Lecture 09
Network Analysis 4

Andrew Sonta

CIVIL 534: Computational systems thinking for sustainable engineering

30 April 2025

Outline

- More on “null” models
  - The random graph model (AKA Erdős–Rényi model)
  - Configuration model
  - Preferential attachment (AKA Barabási–Albert model)
- Network structure
  - Scale-free networks
  - Small-world networks
  - Degree distributions, path lengths, clustering, size of largest con. comp.
- Percolation and resilience (if time)

- Corresponding parts of Newman: 10.1-10.6, 11, 13.1-13.2, 15
  - Book is much more detailed than what we cover in our class
  - Can be treated as an additional resource
- Assignment 2 due Friday
- Exam 2 in two week

Random graph model – why?

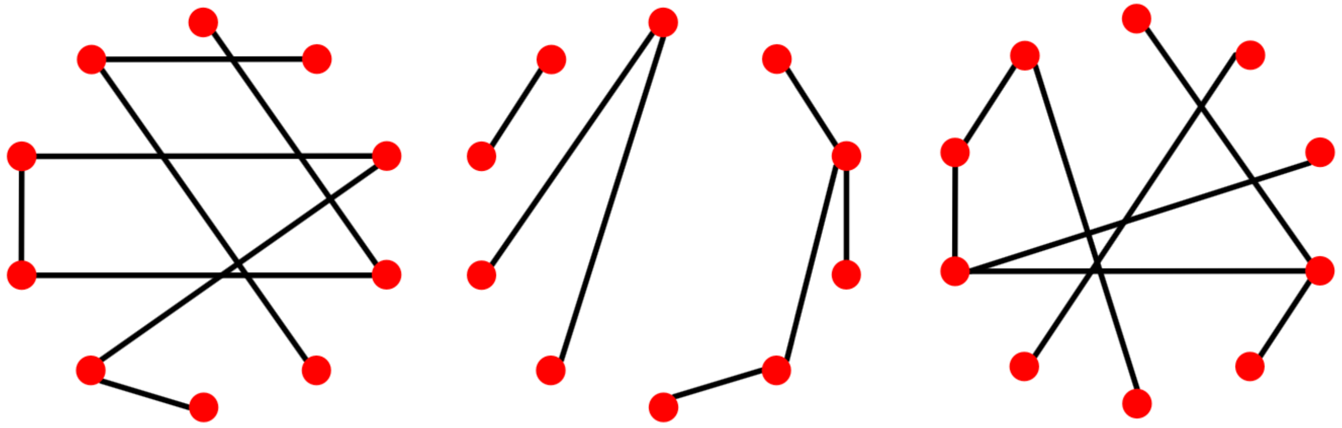
- **Point of comparison** – we have already seen a few examples of comparing a real network to a “random” reference
  - Definition and computation of modularity
  - Motif significance (assignment 4)
- Help us answer the question: to what extent is a particular property of a network the result of some random process?
- “All models are wrong but some are useful”
  - George E.P. Box

G(n,p) AKA Erdős–Rényi model

- Undirected graph
- Fix two things:
  - Number of nodes (n)
  - *Probability* that an edge exists between nodes (p)
- More precisely: each possible edge appears with probability *p*.
- What type of network does this produce?

G(n,p)

- n and p do not uniquely determine a graph
- Since the graph is a result of a random process, we can have many different realizations given the same parameters



n = 10
p= 1/6

Properties of G(n,p)

- We can calculate many different properties to summarize the network
  - Degree distribution
  - Clustering coefficient
  - Size of the connected component
  - Average shortest path length
  - …

Degree distribution of G(n,p)

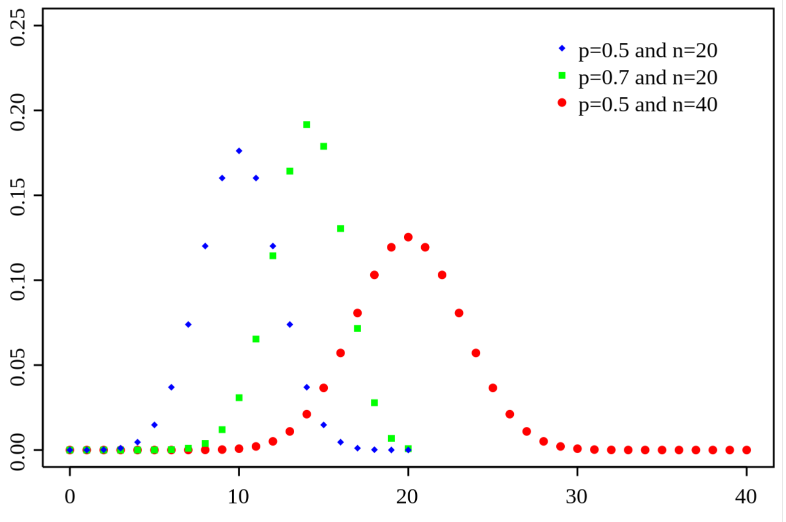
- Each node in the network is connected to the others (n – 1) with probability p
- What is the probability of a node being connected to a particular k other nodes?

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

select k nodes out of n–1

probability of the node having k edges

probability of the node not connecting to the remaining nodes



- Binomial distribution

Clustering coefficient of G(n,p)

- Definition:
  - $c_i = \frac{\text{\# of pairs of neighbors } i \text{ that are connected}}{\text{\# of pairs of neighbors } i}$
  - $c_i = \frac{2e_i}{k_i(k_i-1)}$ where e_i is the number of edges between the neighbors of i
- In G(n,p), the probability that *any* two nodes are connected is *p*

