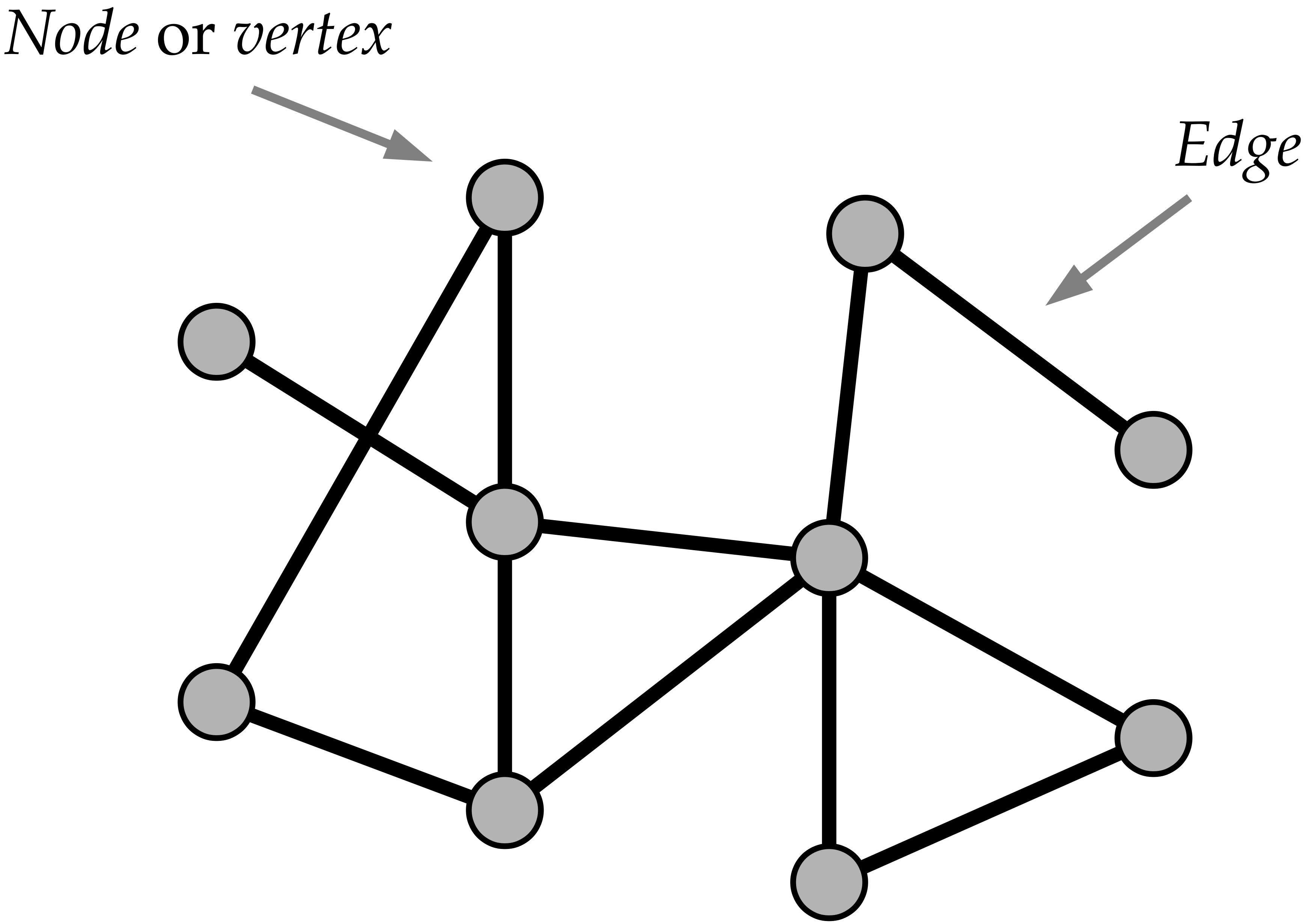
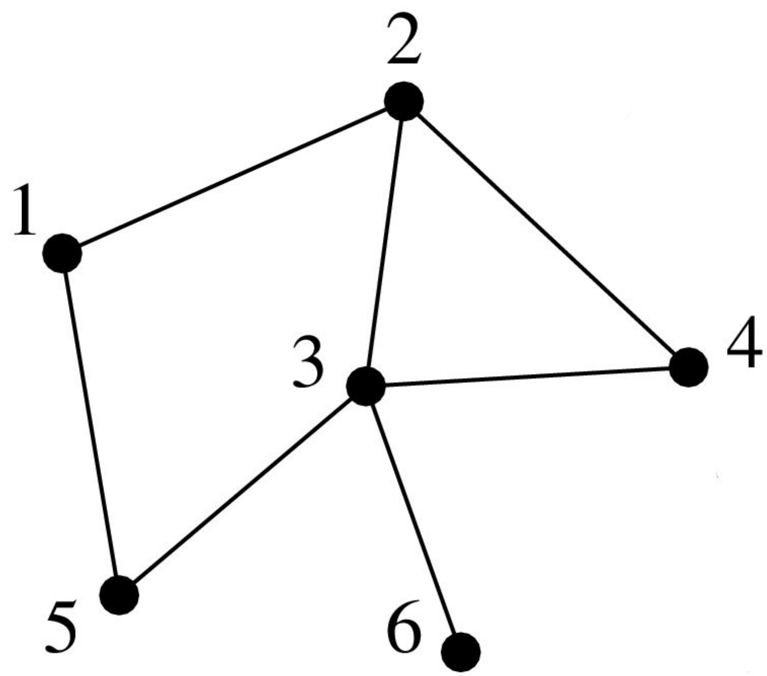

Node or *vertex*
Edge

Adjacency matrix

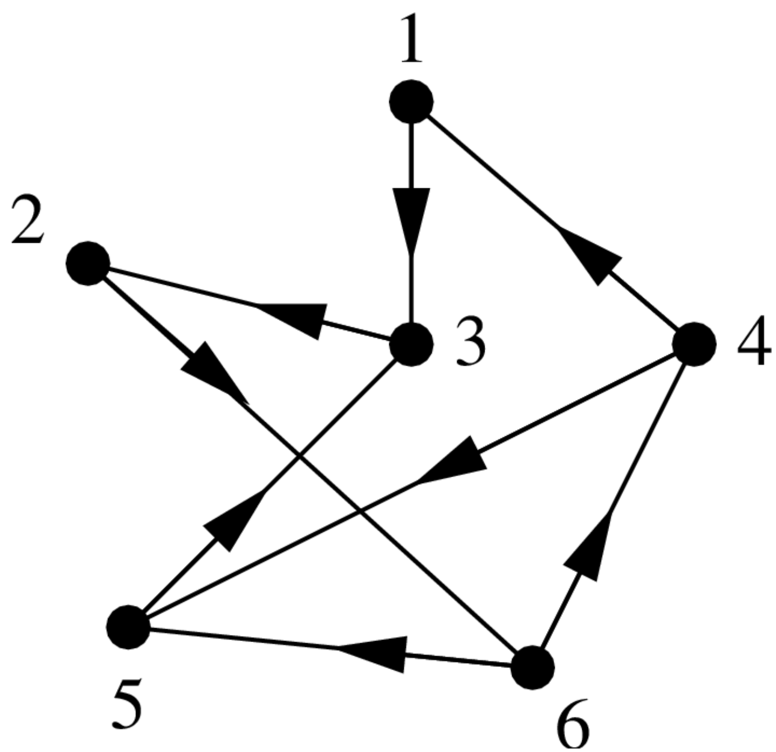
- The basic mathematical representation of a network is its *adjacency matrix*:



