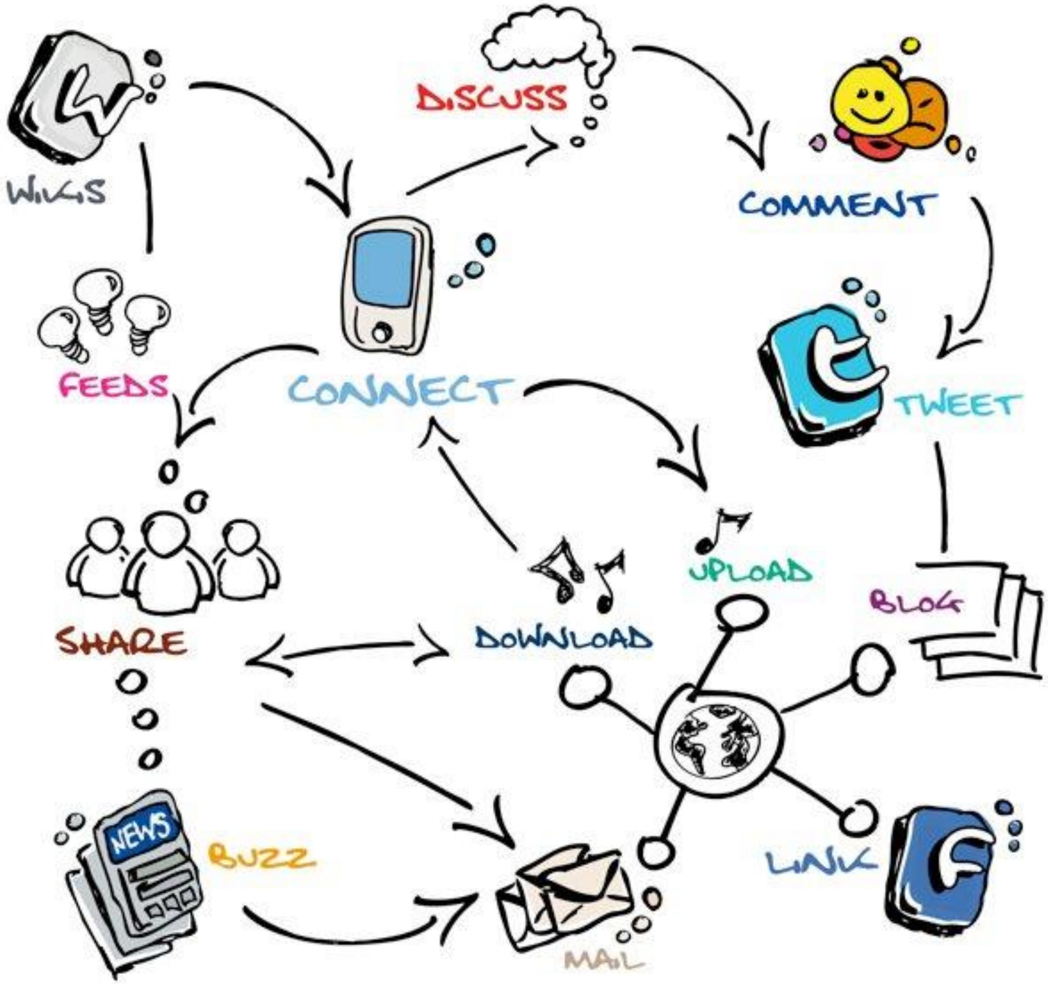


Small-World Phenomena

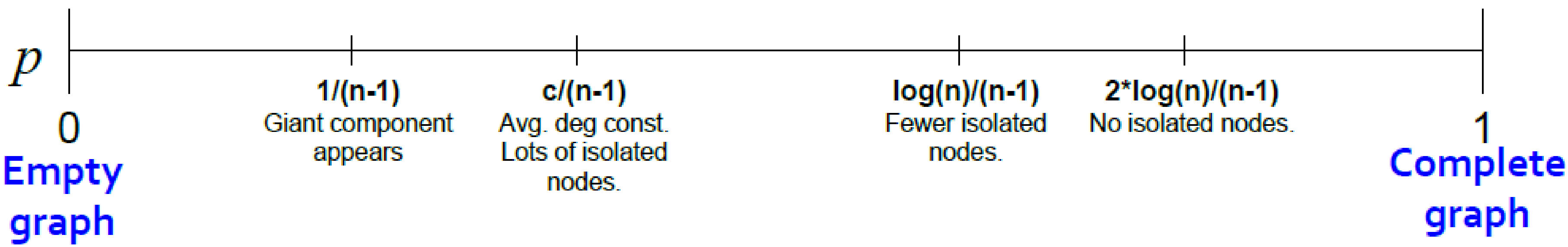
Ahmet Onur Durahim

How the Class Fits Together

Properties	Models	Algorithms
Small diameter, Edge clustering	Small-world model, Erdös-Renyi model	Decentralized search
Scale-free	Preferential attachment, Copying model	PageRank, Hubs and authorities
Strength of weak ties, Core-periphery	Kronecker Graphs	Community detection: Girvan-Newman, Modularity
Densification power law, Shrinking diameters	Microscopic model of evolving networks	Link prediction, Supervised random walks
Complex Graph Structure	Graph Neural Networks	Node Classification Graph Representation Learning
Information virality, Memetracking	Independent cascade model, Game theoretic model	Influence maximization, Outbreak detection, LIM

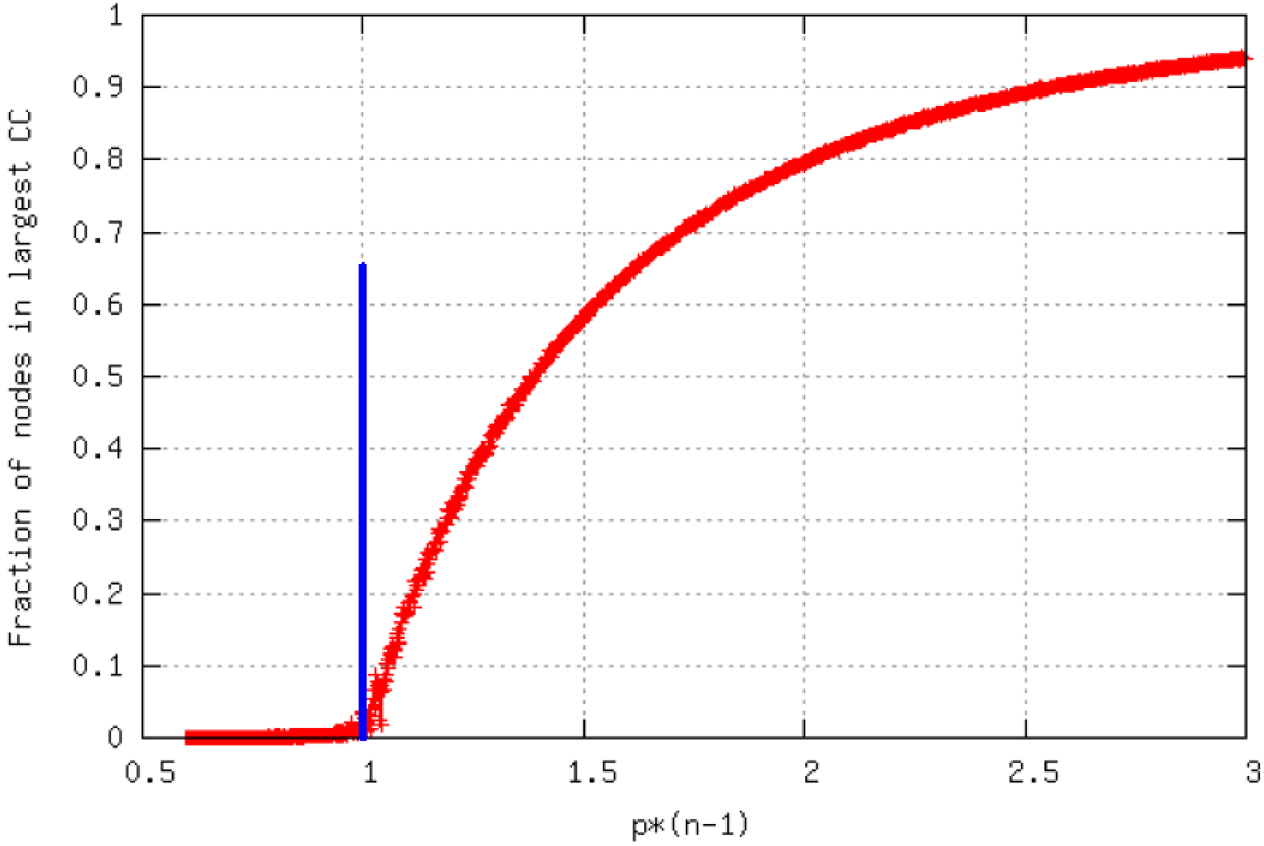
Recap: “Evolution” of a Random Graph

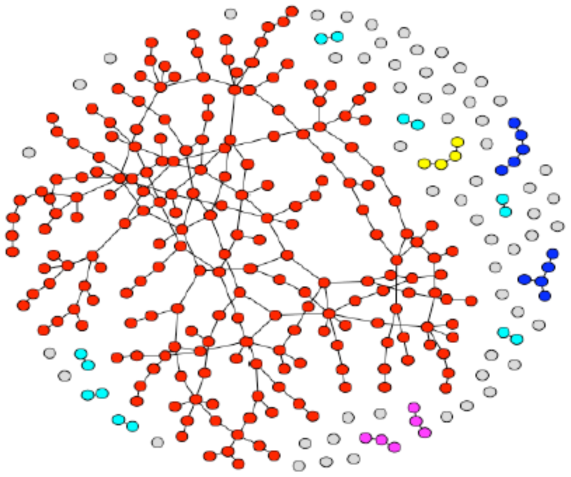
- **Graph structure of G_{np} as p changes:**



