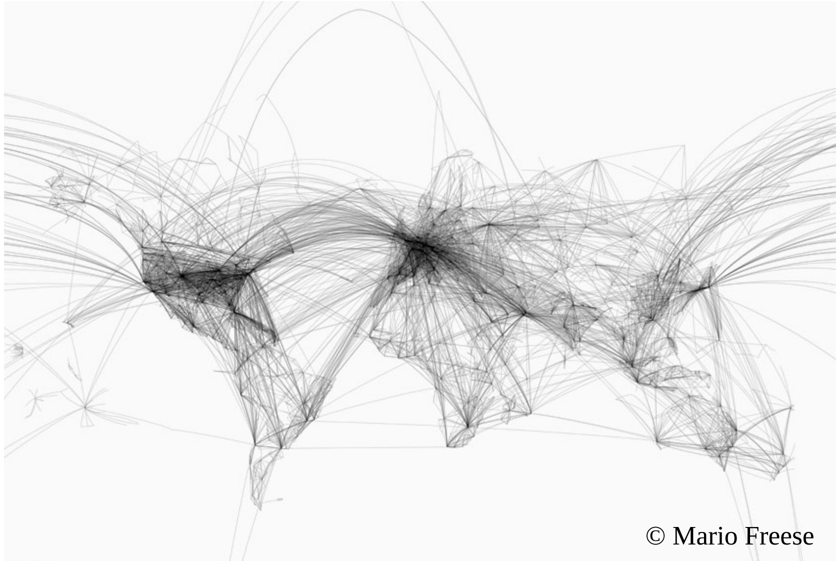
Random Networks



Hiroki Sayama
sayama@binghamton.edu

Small-World Phenomenon

“Bacon Number“

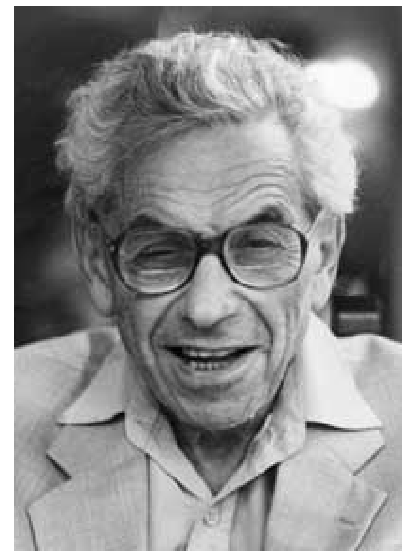
- BN = 1 for those who co-starred with Kevin Bacon in a film
- BN = 2 for those who co-starred with actors/actresses with BN=1

  …

- Mostly BN <= 3 !!
- The largest finite BN = 8 !!

"Erdös Number"

- EN = 1 for those who co-authored a paper with a Hungarian mathematician Paul Erdös (1913-1996)
- EN = 2 for those who co-authored a paper with authors with EN=1



…

- Mostly EN <= 7 !!
- The largest finite EN = 13 !!

FYI - Hiroki's EN=4 (by Bing/Microsoft Academic Search)

