .

If we increase  $p$  a random network becomes denser:

The average number of links increase linearly from  $\langle L \rangle = 0$  to  $L_{\max}$

The average degree of a node increases from  $\langle k \rangle = 0$  to  $\langle k \rangle = N-1$ .

# The random network model underestimates the size and frequency of the high degree nodes, and the number of low degree nodes.



**We will come back and improve on this later.  
But for now, at least let's explain why you'll be out of good cake fast!**

# Let's inspect how the size of the largest connected component within the network, $N_G$ , varies with $\langle k \rangle$

For  $p = 0$  we have  $\langle k \rangle = 0$ , hence all nodes are isolated. Therefore the largest component has size  $N_G = 1$  and  $N_G/N \rightarrow 0$  for large  $N$ .

For  $p = 1$  we have  $\langle k \rangle = N-1$ , hence the network is a complete graph and all nodes belong to a single component. Therefore  $N_G = N$  and  $N_G/N = 1$ .

The average degree of a random network is:

$$\langle k \rangle = \frac{2\langle L \rangle}{N} = p(N-1)$$

# Let's inspect how the size of the largest connected component within the network, $N_G$ , varies with $\langle k \rangle$

For  $p = 0$  we have  $\langle k \rangle = 0$ , hence all nodes are isolated. Therefore the largest component has size  $N_G = 1$  and  $N_G/N \rightarrow 0$  for large  $N$ .

For  $p = 1$  we have  $\langle k \rangle = N-1$ , hence the network is a complete graph and all nodes belong to a single component. Therefore  $N_G = N$  and  $N_G/N = 1$ .

One would expect that the largest component grows gradually from  $N_G = 1$  to  $N_G = N$  if  $\langle k \rangle$  increases from 0 to  $N-1$ . Right?

The average degree of a random network is:

$$\langle k \rangle = \frac{2\langle L \rangle}{N} = p(N-1)$$



# Let's inspect how the size of the largest connected component within the network, $N_G$ , varies with $\langle k \rangle$

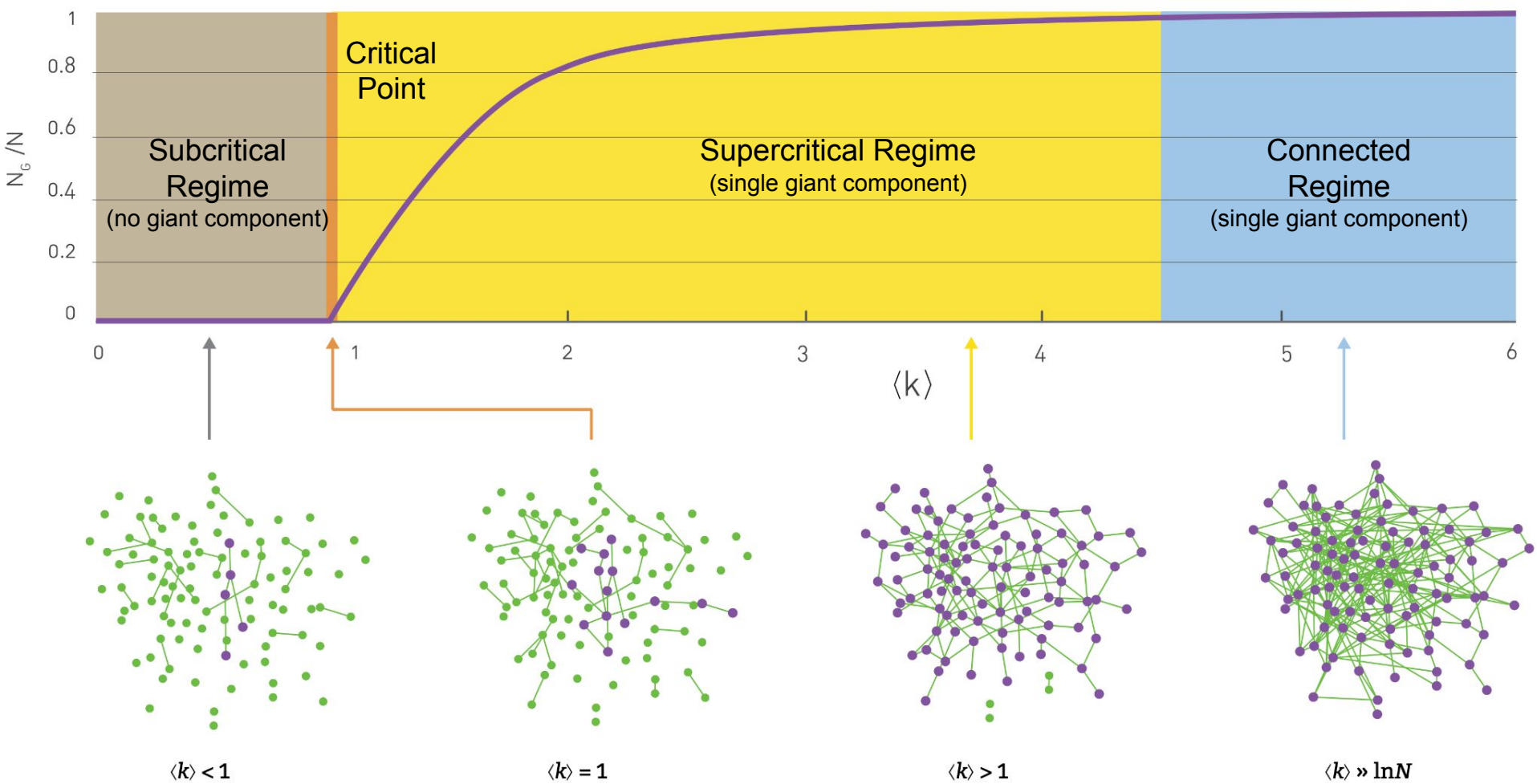
For  $p = 0$  we have  $\langle k \rangle = 0$ , hence all nodes are isolated. Therefore the largest component has size  $N_G = 1$  and  $N_G/N \rightarrow 0$  for large  $N$ .

