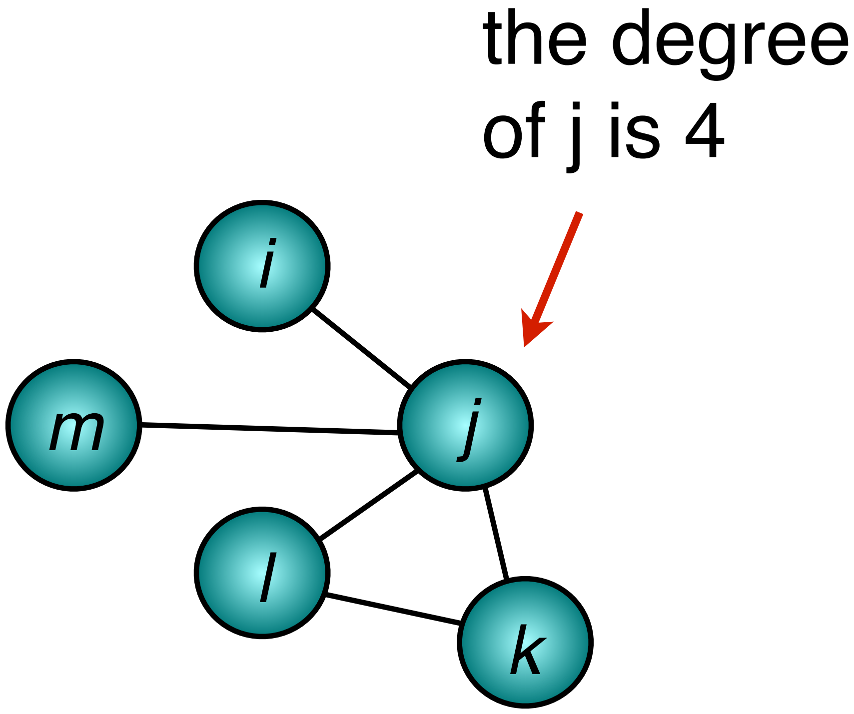
CS-E5740 Lecture II
Static network models: random and small-world networks

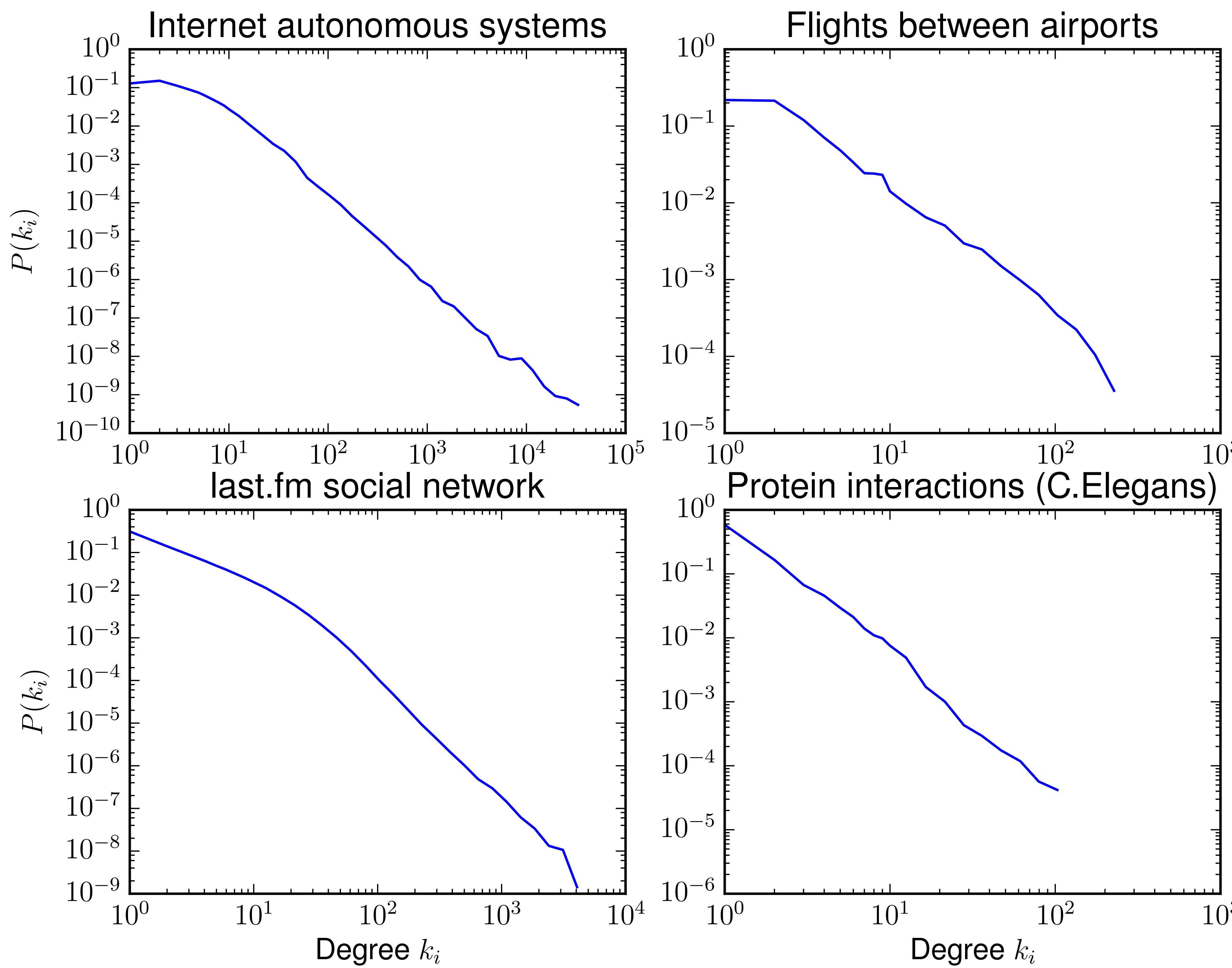
Course outline

1. Introduction (motivation, definitions, etc.)
2. Static network models: random and small-world networks
3. Growing network models: scale-free networks
4. Percolation, error & attack tolerance of networks, epidemic models
5. Network analysis
6. Social networks & (socio)dynamic models
7. Weighted networks
8. Clustering, sampling, inference
9. Temporal networks & multilayer networks

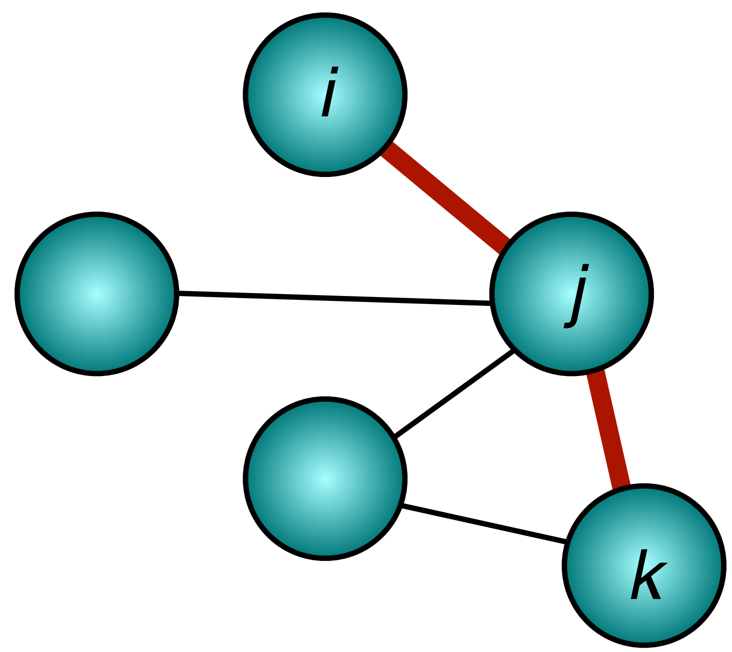
Degree



Degree examples

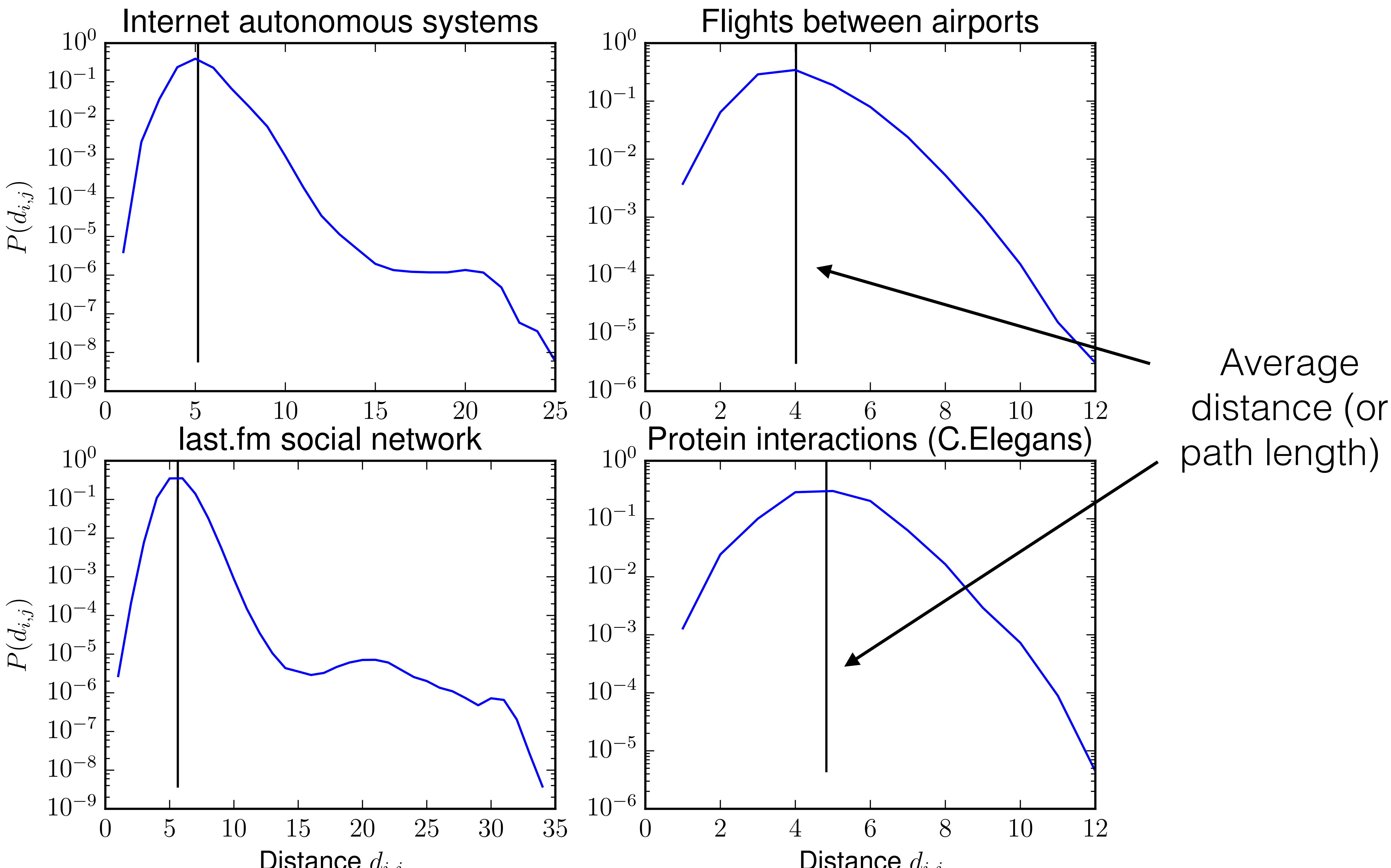


Walks, paths, distances



Path {i,j,k} has length 2. This is the distance between i and k, and also happens to be the diameter of this network