3.5 degrees of separation in FB

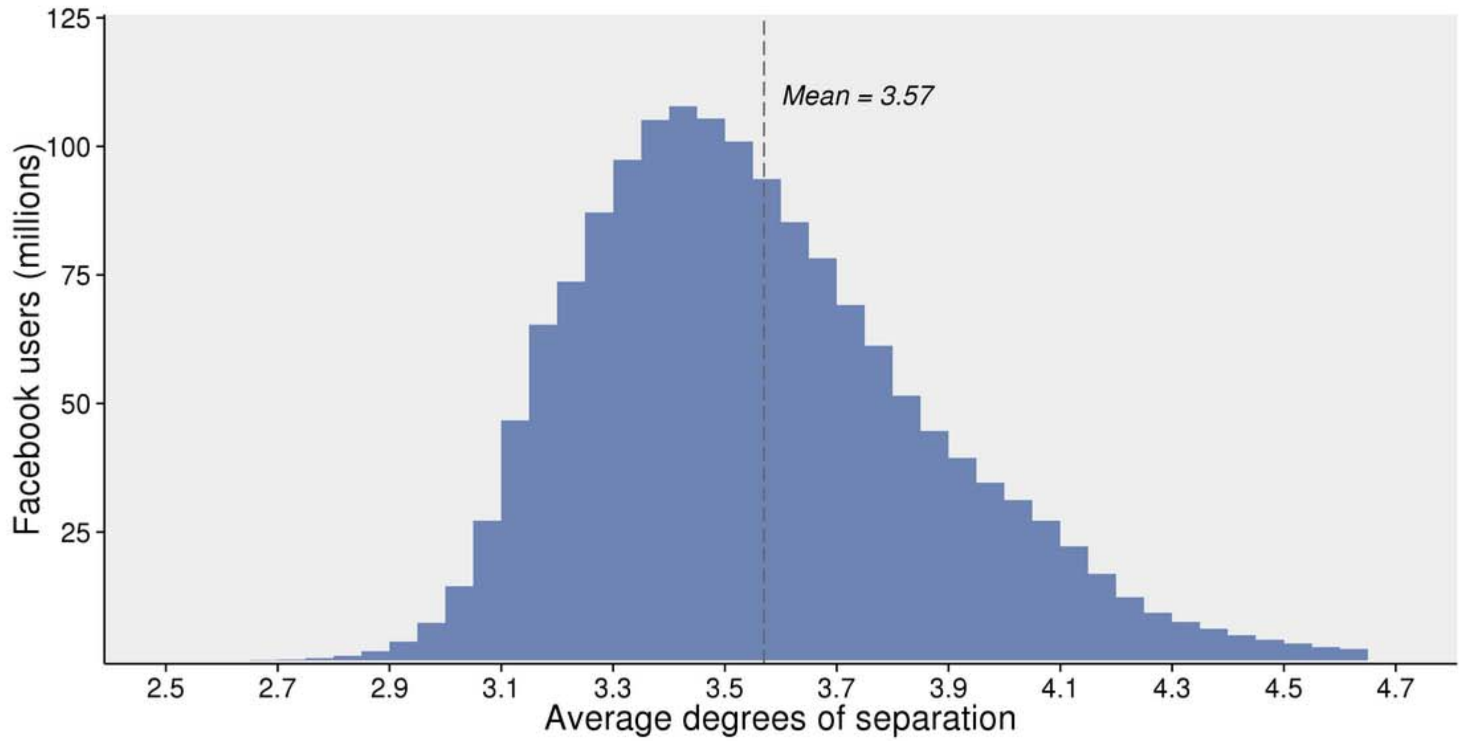


Figure 1. Estimated average degrees of separation between all people on Facebook. The average person is connected to every other person by an average of 3.57 steps. The majority of people have an average between 3 and 4 steps.

- https://research.facebook.com/blog/three-and-a-half-degrees-of-separation/

“Small-world” phenomenon

- Most real-world networks are remarkably “small”
  - Despite a huge number of nodes involved
  - Even if connections are relatively sparse
- Why?

Random Networks

Classical explanation: Erdös-Rényi random network model

- A network made of N nodes
- Each node pair is connected randomly and independently with probability p
- A small characteristic path length is realized because of randomness
  - Number of nodes reachable from a single node within k steps increases exponentially with k

Exercise

- Create and plot a few ER random networks using NetworkX
- Measure their properties
  - Network density
  - Characteristic path length
  - Clustering coefficient
  - Degree distribution
  - etc.

Limitation of ER networks

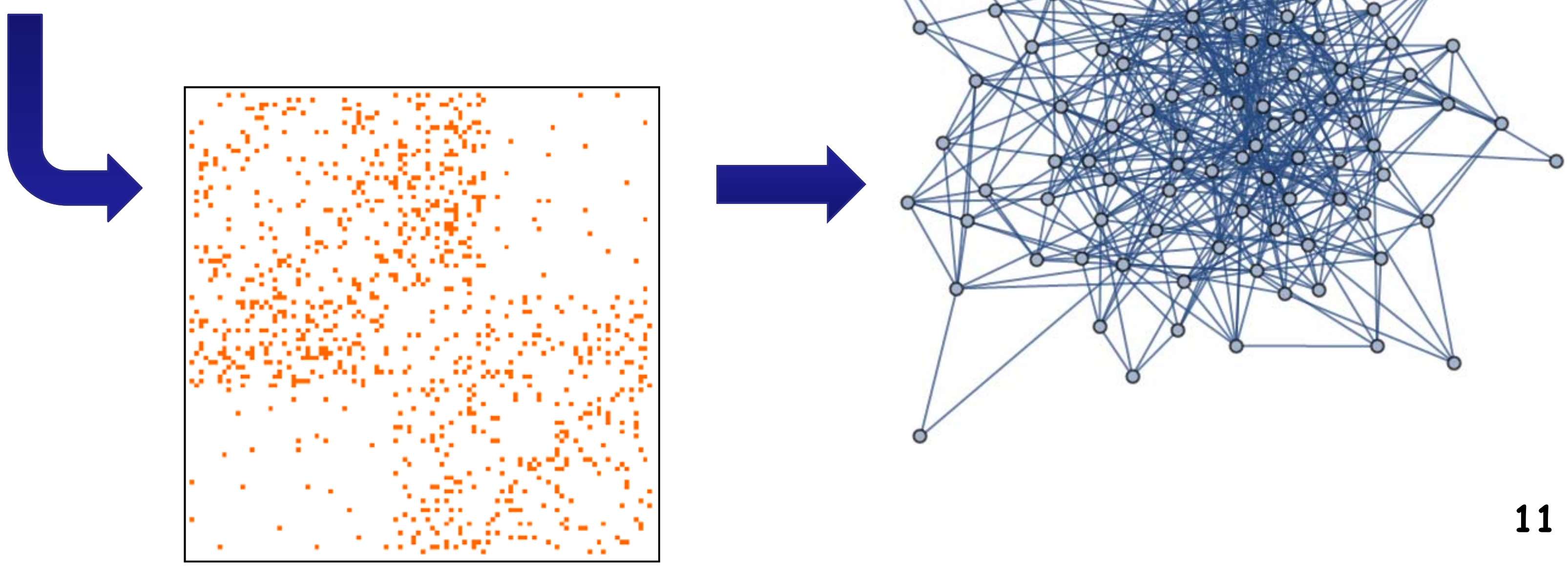
- ER random networks have very few loops or local clusters if connection probability is small

- Real-world networks are often clustered with a lot of local connections, forming "cliques", while maintaining very small characteristic path lengths

ER networks with partitions: Stochastic block models

- **Generates random networks from the connection density matrix for blocks**

$$\begin{pmatrix} 0.05 & 0.1 & 0.01 \\ 0.1 & 0.05 & 0.05 \\ 0.01 & 0.05 & 0.05 \end{pmatrix}$$



Exercise

- See the community information in the Karate Club network data
- Create its block model using the blockmodel() function
- Construct a stochastic block model using the connection probabilities obtained above (this needs coding)
- Compare the original network and the randomly generated one