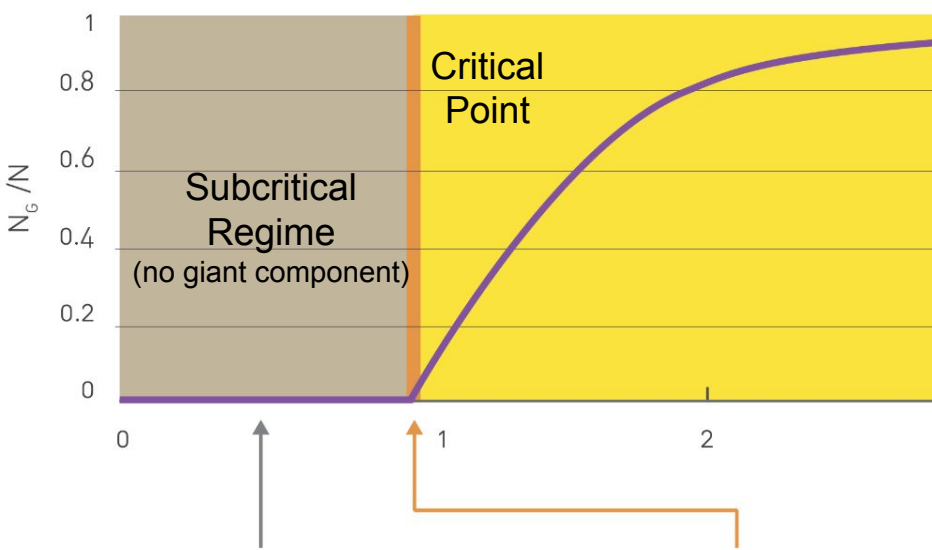
For  $p = 1$  we have  $\langle k \rangle = N-1$ , hence the network is a complete graph and all nodes belong to a single component. Therefore  $N_G = N$  and  $N_G/N = 1$ .

One would expect that the largest component grows gradually from  $N_G = 1$  to  $N_G = N$  if  $\langle k \rangle$  increases from 0 to  $N-1$ . ~~Right?~~ **Wrong**

The average degree of a random network is:

$$\langle k \rangle = \frac{2\langle L \rangle}{N} = p(N-1)$$

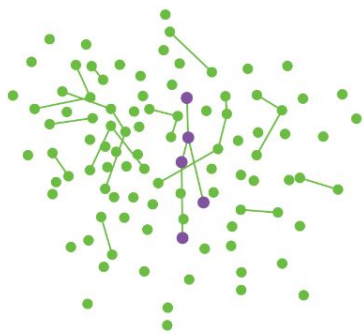




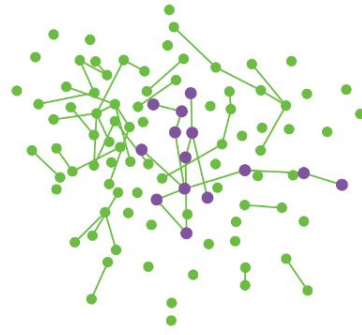
We have one giant component iff each node has on average more than one link.

That we need *at least* one link per node to observe a giant component is not unexpected.

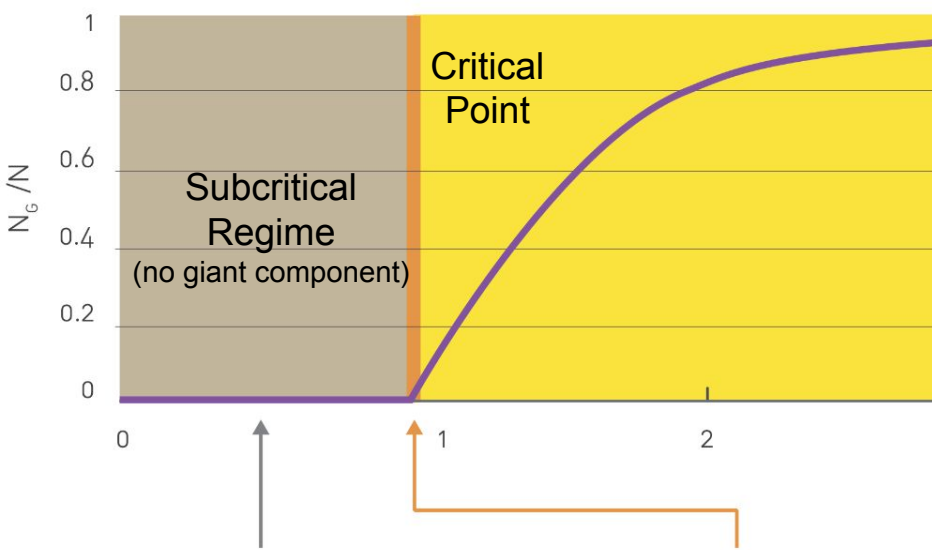
But it is arguably counter-intuitive that one link is *sufficient* for its emergence.



