CS-E5740
Complex Networks
Scale-free 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

From last week

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)

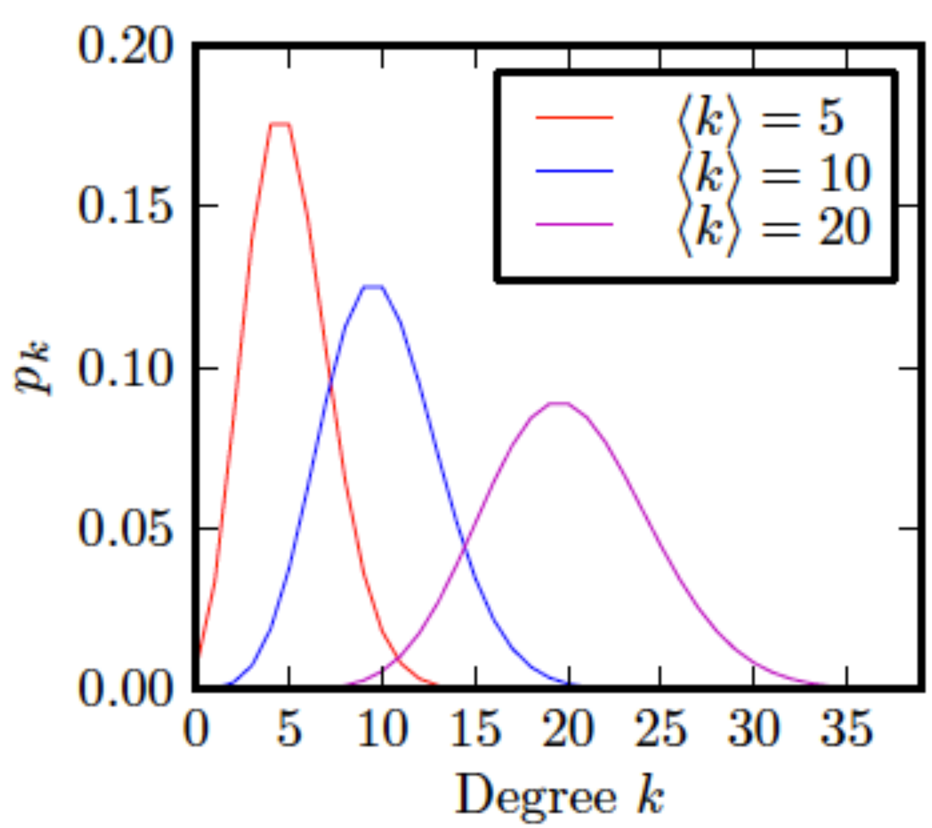
On Network Models

- “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

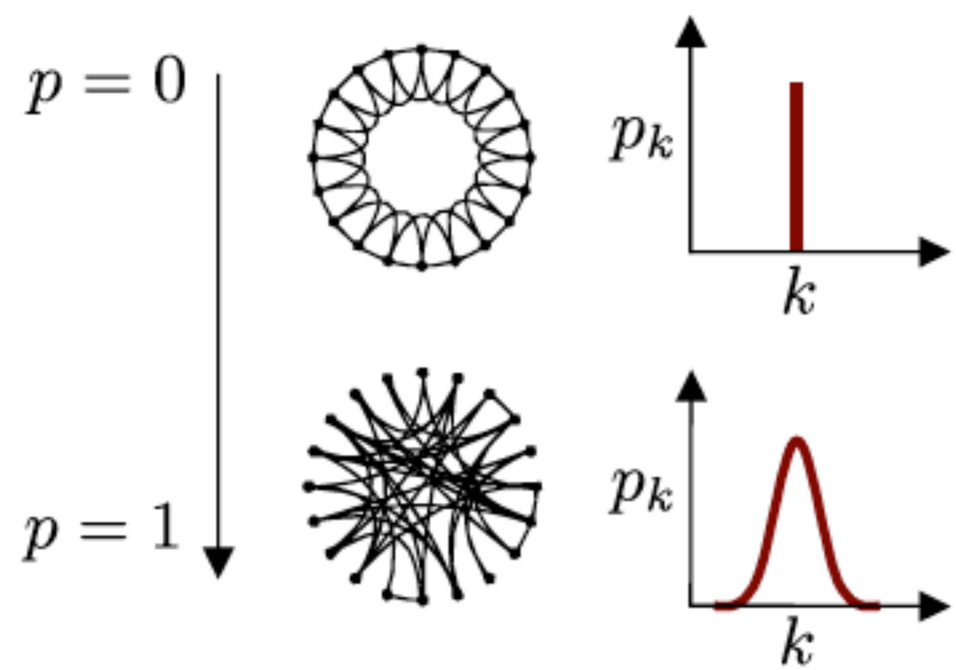
Degree distributions in model networks so far...