Small-World Networks

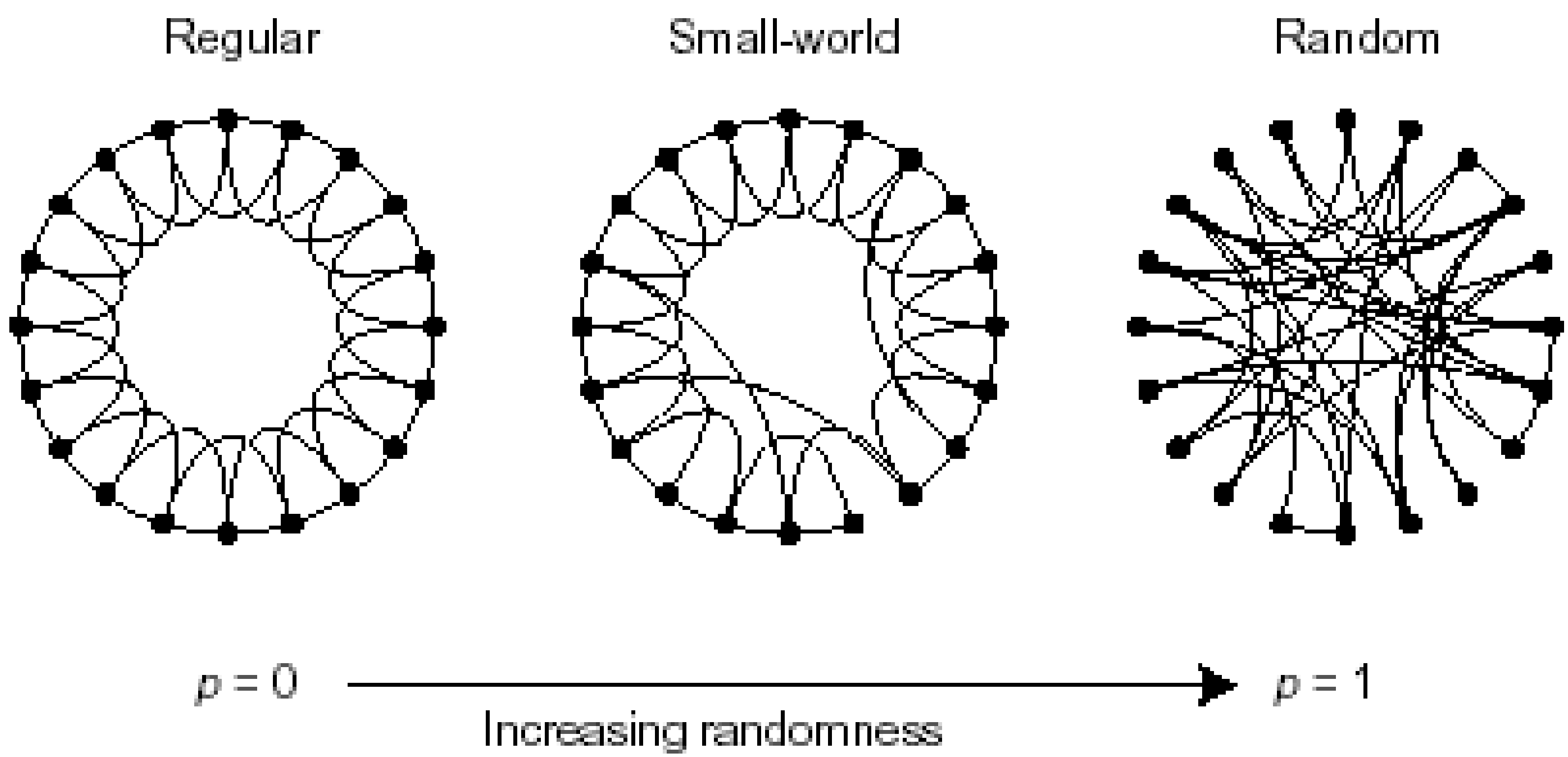
Explanation (1): Small-world network

- D. J. Watts & S. H. Strogatz, Collective dynamics of 'small-world' networks, Nature 393:440-442, 1998.

- A network that is mostly locally connected but with a few global connections
- A SW network generally has a very small characteristic path length

Experiment by Watts & Strogatz

- Moving from a regular, locally connected graph to a random, globally connected graph