- **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 $\Omega(\log n)$

Recap: G_{np} Simulation Experiment

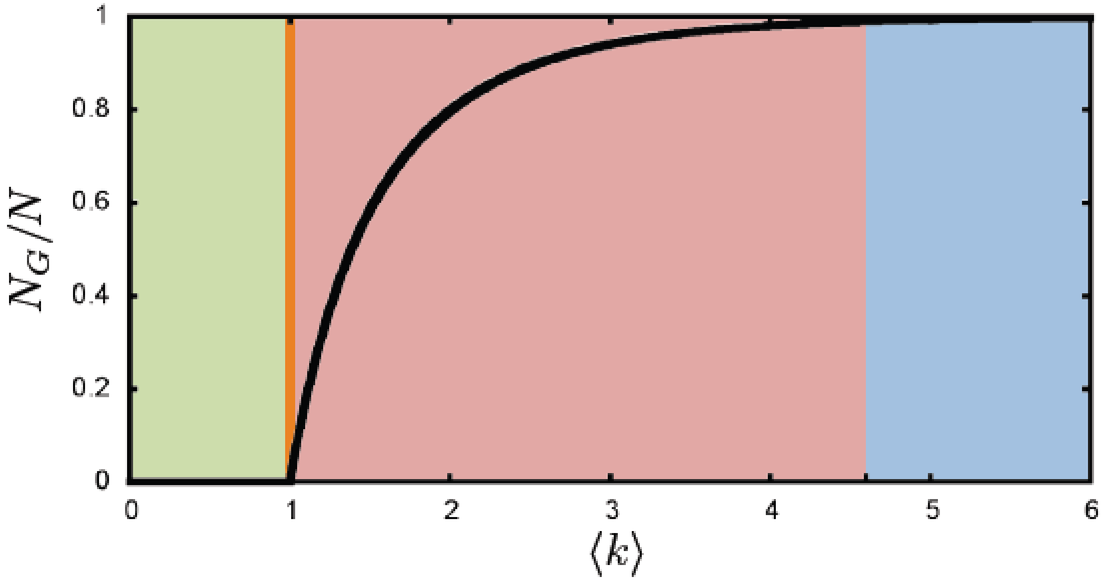




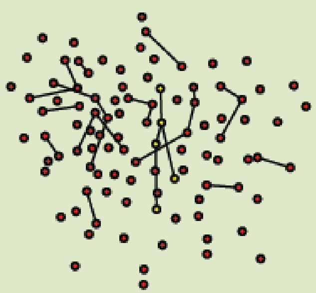
Fraction of nodes in the largest component

- **G_{np}, n=100,000, *$k = p(n-1)$* = 0.5 … 3**

Recap: Evolution of a Random Network



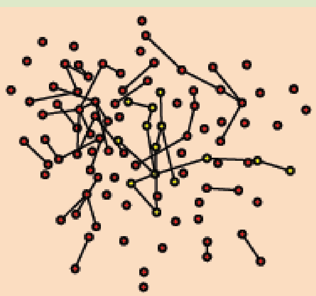
Subcritical regime: 0 < ‹k› < 1, (p < $\frac{1}{N}$)



$\langle k \rangle < 1$

Subcritical regime

- No giant component.
- Cluster size distribution: $P(s) \sim e^{-\alpha s}$
- The size of the largest cluster: $N_G \sim \ln N$
- The clusters are trees.



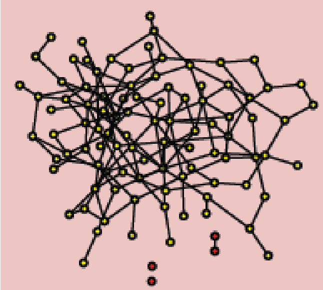
$\langle k \rangle = 1$

Critical point

- No giant component.
- Cluster size distribution: $P(s) \sim s^{-3/2}$
- Size of the largest cluster: $N_G \sim N^{2/3}$
- The clusters may contain loops.

Critical Point: ‹k› = 1, (p = $\frac{1}{N}$)

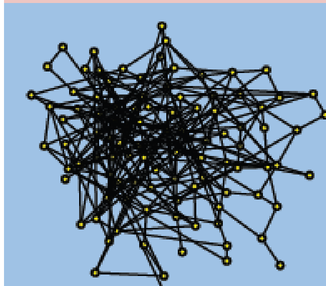
Supercritical regime: ‹k› > 1, (p > $\frac{1}{N}$)



$\langle k \rangle > 1$

Supercritical regime

- Single giant component.
- Cluster size distribution: $P(s) \sim e^{-\alpha s}$
- Size of the giant component: $N_G \sim (p - p_c)N$
- The small clusters are trees.
- GC has loops.



$\langle k \rangle \geq \ln N$

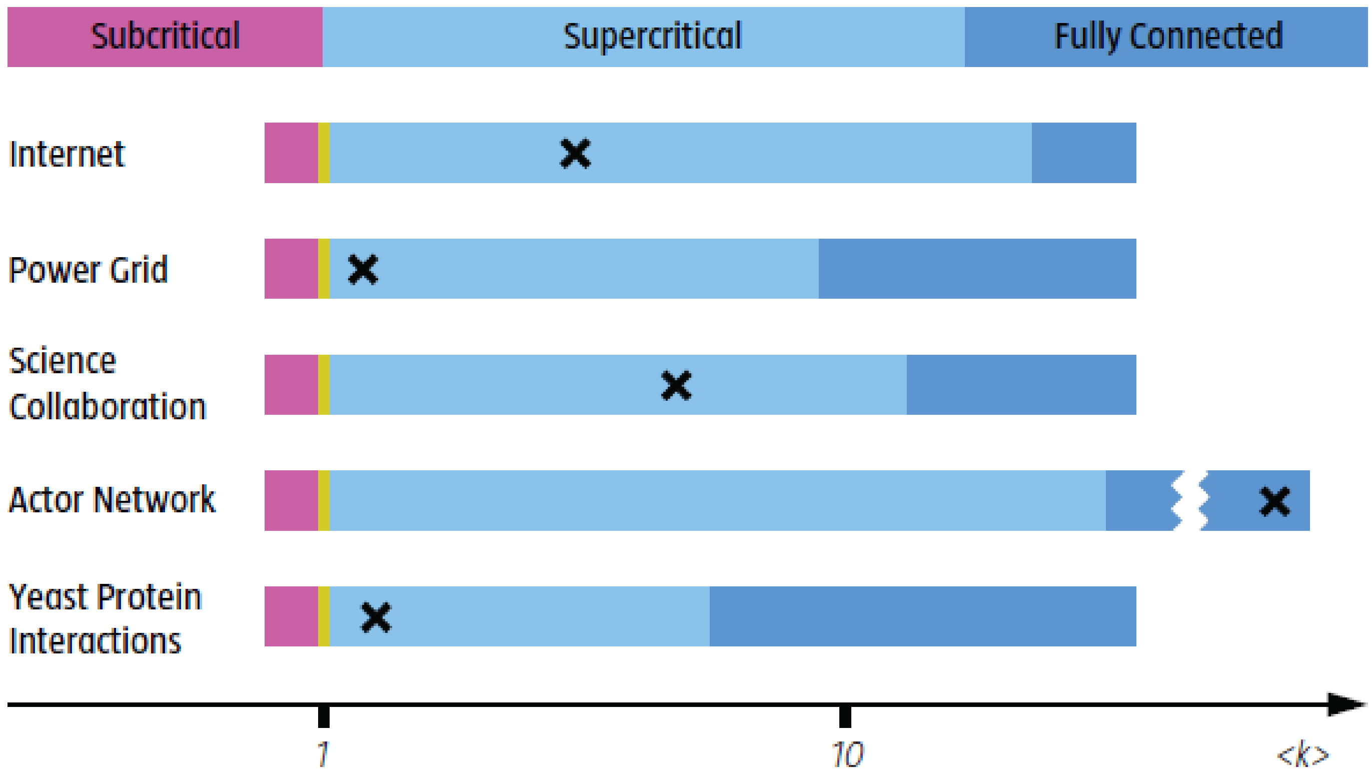
Fully connected regime

- Single giant component.
- No isolated nodes or clusters.
- Size of the giant component: $N_G = N$
- GC has many loops.