$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

Directed network

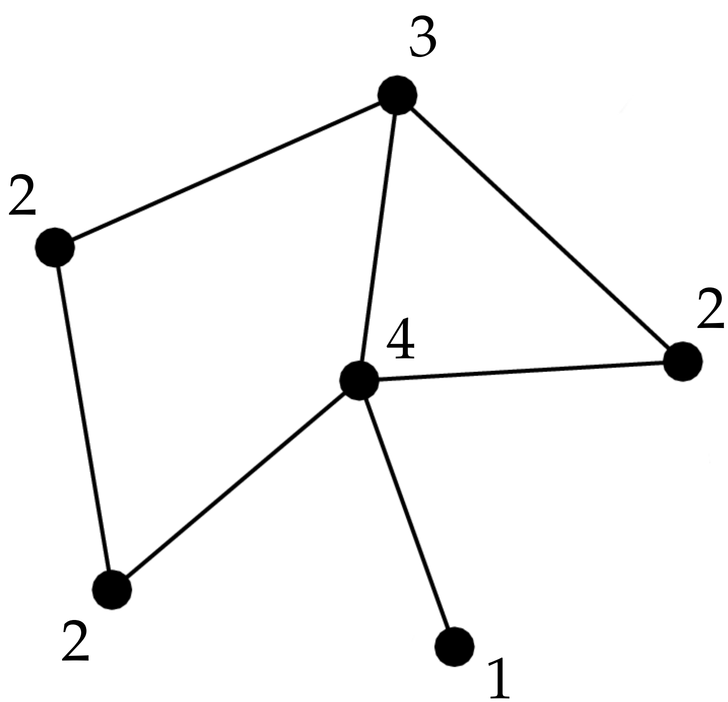
- For a directed network $A_{ij} = 1$ if there is an edge *from j to i*



$$\mathbf{A} = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Degree

- The degree of a node is just the number of connections it has

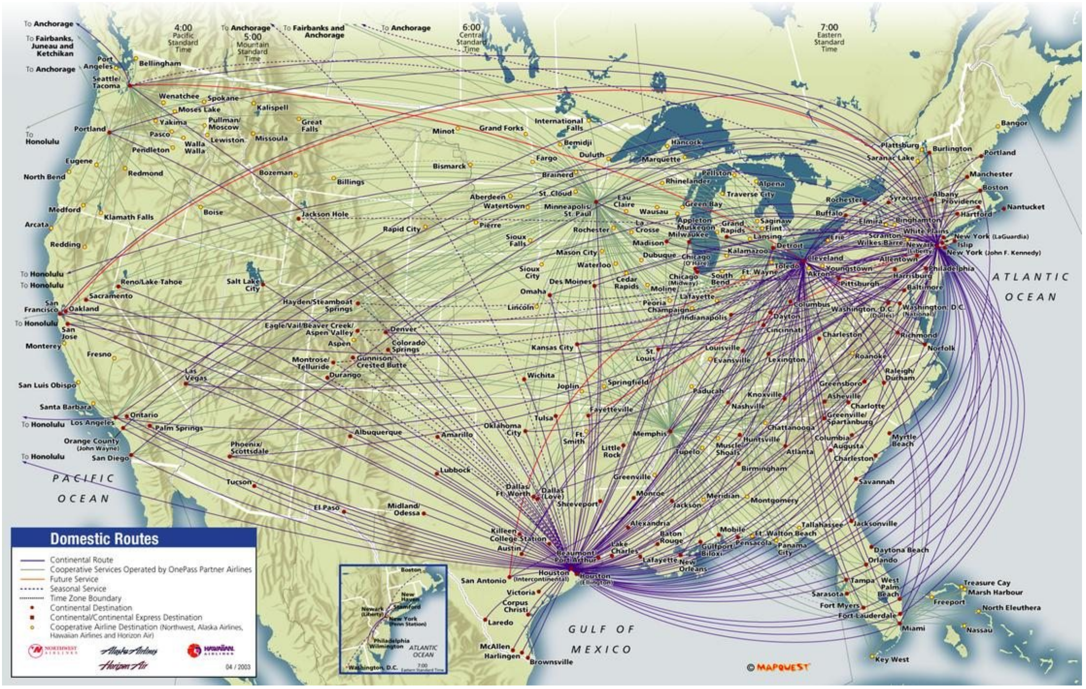


$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

- Degree: $$k_i = \sum_{j=1}^{n} A_{ij}$$

- Sum of degrees: $$\sum_{i=1}^{n} k_i = \sum_{ij} A_{ij} = 2m$$

- Average degree: $$c = \frac{1}{n} \sum_{i=1}^{n} k_i = \frac{2m}{n}$$



- Houston
- Detroit
- New York

- The degree can be used as a *centrality measure*
- A node is arguably more important or influential if it has more connections
  - An airport with more flights
  - A person with more acquaintances
  - A metabolite that participates in more reactions

Eigenvector centrality

- But it's not just how many people you know that counts. It's also who you know.
- Define centrality of vertex *i* to be proportional to the sum of the centralities of its nieghbors:

$$x_i = \frac{1}{\lambda} \sum_{j \in \mathcal{N}(i)} x_j = \frac{1}{\lambda} \sum_{j=1}^{n} A_{ij} x_j$$

- Or

$$\mathbf{A}\mathbf{x} = \lambda \mathbf{x}$$

Eigenvector centrality

- Which eigenvector should we choose?
  - The *Perron-Frobenius theorem* says that only one eigenvector has all elements positive—the eigenvector corresponding to the largest eigenvalue
- This defines the *eigenvector centrality*
- A version of this idea, under the tradename *PageRank*, lies at the heart of the Google search engine
  - Find a set of pages that match your query
  - List them in order of rank

Google

pagerank

Web Images News Videos Books More Search tools

About 33,900,000 results (0.37 seconds)

PageRank - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/**PageRank** Wikipedia

PageRank is an algorithm used by Google Search to rank websites in their search engine results. **PageRank** was named after Larry Page, one of the founders of ...

Panda - Googlebot - Google Toolbar - EigenTrust

PageRank Checker - Instantly Check Google **PageRank**!
check**pagerank**.net/

CheckPageRank.net is the original and most used **pagerank** checker worldwide. Check Google **PageRank** and other SEO statistics for free!