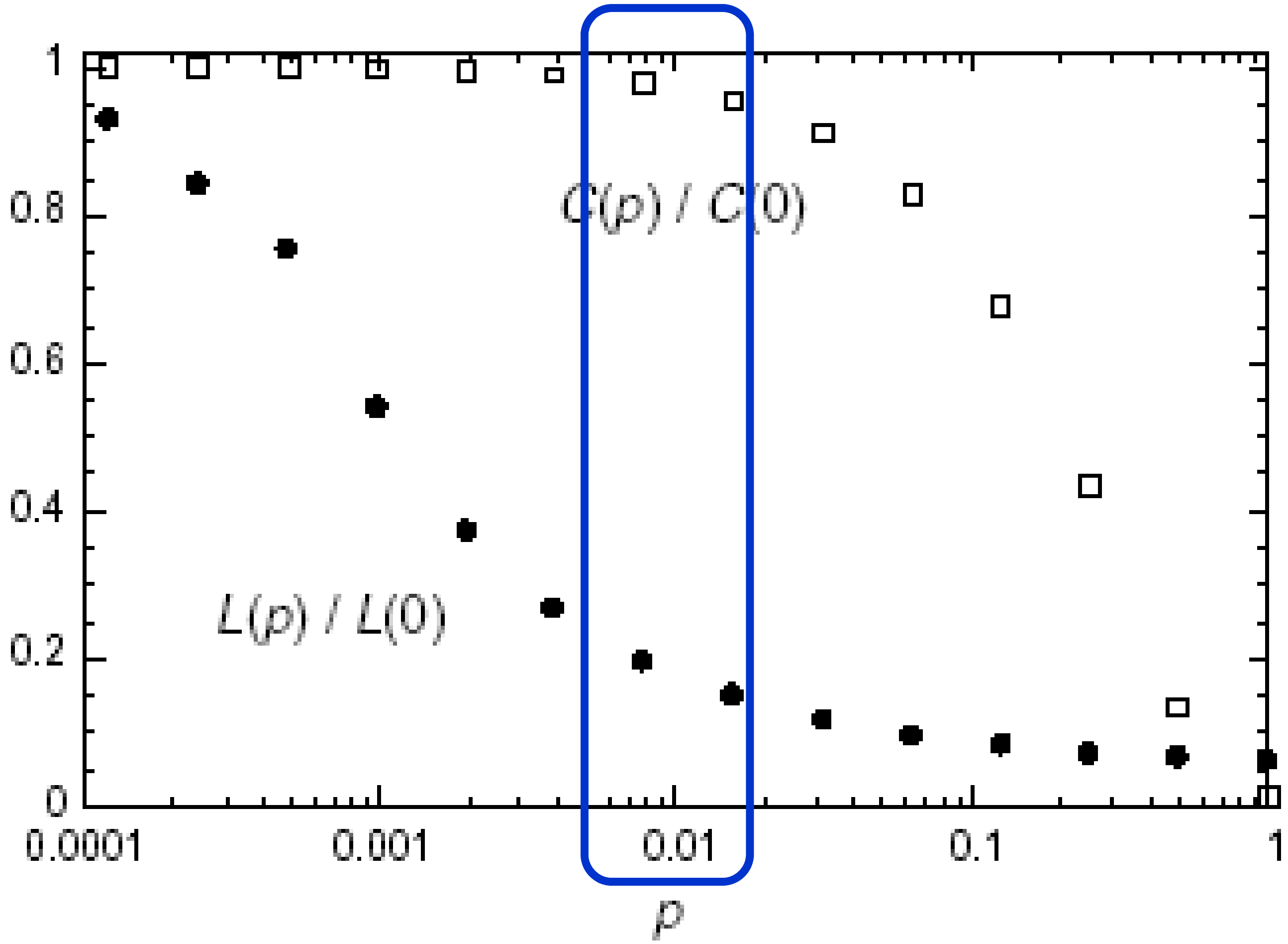


Exercise

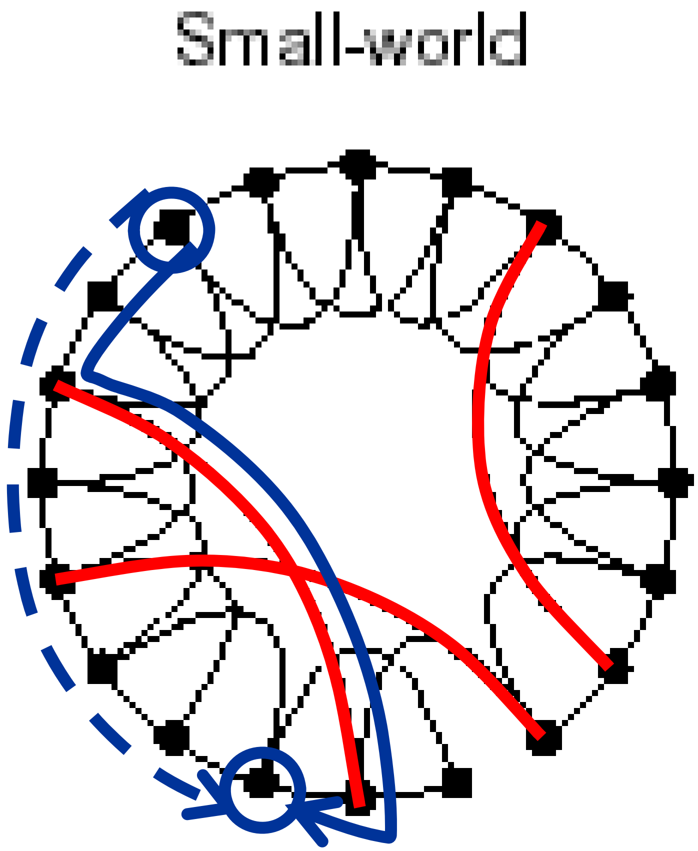
- Create a ring-shaped network made of n nodes
- Connect each node to k nearest neighbors
- Randomly rewire edges one-by-one
- Monitor what happens to the characteristic path length and the average clustering coefficient

The "small-world" property

- This network is small, though still locally connected



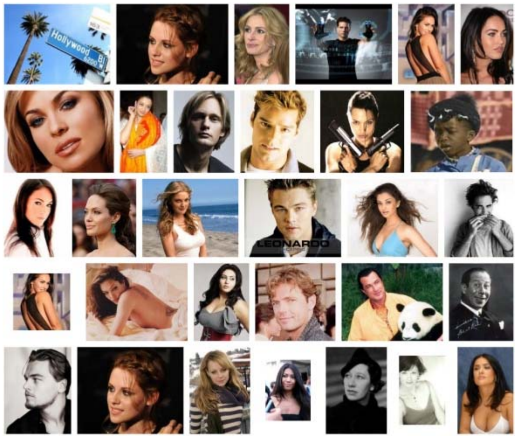
Why such a small world?