Distances in data

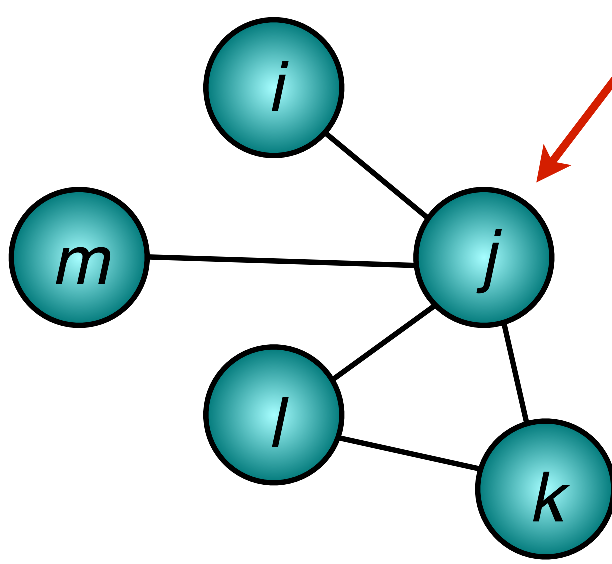


Clustering coefficient

Clustering coefficient defined for node i as the fraction of edges between its neighbours out of possible edges between its neighbours:

$$c_i = \frac{E_i}{\binom{k_i}{2}} = \frac{2E_i}{k_i(k_i-1)}$$

$$c_j = \frac{2\times 1}{4\times 3} = 0.1666$$



$$c_j = \frac{2\times 1}{2\times 1} = 1$$

Examples of values of density and clustering coefficient

Network	ρ	c
last.fm social net	$3.2 \ 10^{-6}$	0.19
Airports & flights	0.0034	0.49
Protein interaction	0.0012	0.02
Internet AS	$7.7 \ 10^{-6}$	0.25

Some properties of many real-wold networks

Network	Degrees	Paths	Clustering
Real-world	Fat-tailed	**short**	**high**

There are many nodes with low number of connections and some with extremely many connections

Nodes are mostly very close to each other

Networks have many relatively dense parts (i.e., dense subgraphs)

“It’s a small world”

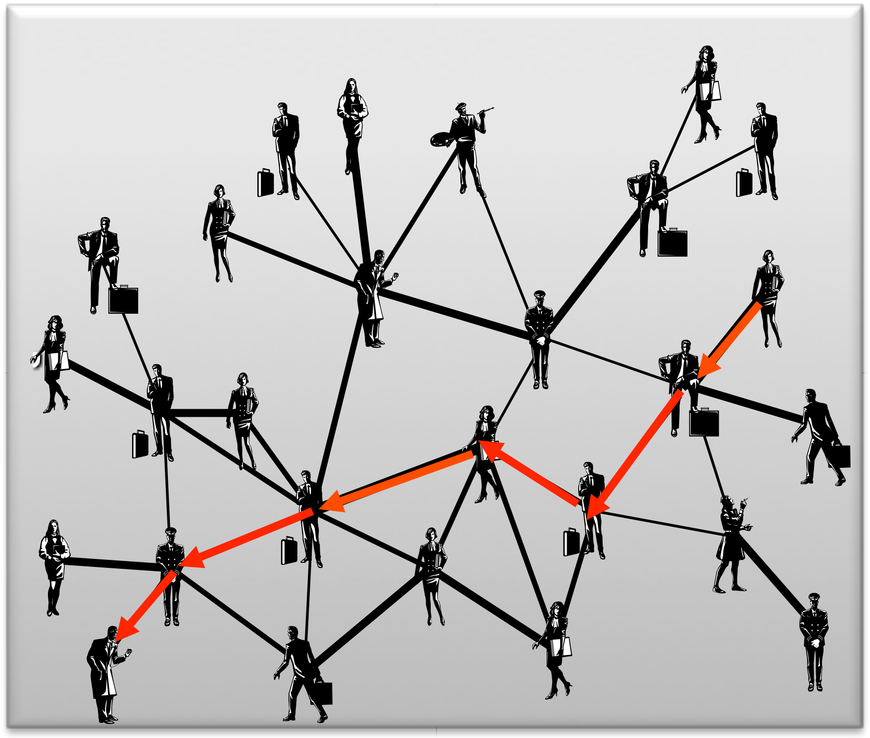
Frigyes Karinthy, short story “Chains”, 1929

- Karinthy believed that “the world is getting smaller” because people move, travel, and communicate more
- The characters of the story engaged in a game, where they have to figure out how people are connected:
- “A fascinating game grew out of this discussion. One of us suggested performing the following experiment to prove that the population of the Earth is closer together now than they have ever been before. We should select any person from the 1.5 billion inhabitants of the Earth—anyone, anywhere at all. He bet us that, using no more than five individuals, one of whom is a personal acquaintance, he could contact the selected individual using nothing except the network of personal acquaintances.”
- These ideas started resonating with social scientists...

The Milgram Experiment

Stanley Milgram: *The Small World Problem*, Psychology Today, 1967

- Milgram picked a number of persons in the central US at random
- Each got a letter that was addressed to a target person in Boston
- “If you know the target person, give the letter to him, otherwise give it to someone who you think is closer to the target person”.
- ~20% of the letters reached the target
- For these, there were on average 5.5 inter-mediaries
- Conclusions:
  - Short chains exist
  - ...and people somehow manage to find them!



Why do real-world networks have similar properties?

- Is there a simple explanation for the structural similarities of real-world networks?
- Out of all possible imaginable networks, what are the properties that most of them have, and what are specific to real-world networks?
- Can we create artificial networks that resemble real-world networks?

→ Network models

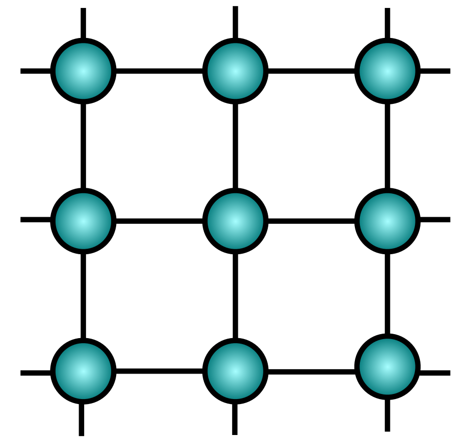
On Network Models

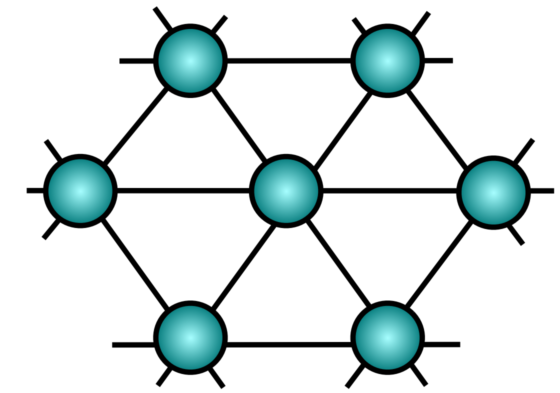
- “Toy models” of networks are common in network science
- These do not attempt to capture everything there is to networks
- Rather, the target is to design simplified models that capture some aspects of reality
- Such models may:
  - Tell something about the origins of networks or their characteristics
  - Allow for simulating processes on networks under controlled circumstances

Regular lattices

- Nodes placed regularly in d-dimensional Euclidean space, connect each node to k closest nodes
  - Often periodic boundaries
- Large number of local loops
  - Triangles, 4-loops, etc. depending on exact structure
- For regular d-dimensional lattices:

$$l \propto N^{1/d}$$





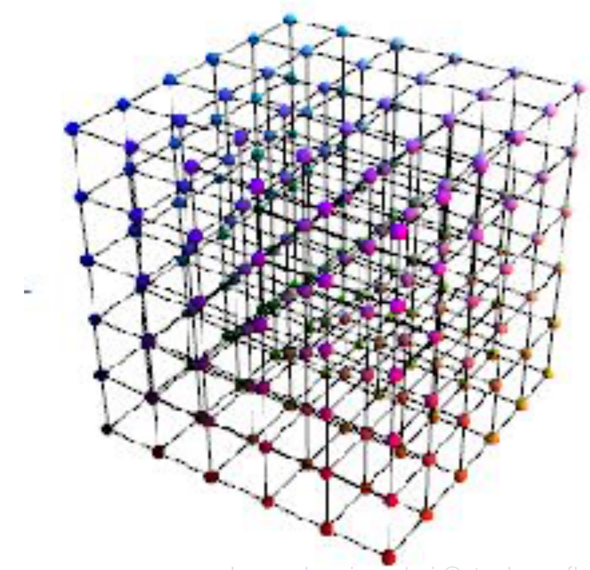
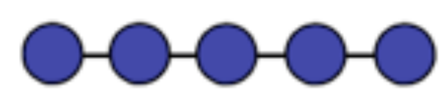
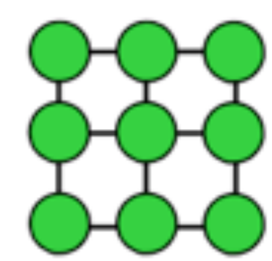