Check PageRank - What is PageRank? - SEO Reporting - PageRank Blog

PageRank - Toolbar Help - Google Help
https://support.google.com/toolbar/answer/79837?hl=en Google

See Google's view of the importance of a webpage Pause your cursor over the **PageRank** button to display the importance of the webpage you're viewing, ...

PageRank » **PageRank** Checker
www.**pagerank**.net/**pagerank**-checker/

PageRank Checker. Use the **PageRank** Checker to check the **PageRank** of any web page. Enter up to 20 URLs to check **PageRank**. One URL per line. URL(s) *.

What Is **PageRank**? - **PageRank** Explained
google.about.com › ... › Search Engine Optimization (SEO)

PageRank is what Google uses to determine the importance of a web page. It's one of many factors used to determine which pages appear in search results.

Feature Column from the AMS
www.ams.org/samplings/.../fcarc-**pageran**... American Mathematical Society

by D Austin - Related articles

Google's **PageRank** algorithm assesses the importance of web pages without human evaluation of the content. In fact, Google feels that the value of its service is ...

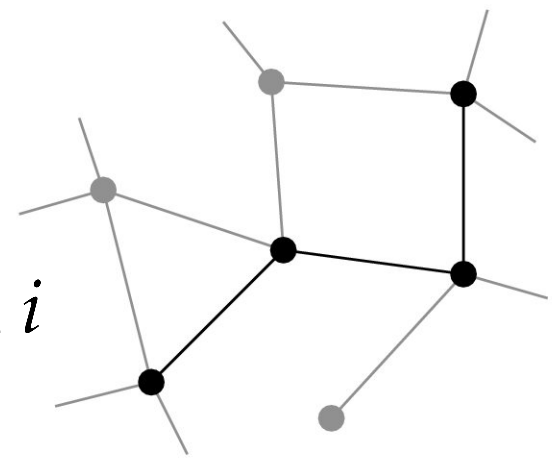
Google **PageRank** - Algorithm
pr.efactory.de/e-**pagerank**-algorithm.shtml

So, first of all, we see that **PageRank** does not rank web sites as a whole, but is determined for each page individually. Further, the **PageRank** of page A is ...

Google **PageRank** - Introduction

Paths

- A *path* or *walk* in a network is a set of nodes and edges traversed in order
- How many paths are there from vertex *i* to vertex *j*?



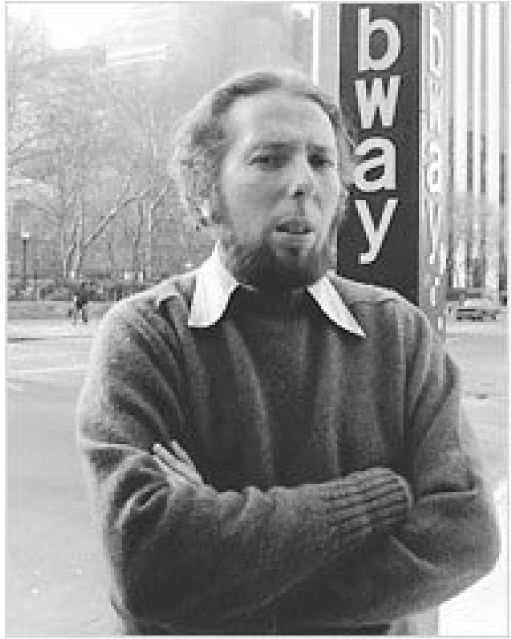
$$A_{ik}A_{kj}$$

- Similarly:

$$N_{ij}^{(r)} = \sum_k N_{ik}^{(r-1)} A_{kj} = [\mathbf{A}^r]_{ij}$$

The small-world effect

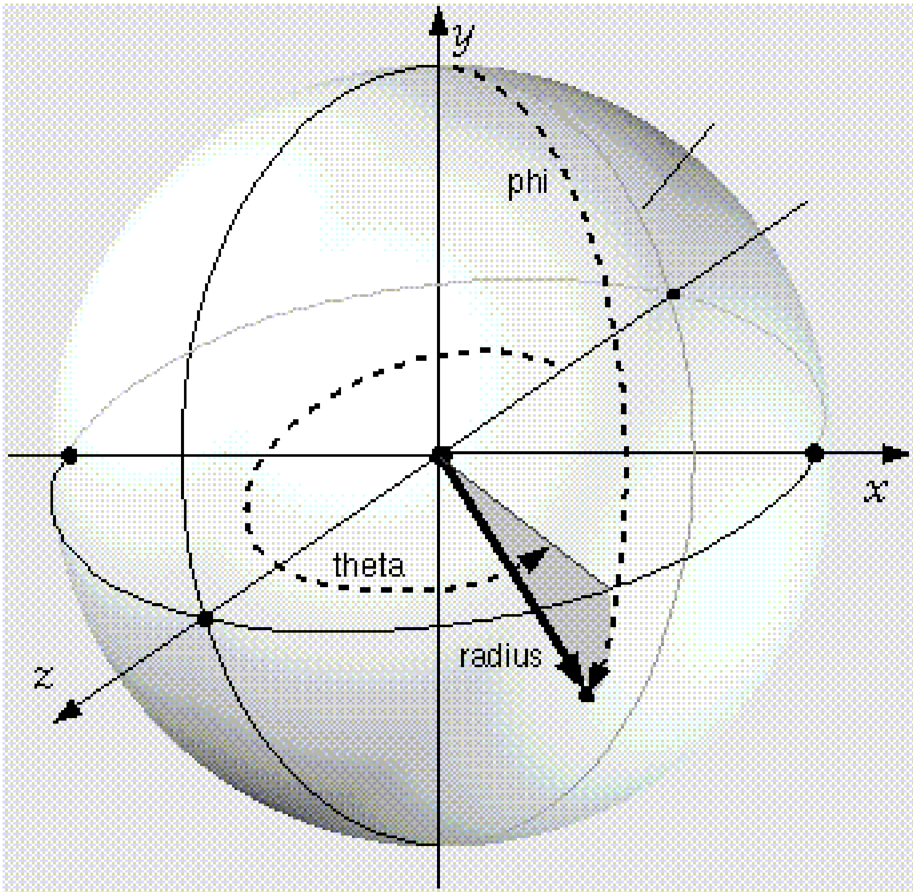
- A *geodesic path* is the shortest path between two vertices
- Stanley Milgram's 1967 experiment
  - 296 people were asked to get a letter to a target person in Boston (196 from Nebraska and 100 from Boston)
  - Letters could only be passed along a chain of first-name acquaintances
  - They took an average of 6.2 steps to get there



New York Times

The small-world effect

- If each person knows 100 people then:
  - Number of people 1 step away from you is 100
  - People 2 steps away is 100 x 100 = 10,000
  - People 3 steps away is 100 x 100 x 100 = 1,000,000
  - People 4 steps away is 100,000,000
  - People 5 steps away is 10,000,000,000
- But 10 billion is more than the total number of people in the world