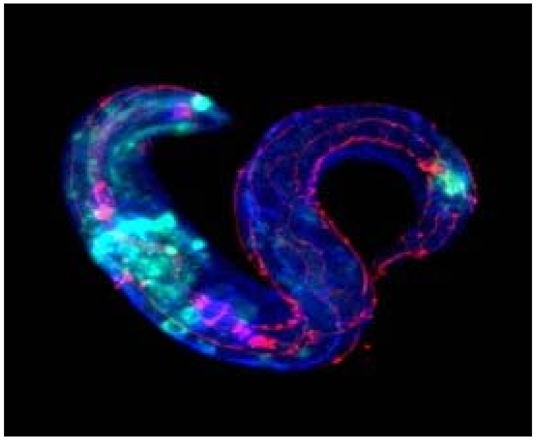


The existence of a few “far leaping” links significantly decreases the length of shortest paths for most pairs of nodes

Small-world property found in real-world networks

	L_{actual}	L_{random}	C_{actual}	C_{random}
Film actors	3.65	2.99	0.79	0.00027
Power grid	18.7	12.4	0.080	0.005
C. elegans	2.65	2.25	0.28	0.05





Exercise

- Create and plot several WS small-world networks using NetworkX
- Measure their properties
- Study how the characteristic path length and the clustering coefficient of WS networks change with increasing rewiring probability (for the same number of nodes, e.g. n=100)

Degree Distribution

Degree distribution

P(k) = Prob. (or #) of nodes with degree k

- Gives a rough profile of how the connectivity is distributed within the network

$\Sigma_k P(k) = 1$ (or total # of nodes)

Degree distribution of ER networks

- Degree distribution of an ER random network is given by a binomial distribution:

$$P(k) = {}_{N-1}C_k \, p^k \, (1-p)^{N-1-k}$$

- With large N (with fixed Np), it approaches a Poisson distribution:

$$P(k) \sim (Np)^k \, e^{-Np} \, / \, k!$$

Exercise

- Obtain the degree distribution of the Supreme Court Citation network (after making it into undirected)

- Plot the distribution in a linear scale
- Plot the distribution in a log-log scale

Exercise

- Create an arbitrary complex network of your choice, with at least 10,000 nodes in it

- Plot its degree distribution

Scale-Free Networks

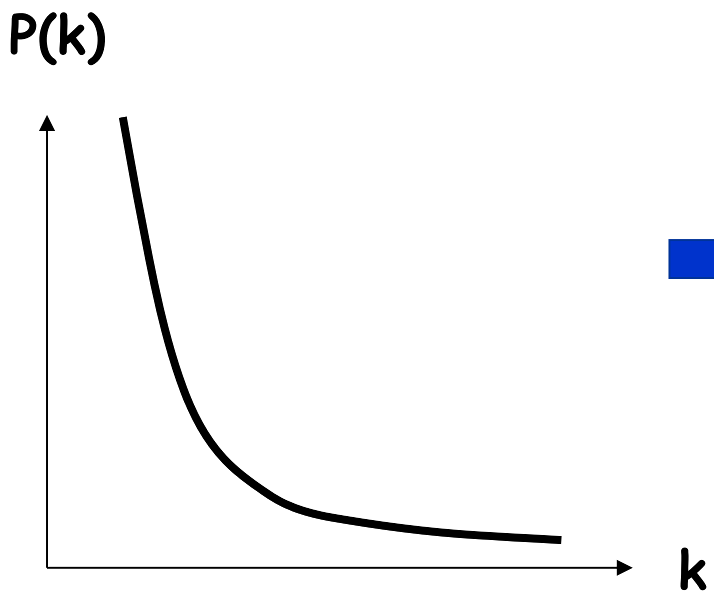
Explanation (2): Scale-free network

- A network whose degree distribution obeys a power law
- More general and natural than the small-world network model