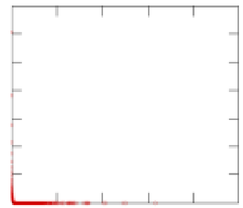
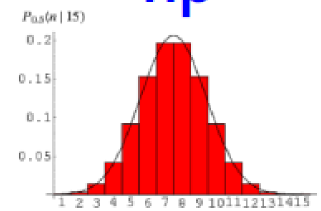
Connected regime: ‹k› ≥ lnN, (p ≥ $\frac{\ln N}{N}$)

Real Networks are Supercritical

Network	N	L	$\langle k \rangle$	$\ln N$
Internet	192,244	609,066	6.34	12.17
Power Grid	4,941	6,594	2.67	8.51
Science Collaboration	23,133	186,936	8.08	10.04
Actor Network	212,250	3,054,278	28.78	12.27
Yeast Protein Interactions	2,018	2,930	2.90	7.61



Recap: MSN vs. G_{np}

	MSN	G_{np}	n=180M
Degree distribution:			
Avg. path length:	**6.6**	$O(\log n)$ $h \approx 8.2$	
Avg. clustering coef.:	0.11	$\bar{k} / n$ $C \approx 8 \cdot 10^{-8}$	
Largest Conn. Comp.:	99%	GCC exists when $\bar{k}>1$. $\bar{k} \approx 14$.	

Recap: Real Networks vs. G_{np}

- **Are real networks like random graphs?**
  - Giant connected component: ☺
  - Average path length: ☺
  - Clustering Coefficient: ☹
  - Degree Distribution: ☹
- **Problems with the random networks model:**
  - Degree distribution differs from that of real networks
  - Giant component in most real network does NOT emerge through a phase transition
  - No *local structure* – clustering coefficient is too low
- **Most important: Are real networks random?**
  - **The answer is simply: NO!**

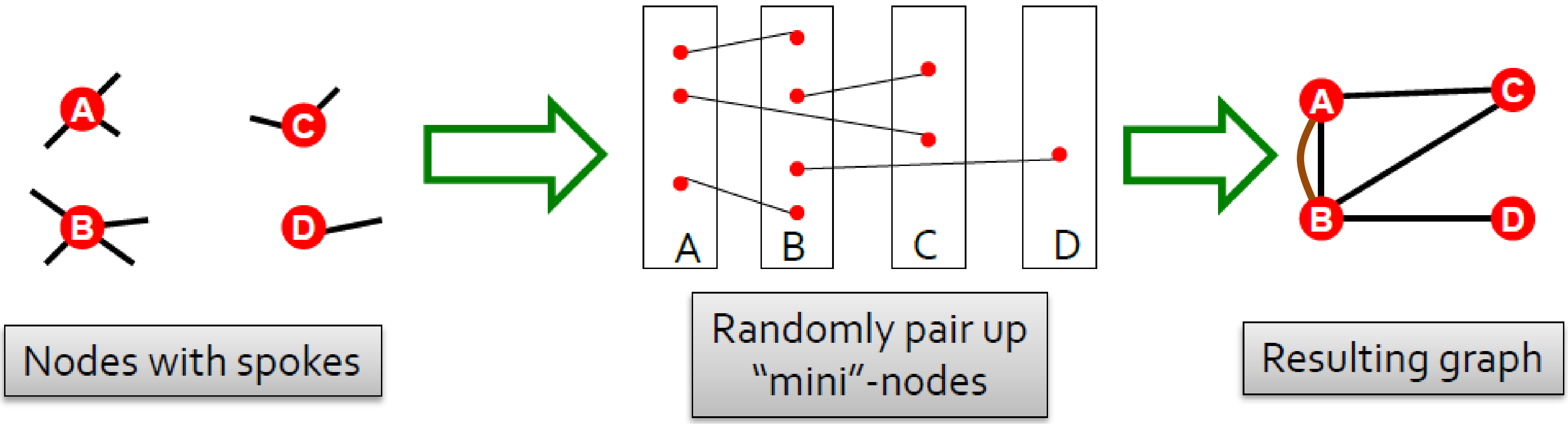
Recap: Real Networks vs. G_{np}

- **If G_{np} is wrong, why did we spend time on it?**
  - It is the ***reference model*** for the rest of the class
  - It will help us calculate many ***quantities***, that can then be ***compared*** to the real data
  - It will help us understand to what degree is a particular property the result of some ***random process***

So, while G_{np} is WRONG, it will turn out to be extremely USEFUL!

Intermezzo: Configuration Model

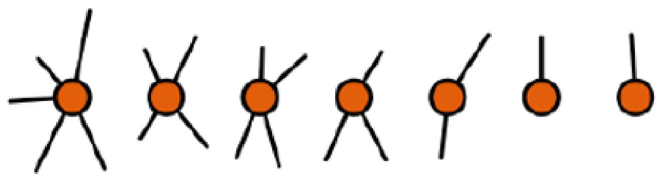
- **<u>Goal:</u>** Generate a random graph with a given degree sequence $k_1, k_2, \ldots k_N$
- **Configuration model:**



- **Useful as a *"null"* model of networks**
  – We can compare the real network ***G*** and a "random" ***G'*** which has the same degree sequence as ***G***

Configuration Model

- Random graph with n nodes with a ***given degree sequence*** D={k_1, k_2, … k_N} and m = ½ $\sum k_i$ edges
- Construct by randomly matching two stubs and connecting them by an edge



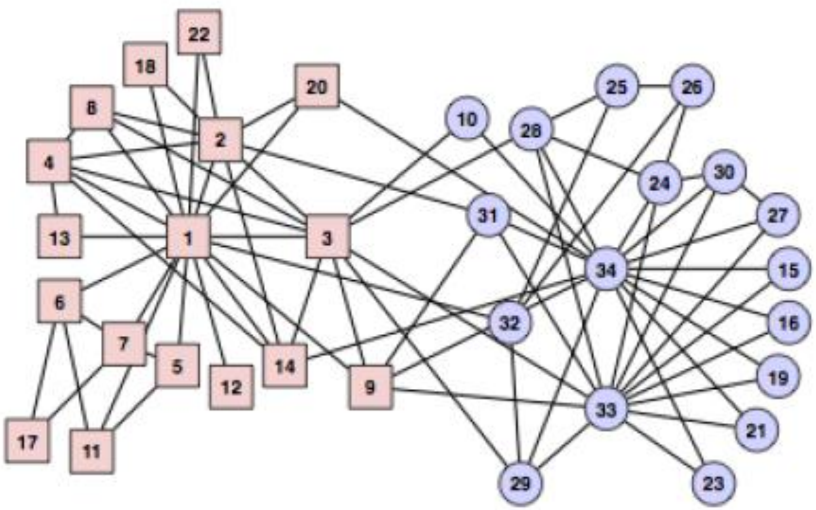
- Probability that two nodes i and j are connected

$$p_{ij} = \frac{k_i k_j}{2m - 1}$$

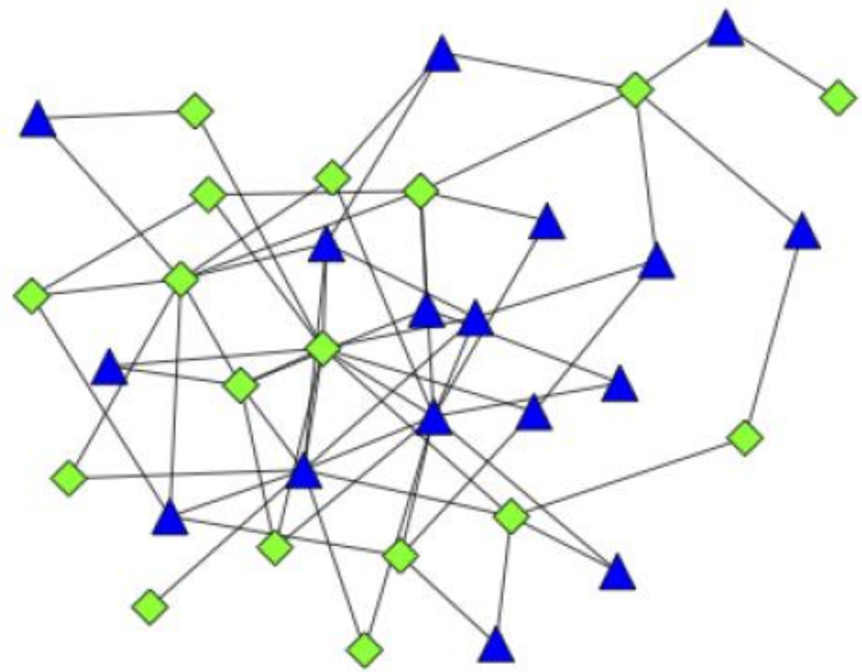
- Can contain ***self loops*** and ***multiple edges***
- Will be a simple graph for special ***graphical degree sequence***