$\langle k \rangle < 1$

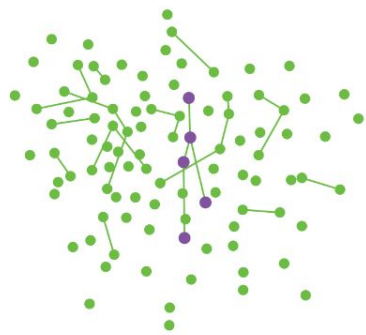


$\langle k \rangle = 1$

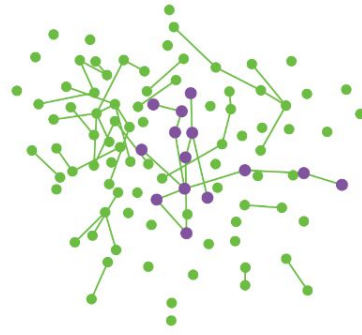


What's the average degree  $\langle k \rangle$  in the HW1 networks?

- Is  $\langle k \rangle > 1$ ? Implying that they have a giant component.
- Is  $\langle k \rangle > \ln N$ ? Implying that they have a *single* giant component.



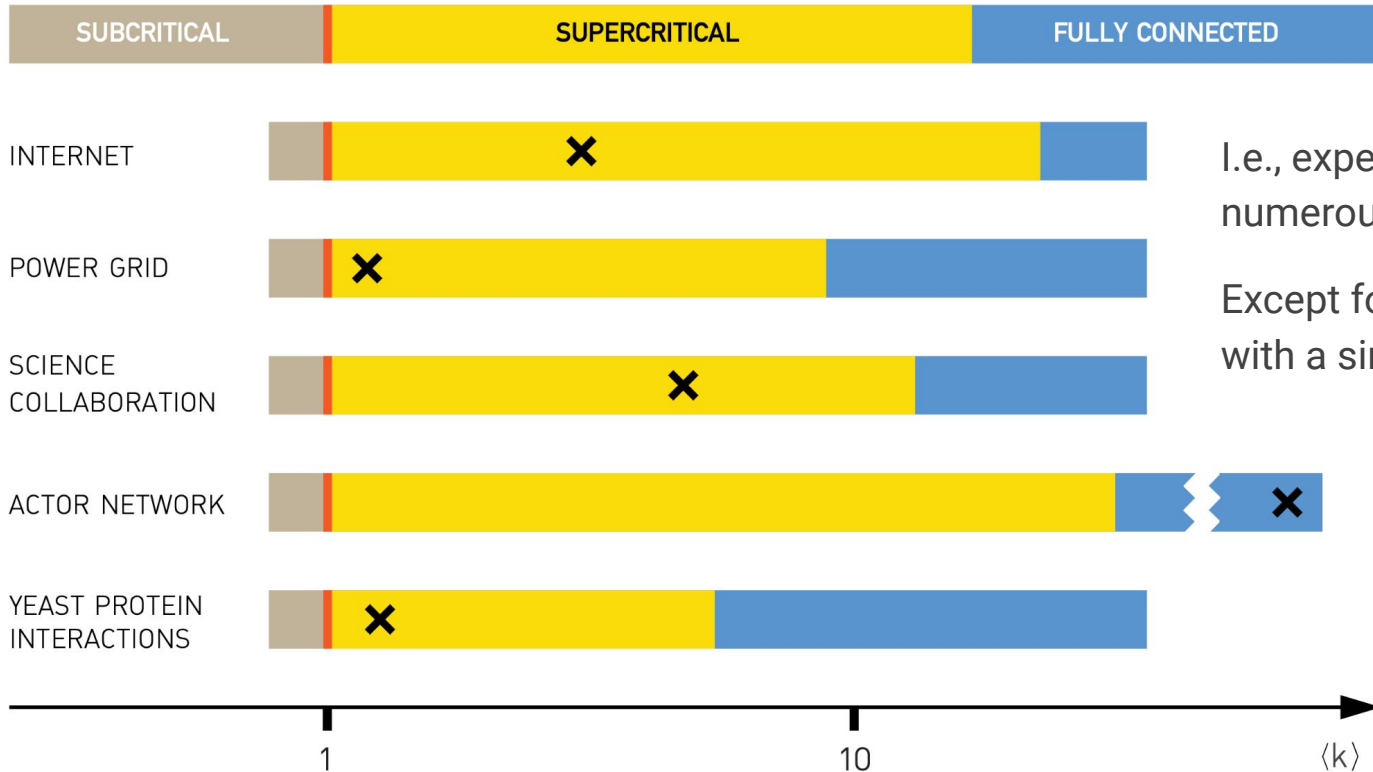
$\langle k \rangle < 1$



$\langle k \rangle = 1$

(If the average individual in a world social network has more than  $\ln(7 \times 10^9) \approx 22.7$  acquaintances, then a random society must have a single component)

# Most real networks are supercritical



I.e., expected to be broken into numerous isolated components.