$$c_{avg} = p = \frac{\bar{k}}{n-1}$$

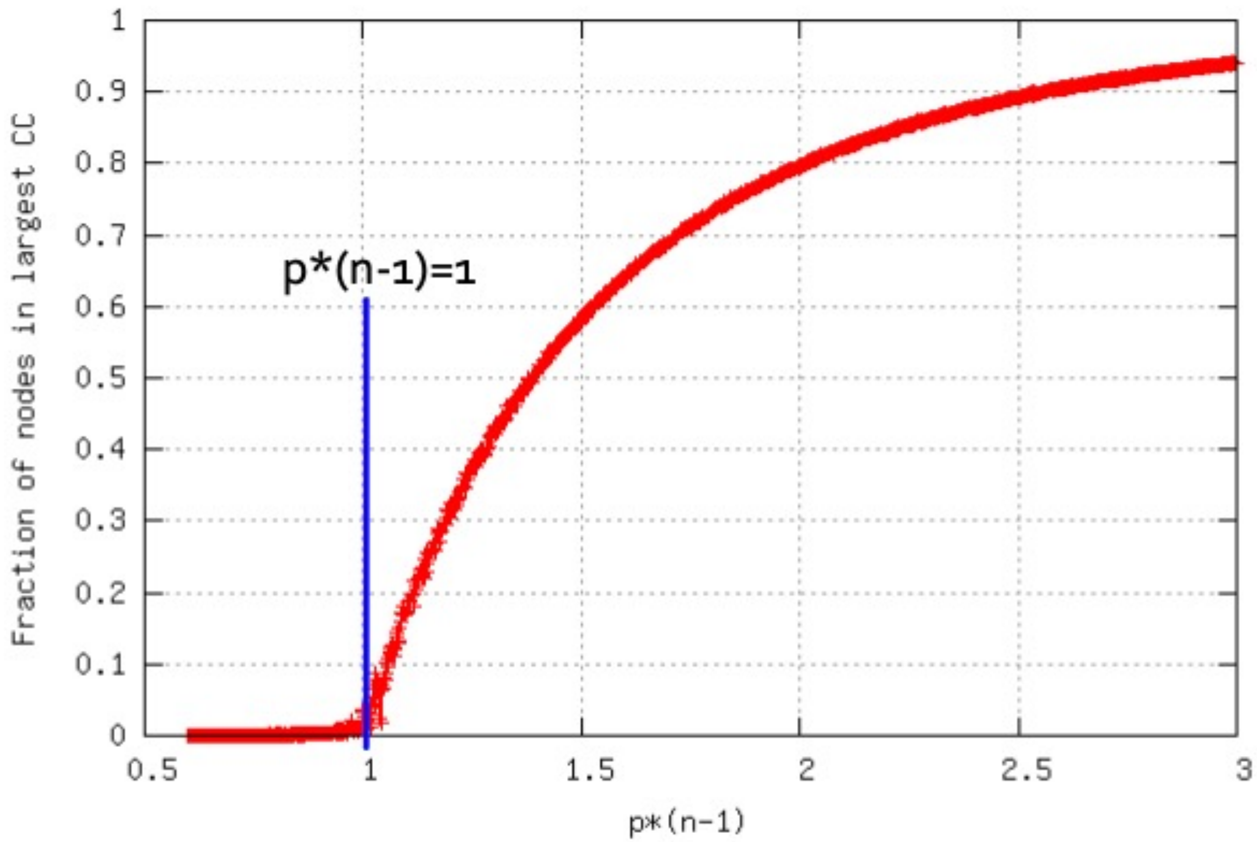
- If we hold k constant, as the graph grows large *p* must remain small and we see a small clustering coefficient

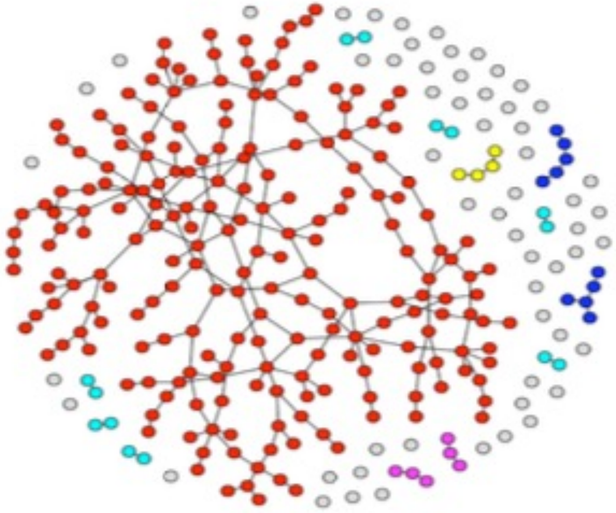
Size of the largest connected component of G(n,p)

- Let’s simulate it!

Size of the largest connected component of G(n,p)

- Much larger simulation (n=100,000)

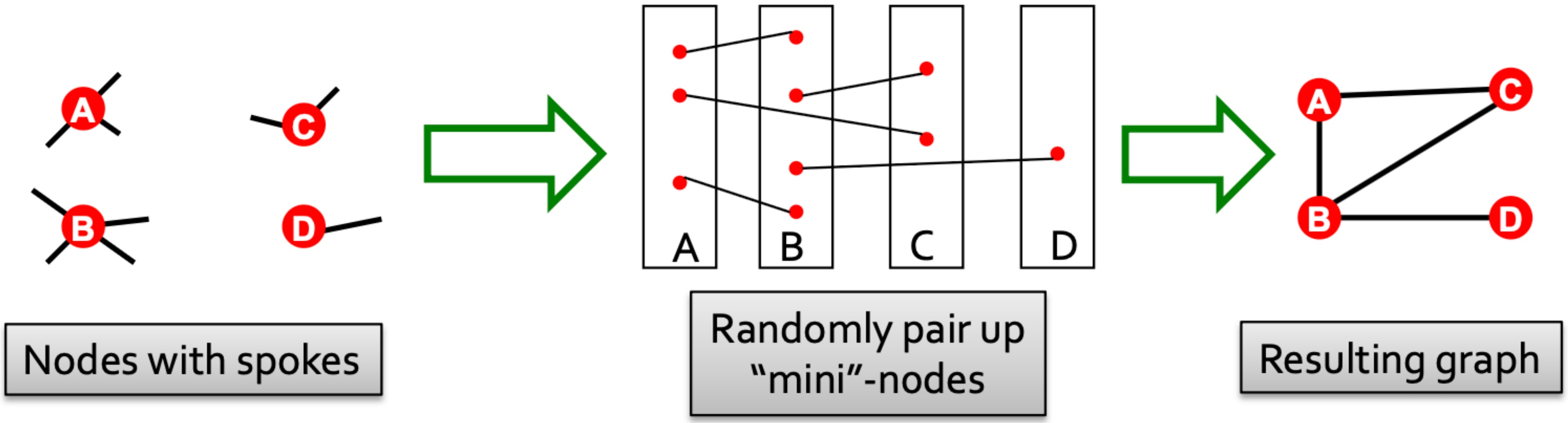




Fraction of nodes in the largest component

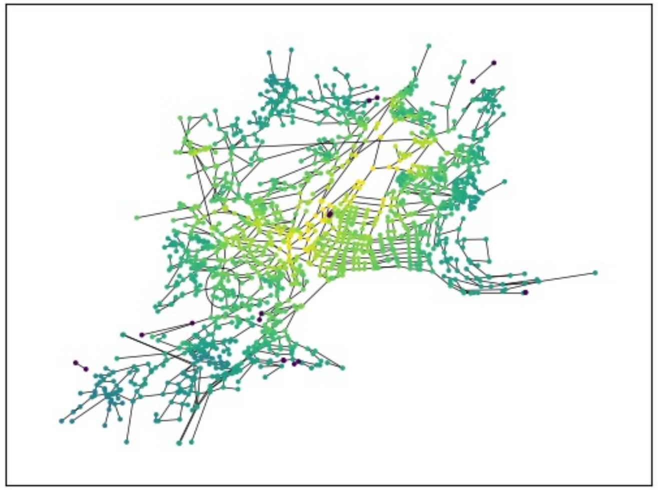
Side note: Configuration model (Alternative random graph model)

- Generate a random graph with a fixed degree sequence (k_1, k2, …, k_N)