Configuration Model

- Can be used as a "null model" for ***comparative network analysis***



karate club

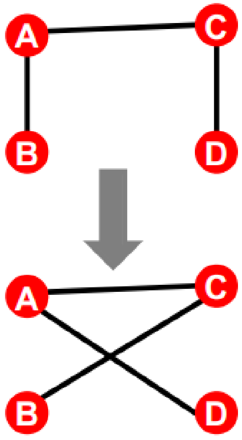


configuration model

Clauset, 2014

Alternative for Spokes: ***Switching***

- Start from a **given graph** G
- Repeat **the switching step** $Q \cdot |E|$ times:
  - Select a pair of edges A→B, C→D at random
  - **Exchange** the endpoints to give A→D, C→B
    - Exchange edges only if no multiple edges or self-edges are generated



- **Result:** A randomly rewired graph:
  - Same node degrees, randomly rewired edges
- Q is chosen large enough *(e.g.,* $Q = 100$*)* for the process to converge

Network Models

- ***Empirical network features:***
  - *Power-law* (heavy-tailed) degree distribution
  - Small average distance (graph *diameter*)
  - Large clustering coefficient (*transitivity*)
  - *Giant connected component*, hierarchical structure, etc
- ***Generative models:***
  - Random graph model (Erdos & Renyi, 1959)
  - ***"Small world" model*** **(Watts & Strogatz, 1998)**
  - Preferential attachment (Scale-free) model (Barabasi & Albert, 1999)

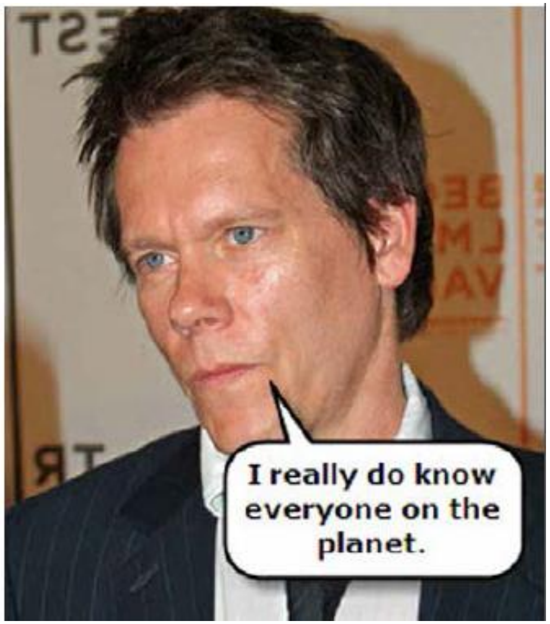
The Small-World Model

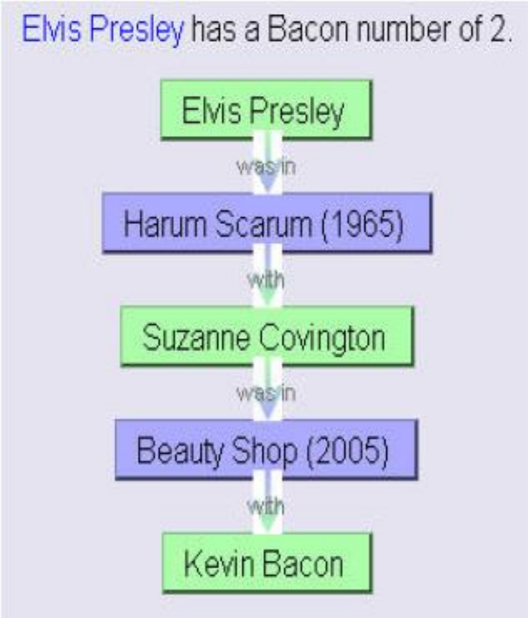
Can we have high clustering while also having short paths?

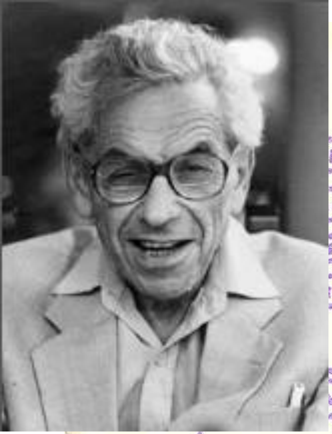
Six Degrees of Kevin Bacon

Origins of a small-world idea:

- **The Bacon number:**
  - Create a network of Hollywood actors
  - Connect two actors if they co-appeared in the movie
  - **Bacon number:** number of steps to Kevin Bacon
- As of Dec 2007, the highest (finite) Bacon number reported is 8
- Only approx. 12% of all actors cannot be linked to Bacon







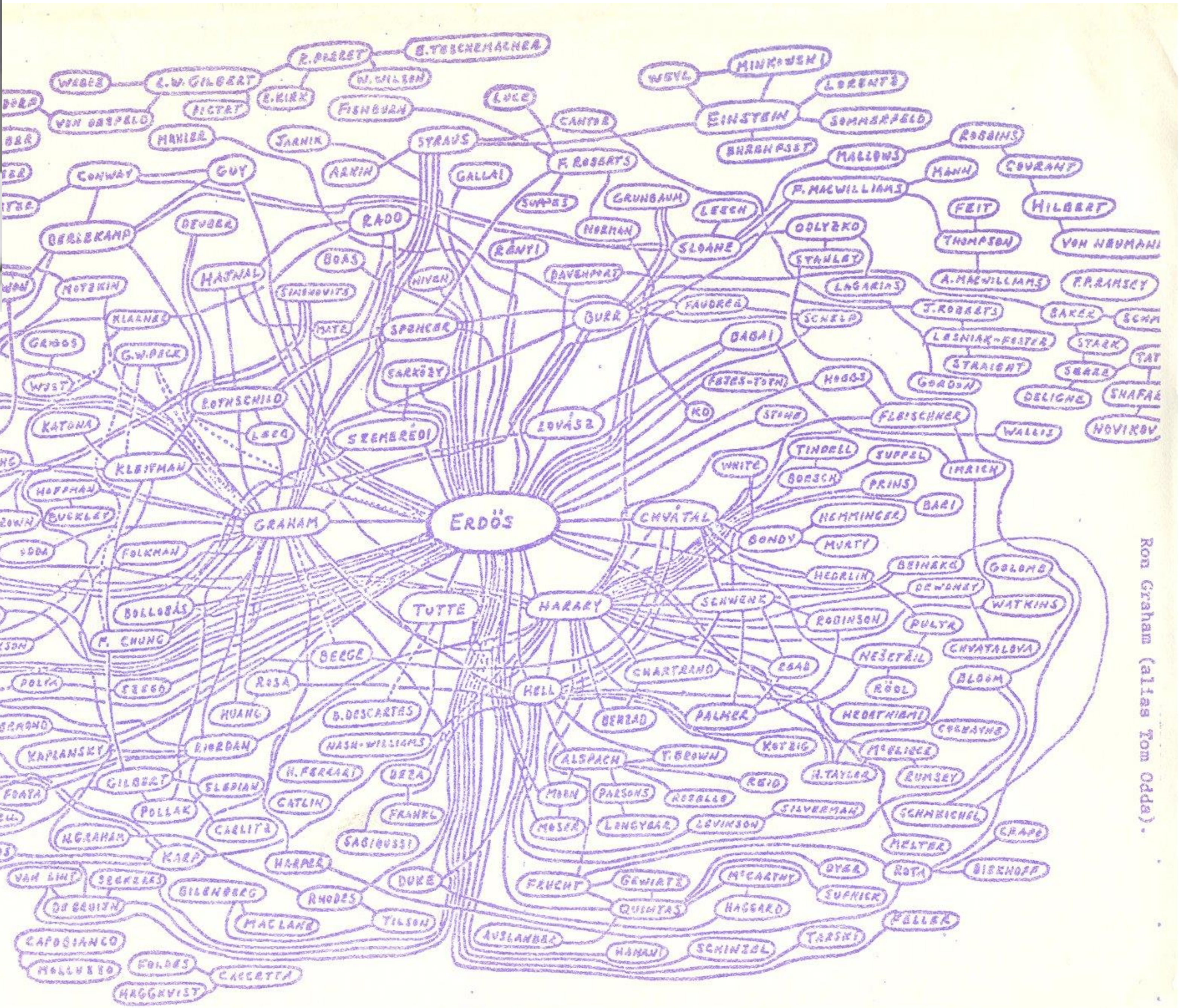


Figure 1

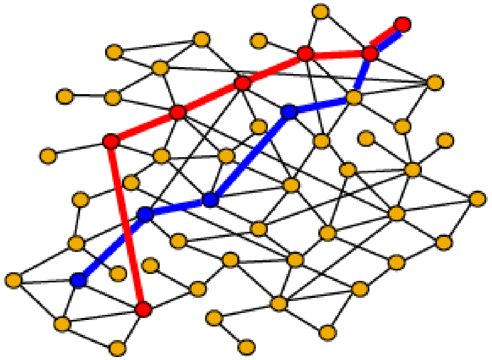
To appear in Topics in Graph Theory (F. Harary, ed.) New York Academy of Sciences (1979).