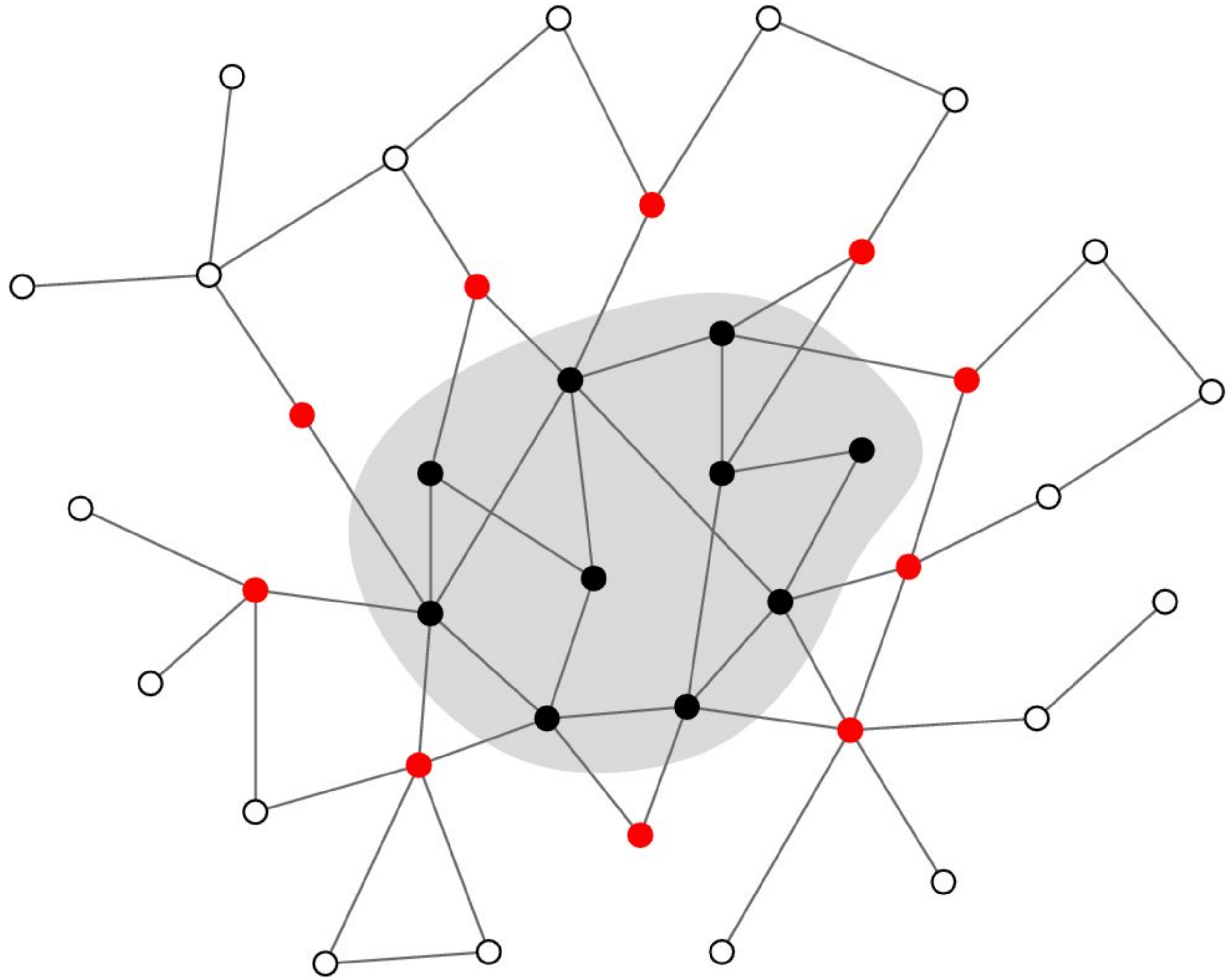
$$A = 4\pi r^2$$

$$V = \frac{4}{3}\pi r^3$$

$$A = \sqrt[3]{36\pi}\, V^{2/3}$$

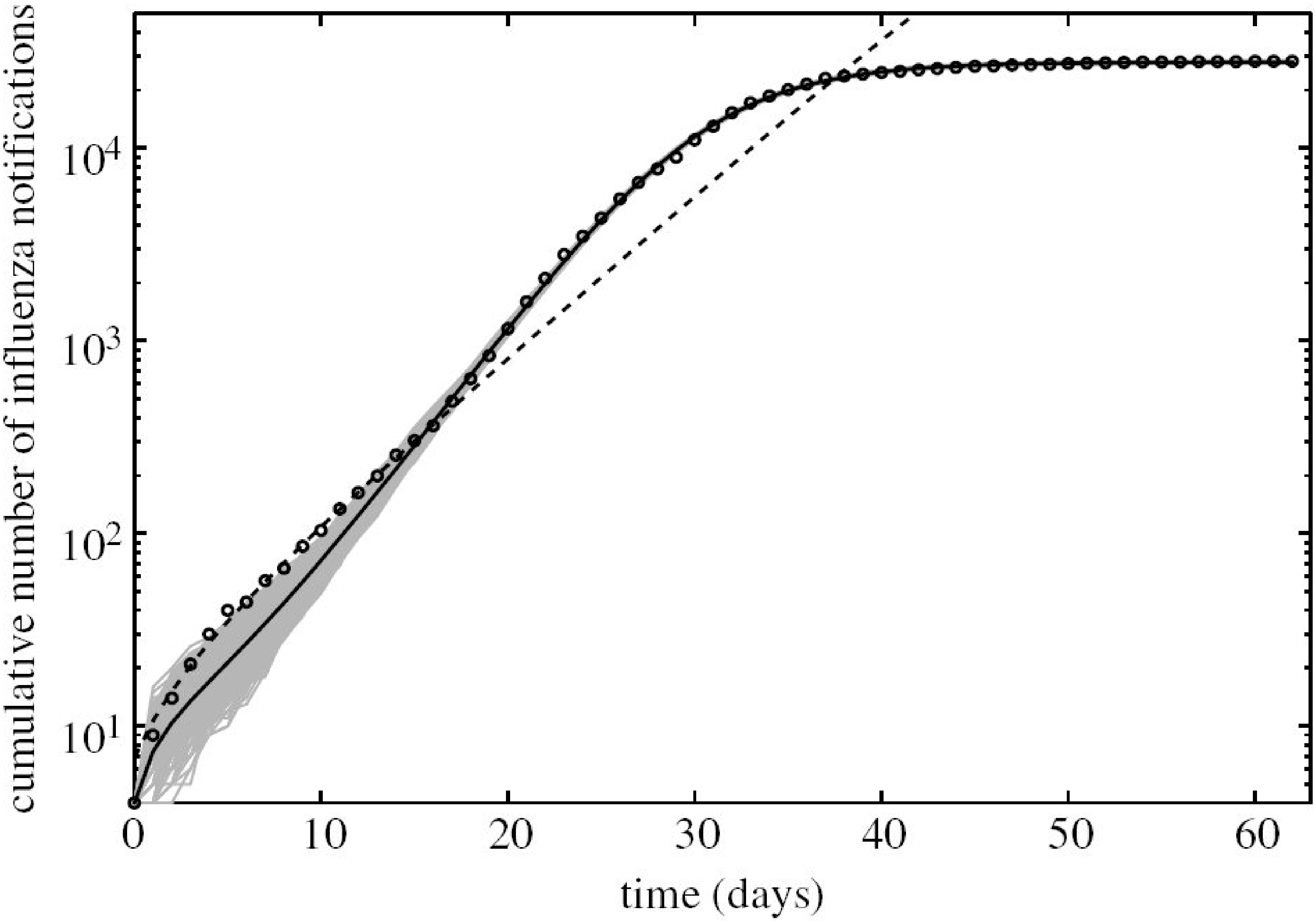
And in general dimension d