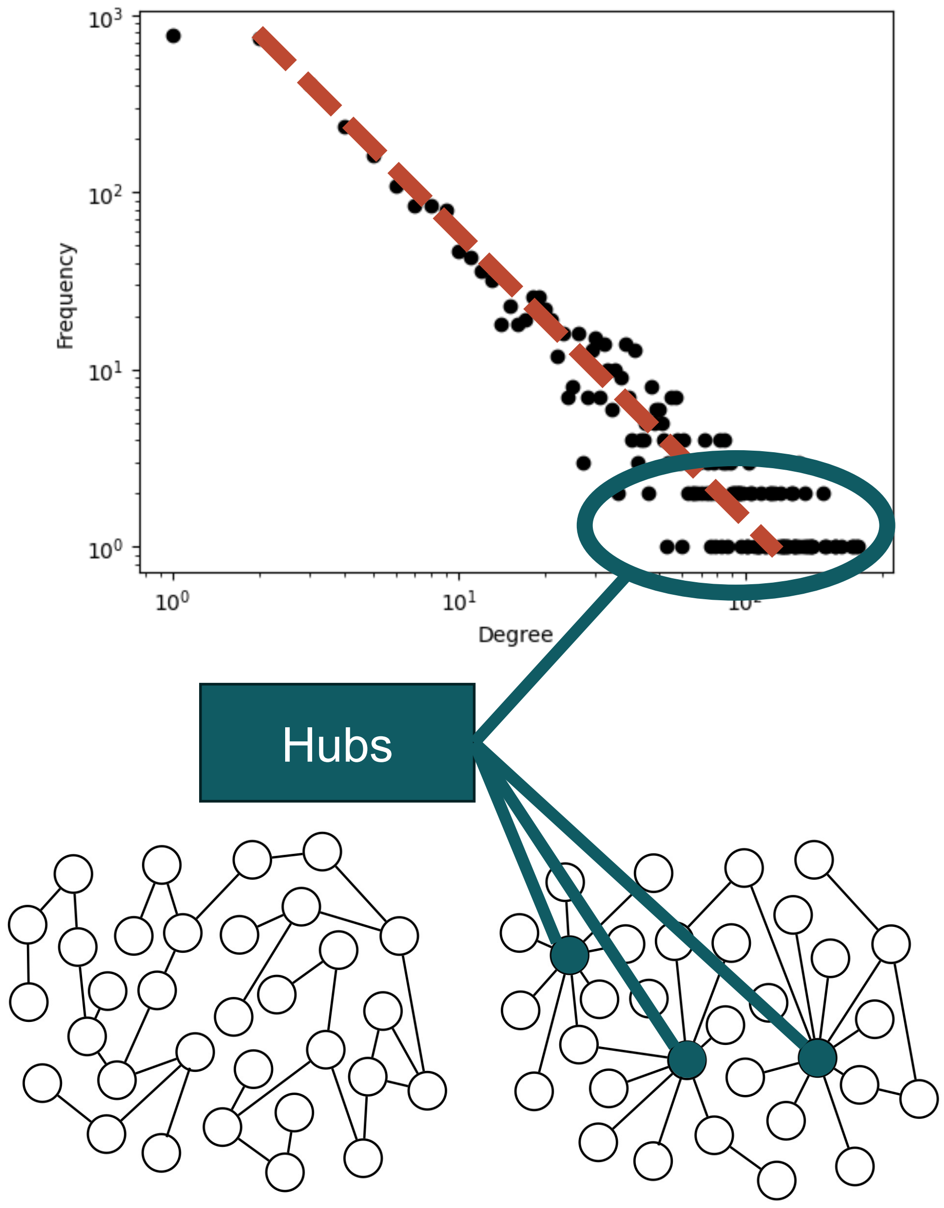
How does Erdős–Rényi compare to a real network?

- Let's look at two mobility infrastructure networks together
  - Network 1: Small segment of the Switzerland road network
    - Nodes: intersections
    - Edges: roads
  - Network 2: Global network of air transportation
    - Nodes: airports
    - Edges: if there exists a flight between them



“Scale free” networks

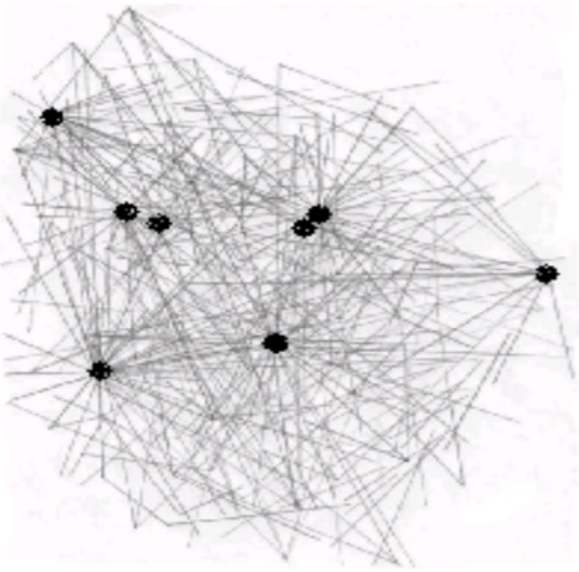
- Degree distribution follows a power law
  $$P(k) \sim k^{-\alpha}$$
  Typically, α is between 2 and 3
- Common in certain types of networks:
  - Social networks
  - Internet
  - Airline networks
  - Anywhere we might find “hubs”

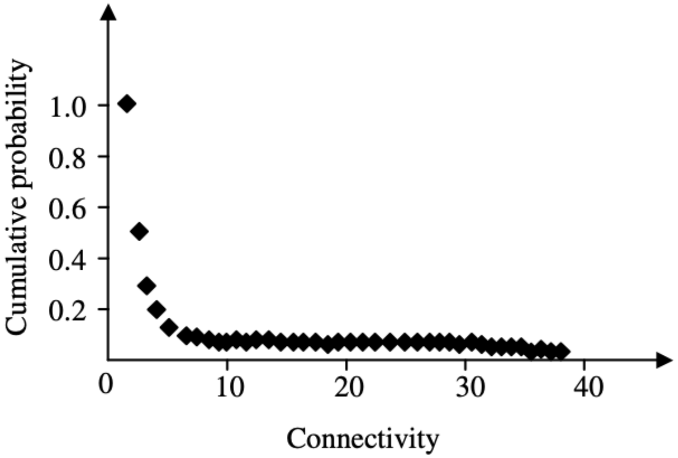


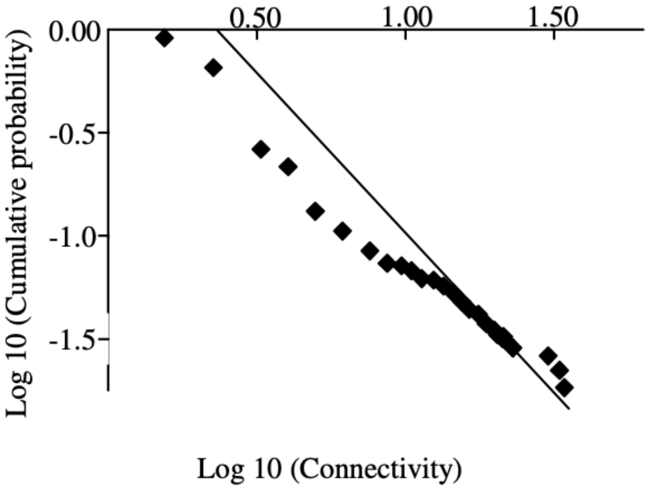
(a) Random network (b) Scale-free network