Ron Graham (alias Tom Odda).

The Small-World Experiment

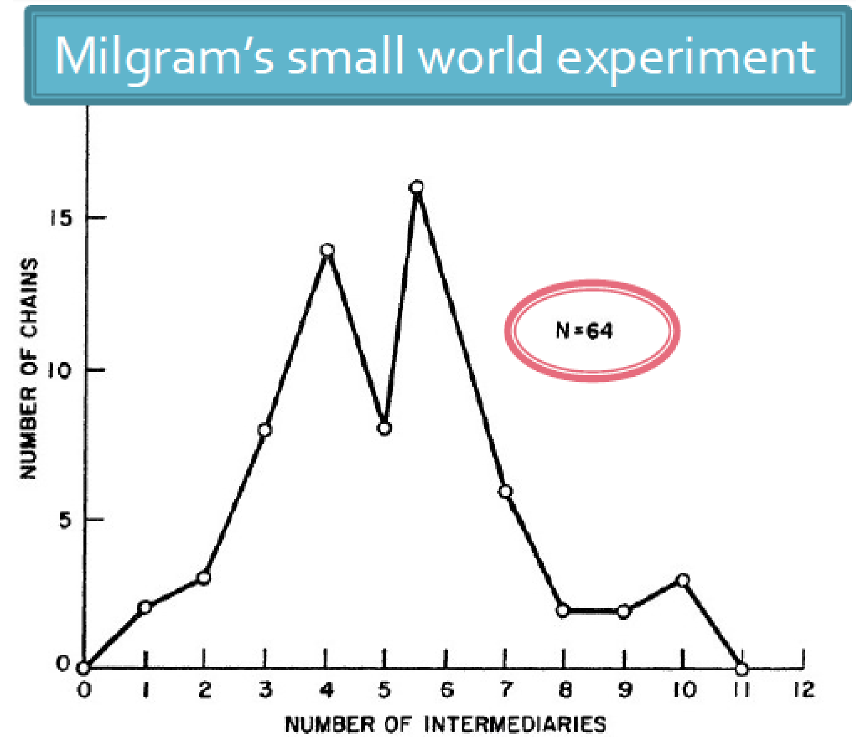
[Milgram, ‘67]

- **What is the typical shortest path length between any two people?**
  - *Experiment on the global friendship network*
    - Can’t measure, need to probe explicitly
- **Small-world experiment** [Milgram ’67]
  - Picked 300 people in Omaha, Nebraska and Wichita, Kansas
  - Ask them to get a letter to a stock-broker in Boston by passing it through friends
- **How many steps did it take?**



The Small-World Experiment

[Milgram, ’67]



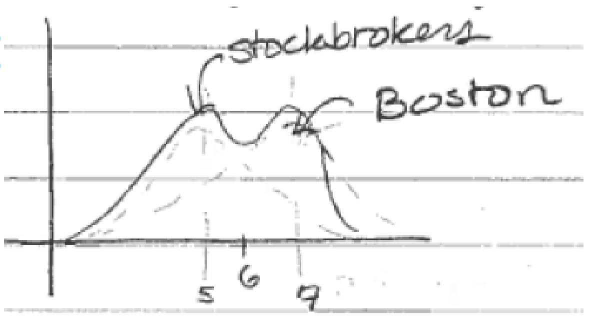
- **64 chains completed:**
  (i.e., 64 letters reached the target)
  – It took 6.2 steps on the average, thus
  “6 degrees of separation”

- **Further observations:**
  – People who owned stock had shorter paths to the stockbroker than random people: 5.4 vs. 6.7
  – People from the Boston area have even closer paths: 4.4

Occupation and ***Geographic Proximity***
also have effect on the degree of separation

Milgram: Further Observations

- **Boston vs. Occupation networks:**



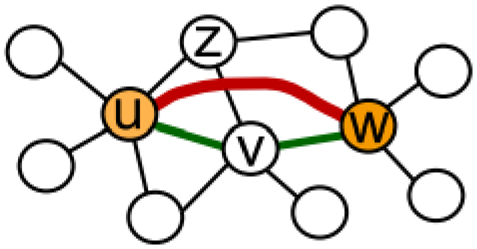
- **Criticism:**
  - **Funneling:**
    - 31 of 64 chains passed through 1 of 3 people as their final step => ***Not all links/nodes are equal***
  - Starting points and the target were ***non-random***
  - There are ***not many samples*** (only 64)
  - People refused to participate (25% for Milgram)
    - Not all searches finished (only 64 out of 300)
  - **Some sort of social search:** People in the experiment follow some strategy instead of forwarding the letter to everyone. **They are not finding the shortest path!**
  - People might have used ***extra information*** resources

6-Degree: Should We Be Surprised?

- **Assume each human is connected to 100 other people**

  Then:
  - Step 1: reach 100 people
  - Step 2: reach 100*100 = 10,000 people
  - Step 3: reach 100*100*100 = 1,000,000 people
  - Step 4: reach 100*100*100*100 = 100M people
  - In 5 steps we can reach 10 billion people
- **What's wrong here? => we ignore *clustering* !!!**
  - ***92% of new FB friendships are to a friend-of-a-friend***

    [Backstom-Leskovec '11] [Leskovec '08]



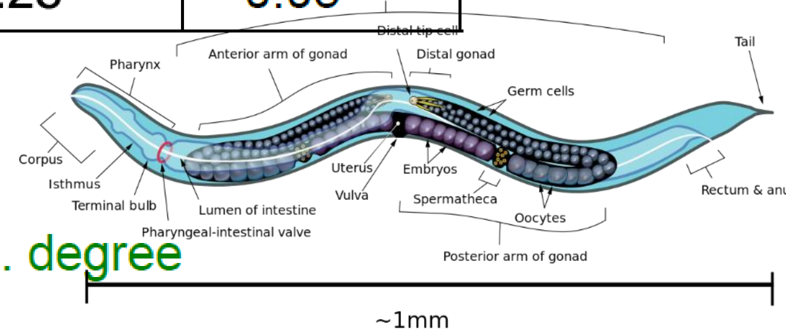
Clustering Implies Edge Locality

- **MSN network has 7 orders of magnitude larger clustering than the corresponding G_{np}!**
- **Other examples:**

  Actor Collaborations (IMDB): N = 225,226 nodes, avg. degree $\overline{k}$ = 61
  Electrical power grid: N = 4,941 nodes, $\overline{k}$ = 2.67
  Network of neurons: N = 282 nodes, $\overline{k}$ = 14

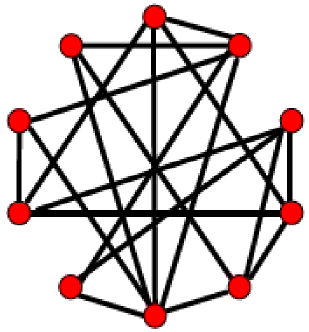
Network	h_{actual}	h_{random}	C_{actual}	C_{random}
Film actors	3.65	2.99	0.79	0.00027
Power Grid	18.70	12.40	0.080	0.005
C. elegans	2.65	2.25	0.28	0.05

h ... Average shortest path length
C ... Average clustering coefficient
“actual” … real network
“random” … random graph with same avg. degree

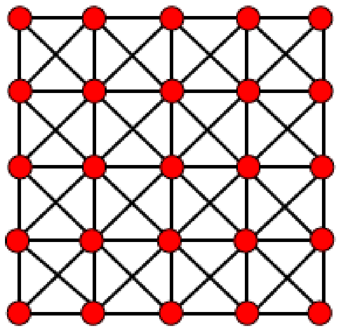


The "Controversy"

- **Consequence of expansion:**
  - **Short paths: *O(log n)***
    - This is "best" we can do if we have a constant degree
  - But ***clustering is low*** !
- **But networks have "local" structure:**
  - ***Triadic closure:*** Friend of a friend is my friend
  - ***High clustering*** but ***diameter is also high***
- ***How can we have both?***



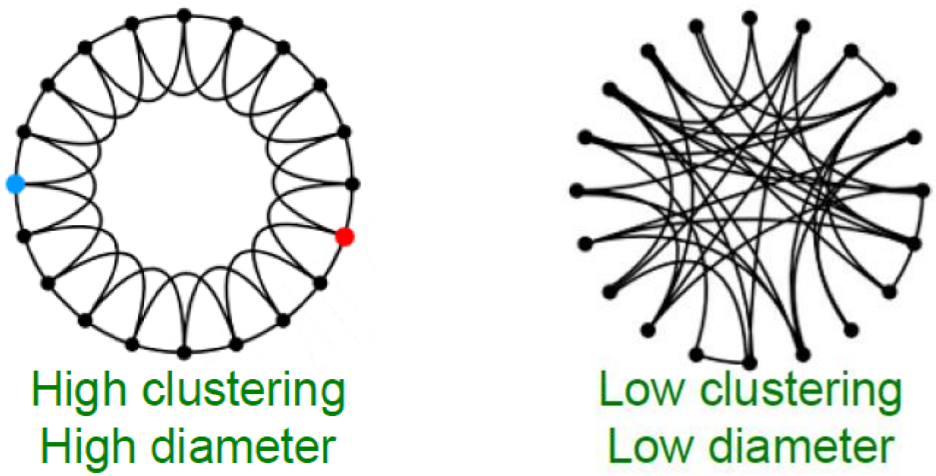
Low diameter
Low clustering coefficient



High clustering coefficient
High diameter

Small-World: How ?