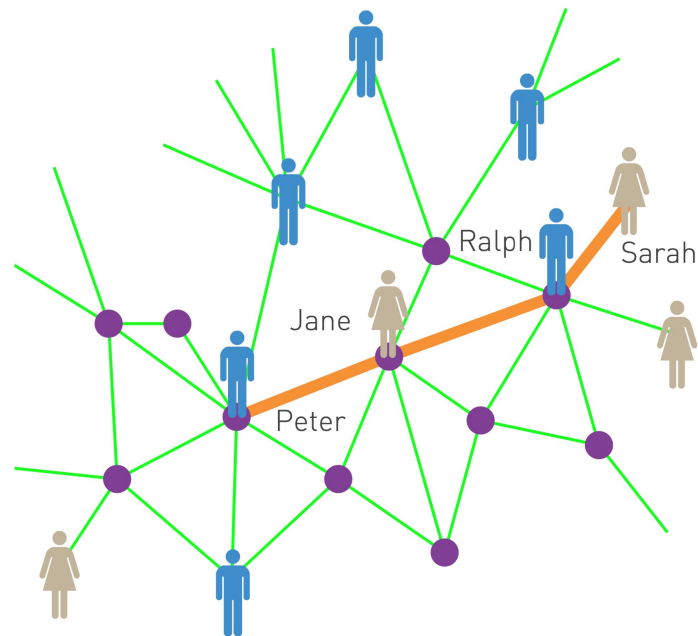
Except for the actor network, with a single giant component.

# Back to Six Degrees of Kevin Bacon (Aka the “Small world” phenomenon)

# Small world property: The distance between any two nodes in a network is small.

Consider a random network with average degree  $\langle k \rangle$ . A node in this network has on average:

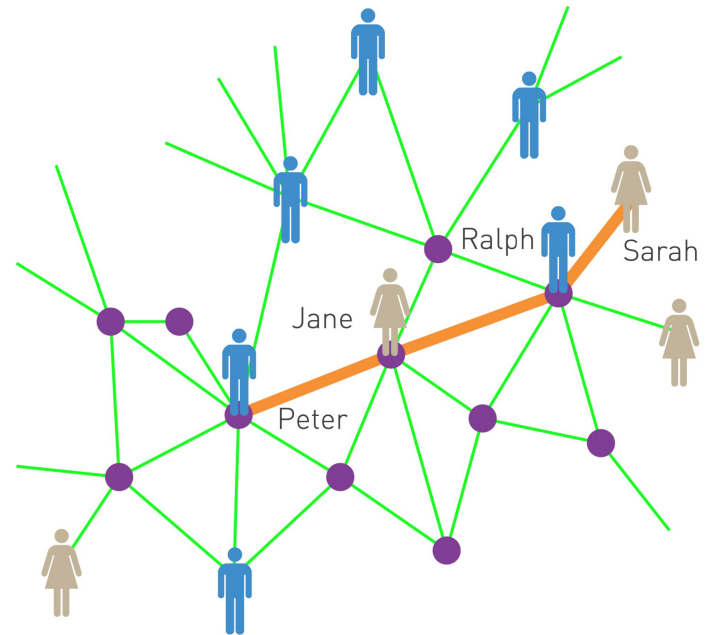
- How many nodes at distance one ( $d=1$ )?



# Small world property: The distance between any two nodes in a network is small.

Consider a random network with average degree  $\langle k \rangle$ . A node in this network has on average:

- $\langle k \rangle$  nodes at distance one ( $d=1$ )
- How many nodes at distance two ( $d=2$ )?

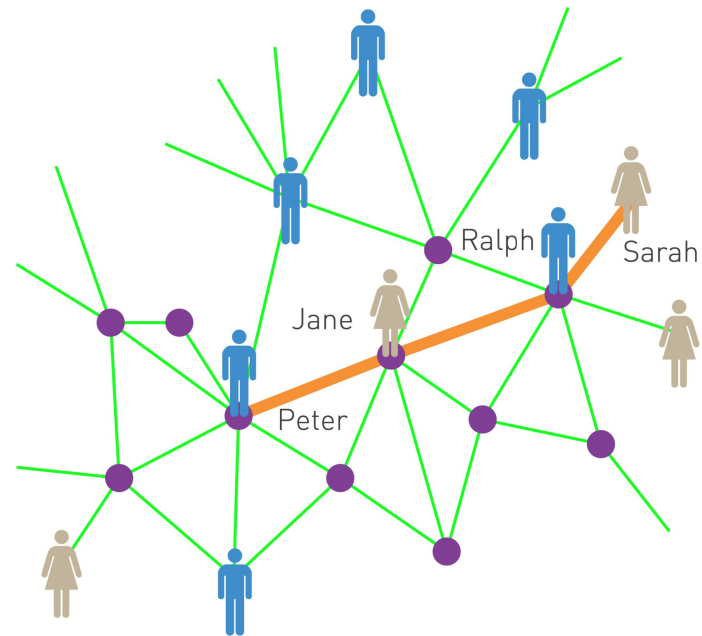




# Small world property: The distance between any two nodes in a network is small.

Consider a random network with average degree  $\langle k \rangle$ . A node in this network has on average:

- $\langle k \rangle$  nodes at distance one ( $d=1$ )
- $\langle k \rangle^2$  nodes at distance two ( $d=2$ )

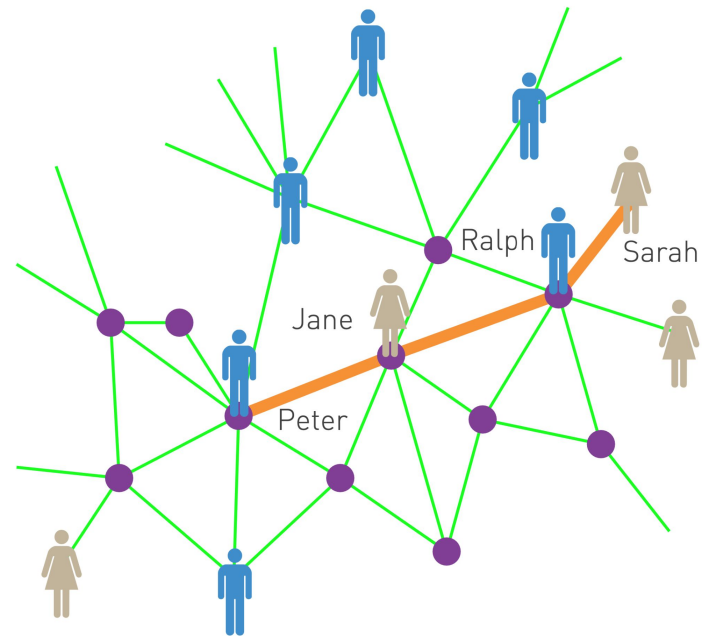


# Small world property: The distance between any two nodes in a network is small.

Consider a random network with average degree  $\langle k \rangle$ . A node in this network has on average:

- $\langle k \rangle$  nodes at distance one ( $d=1$ )
- $\langle k \rangle^2$  nodes at distance two ( $d=2$ )
- $\langle k \rangle^3$  nodes at distance three ( $d=3$ )
- ...
- $\langle k \rangle^d$  nodes at distance  $d$

E.g., if  $\langle k \rangle \approx 1,000$  (the estimated number of acquaintances an individual has), we expect  $10^6$  individuals at  $d=2$  and about a billion, i.e. almost the whole earth's population, at  $d=3$  from us.



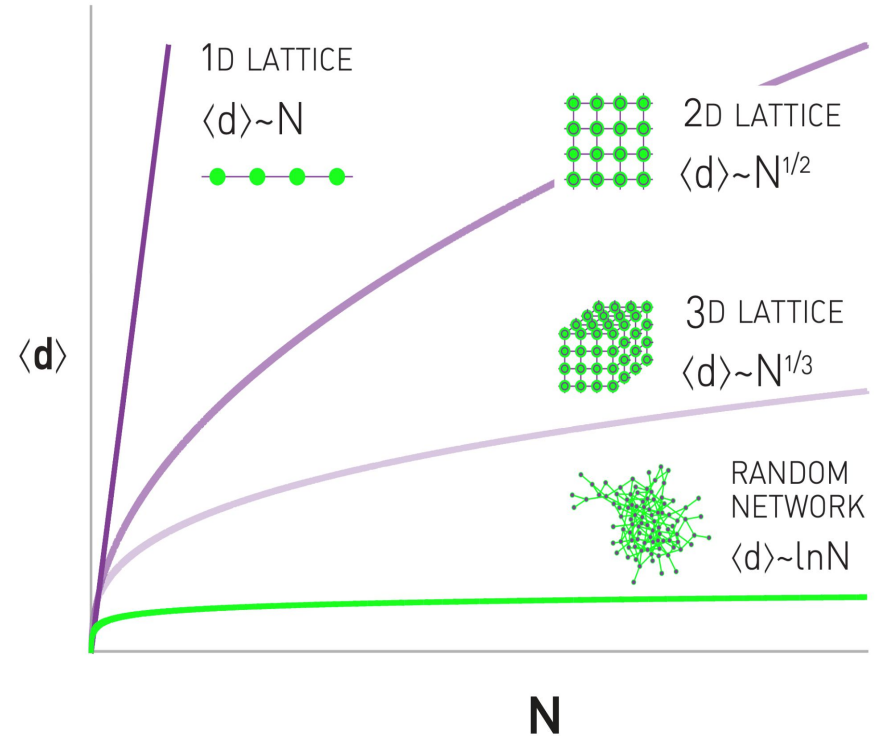
# “Small” as in proportional to $\ln N$ , rather than $N$ (or a power of $N$ )

The dependence of the average distance in a random network on  $N$  and  $\langle k \rangle$ :