- Erdős-Rényi networks:

$$p_k = \frac{\langle k \rangle^k}{k!} e^{-\langle k \rangle}$$

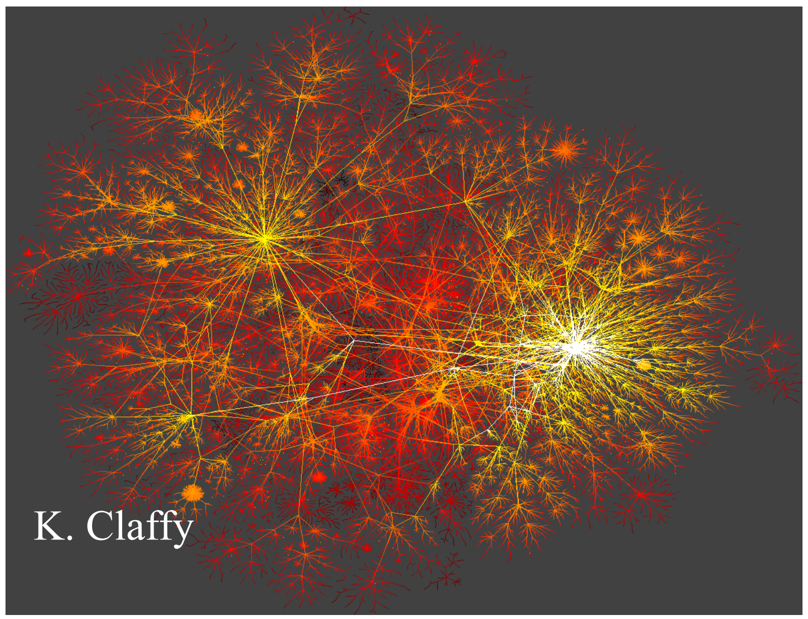


- Watts & Strogatz small-world model

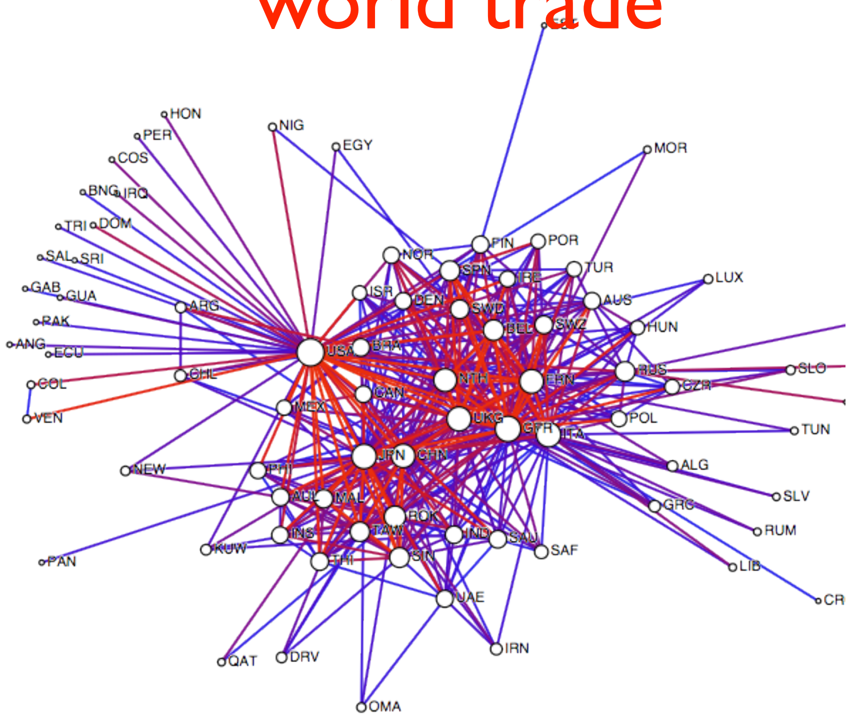


...however, real-world networks look like this:

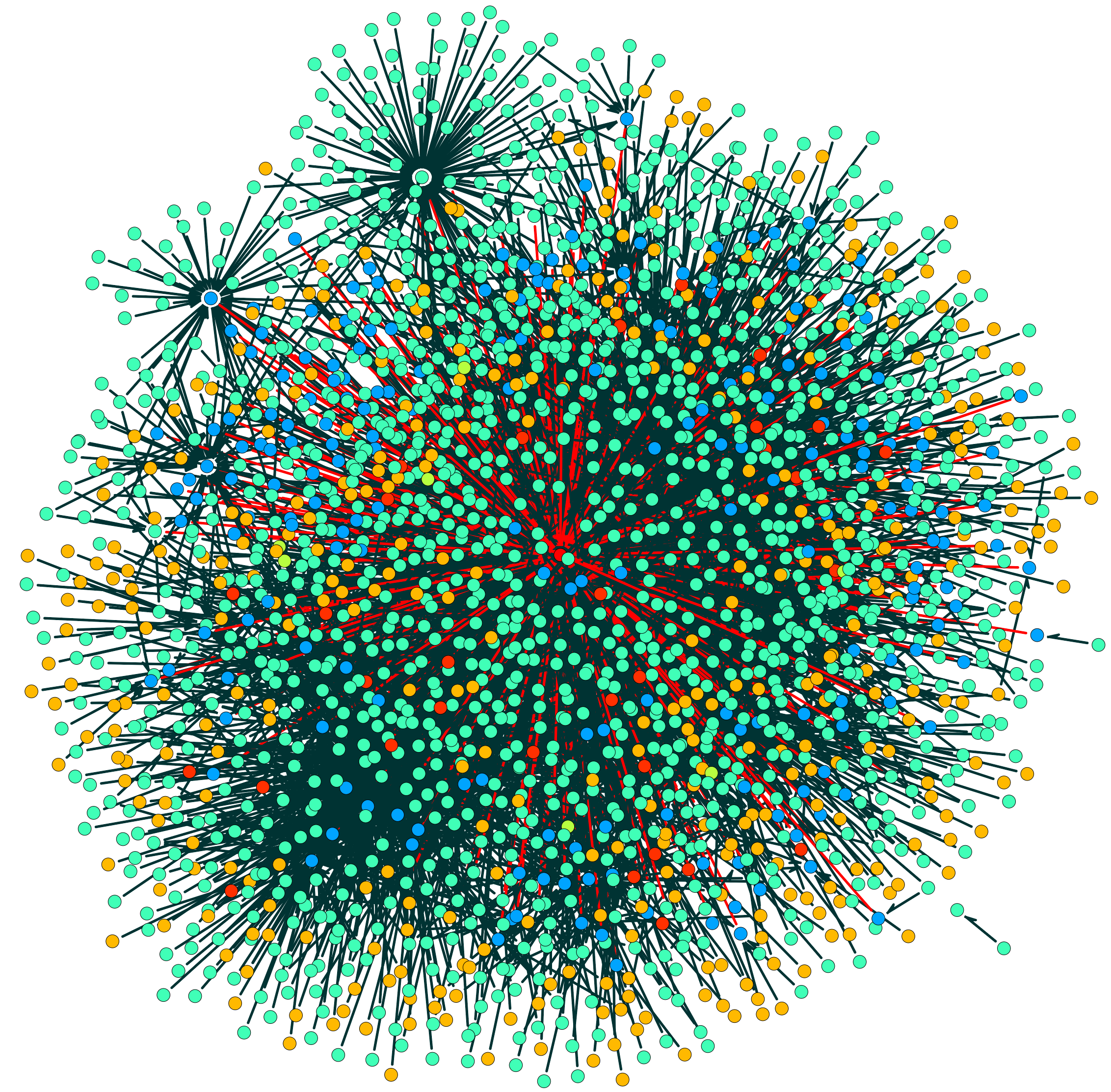
the Internet



world trade

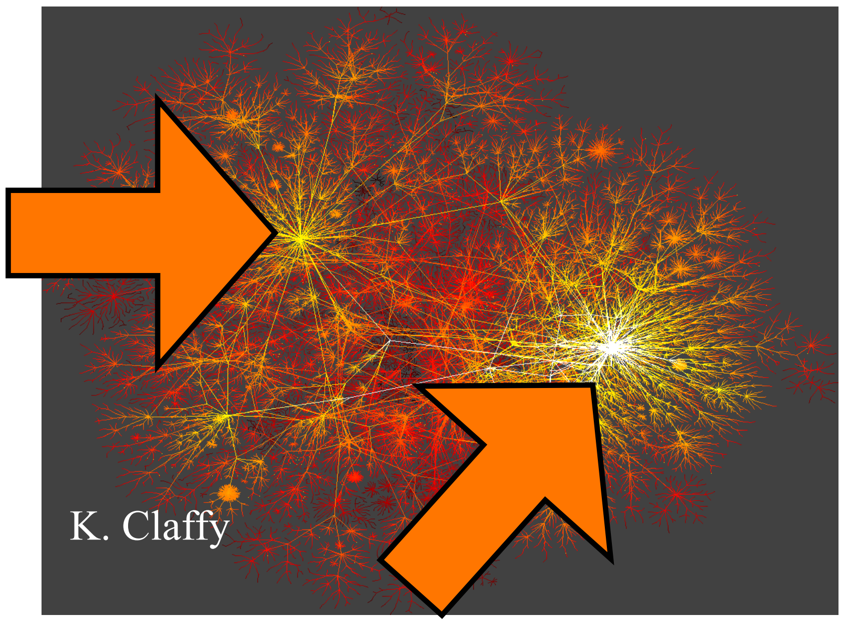