- **Could a network with high clustering be at the same time a small world (log n diameter)?**
  - How can we at the same time have ***high clustering*** and ***small diameter?***



  - **Clustering** implies *“**edge locality**”*
  - **Randomness** enables *“**shortcuts**”*

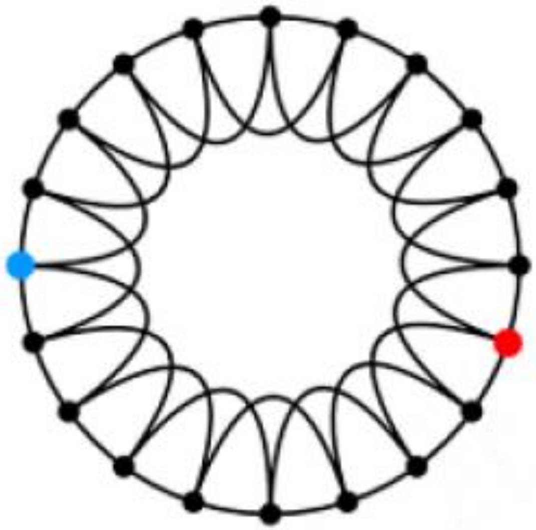
[Watts-Strogatz, '98]

Solution: The Small-World Model

- **Small-World Model** [Watts-Strogatz '98]
- Two components to the model:
  - (1) **Start with a *low-dimensional regular lattice***
    - in our case we using a ring as a lattice
    - Has ***high clustering coefficient***

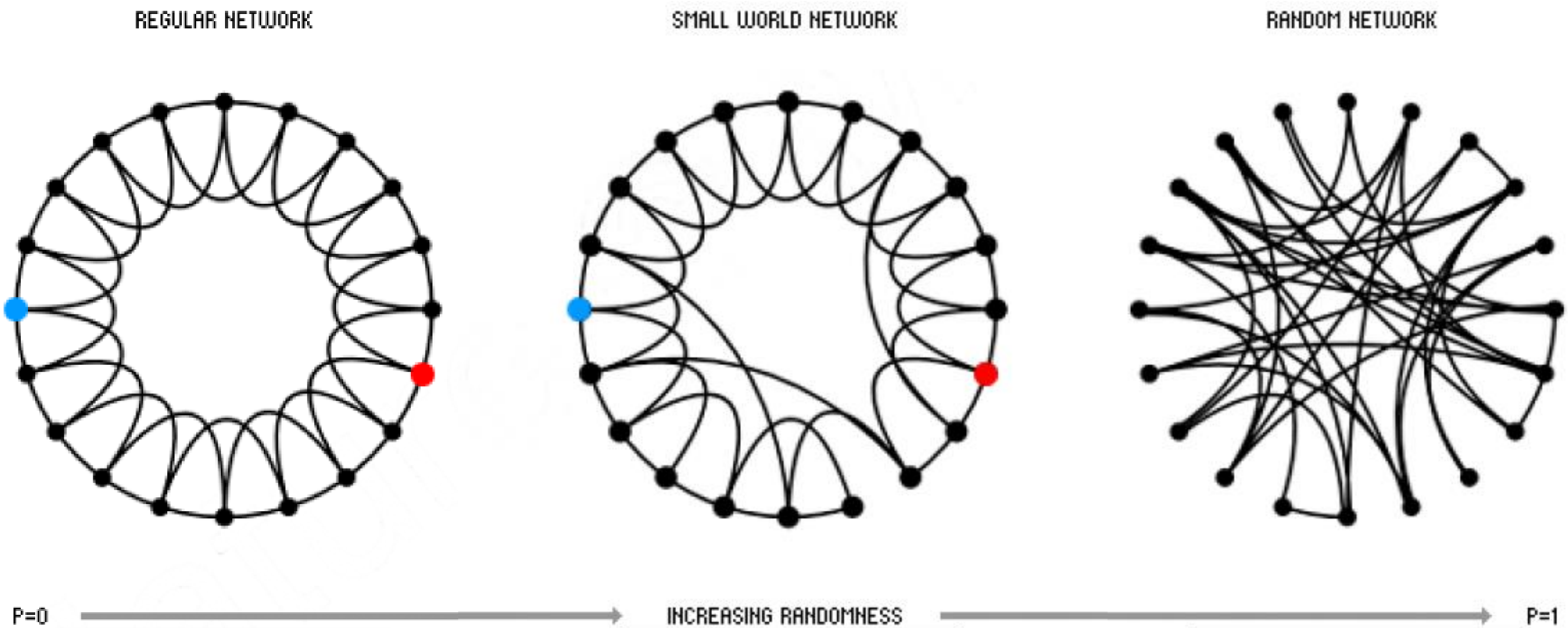
  Now introduce ***randomness*** ("***shortcuts***")

  - (2) ***Rewire:***
    - Add/remove edges to create shortcuts to join remote parts of the lattice
    - For each edge with ***probability p*** move the other end to a random node



[Watts-Strogatz, '98]

The Small-World Model

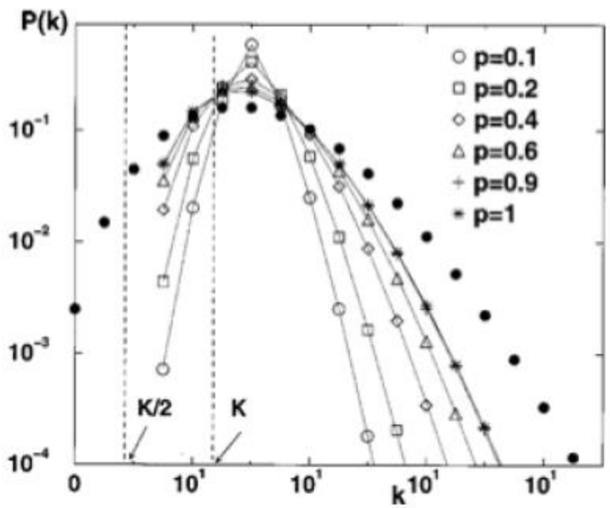


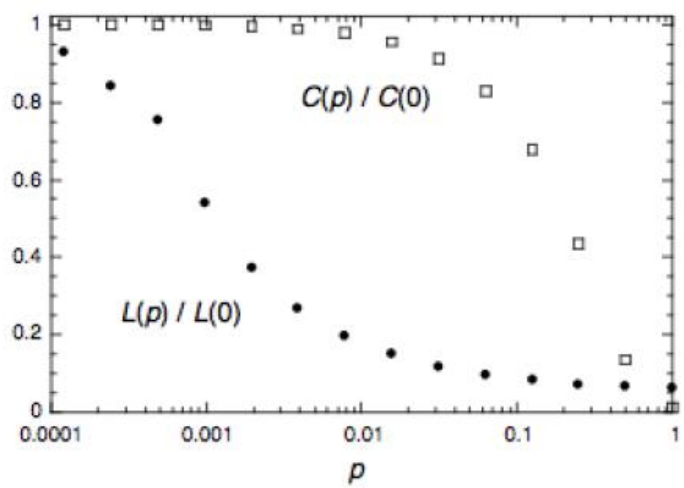
High clustering High diameter	High clustering Low diameter	Low clustering Low diameter
$h = \frac{N}{2\bar{k}}$ $\quad C = \frac{3}{4}$		$h = \frac{\log N}{\log \alpha}$ $\quad C = \frac{\bar{k}}{N}$