Image by sinandrei @stackoverflow



$l \propto N$



$l \propto N^{1/2}$

Network	Degrees	Paths	Clustering
Real-world	Fat-tailed	short	high
Regular lattices	Fixed	long	high*

* (depending on layout & clustering measure)

Erdős-Rényi networks

Erdős-Rényi networks:

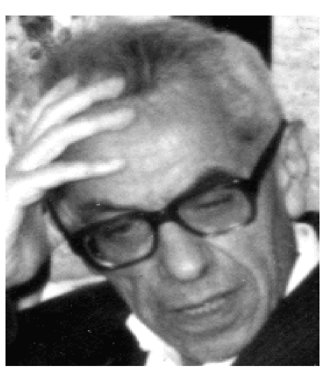
- A maximally random ensemble of networks of given size

Construction:

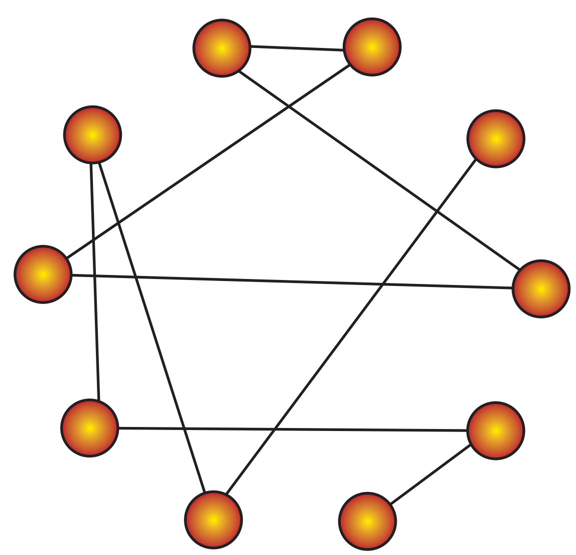
- Connect N vertices randomly

Two versions:

- $\mathrm{G}(N,p)$: connect each pair of vertices with probability p
- $\mathrm{G}(N,m)$: place m edges randomly on the network
- these define *ensembles* of networks



Pál Erdős
(1913-1996)



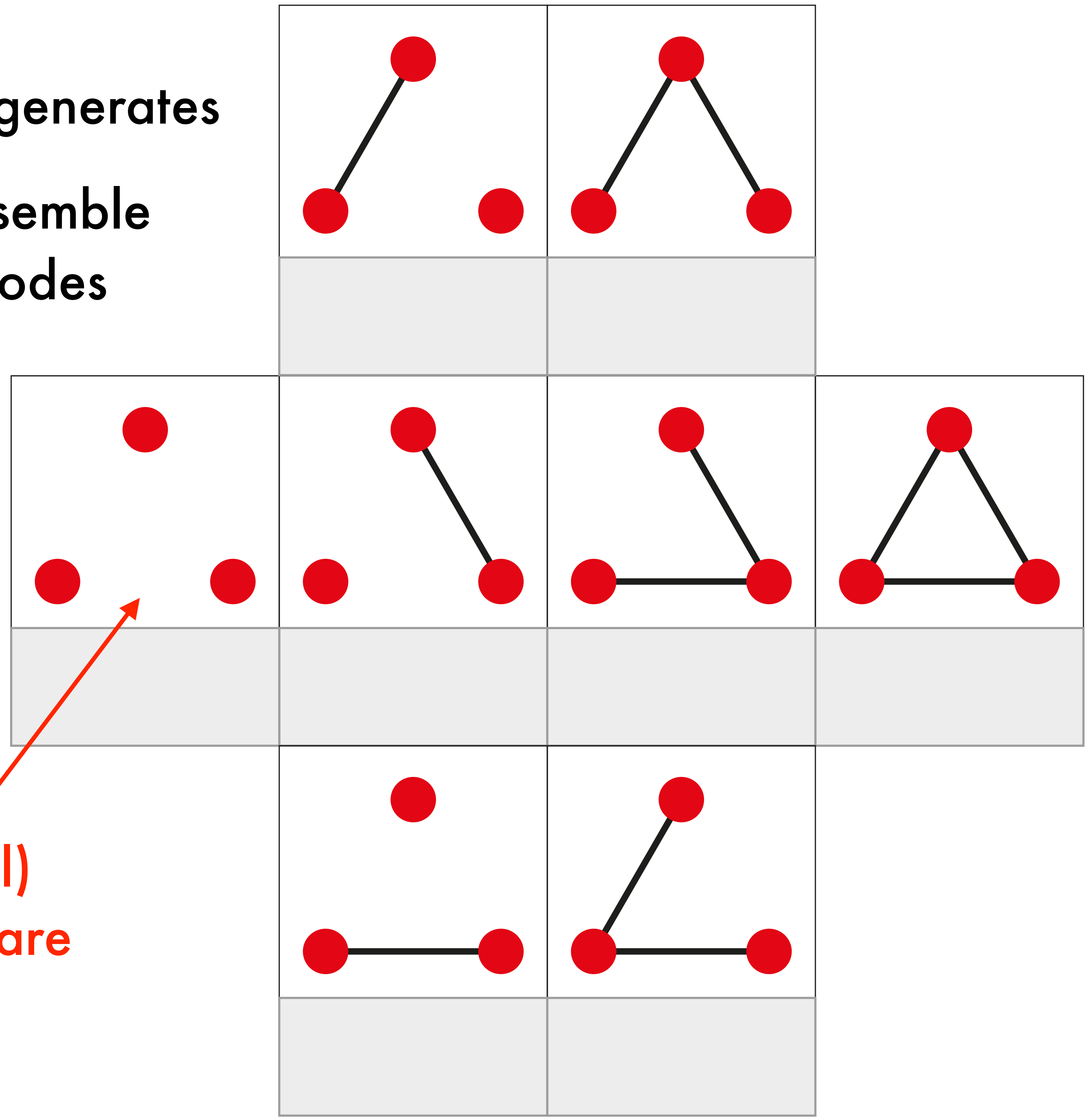
$$N = 10$$
$$p = 1/5$$
$$\langle k \rangle = 1.8$$

Erdős, P.; Rényi, A. (1959). "On Random Graphs. I.". Publicationes Mathematicae 6: 290–297.

Ensemble $G(N,p)$ with $N = 3$

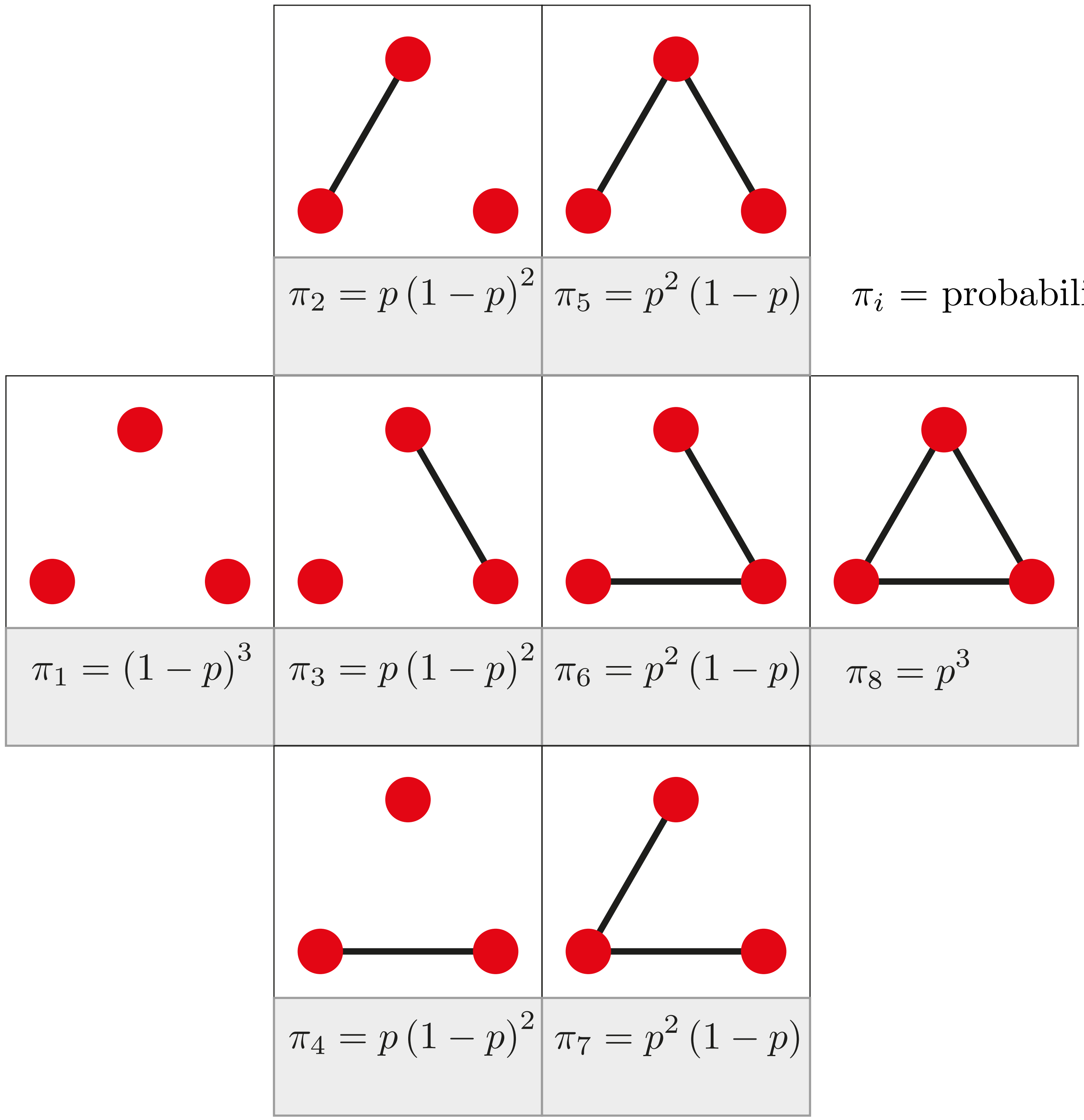
Ensemble of model
= set of all graphs it generates