Urban systems example

- Example: Urban transit network in Beijing
- Nodes: transit sites
- Edges: connections between sites
- Found that node degree followed a scale-free pattern



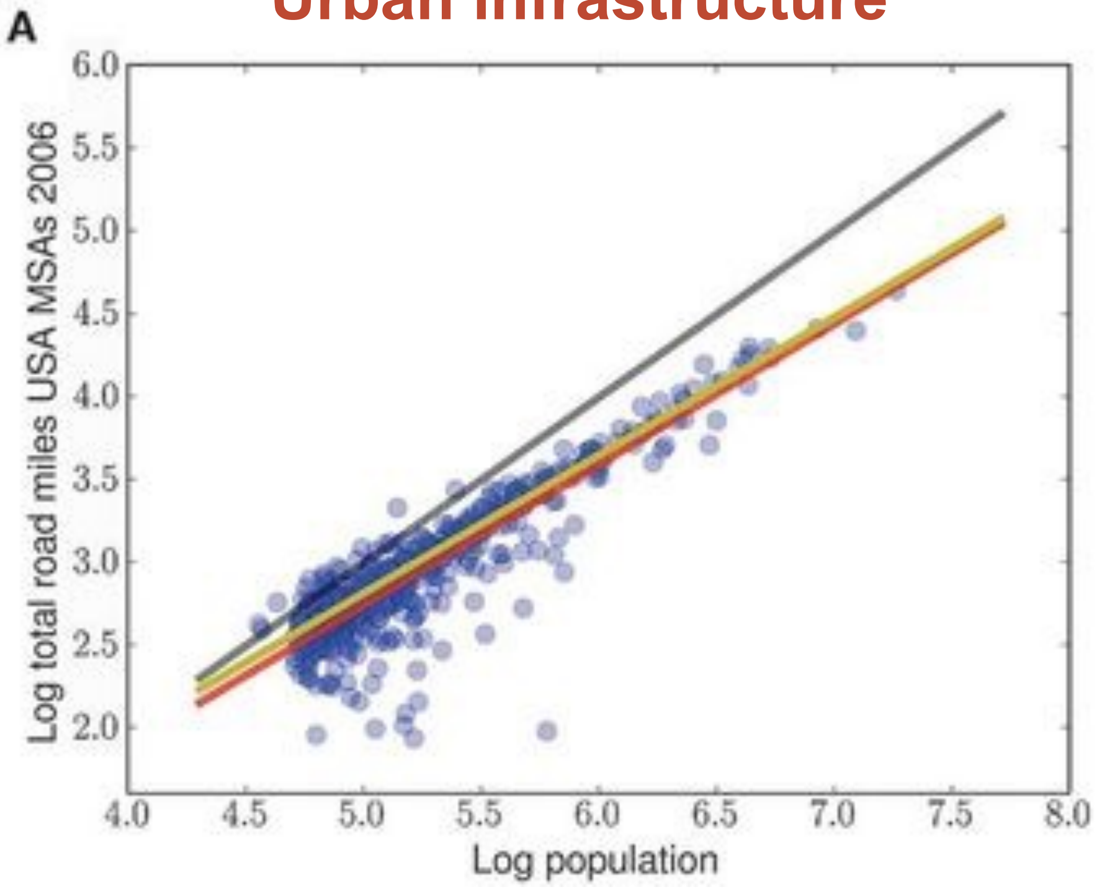




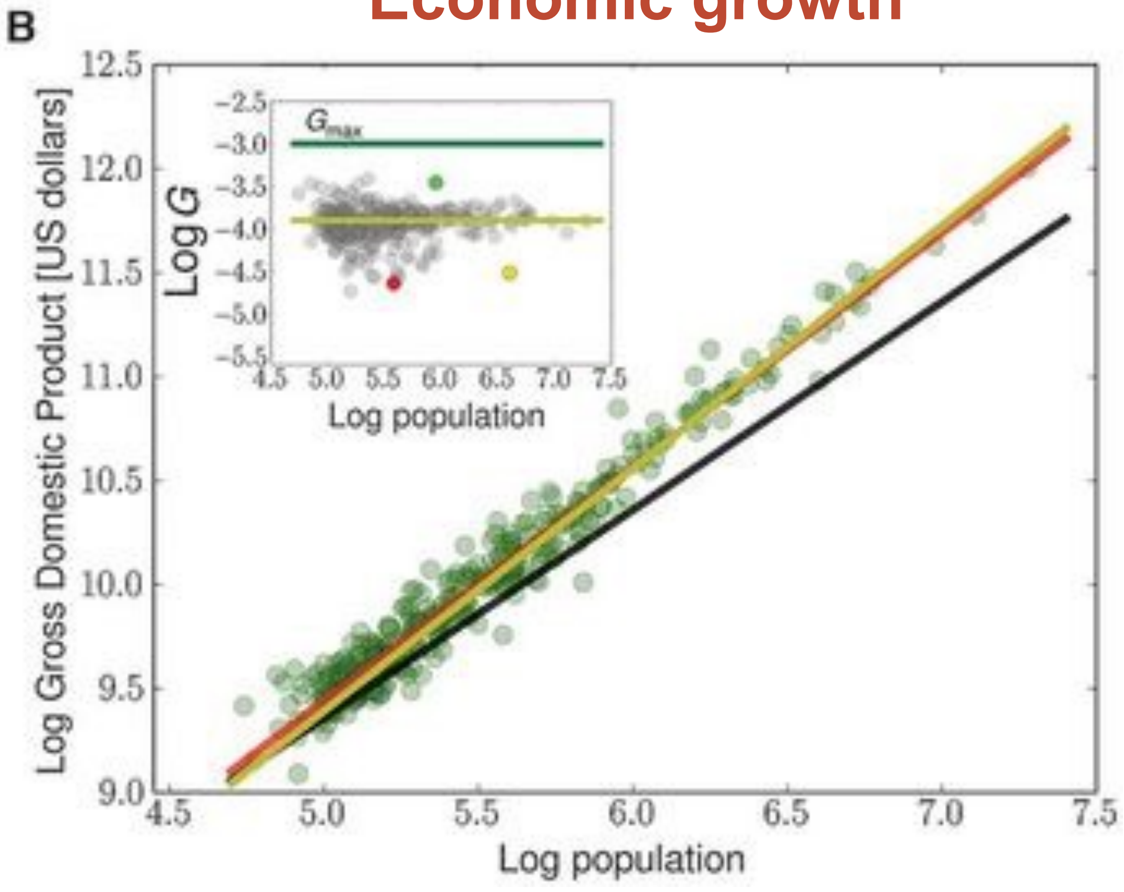
https://doi.org/10.1142/S021798490400758X