Rewiring allows us to "interpolate" between a regular lattice and a random graph

Small World Model

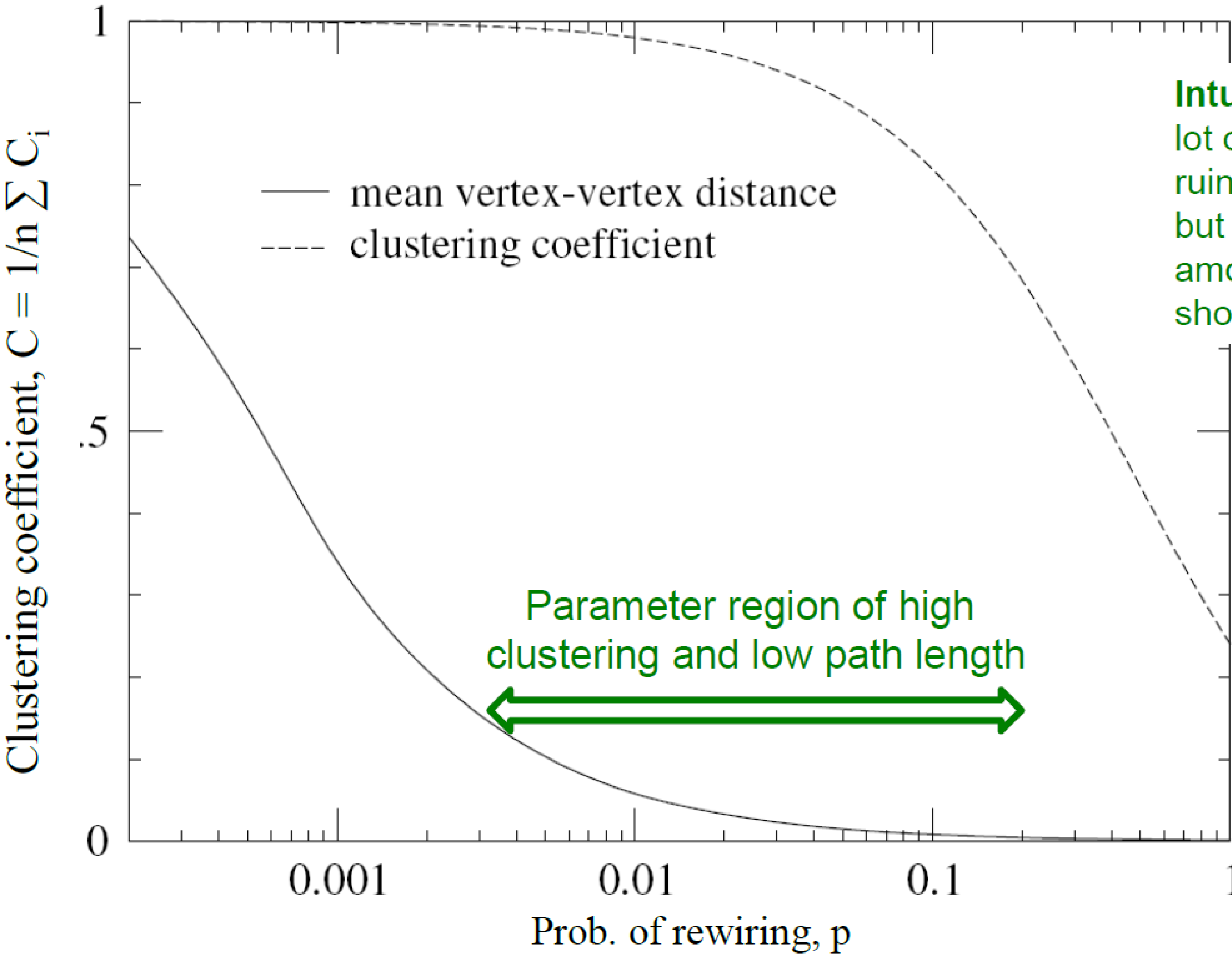
- ***Node degree distribution:***
  - Poisson like
- ***Average path length $\langle L(p) \rangle$:***
  - $p \to 0$, ring lattice. $\langle L(0) \rangle = 2n/k$
  - $p \to 1$, random graph. $\langle L(1) \rangle = \log(n)/\log(k)$
- ***Clustering coefficient $C(p)$:***
  - $p \to 0$, ring lattice. C(0) = ¾ = constant
  - $p \to 1$, random graph. C(1) = k/n





[Watts-Strogatz, '98]

The Small-World Model



Intuition: It takes a lot of randomness to ruin the clustering, but a very small amount to create shortcuts.

Small World Model

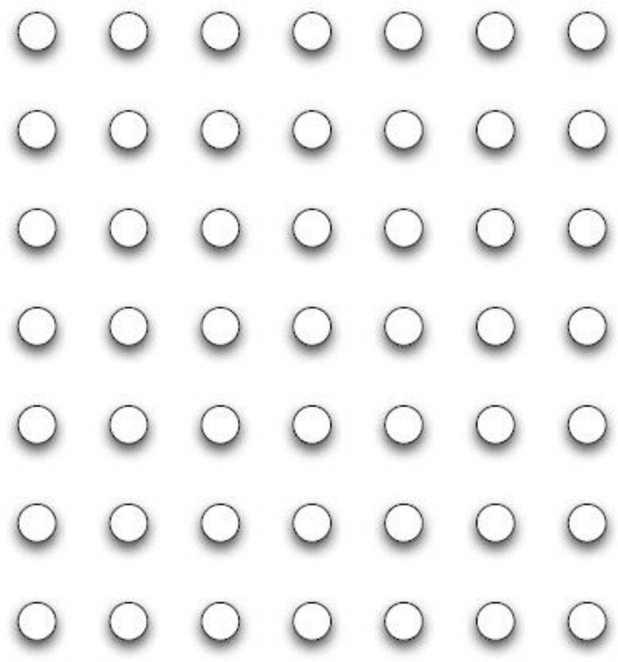
Model follows naturally from a combination of two basic social-network ideas

- ***Homophily***: the principle that we connect to others who are like ourselves
  - creates many triangles
- ***Weak ties***: the links to acquaintances that connect us to parts of the network that would otherwise be far away
  - produce the kind of widely branching structure that reaches many nodes in a few steps

Alternative Formulation of the Model

Suppose that everyone lives on a two-dimensional grid

- Imagine the grid as a model of ***geographic proximity***
  - or more abstract kind of ***social proximity***
- In any case, a notion of ***similarity*** that guides the ***formation of links***



Nodes arranged in a grid

Alternative Formulation of the Model

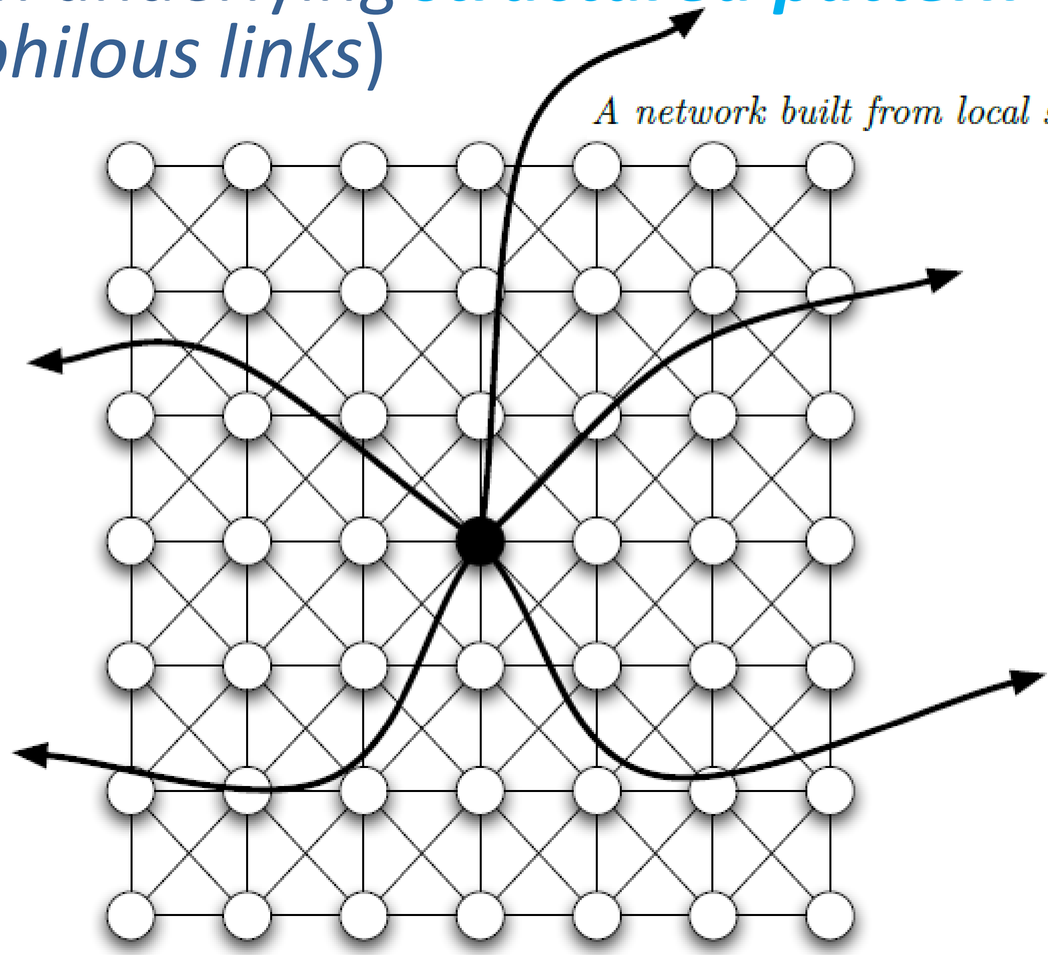
Give each node two kinds of links

- Those explainable purely by ***homophily***
  - captured by having each node form a link to all other nodes that lie within a radius of up to ***r*** grid steps away, for some constant value of ***r***
    - these are the links you form to people because you are ***similar*** to them
- Those that constitute ***weak ties***
  - each node also forms a link to ***k*** (some constant value) other nodes selected uniformly at random from the grid
    - connecting nodes who lie very ***far apart*** on the grid