subcontracting network of a car manufacturer

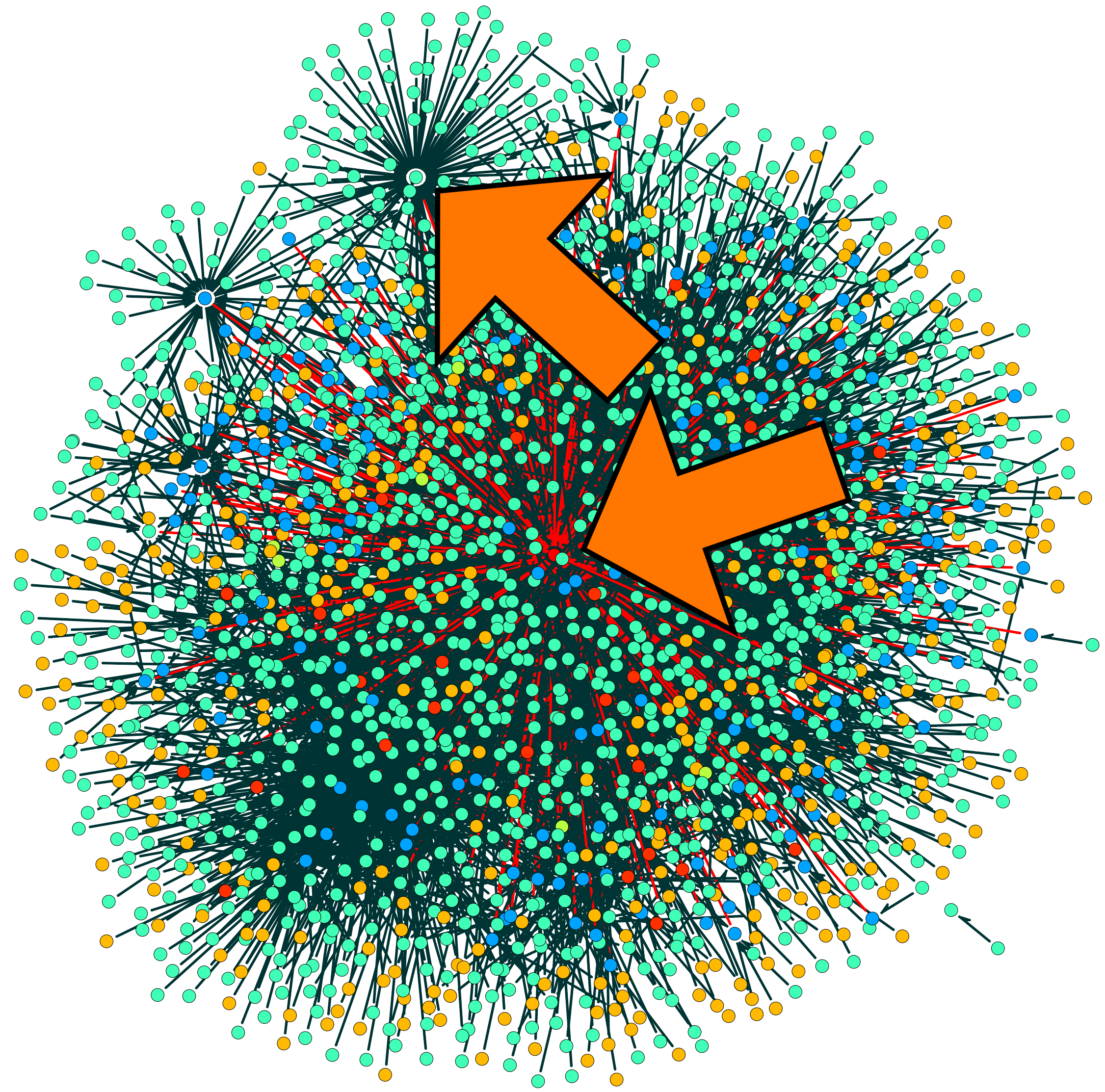