$$A \propto V^{(d-1)/d}$$



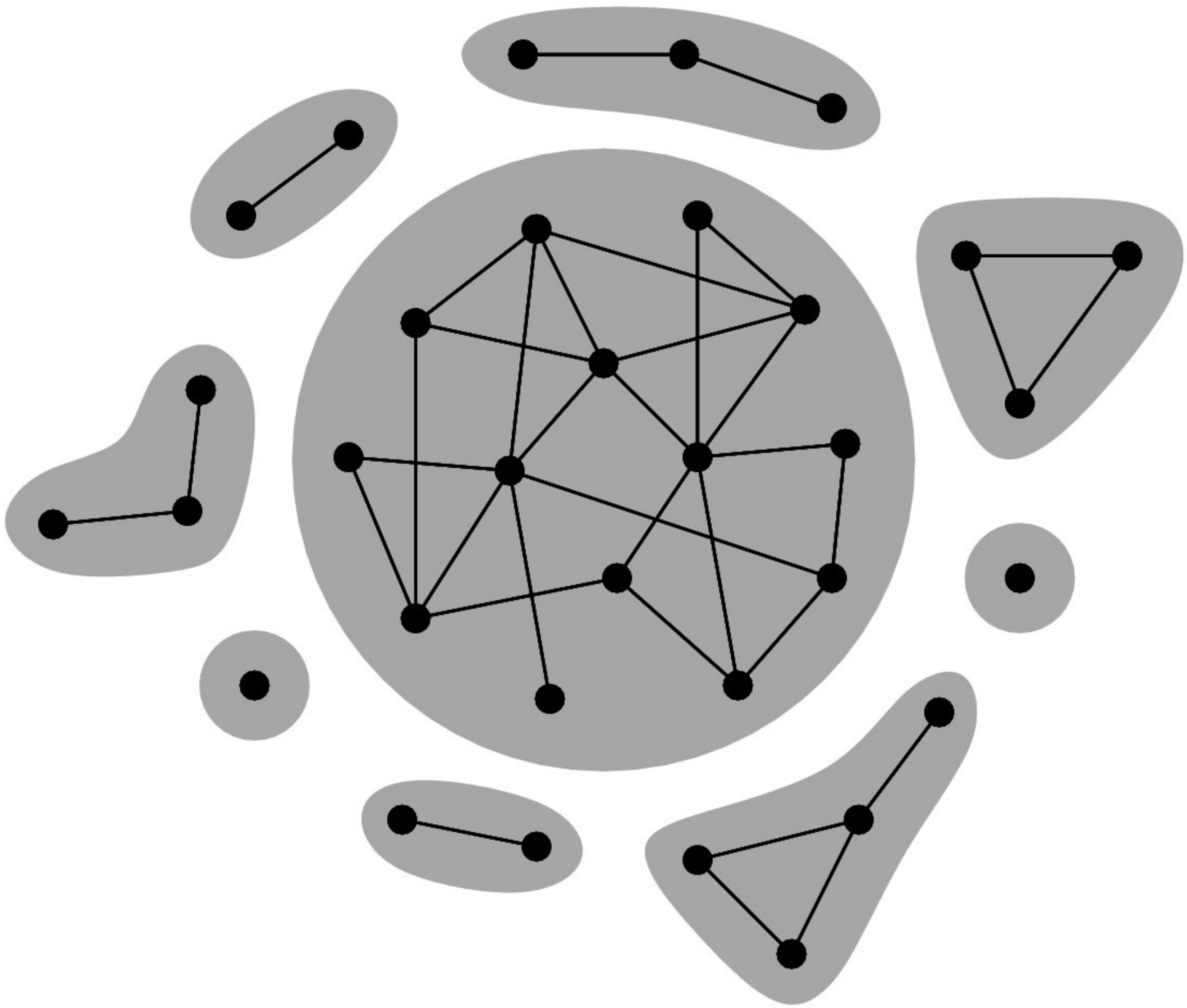
- In a network the surface can be proportional to volume
- Equivalent to choosing *d* infinite