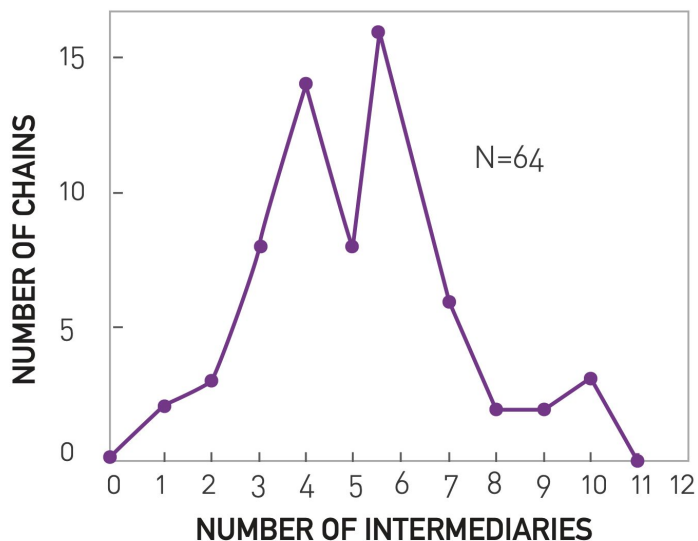
$$\langle d \rangle \approx \frac{\ln N}{\ln \langle k \rangle}$$

The distances in a random network are orders of magnitude smaller than the size of the network.

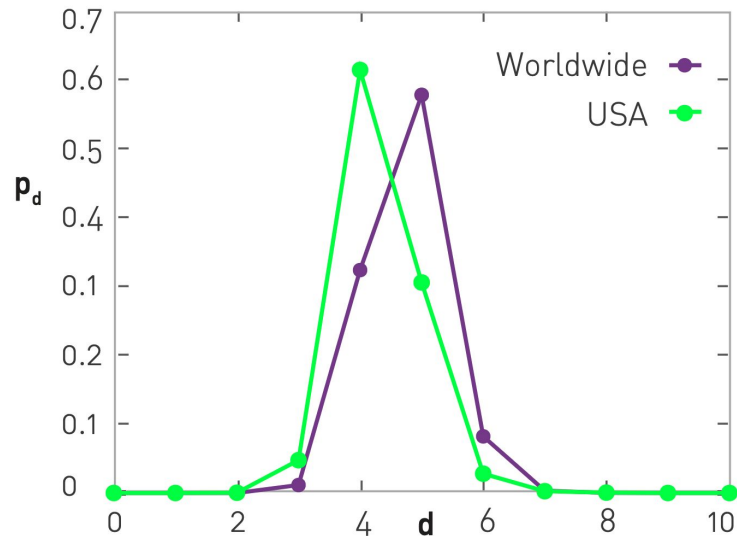
(For our world social network, if  $N \approx 7 \times 10^9$  and  $\langle k \rangle \approx 10^3$ , we get  $\langle d \rangle \approx 3.28$ .)



# Six degrees: Experimental confirmation



Recall (Milgram, 1967) – the letter forwarding study: median 5.2 hops



Facebook 2011 network (721M active users, 68B symmetric friendship links): average distance 4.74

(Backstrom et al, 2012)

# Larger Building Blocks: The Triad

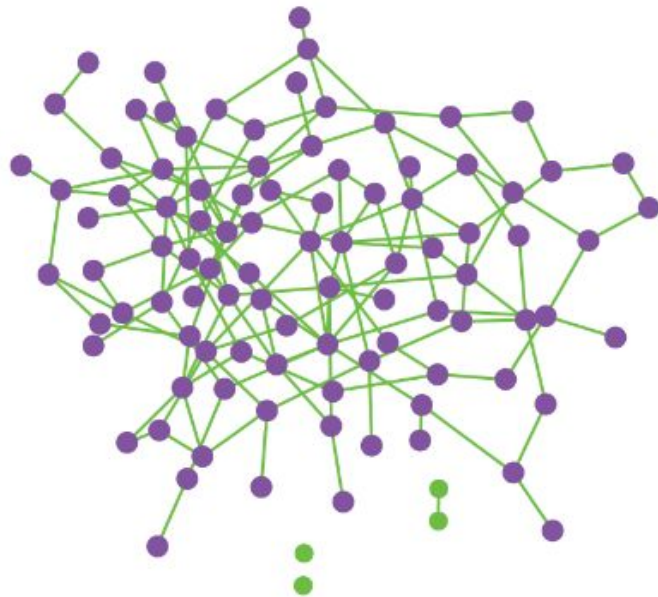
# The Triad

Random graphs are useful as a baseline model

But real-world social networks differ from random graphs in an important way

**They contain more triangles**

These triangles, or triads, are the telltale sign of social groups



Probability of a triangle is close to 0 in random graphs

# The Triad

How is a trio different from a duo?