E-R model $G(N,p)$: ensemble
= all graphs with N nodes



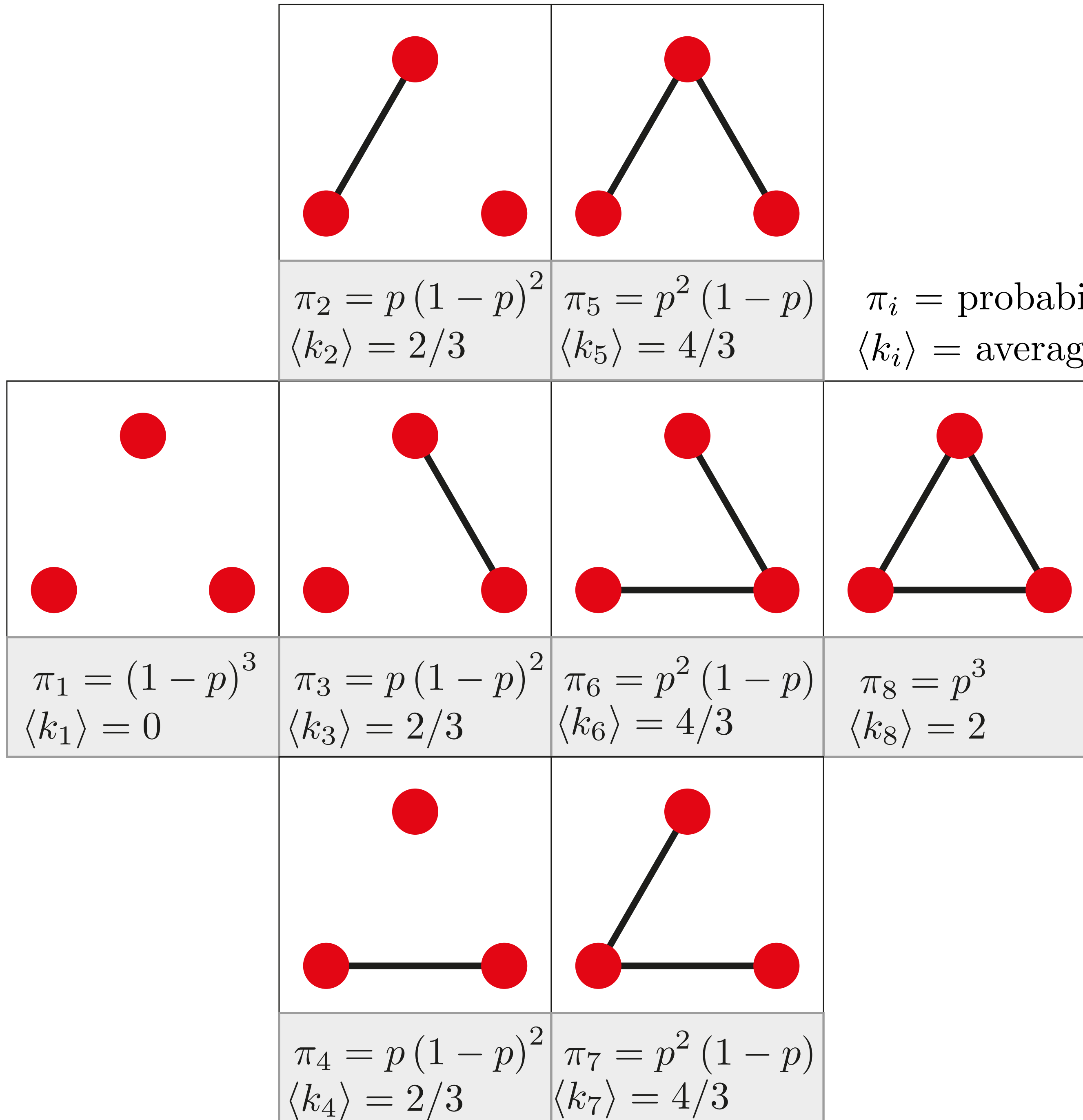
empty (or full)
networks are rare

Ensemble G(N,p) with N=3



π_i = probability of obtaining graph i

Ensemble G(N,p) with N=3



π_i = probability of obtaining graph i
$\langle k_i \rangle$ = average degree of graph i

Ensemble G(N,p) with N=3

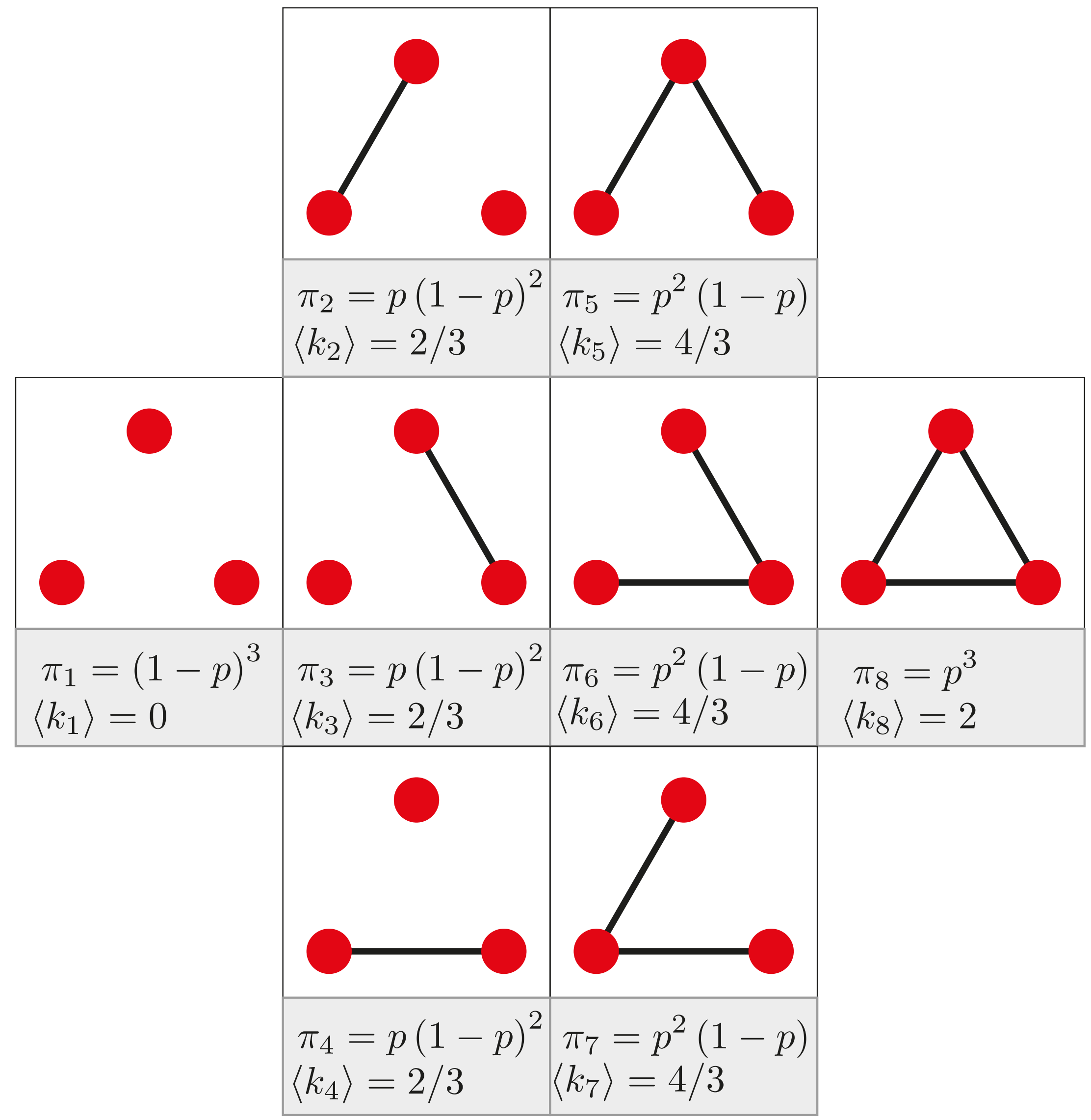
- The properties of random graphs are defined as ensemble averages, e.g. mean degree is

$$\langle k \rangle = \sum_{i=1}^{8} \pi_i \langle k_i \rangle$$

m= # of links.

- The probability of each graph is $\pi_i = p^m(1-p)^{3-m}$:

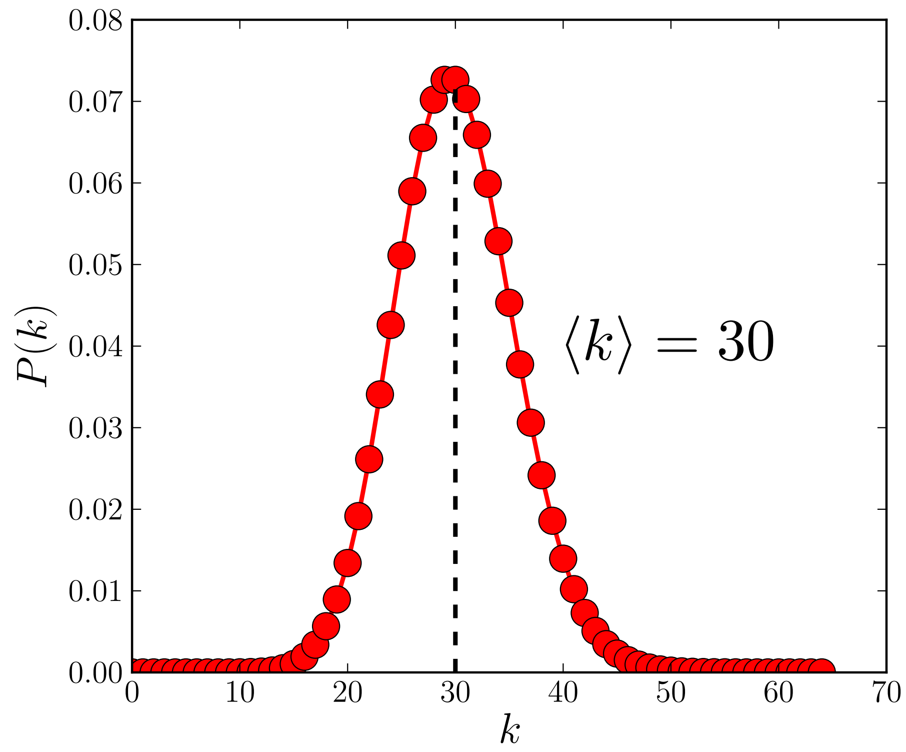
$$\pi_1 = (1-p)^3$$
$$\pi_2 = \pi_3 = \pi_4 = p(1-p)^2$$
$$\pi_5 = \pi_6 = \pi_7 = p^2(1-p)$$
$$\pi_8 = p^3$$