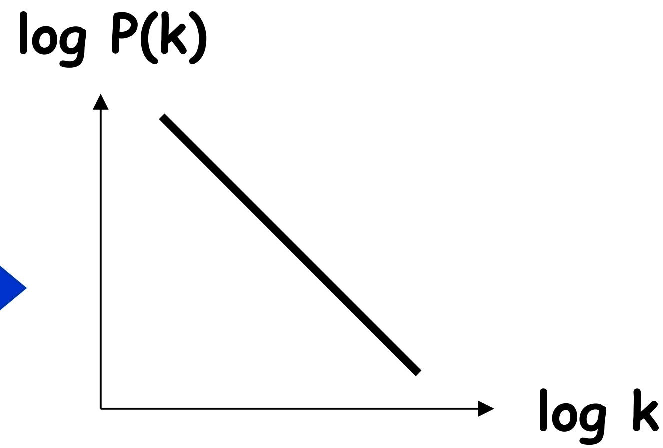
Power law degree distribution

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

A few well-connected nodes,
a lot of poorly connected nodes



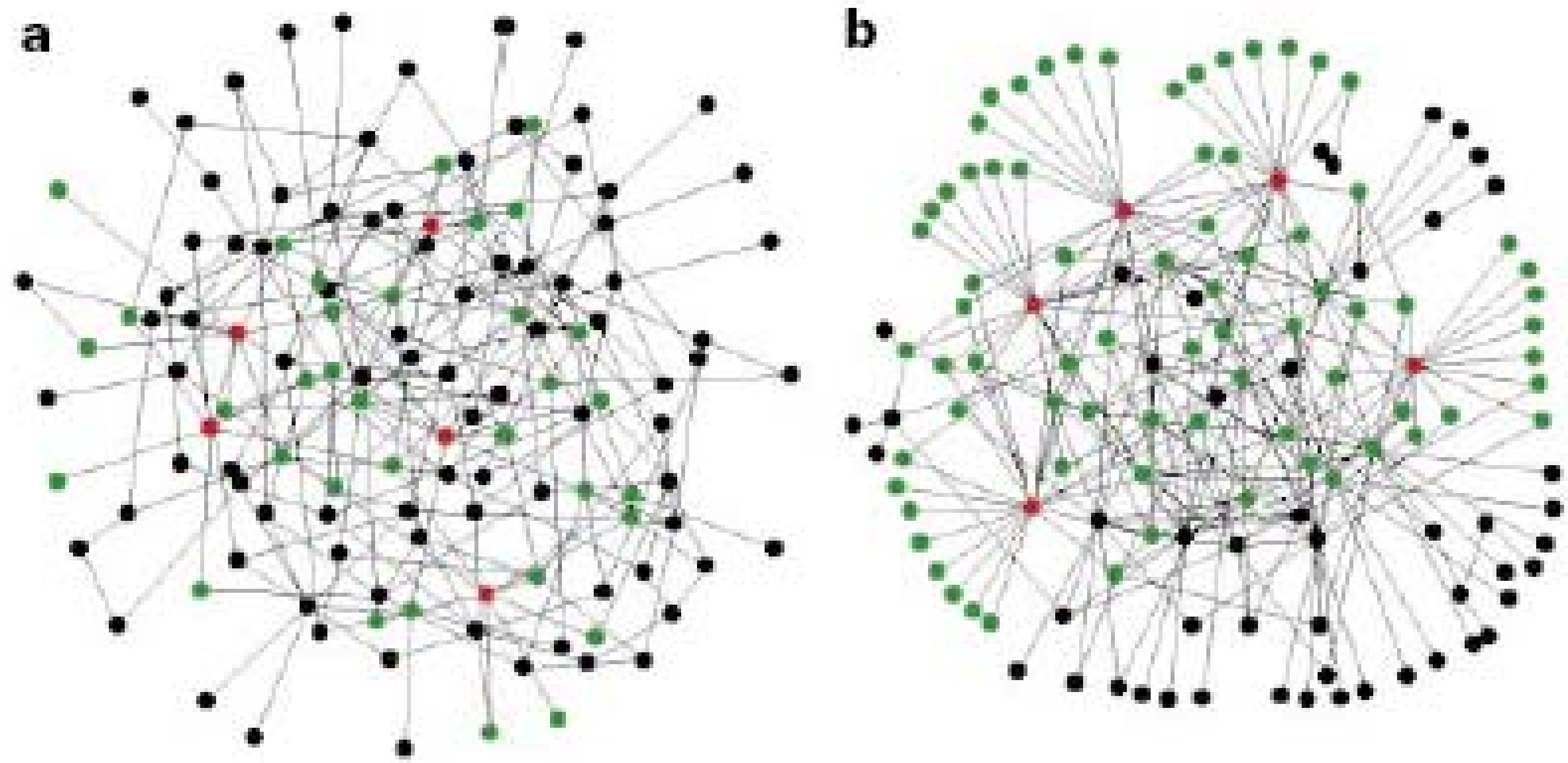
Scale-free network



Linear in log-log plot

-> No characteristic scale
(Scale-free networks)

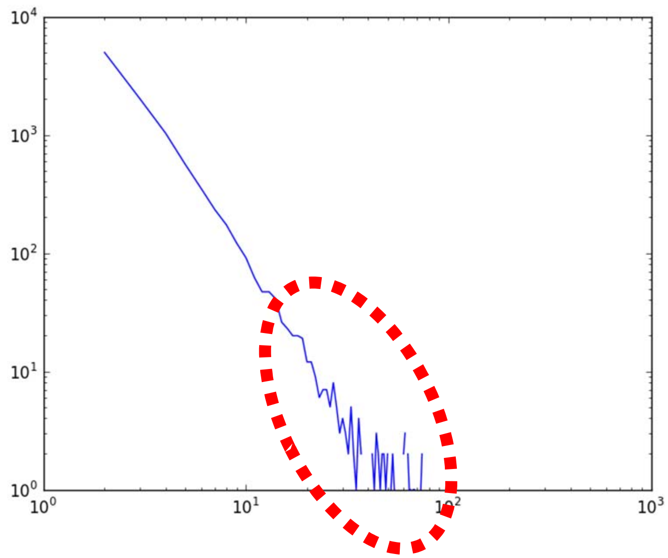
How it appears

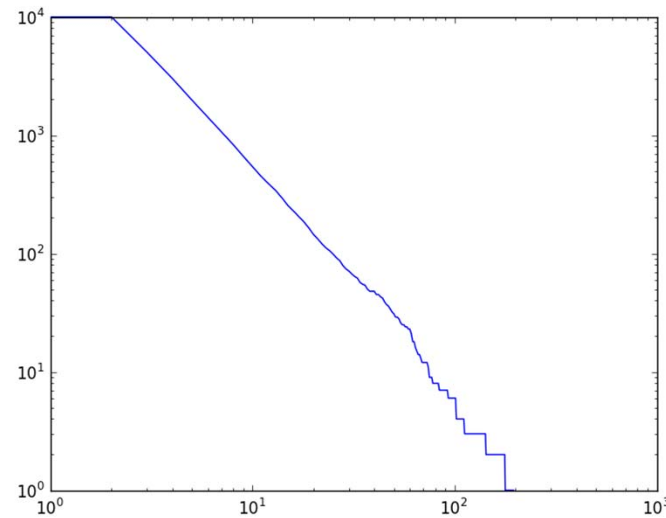


Random　　　　　　　　Scale-free

Complementary Cumulative Distribution Function (CCDF)