there are HUBS, nodes of very high degree

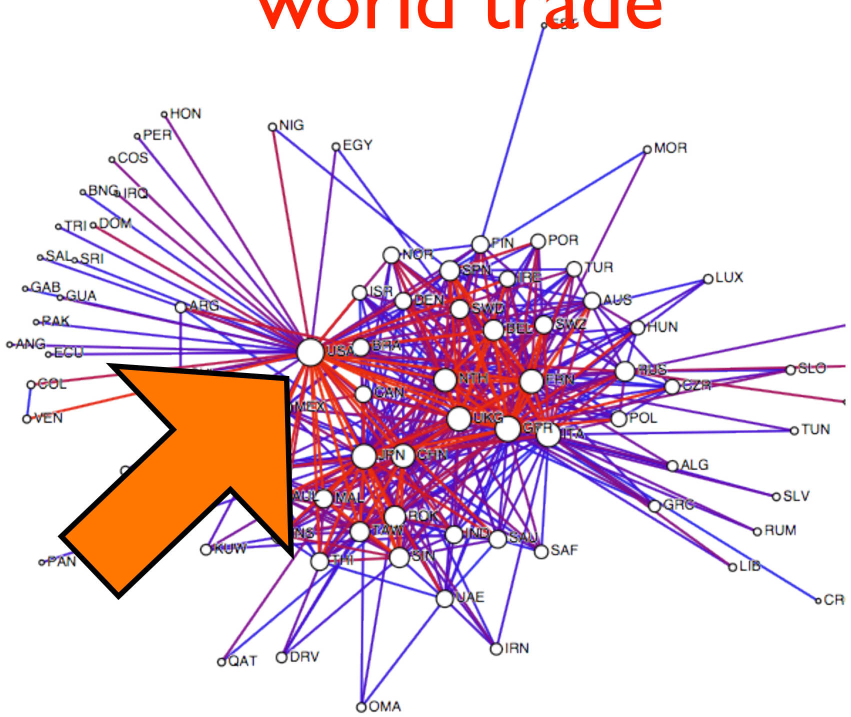