Properties of G(N,p)

Edges & degrees

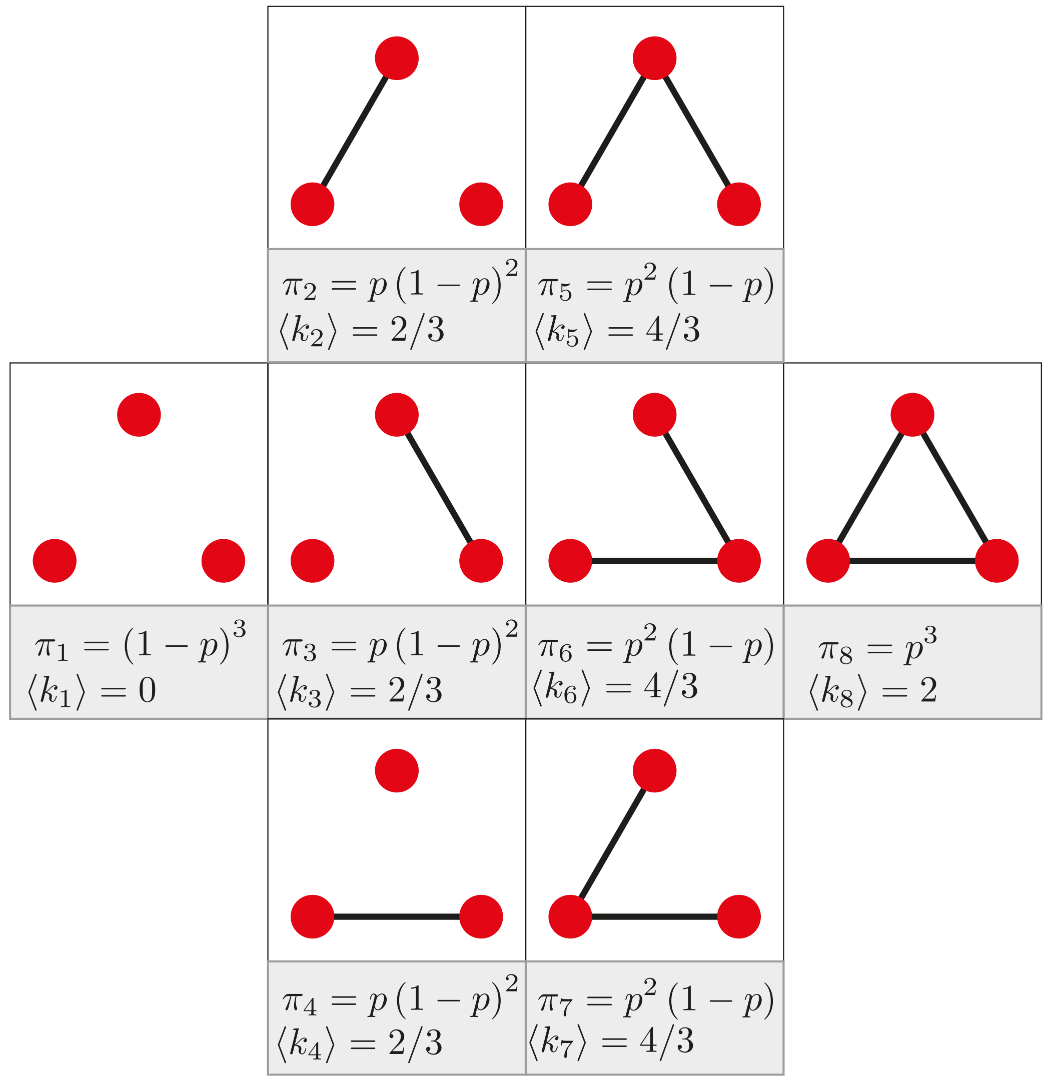
- On average, the number of edges is $\langle m \rangle = \binom{N}{2} p = p \times N(N-1)/2$.
- Hence the average degree is $\langle k \rangle = 2\langle m \rangle / N = (N-1)p \approx Np$.



Degree distribution

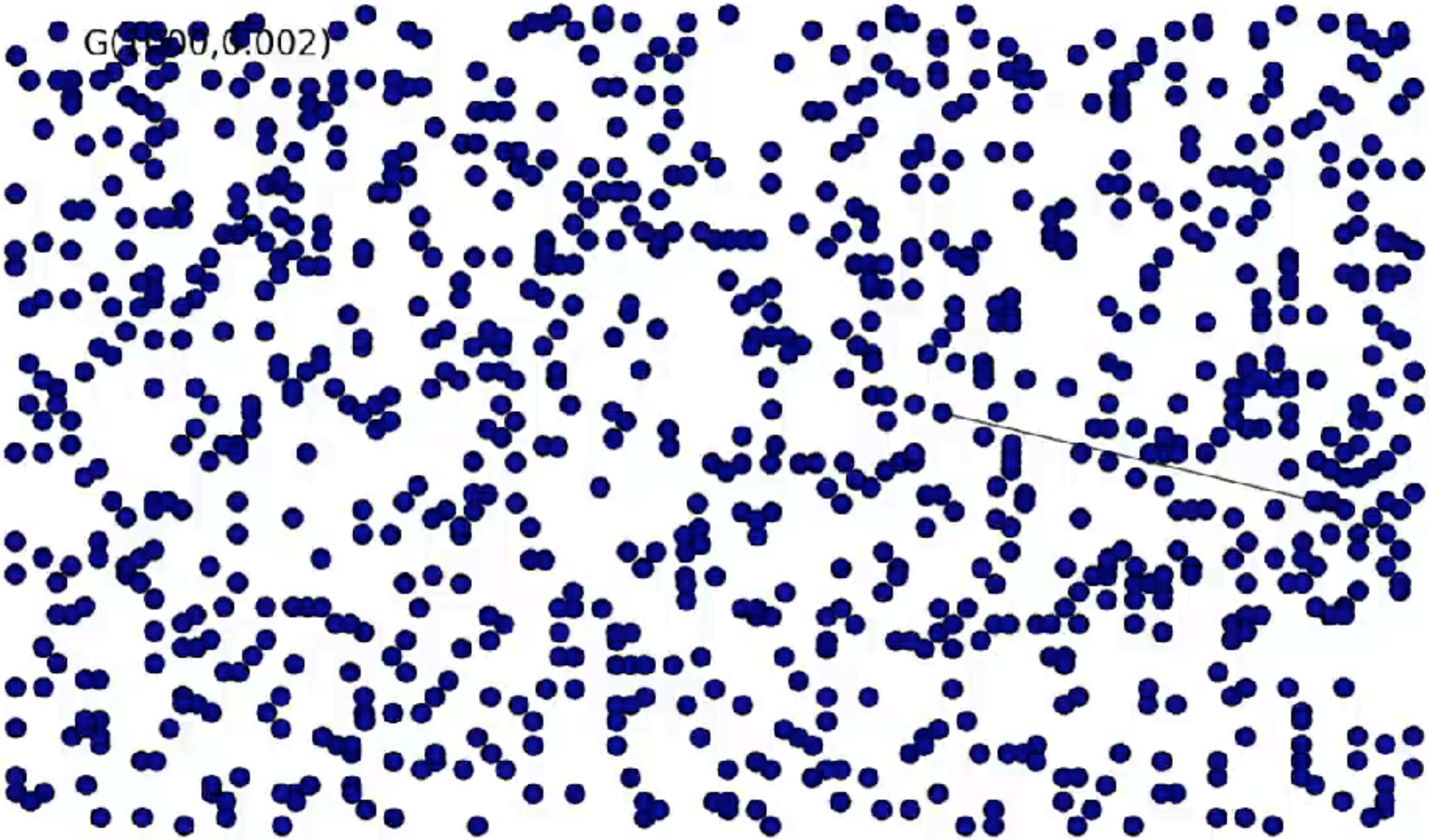
- Each node's number of links comes from $N-1$ independent trials with probability p.
- Hence $P(k) = \mathrm{Bin}\left((N-1), p\right)$
  $= \binom{n-1}{k} p^k \left(1-p\right)^{n-1-k}$
- For $N \to \infty$ with $\langle k \rangle$ constant,
  $P(k) \to \frac{\langle k \rangle^k}{k!} e^{-\langle k \rangle}$,
  that is, $P(k) = \mathrm{Poisson}(\langle k \rangle)$.