Alternative Formulation of the Model

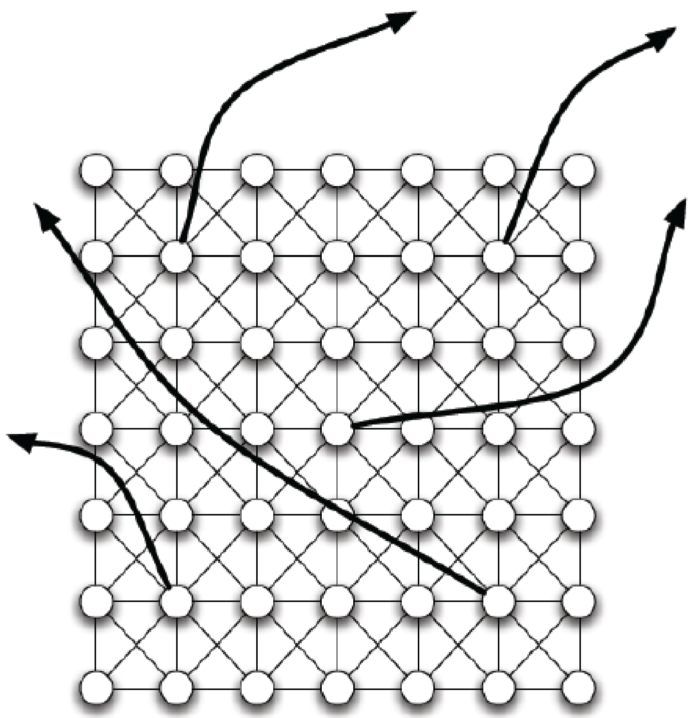
- ***Network is a hybrid structure*** consisting of a ***small amount of randomness*** (*the weak ties*) sprinkled onto an underlying ***structured pattern*** (*the homophilous links*)

A network built from local structure and random edges



Diameter of the Watts-Strogatz

- **Alternative formulation of the model:**
  - Start with a square grid
  - Each node has 1 random long-range edge
    - Each node has 1 spoke. Then randomly connect them.



$$C_i = \frac{2 \cdot e_i}{k_i(k_i - 1)} = \frac{2 \cdot 12}{9 \cdot 8} \geq 0.33$$

There are already 12 triangles in the grid and the long-range edge can only close more.

What's the diameter?

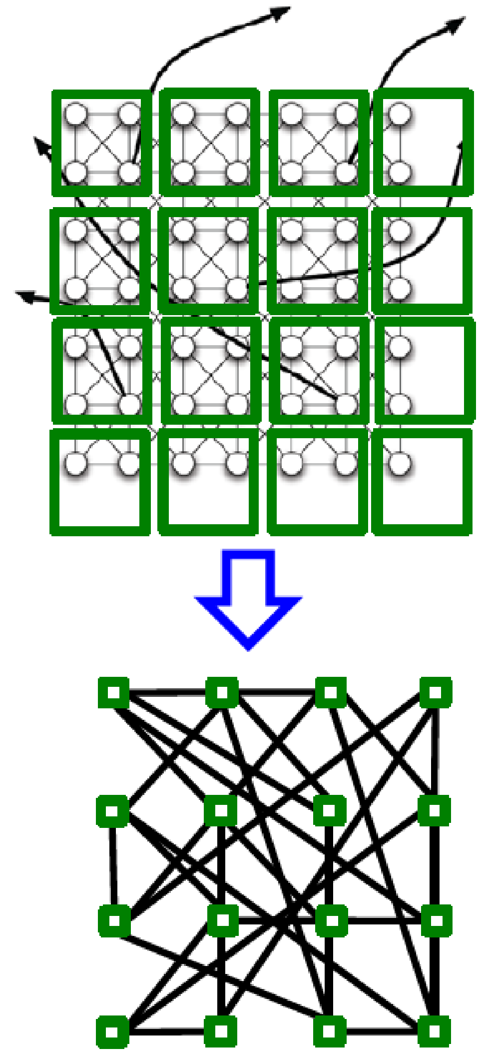
It is $O(log(n))$

Why?

Diameter of the Watts-Strogatz

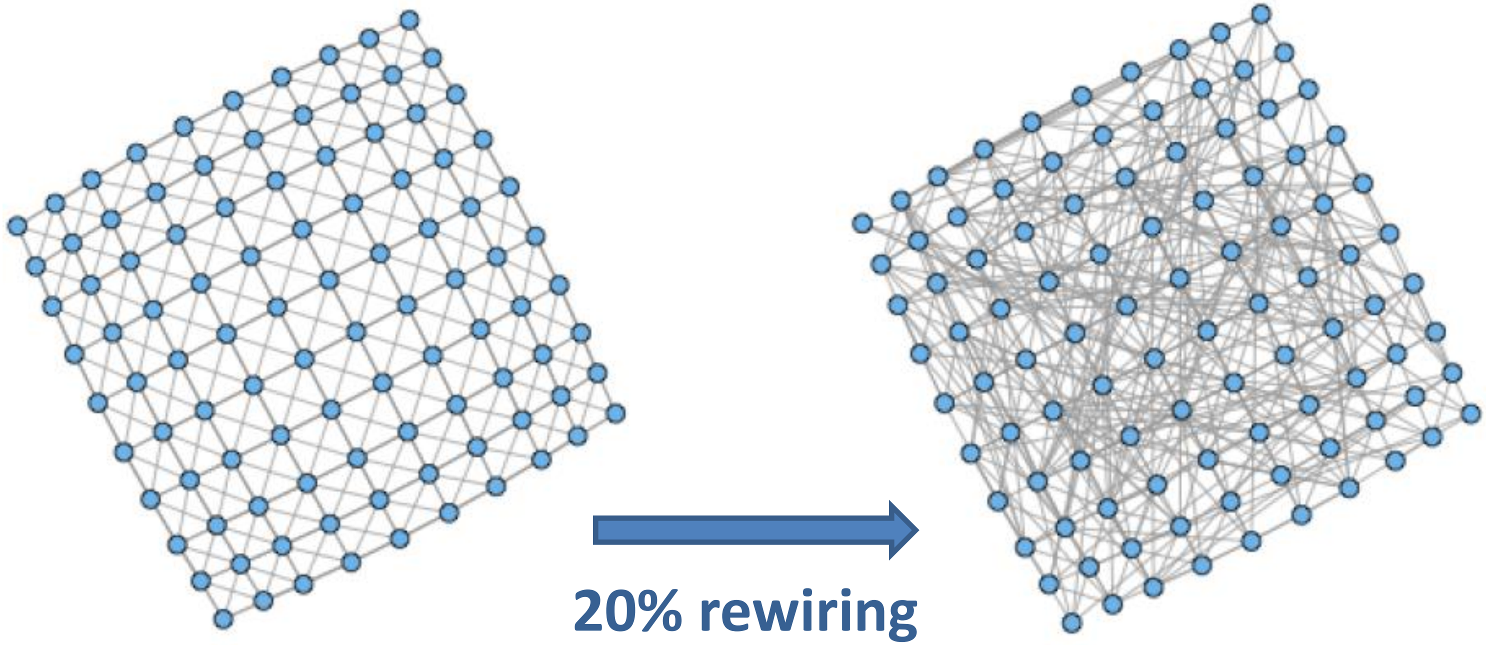
- **Proof:**
  - Consider a graph where we contract 2x2 *subgraphs* into supernodes
  - Now we have 4 edges sticking out of each supernode
    - ***4-regular random graph!***
  - From Thm. we have short paths between super nodes
  - We can turn this into a path in a real graph by adding at most 2 steps per hop

$\Rightarrow$ **Diameter of the model is O(2 log n)**



4-regular random graph

Small World Model – Simulation



Avg. path length = 3.58 → **Avg. path length = 2.32**

Clustering Coeff. = 0.49 → **Clustering Coeff. = 0.19**

Crux of the Watts-Strogatz model

Introducing a tiny amount of randomness,
in the form of long-range weak ties,
is enough to make the world “small,”
with short paths between every pair of nodes

Small-World: Summary

- **Could a network with high clustering be at the same time a small world?**
  - Yes! You don't need more than a few random links
- **The Watts Strogatz Model:**
  - Provides insight on the interplay between ***clustering*** and the ***small-world***
  - Captures the *structure* of many *realistic networks*
  - Accounts for the ***high clustering*** of real networks
  - **Does not** lead to the correct ***degree distribution***
  - **Does not** enable ***navigation***

Model Comparison

	Random	BA model	WS model	Empirical networks
P(k)	$\frac{\lambda^k e^{-\lambda}}{k!}$	k^{-3}	poisson like	power law
C	$\langle k \rangle / N$	$N^{-0.75}$	const	large
$\langle L \rangle$	$\frac{\log(N)}{\log(\langle k \rangle)}$	$\frac{\log(N)}{\log \log(N)}$	$\log(N)$	small

Two Questions

- ***(This Week)* What is the structure of a social network?**
- ***(Next)* What kind of mechanisms do people use to route and find the target?**