Dyad is not a group:

- Not perceived as a social group
- Variable and capricious
- Individuality and emotion of the two dictate the relationship

Triad is a group:

- The smallest objectified social group (“me” and “them”)
- The other two can pressure an individual to comply
- Can reduce caprice and uncertainty
- Individual fears exclusion and complies
- Individuality is suppressed and the group is cohesive



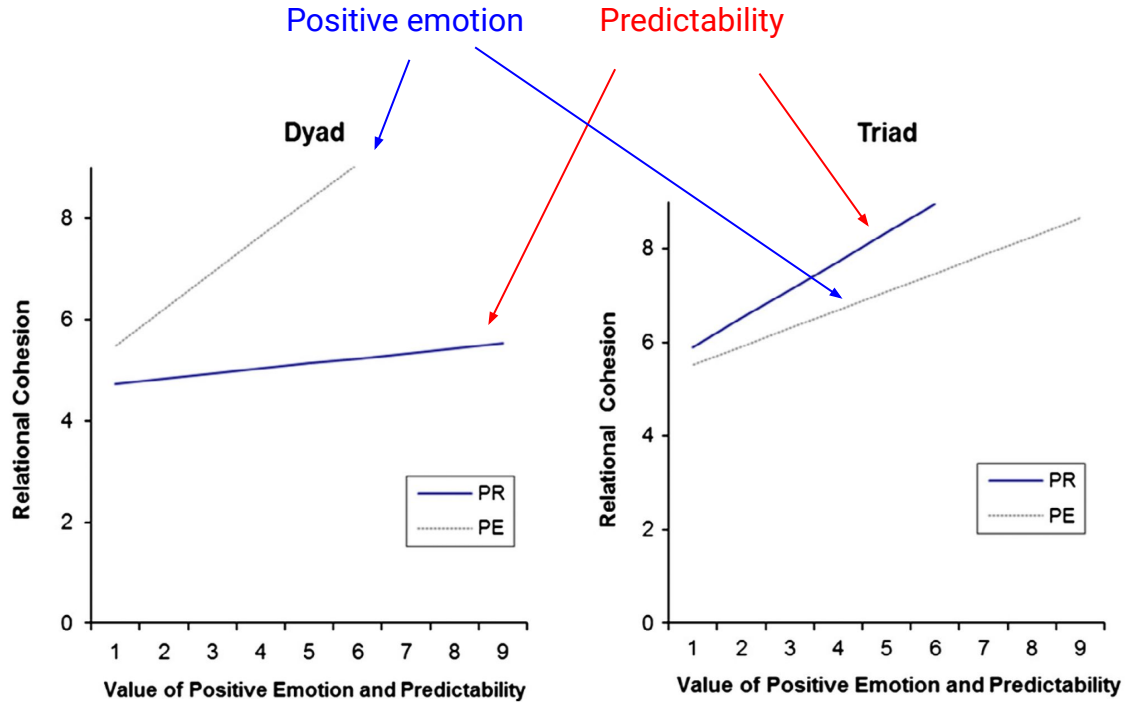
*“each of the two [in a dyad] feels himself confronted only by the other, not by a collectivity above him” -Georg Simmel*

# The Triad

Experimental subjects were divided into either a 2-person exchange or a 3-person exchange condition

**Positive emotion** was more important for group cohesion in **dyads**

**Predictability** of exchange partners was more important for group cohesion in **triads**



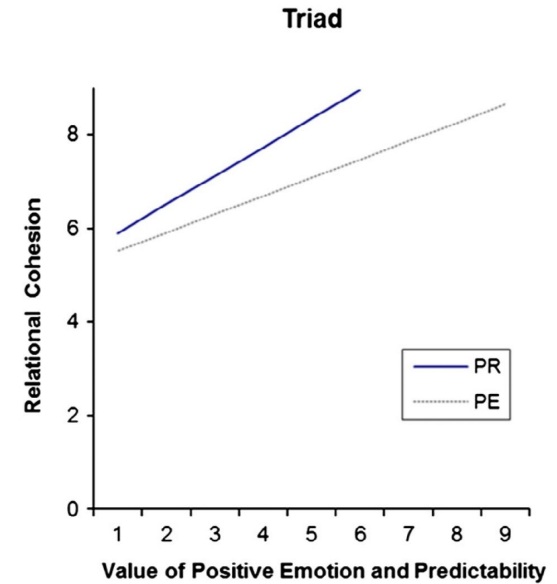
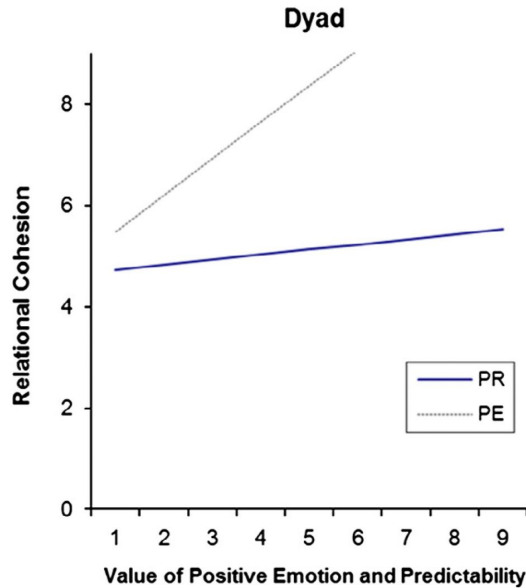
(Yoon et al, 2013)



# The Triad

In short, triads are the smallest social group

Our aim is to understand the properties of a triad as a starting point for understanding the network properties of larger social groups



(Yoon et al, 2013)

# Structural Balance: Triads of Friends and Enemies

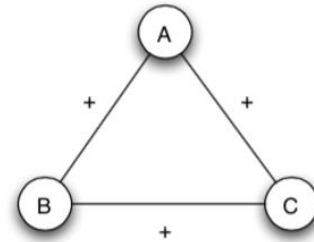
One of the earliest models of triads took inspiration from psychology