Ensemble $\mathrm{G}(N,p)$ with $N=3$



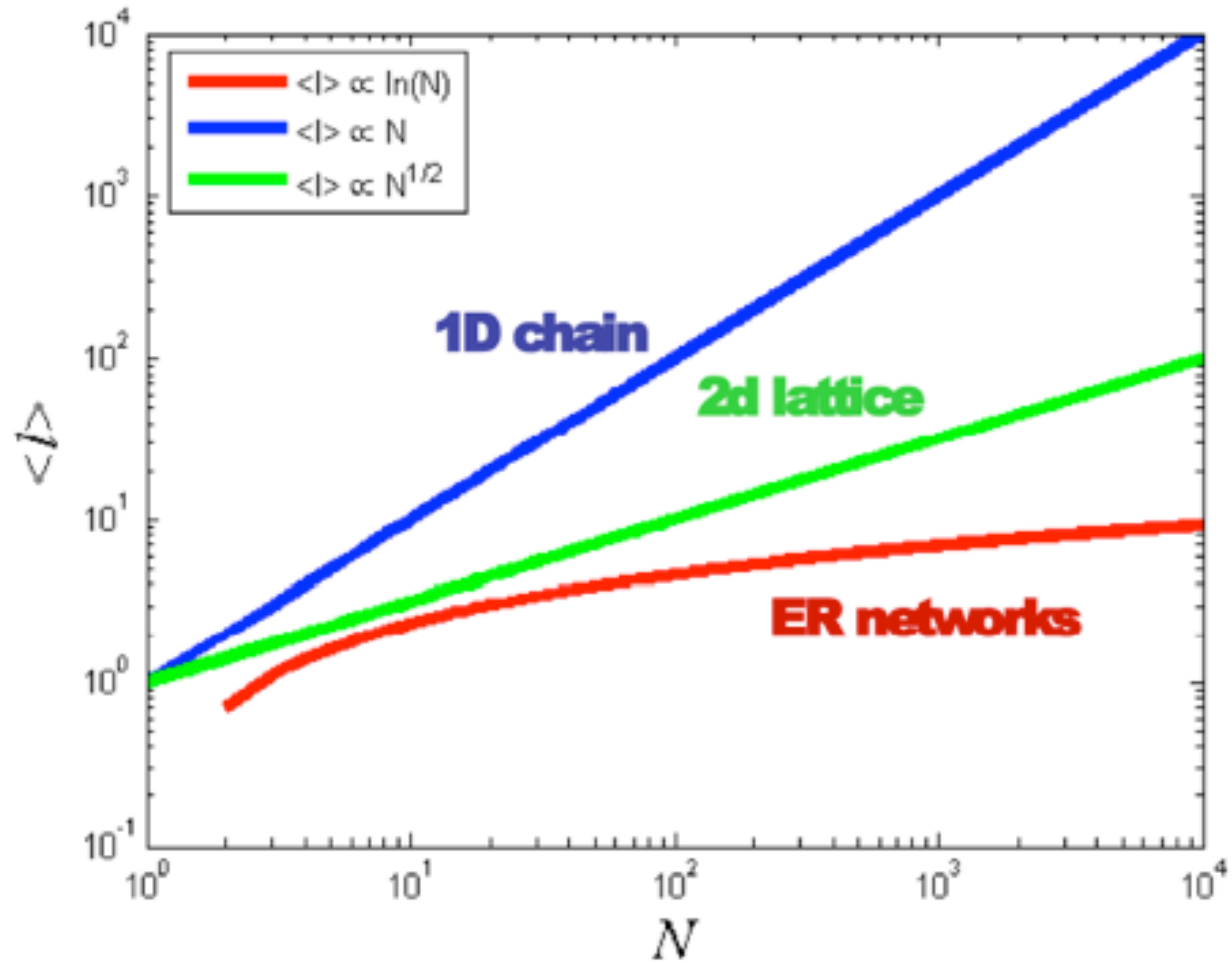
$$\text{if } p = 1/3, \text{ then } \langle k \rangle = \sum_{i=1}^{8} \pi_i \langle k_i \rangle = \frac{2}{3} = p(N-1)$$

Evolution of G(n,p) when <k> is increased



Shortest paths and clustering in Erdős-Rényi networks

- Remember: clustering coefficient is the density of the local neighbourhood of a node
- In ER network links are independent:
  $$\langle c \rangle = p = \langle \rho \rangle$$
- For ER networks it turns out that:
  $$\langle l \rangle \propto \ln N$$
- ER networks are said to be infinite dimensional because average path lengths grow slower than in any lattice



The Erdős Number

- Pál Erdős published an impressive number of scientific papers (512!)
- Scientists calculate their Erdős number as the collaboration distance to Erdős
- Authored a paper with Erdős: distance = 1, authored a paper with someone who authored a paper with Erdős: distance = 2, etc.
- The average distance is about 5, and for those who have a finite Erdős number it is usually <8

http://www.ams.org/mathscinet/collaborationDistance.html

Network	Degrees	Paths	Clustering
Real-world	Fat-tailed	short	high
Regular lattices	Fixed	long	high*
Erdős-Renyi	Poissonian	short	low

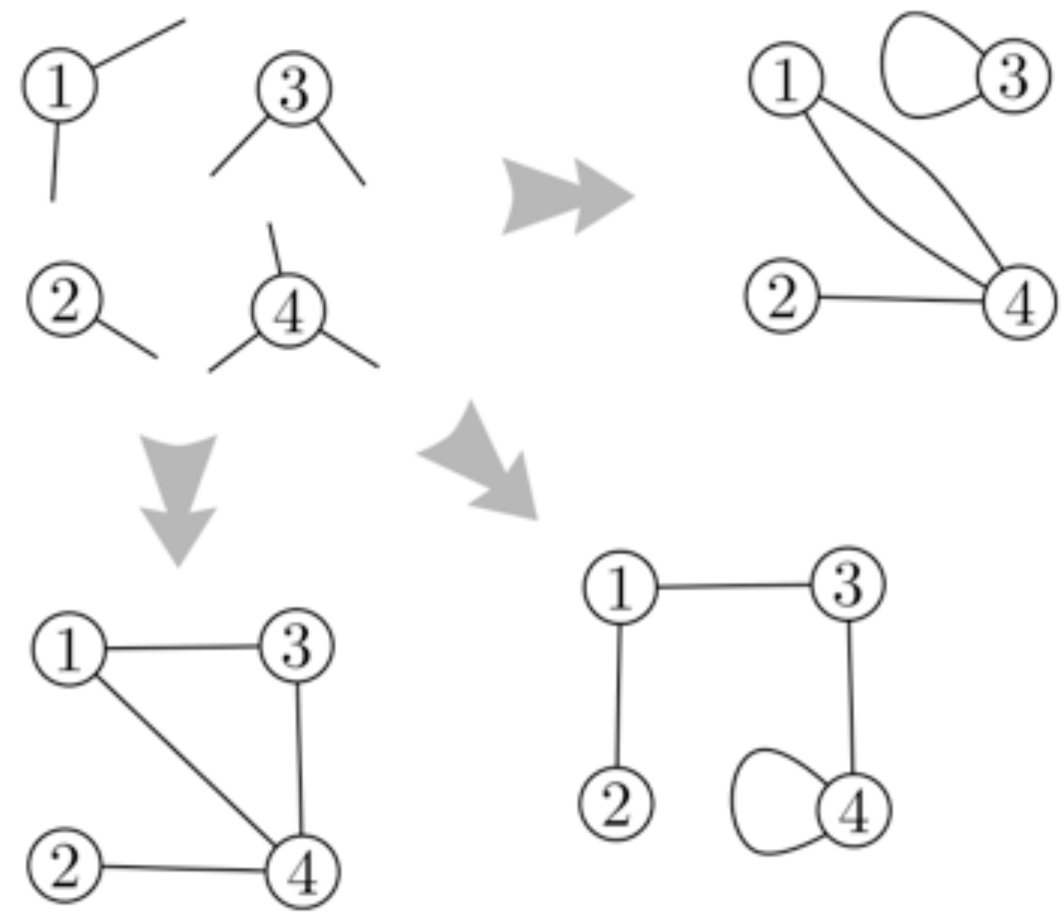
* (depending on layout & clustering measure)

Configuration model

- Motivation:
  - Real-world networks do not have Poissonian degrees
- Target:
  - Generate networks with chosen degrees that are otherwise maximally random

The configuration model algorithm

- Take a degree sequence $[k_1, k_2, \ldots, k_n]$
- Give each vertex k_i "stubs".
- Pick two stubs uniformly at random and create an edge by joining them.
- Repeat until no stubs remain.



[see Fosdick et al, https://arxiv.org/abs/1608.00607]

Network	Degrees	Paths	Clustering
Real-world	Fat-tailed	short	high
Regular lattices	Fixed	long	high*
Erdős-Renyi	Poissonian	short	low
Configuration	Free to choose	short	low

* (depending on layout & clustering measure)

birth of the science of complex networks
Small-World Networks

D.J. Watts and S. Strogatz,
”Collective dynamics of 'small-world' networks",
Nature **393**, 440–442, 1998

- The paper examines the properties of real-world networks
- It also presents a simple network model that reproduces those properties...
- ...and, first and foremost, shows that networks structure matters a lot.


letters to nature

Collective dynamics of 'small-world' networks

Duncan J. Watts & Steven H. Strogatz

Path lengths in real-world networks

Network	n	ℓ	ℓ_{rand}
film actors	22500	3.65	2.99
US power grid	4941	18.7	12.4
C.Elegans	282	2.65	2.25

number of nodes

average shortest path length

average shortest path length in E-R networks (same *n*, same number of links)

Observation: in real-world networks, path lengths resemble those of E-R networks.

Watts, D. J. & Strogatz, S. H., Collective dynamics of ‘small-world’ networks. Nature **393**, 440-442 (1998)

Clustering coefficient in real-world networks

number of nodes

average shortest path length

average shortest path length in E-R networks (same *n*, same number of links)

Network	n	ℓ	ℓ_{rand}	C	C_{rand}
film actors	22500	3.65	2.99	0.79	0.00027
US power grid	4941	18.7	12.4	0.08	0.006
C.Elegans	282	2.65	2.25	0.28	0.05

clustering coefficient

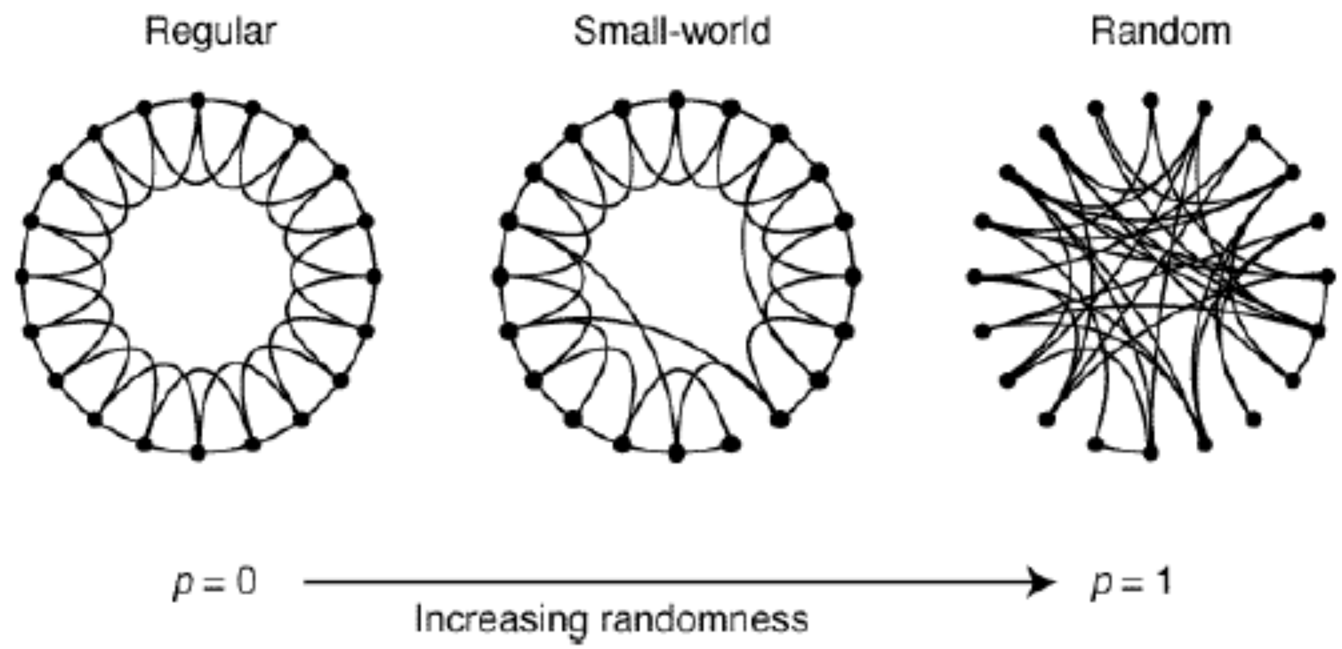
clustering coefficient in E-R networks (same *n*, same number of links)