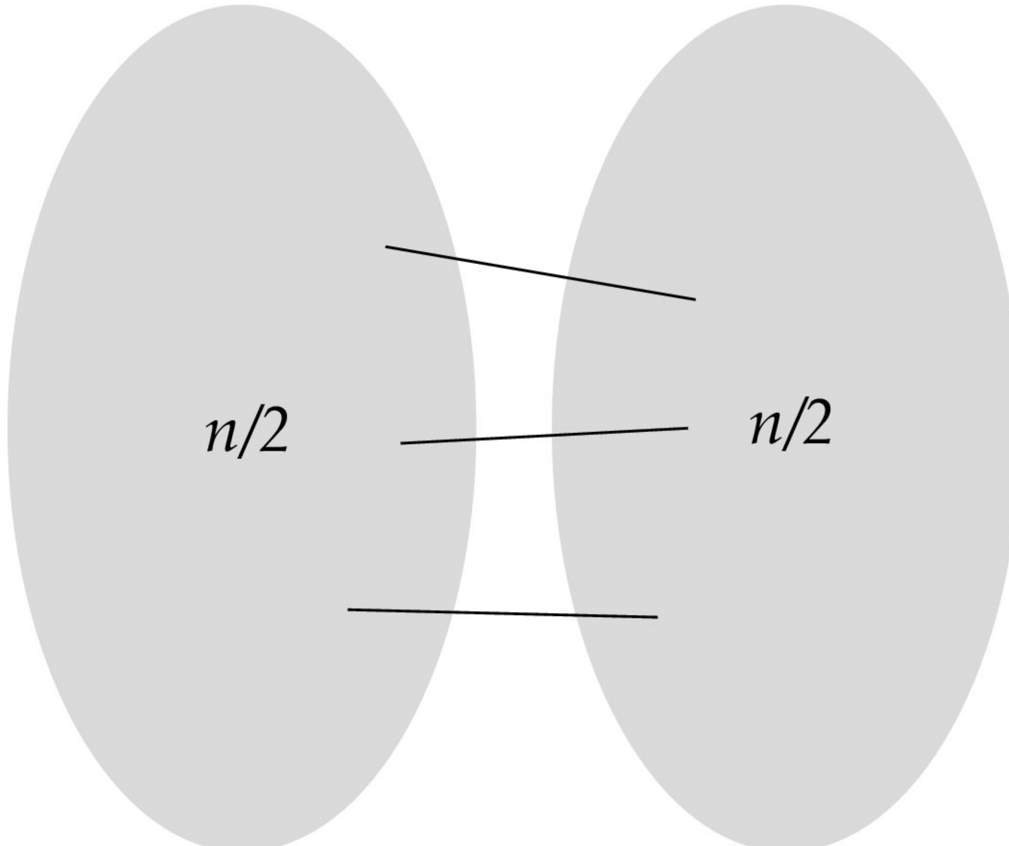
Spanish flu, San Francisco 1918-1919



Chowell *et al.* 2007

Components

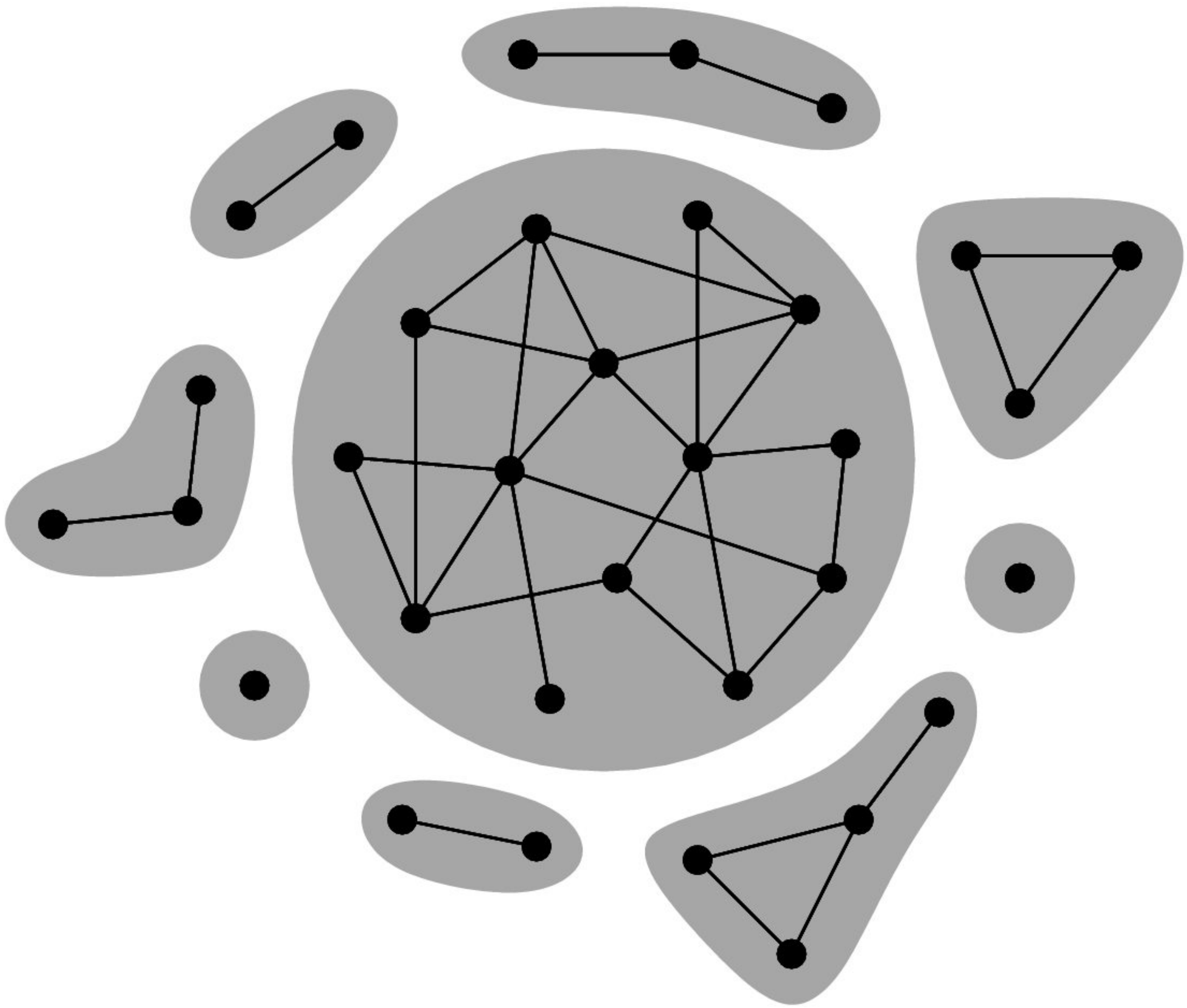




Number of places to put an edge linking the two components:

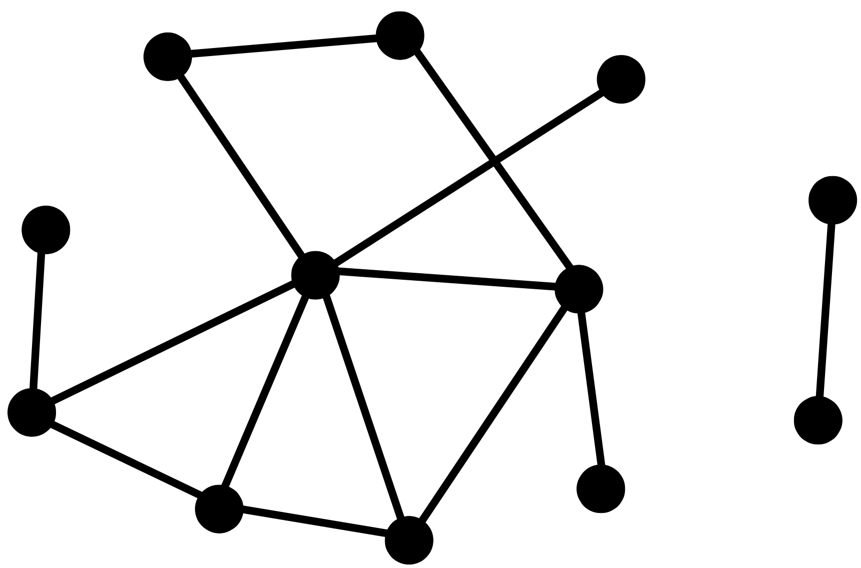
$$\frac{n}{2} \times \frac{n}{2} = \frac{n^2}{4}$$