We have seen this power law property before

Urban infrastructure

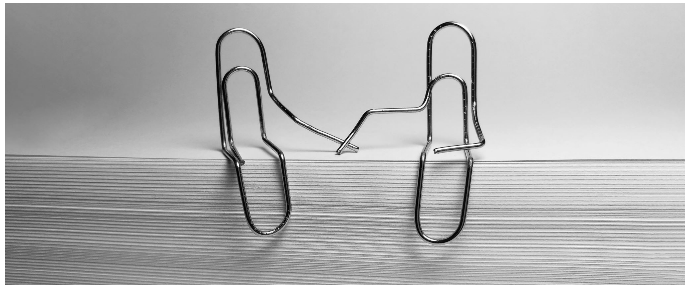


Economic growth



Luís M. A. Bettencourt. The Origins of Scaling in Cities. *Science* **340**,1438-1441(2013). DOI:[10.1126/science.1235823](https://doi.org/10.1126/science.1235823)

Preferential attachment (Barabási–Albert)



- A simple model for how we get scale-free properties in networks
- Algorithm:
  - Start with n_0 nodes
  - Add a node
  - Connect new node to $n \leq n_0$ nodes according to the following probability:

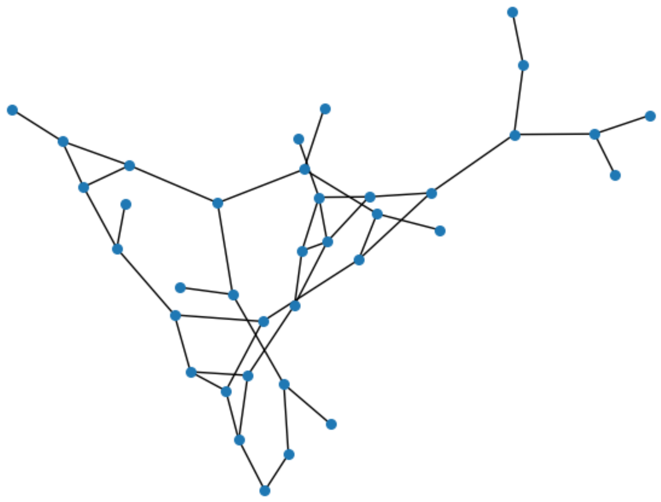
$$p_i = \frac{k_i}{\sum_j k_j}$$

Average shortest path length

- Distance between any two nodes is the number of edges along the **shortest path** between them
- Average shortest path length (average path length) is the average distance between pairs of nodes normalized by the maximum number of edges we could see in a graph - n*(n-1)
- If graph is not connected, we usually measure the average path length of the largest connected component