The clustering coefficient is larger by orders of magnitude than the random reference value

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

Network	l_{data}	$\langle l \rangle_{\mathrm{ER}}$	c_{data}	$\langle c \rangle_{\mathrm{ER}}$
last.fm social net	5.6	7.1	0.19	$3.2\ 10^{-6}$
Airports & flights	4.0	3.6	0.49	0.0034
Protein	4.8	7.0	0.02	0.0012
Internet AS	5.1	5.9	0.25	$7.7\ 10^{-6}$

The Watts-Strogatz Model



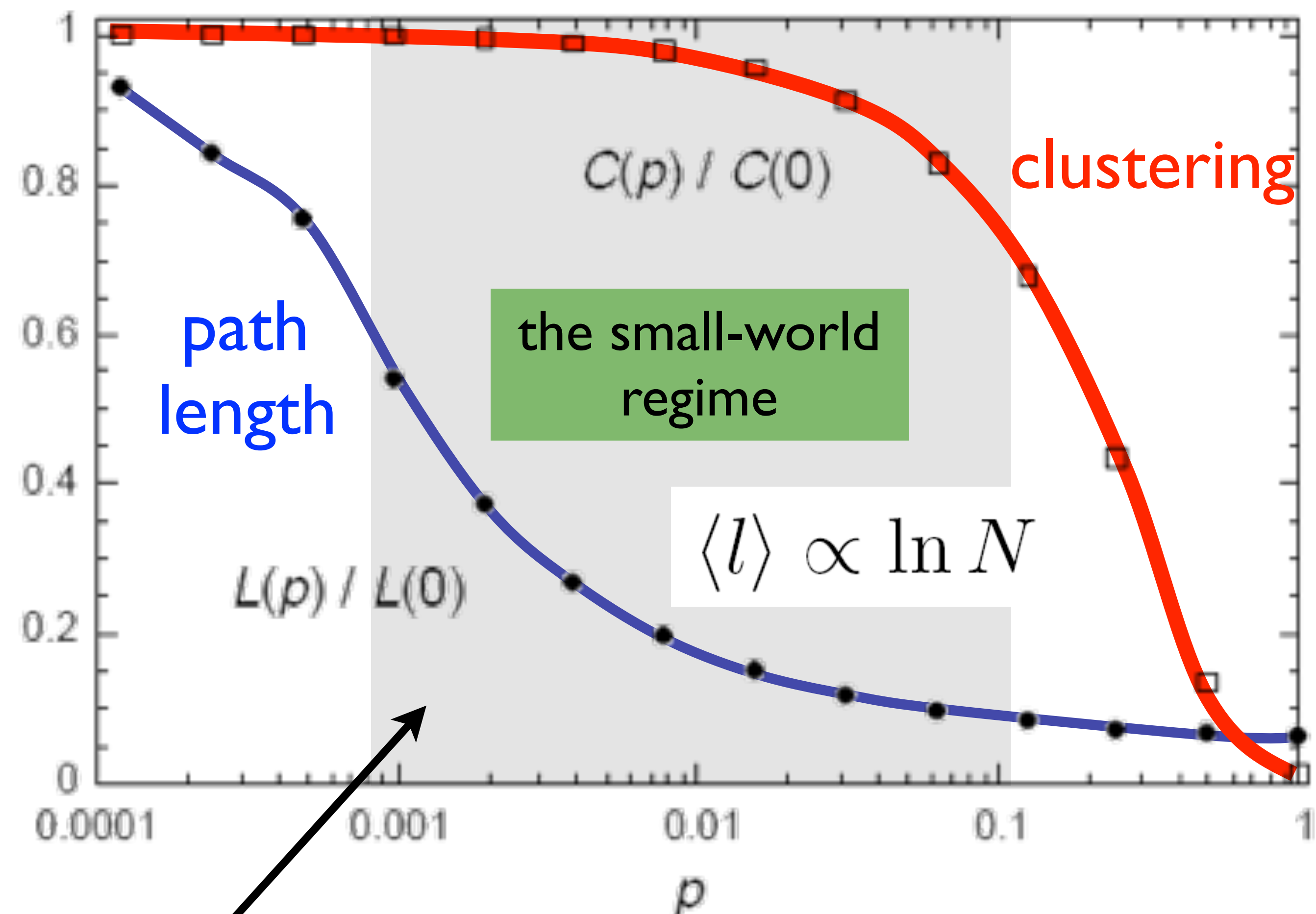
Watts, & Strogatz, Nature 393, 440-442 (1998)

- The model:
  - Take a regular 1D network with triangles (high clustering)
  - With probability p, randomly rewire the endpoint of each link

Path lengths and clustering in the WS model

$C(p)$ = average clustering coefficient

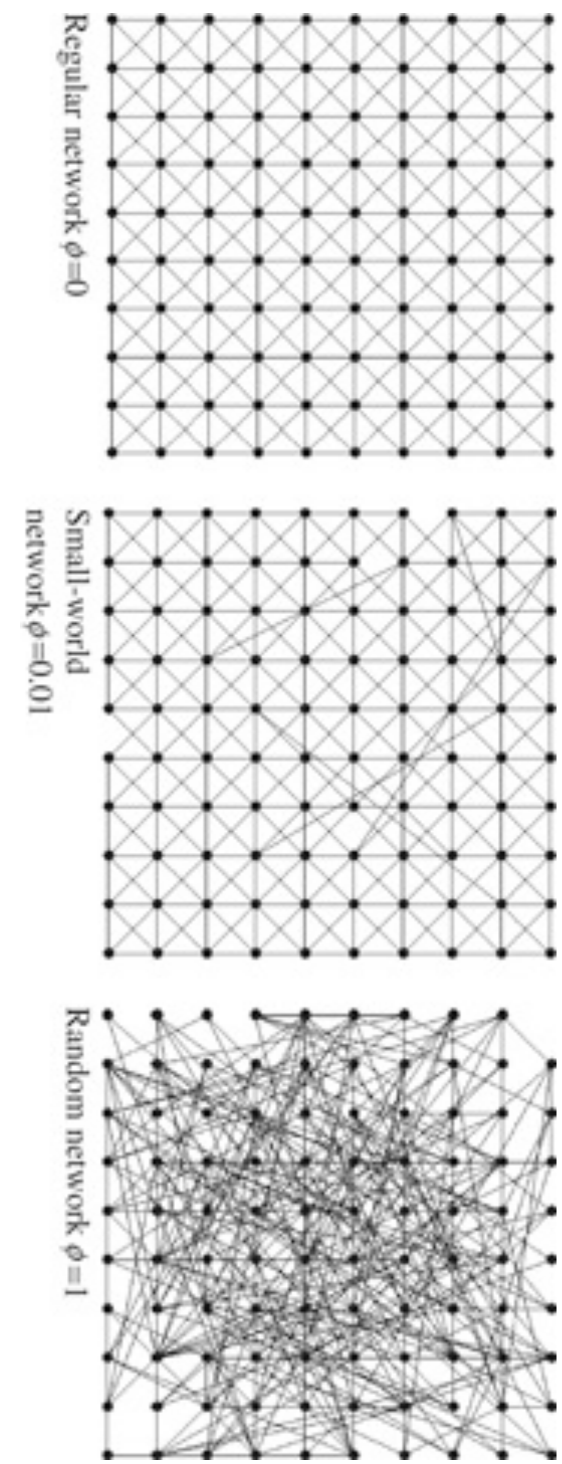
$L(p)$ = average shortest path length



for low p, clustering is high but path lengths rapidly decrease

Random shortcuts & path lengths

- Note: the SW rewiring procedure that produces random "shortcuts" works on any network.
- Shortcuts decrease path lengths.
- If the network is clustered, rewiring some links does not destroy this.
- Only networks where most links are extremely unlikely are not small worlds.
- Because real-world networks contain an element of randomness, they are naturally small worlds.