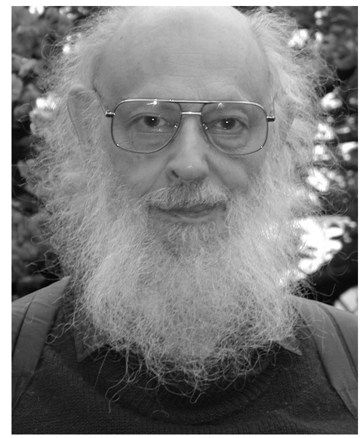
Components



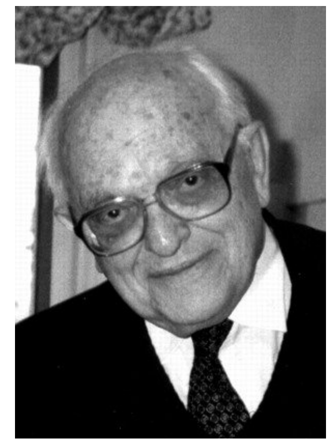
The random graph

- Take n vertices and add m edges at random

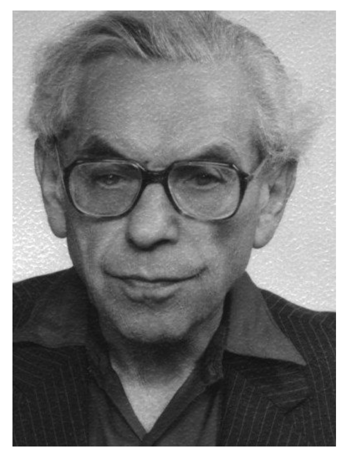




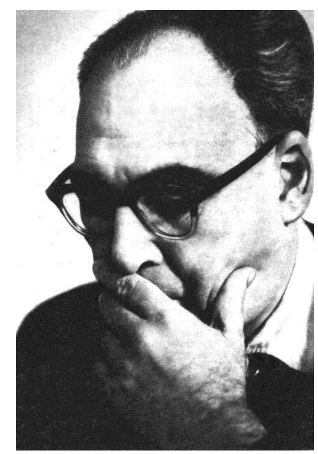
Solomonoff



Rapoport



Erdős



Rényi

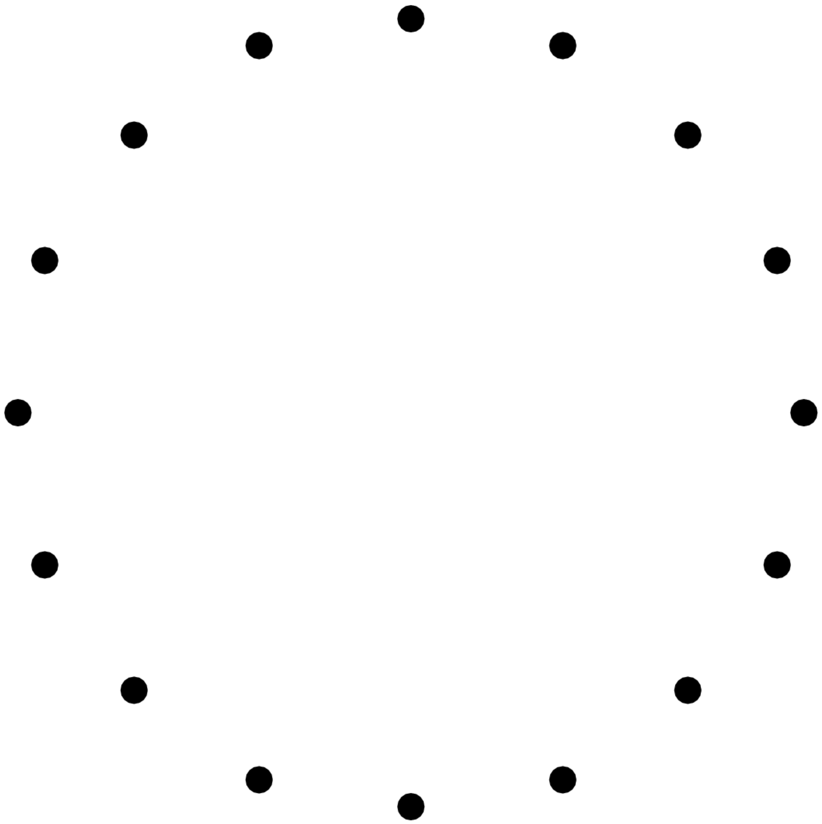
The random graph

- Average degree:

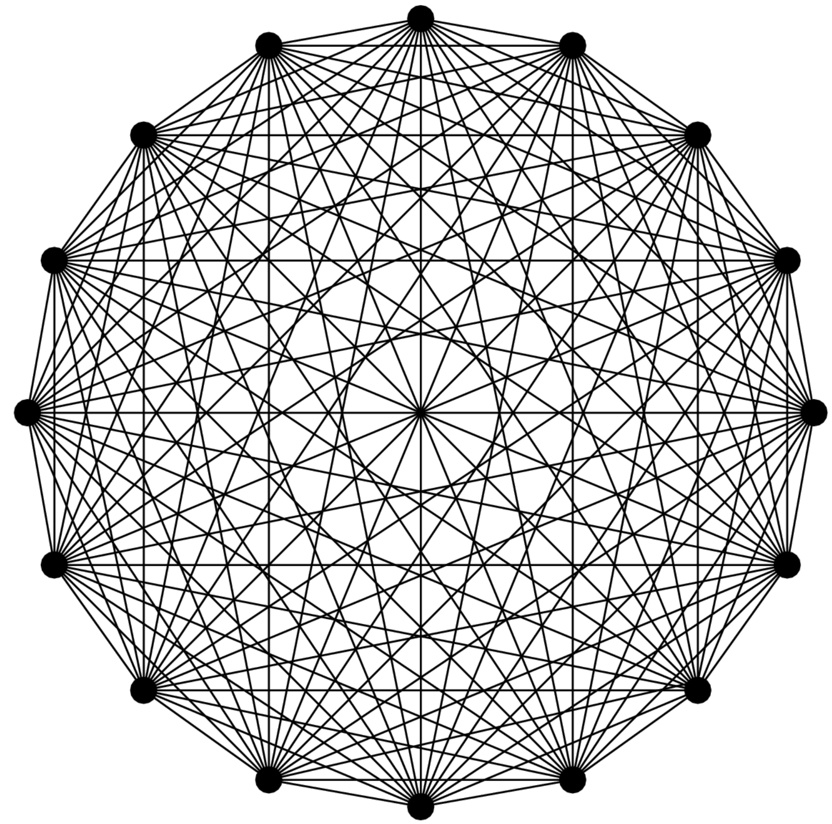
$$c = \frac{2m}{n}$$

- Probability of an edge:

$$p = \frac{m}{\binom{n}{2}} = \frac{2m}{n(n-1)} = \frac{c}{n-1}$$



$p = 0$



$p = 1$

Giant component

- Let u be the probability that a randomly chosen vertex does *not* belong to the giant component:

$$1 - p + pu$$

- But $p = c/(n-1)$ so

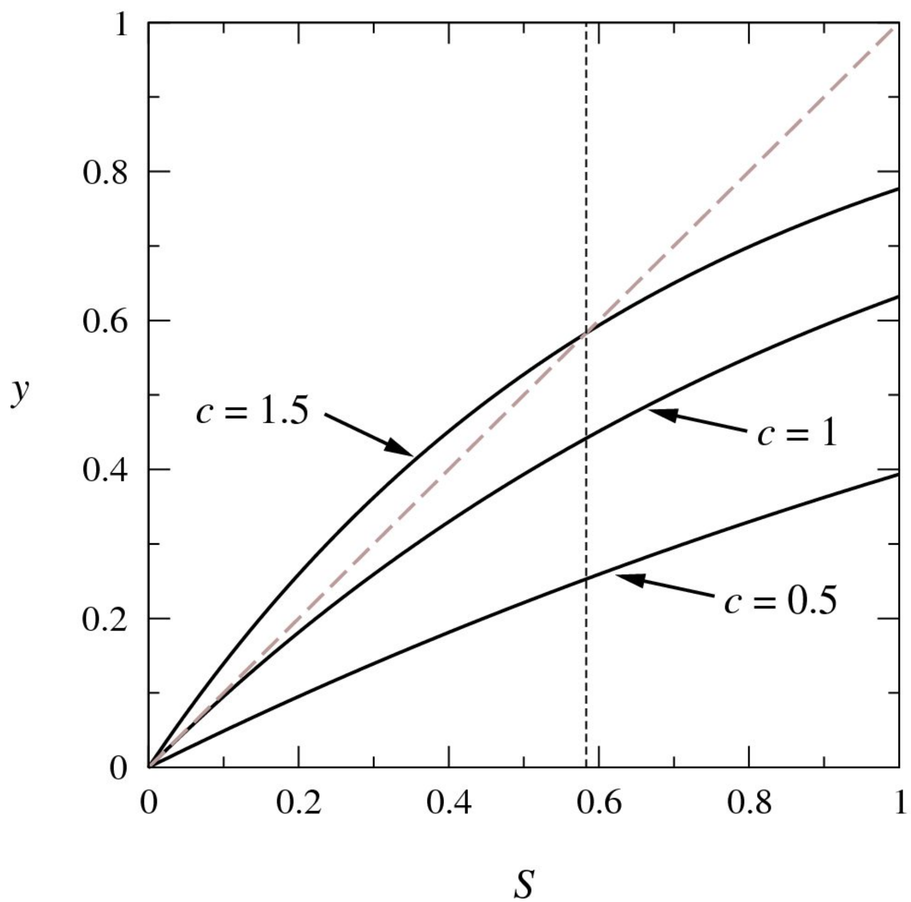
$$u = \left[1 + \frac{c(u-1)}{n-1}\right]^{n-1} = \mathrm{e}^{c(u-1)}$$

- Now define $S = 1 - u$ and

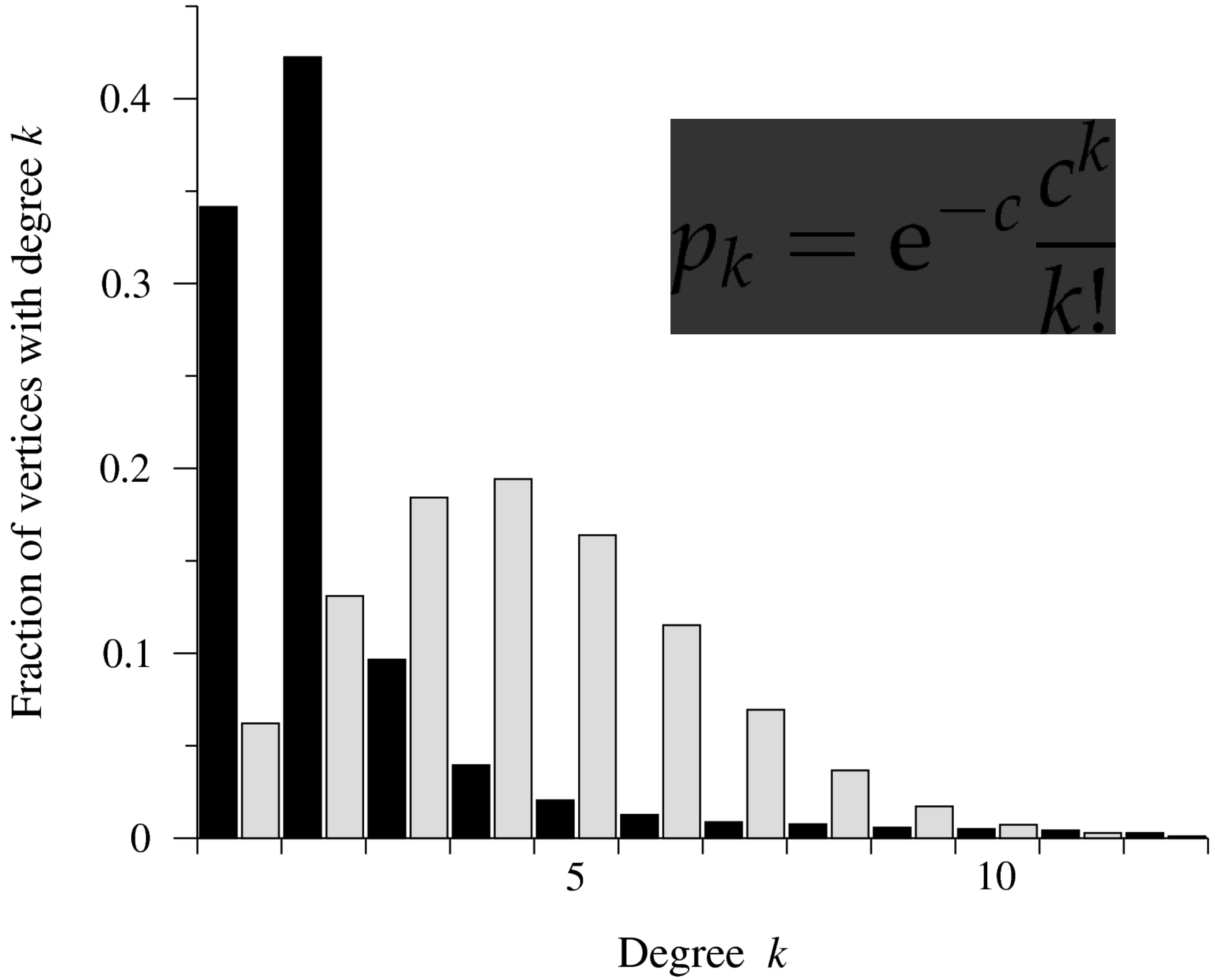
$$S = 1 - \mathrm{e}^{-cS}$$

Giant component

$$S = 1 - e^{-cS}$$



Degree distribution



- Many other properties also don't match, including degree correlations and clustering coefficient