$P(k)$ $\quad\quad$ $CCDF(k) = \Sigma_{k' \geq k} P(k')$





$\sim k^{-\gamma}$ $\quad\quad$ $\sim k^{-(\gamma-1)}$

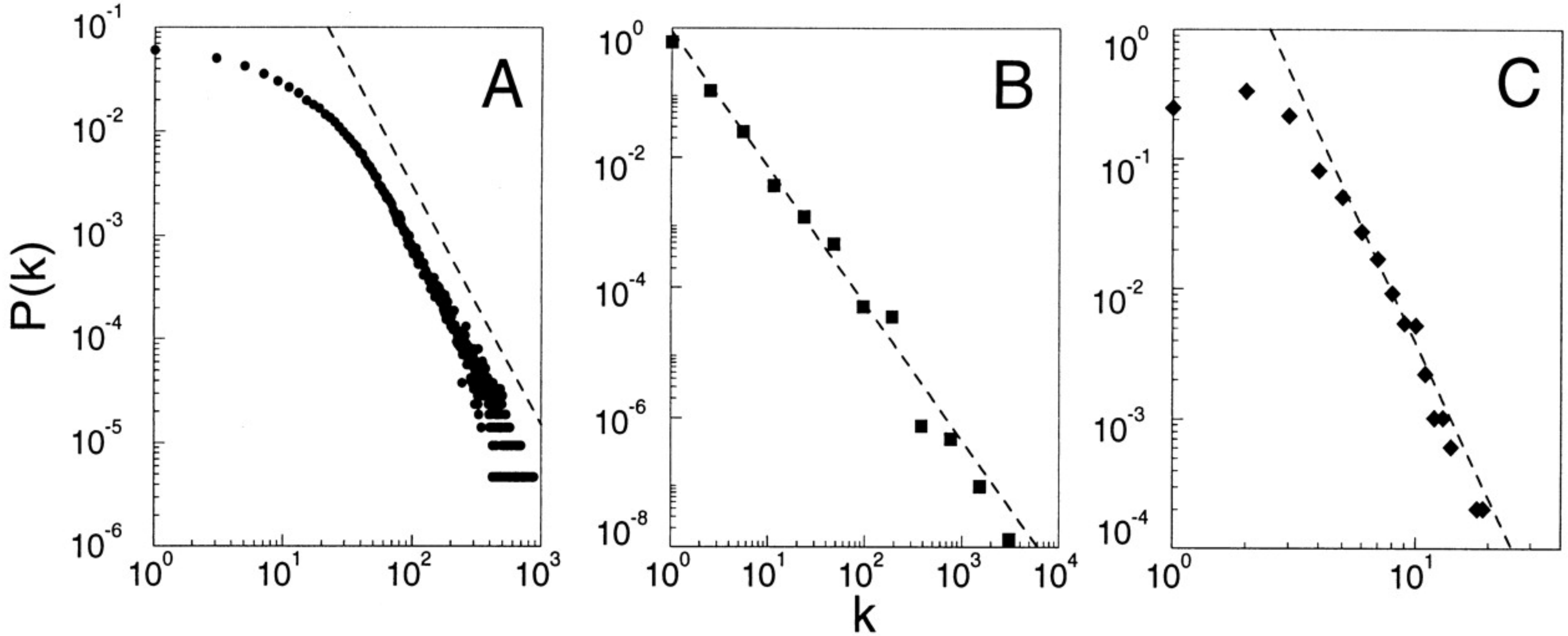
(if P(k) is a power law & γ>1)

Exercise

- Plot the CCDF of the degree distribution of the Supreme Court Citation network, in a log-log scale

- Compare it with the original degree distribution

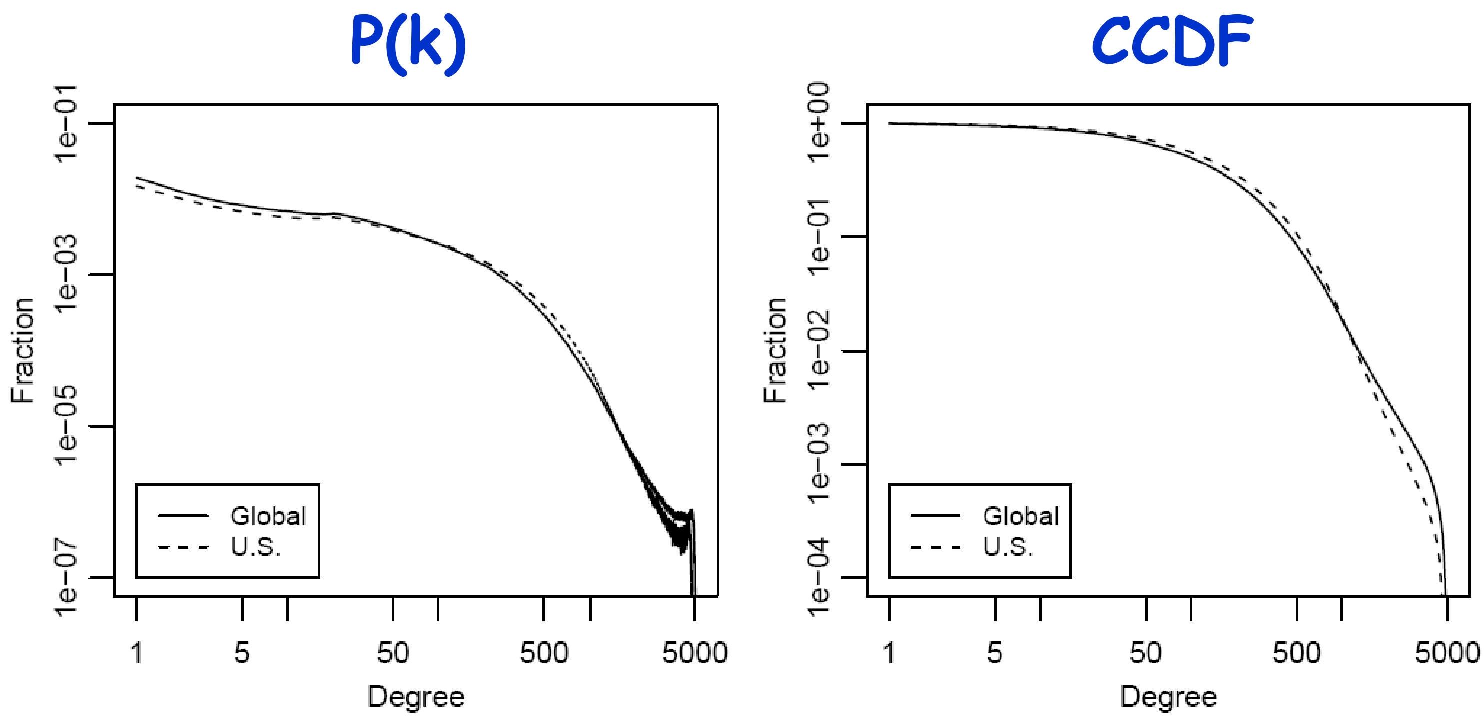
Degree Distributions of Real-World Complex Networks