the Internet



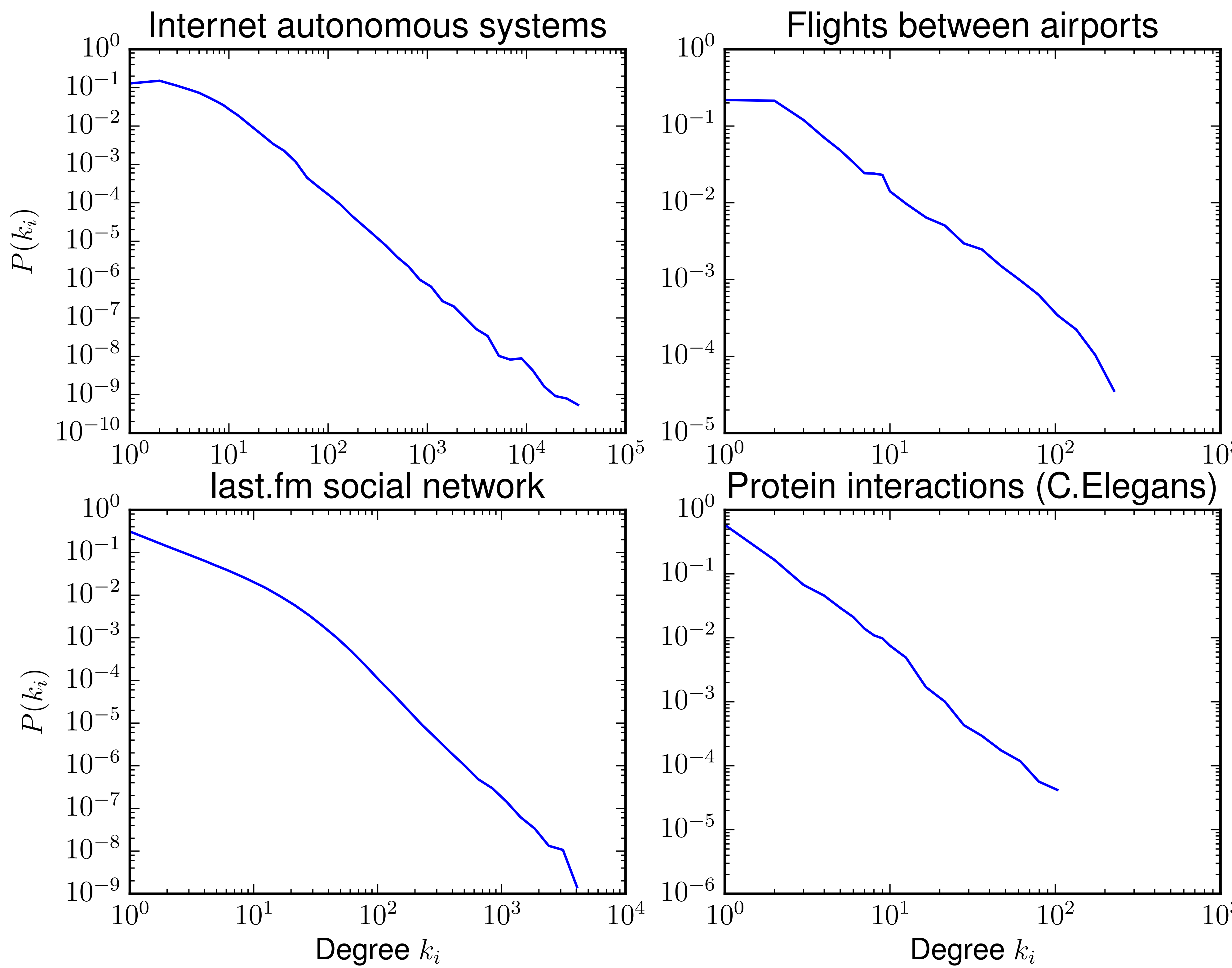
subcontracting network of a car manufacturer