From Xu et al, Geographical Analysis 41(3):263 - 282 · July 2009

Real-wold networks: between randomness and order

Network	Degrees	Paths	Clustering
Real-world	Fat-tailed	short	high
Regular lattices	Fixed	long	high*
Erdős-Renyi	Poissonian	short	low
Configuration	Free to choose	short	low
Watts-Strogatz	Fixed to Poisson	short	high

* (depending on layout & clustering measure)

The deeper meaning of the results of Watts & Strogatz

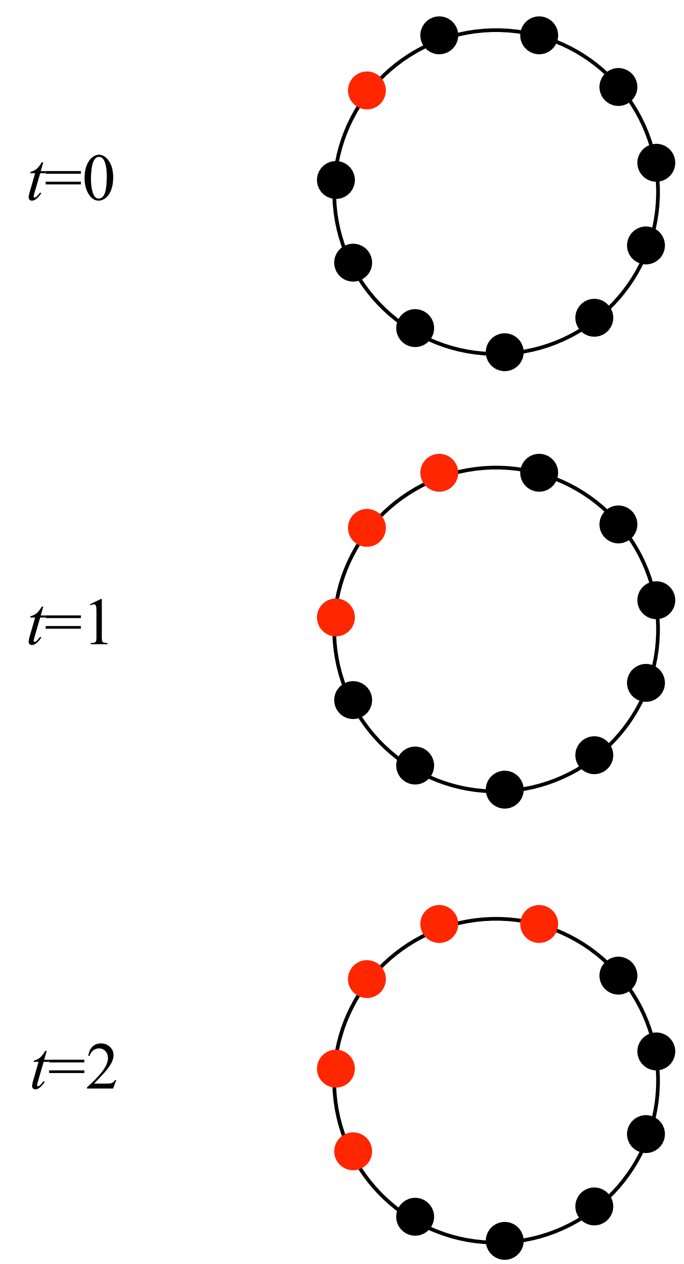
- W & S simulated many processes in small-world model networks while varying the randomness p
- They saw that the network structure is very important for those processes:
  - epidemic spreading
  - synchronization of oscillators
  - game-theoretical models
- **Many real-world processes take place on networks**
- **So to understand these processes, one has to understand the underlying network structure!**

Example: the SI model

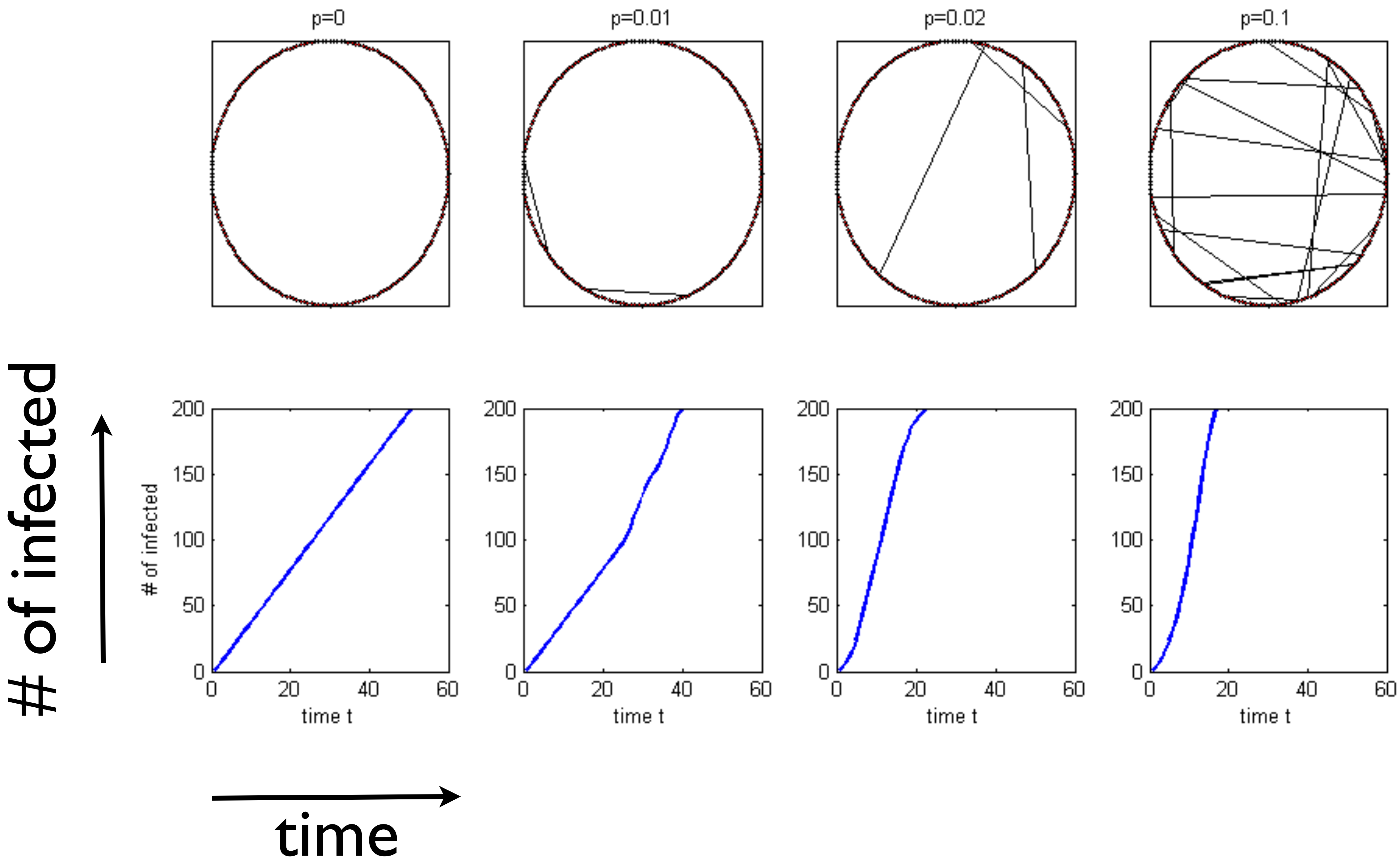
- The simplest possible model of (disease) spreading

- Set all nodes as (S)usceptible, except for one (I)nfected

- At every time step, every infected node infects its susceptible neighbours



The SI model on small worlds

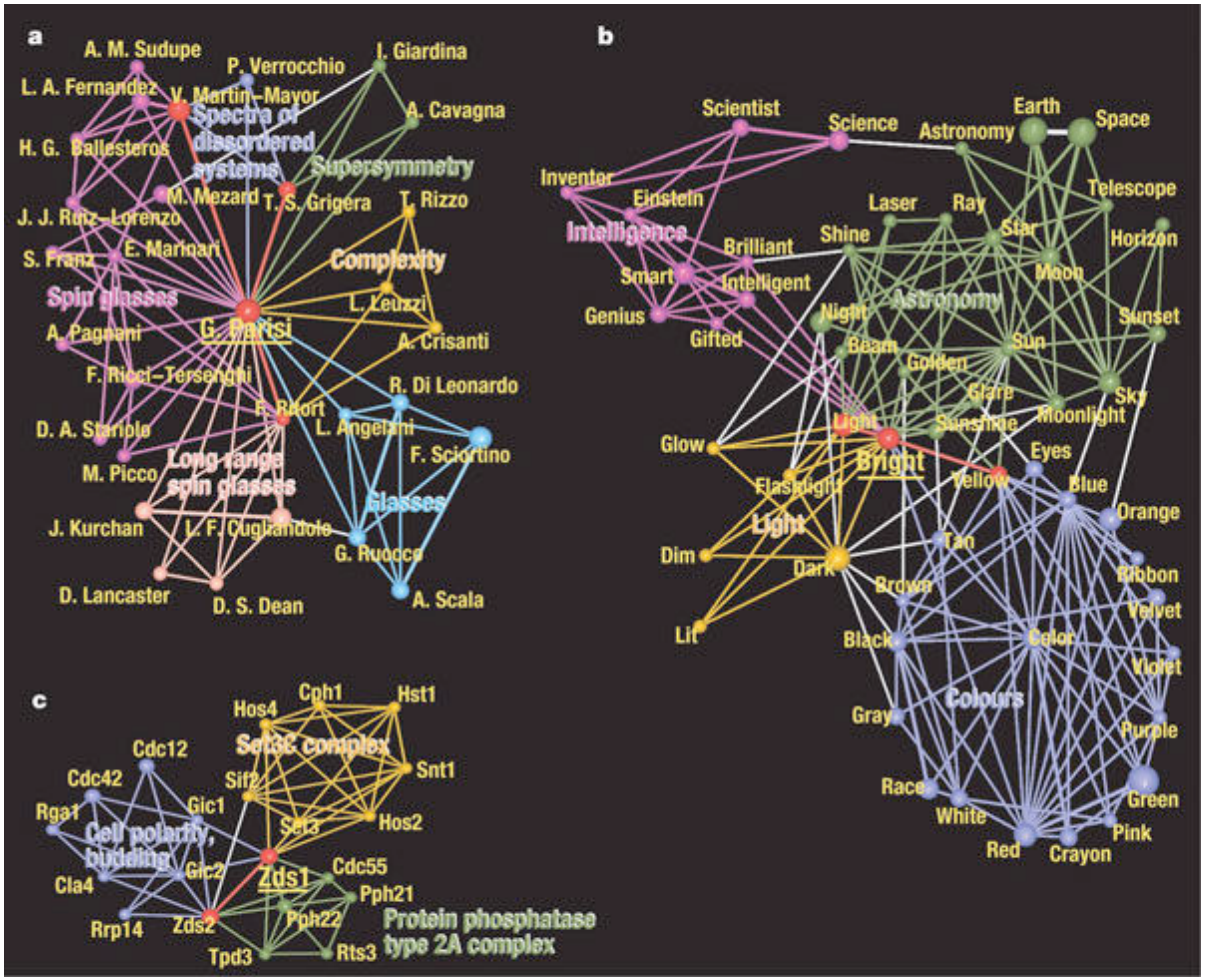


Why are real-world networks small worlds??

- **Path lengths:** because there is randomness in networks
- **Clustering:** a much harder question...
  - The answer depends on the network - and usually tells something important about the network!

- Social networks: clustered because people learn to know new people through their friends
- Biological networks: because genes & proteins participate in functional modules; because of duplication-divergence processes; etc
- Infrastructure networks: because they are embedded in 2D space

Triangles emerge naturally in most real-world networks.



Palla et al, Nature 435 (2005)