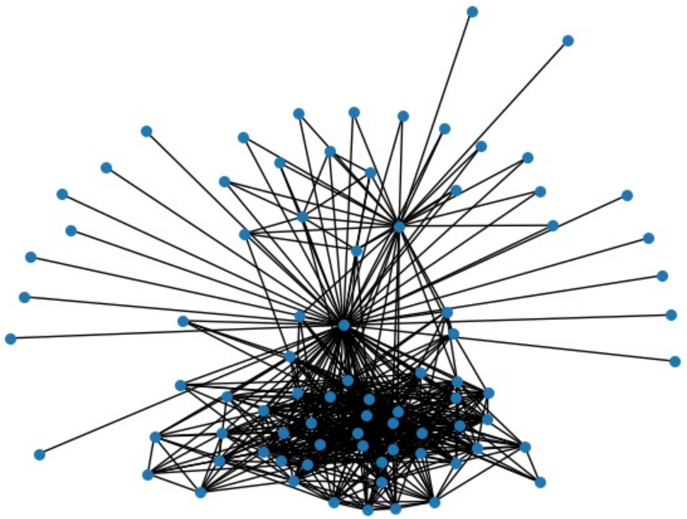
Degree distribution comparison

Non-random degree distribution



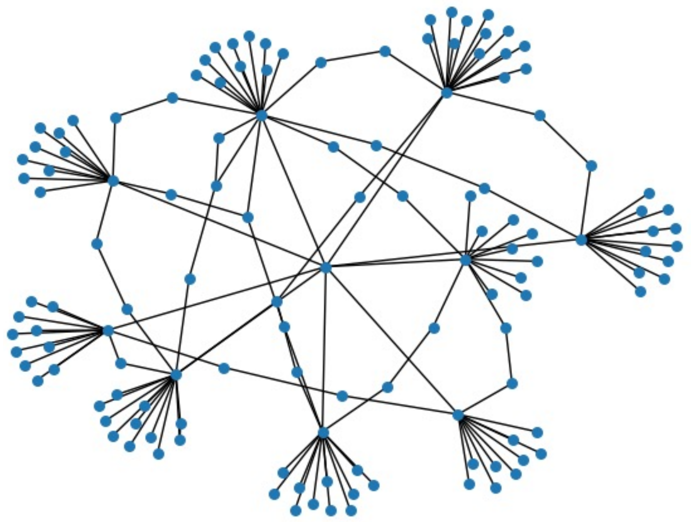
Road subgraph

Power-law degree distribution



Air travel subgraph

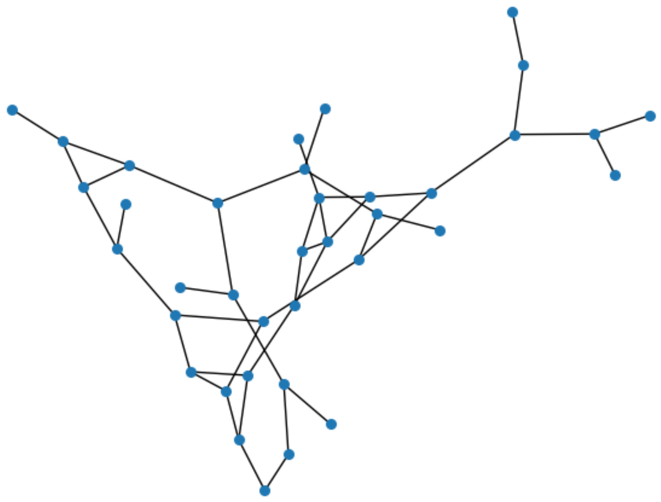
Binomial degree distribution



Random (ER) subgraph

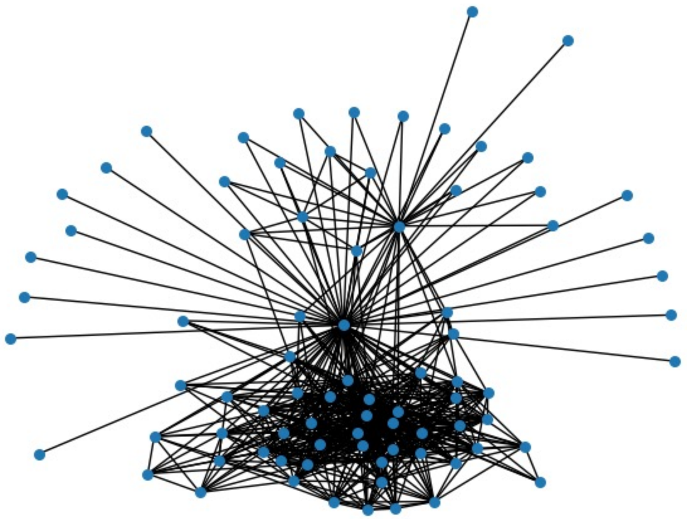
Average path length comparison

Non-random degree distribution
Large average path length



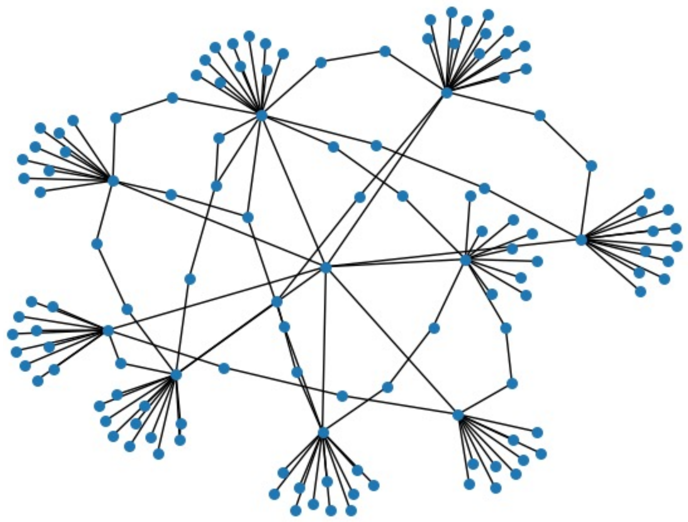
Road subgraph

Power-law degree distribution
Small average path length



Air travel subgraph

Binomial degree distribution
Small average path length



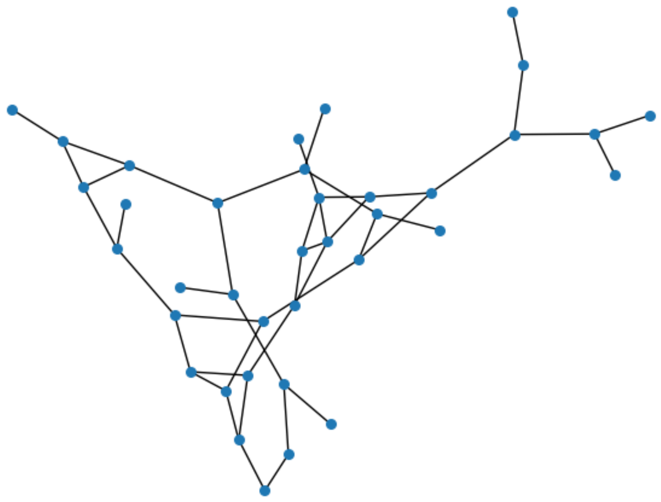
Random (ER) subgraph

Clustering coefficient comparison

Non-random degree distribution

Large average path length

Large clustering coefficient



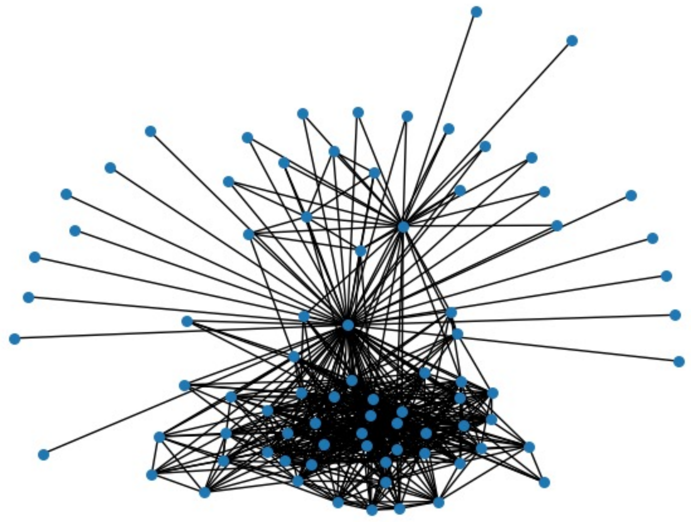
Road subgraph

Power-law degree distribution

Small average path length

Large clustering coefficient

???

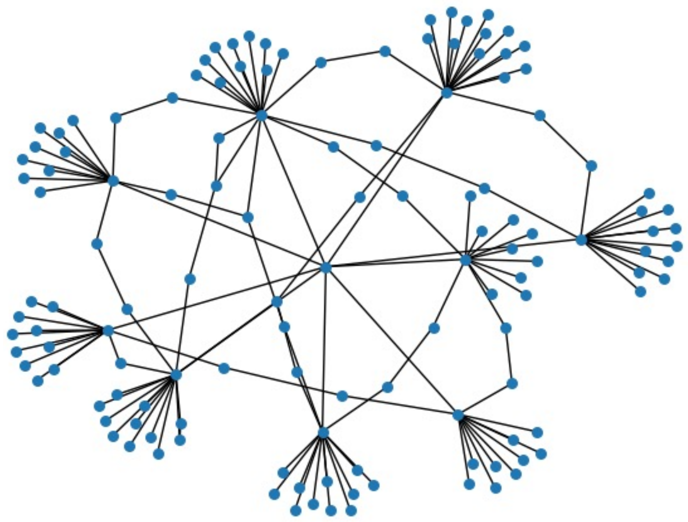


Air travel subgraph

Binomial degree distribution

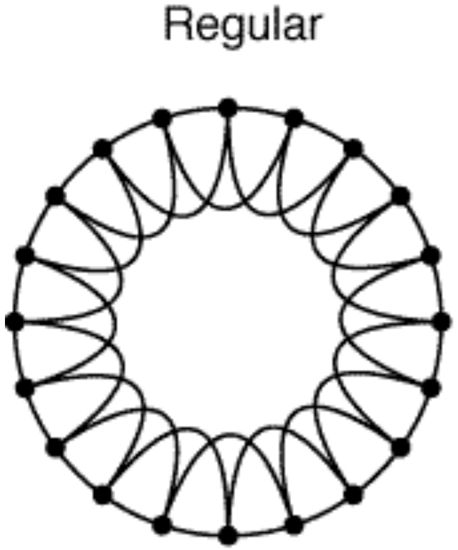
Small average path length

Small clustering coefficient

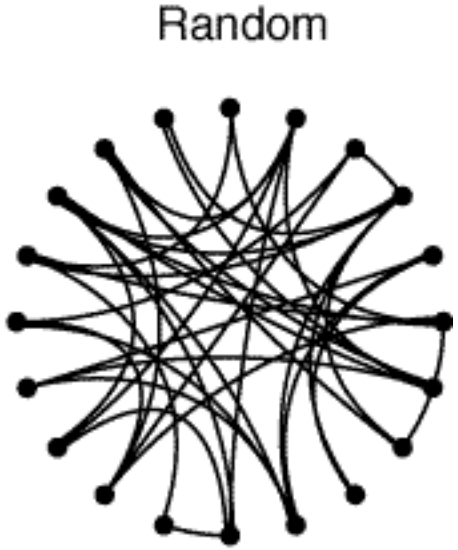


Random (ER) subgraph

How can we have high clustering and small average path distance?