Actors WWW Power grid

A Barabási, R Albert Science 1999;286:509-512

Degree distribution of FB



(a) (b)

- http://www.facebook.com/note.php?note_id=10150388519243859
- http://arxiv.org/abs/1111.4503

Properties of those networks

- A small number of well-connected nodes (hubs) significantly reduce the diameter of the entire networks
- Such degree-distribution seems to be dynamically formed and maintained by quite simple, self-organizing mechanisms

Barabási-Albert scale-free network model (Barabási & Albert 1999)

- Nodes are sequentially added to the network one by one
- When adding a new node, it is connected to m nodes chosen from the existing network
- Probability for a node to be chosen is proportional to its degree:

$$p_u = deg(u) \,/\, \Sigma_v \, deg(v)$$

Exercise

- Plot degree distributions (and their CCDFs) of several different random networks described so far
  - Use a large number of nodes, e.g. 10,000

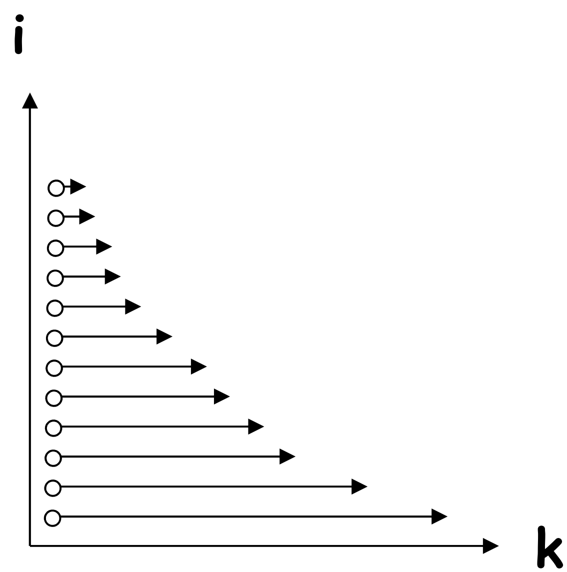
- Compare their properties

Exercise: Obtaining asymptotic degree distribution of the BA model

- Obtain the power law exponent of Barabasi-Albert growing networks analytically
  - Start with one node
  - Repeat adding a node by connecting it to the network by one link, with degree-proportional preferential attachment
  - Analytically show that $P(k) \sim k^{-\gamma}$, and find the value of its exponent γ

Exercise: Obtaining asymptotic degree distribution of the BA model

- Think about how the (expected value of) degree of the i-th node will grow over time



- $k_i(t=i) = m$
- $k_i(t)$ changes at the rate of $m(k_i(t)/2mt)$
- Degree distribution:

$$P(k) \sim -di(k)/dk$$

Degree Correlation

Degree correlation (assortativity)

- Pearson's correlation coefficient of node degrees across links

$$r = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

- X: degree of start node (in / out)
- Y: degree of end node (in / out)

Exercise

- Measure degree correlation (assortativity) for the following networks
  - Erdos-Renyi random networks
  - Watts-Strogatz small-world networks
  - Barabasi-Albert scale-free networks
- Repeat measurements multiple times and plot histograms of assortativity

Assortative/disassortative networks

Network	n	r
Physics coauthorship (a)	52 909	0.363
Biology coauthorship (a)	1 520 251	0.127
Mathematics coauthorship (b)	253 339	0.120
Film actor collaborations (c)	449 913	0.208
Company directors (d)	7 673	0.276
Internet (e)	10 697	−0.189
World-Wide Web (f)	269 504	−0.065
Protein interactions (g)	2 115	−0.156
Neural network (h)	307	−0.163
Marine food web (i)	134	−0.247
Freshwater food web (j)	92	−0.276
Random graph (u)		0
Callaway *et al.* (v)		$\delta/(1+2\delta)$
Barabási and Albert (w)		0

Social networks are assortative

Engineered / biological networks are disassortative (could be just because of “structural cutoffs”)

(from Newman, M. E. J., Phys. Rev. Lett. 89: 208701, 2002)

Exercise

- Measure degree correlations in the Supreme Court Citation Network
  - In-in correlation
  - In-out correlation
  - Out-in correlation
  - Out-out correlation

- Compare the observed results with those of randomized networks