BALANCE THEORY

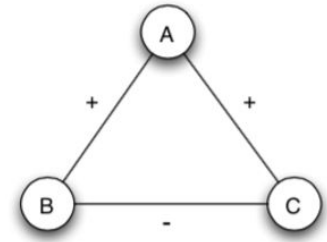


FRITZ HEIDER

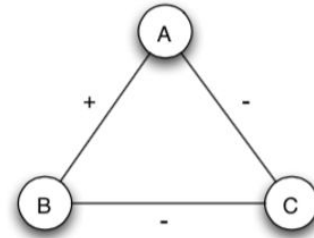
- People try to avoid cognitive inconsistencies in their relationships
- This idea was applied to triads and developed to structural balance theory



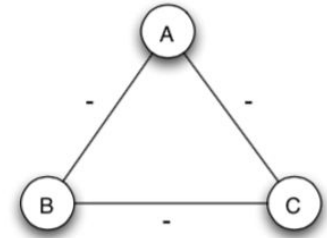
Balanced



Not balanced



Balanced



Not balanced

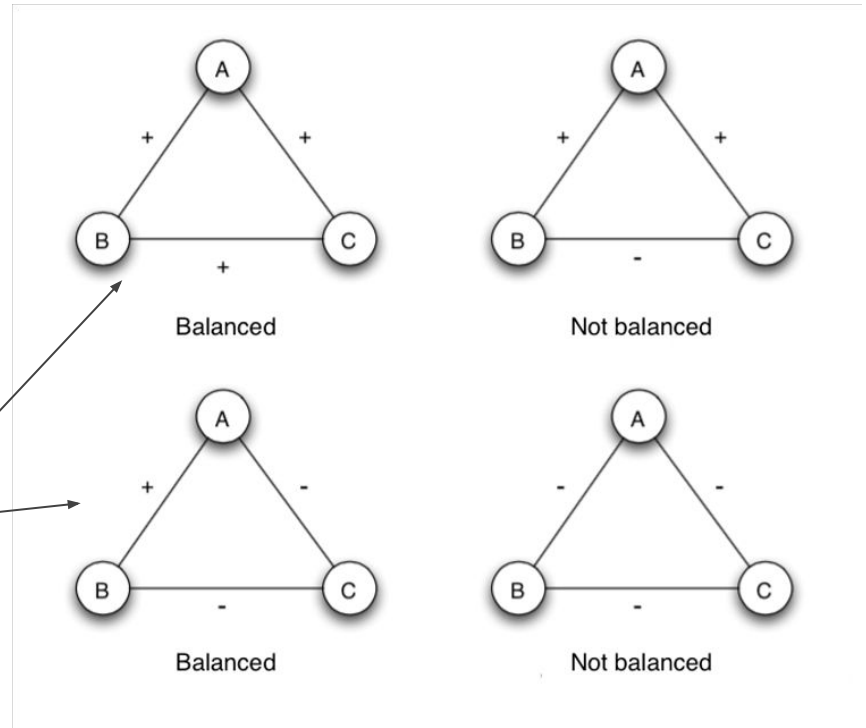
# Structural Balance: Triads of Friends and Enemies

Friendships: “+”

Enemy relations: “-”

Structural balance is achieved when

- A friend of my friend is also my friend
- An enemy of my friend is my enemy

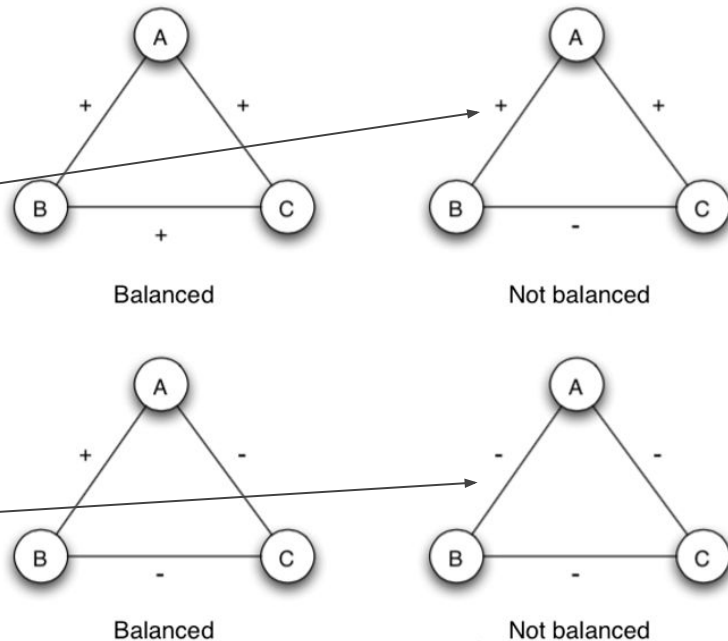


# Structural Balance: Triads of Friends and Enemies

Triad is unbalanced when it consists of an **odd number of negative edges**

One negative edge: My two friends are enemies

Three negative edges: The enemy of my enemy is also my enemy



**... to be continued**

# Today's Summary

Giant components

The random graph model

An explanation for Six Degrees of Kevin Bacon

A teaser for the next smallest building block – triads and their social implications