Clustering implies "local structure"

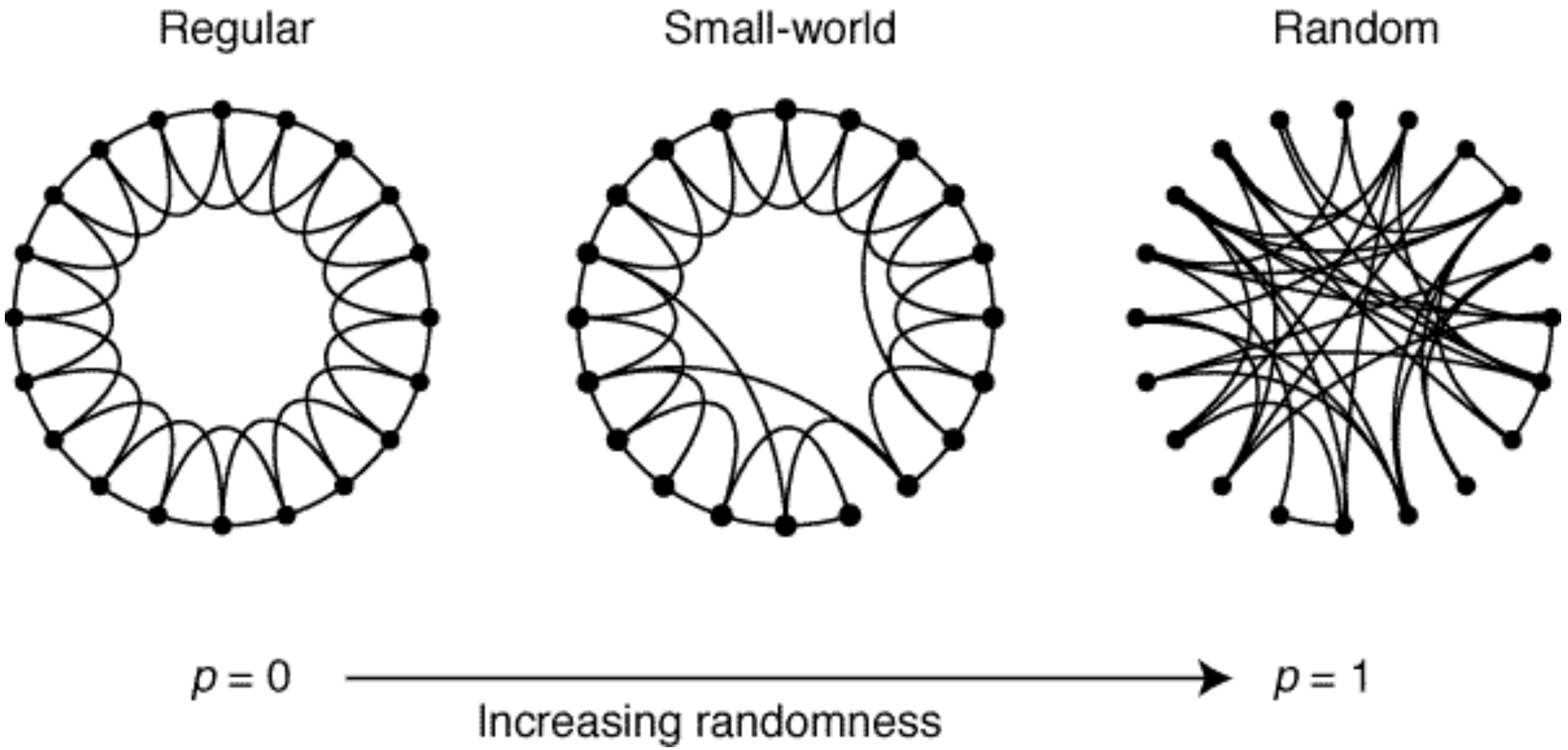


Randomness enables "shortcuts"

Watts, D., Strogatz, S. Collective dynamics of 'small-world' networks. Nature 393, 440–442 (1998). https://doi.org/10.1038/30918

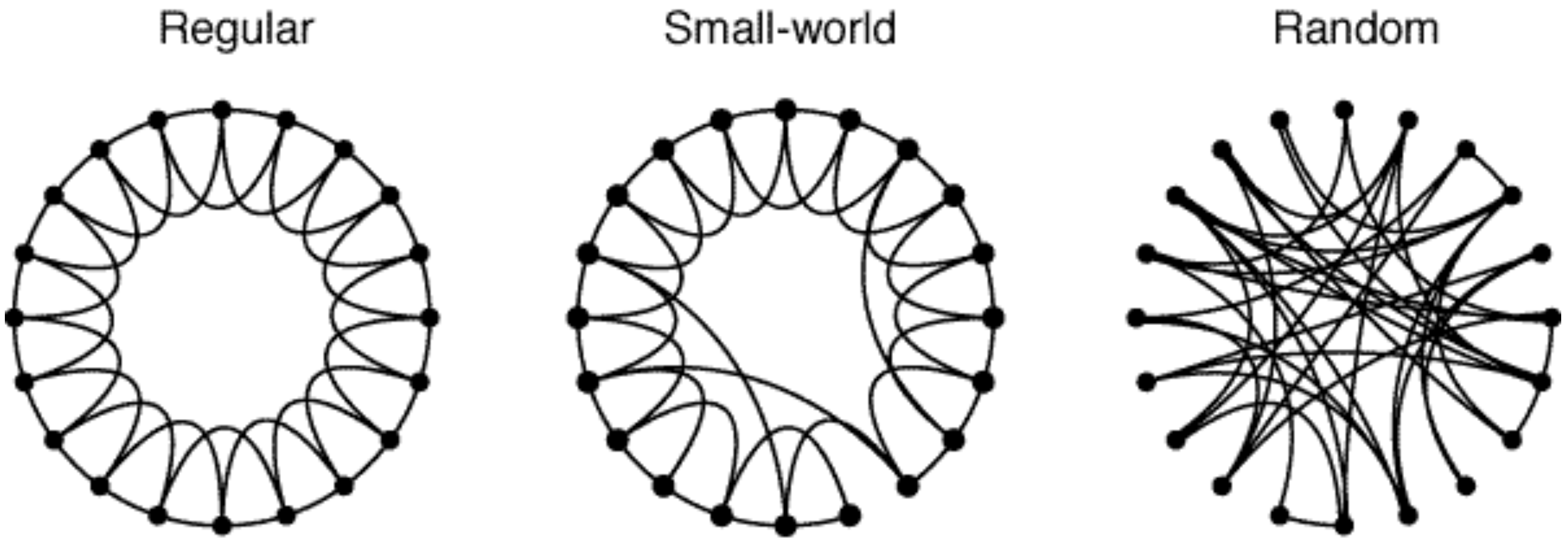
Small world networks (Watts-Strogatz)

Start with regular local structure and randomly rewire



Watts, D., Strogatz, S. Collective dynamics of 'small-world' networks. Nature 393, 440–442 (1998). https://doi.org/10.1038/30918

Small world networks (Watts-Strogatz)



High clustering
High path lengths

High clustering
Low path lengths

Low clustering
Low path lengths

Watts, D., Strogatz, S. Collective dynamics of 'small-world' networks. Nature 393, 440–442 (1998). https://doi.org/10.1038/30918

Small world effects

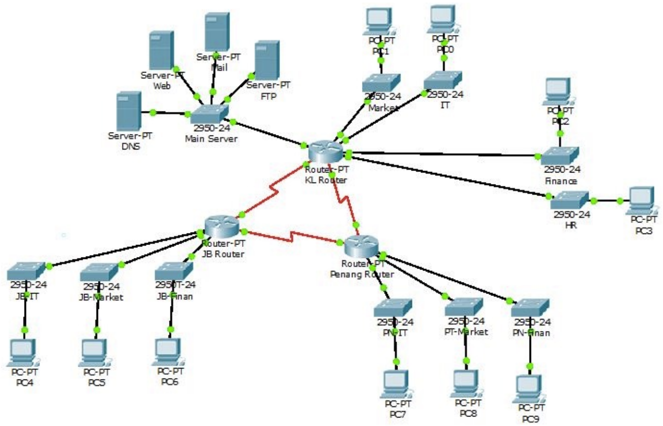
- Milgram's famous 1967 experiment:
  - Picked 300 people in Nebraska and Kansas (middle of nowhere)
  - Asked them to get a letter to a stock-broker in Boston by passing the letter through friends
  - This was successful 64 times
  - On average, it took 6.2 steps → "six degrees of separation"
  - Surprisingly small path length given that social networks are known to have high clustering
- Another example:
  - My Erdős number is 4

Scale free vs. Small world

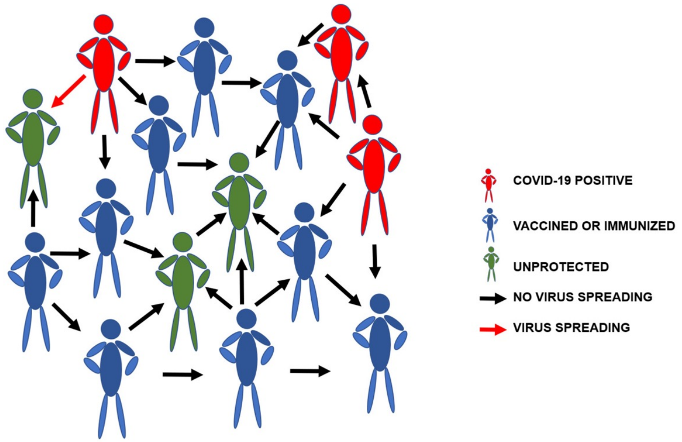
- Important to remember that these are not the same thing
- The primary properties of scale-free networks are:
  - **Power-law distribution of node degrees**
  - Often, we see a power-law distribution of clustering coefficients
  - Often, we see a small average path length
- The primary properties of small world networks are:
  - **High clustering coefficient**
  - **Small average path length**
  - No specific expectation on degree distributions
- Networks can be both at the same time!

Percolation

- The process of removing nodes from the network and the associated edges
- Helps us understand "what-if" scenarios
- Examples:



Internet



"Herd immunity"

Lots of urban systems!

Simulating percolation

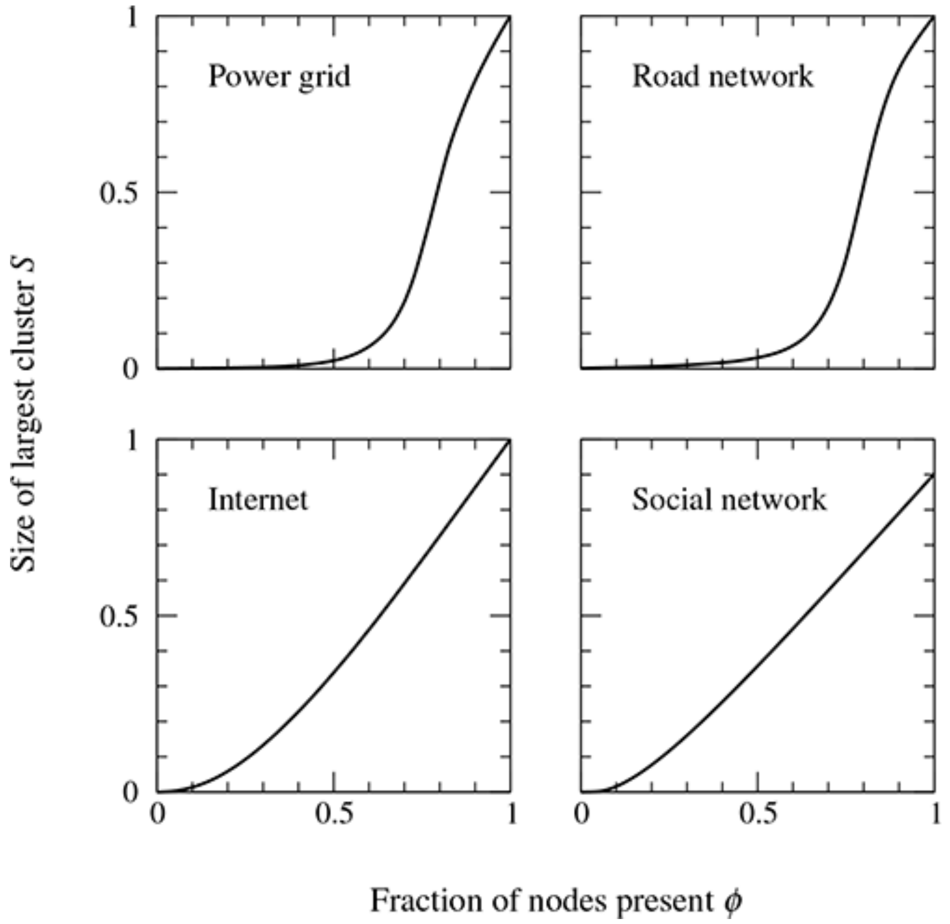
- Key parameter: ϕ (occupational probability)
  - $\phi = 1$ → all nodes are present
  - $\phi = 0$ → all nodes have been removed
- Process of removing nodes
  - Uniform: all nodes have same probability of removal
  - Non-uniform: based on node properties (e.g., different forms of centrality)
  - The choice is normally driven by the domain

Percolation theory

- Newman goes into detail in Chapter 15
  - (And there can be a lot of mathematical detail)
- Percolation is an entire subfield of network science
- Offers an understanding of network resilience and robustness

Qualitative percolation analysis example

Uniform removal of edges



What does this signify?

Qualitative percolation analysis example

Targeted edge removal (target largest degree first)