world trade

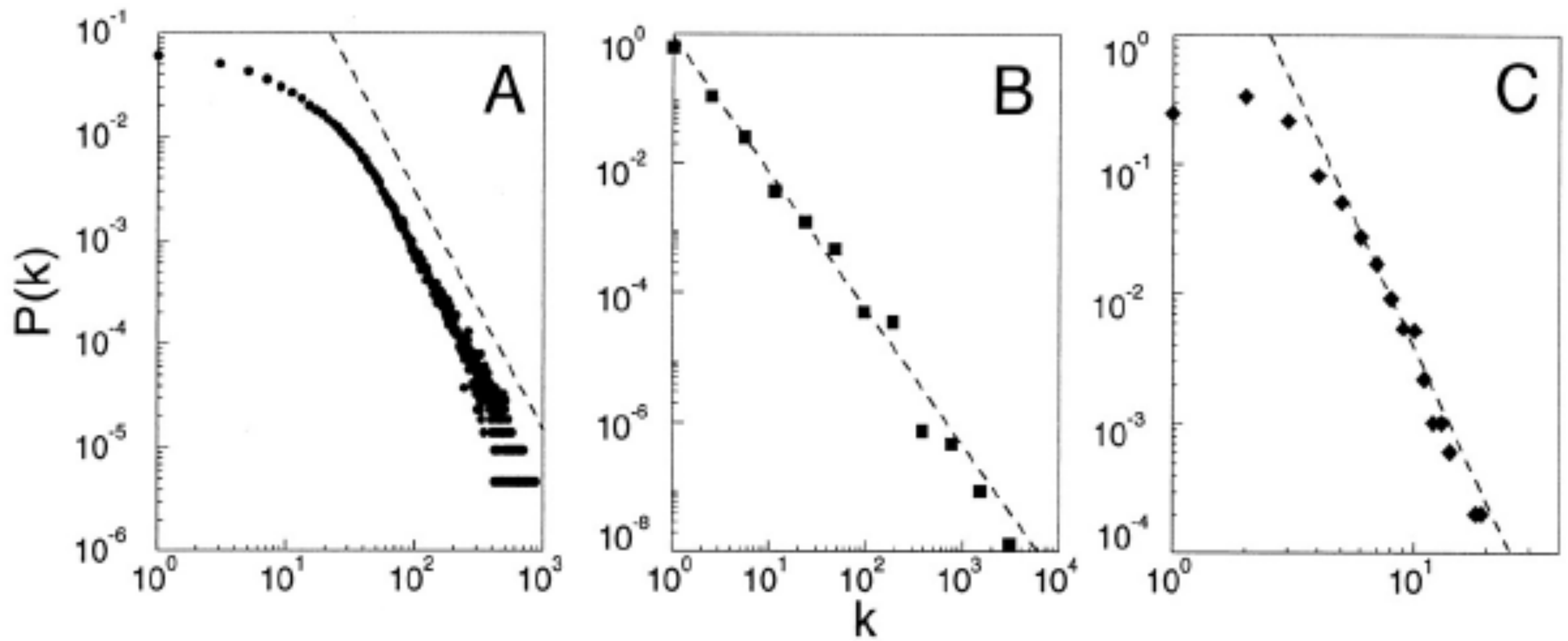


Degree examples



Degree distributions in real-world networks

Barabási, A.-L. & Albert, R., Science 286, 509-512 (1999)



A Actor collaboration network

B WWW

C Power grid data

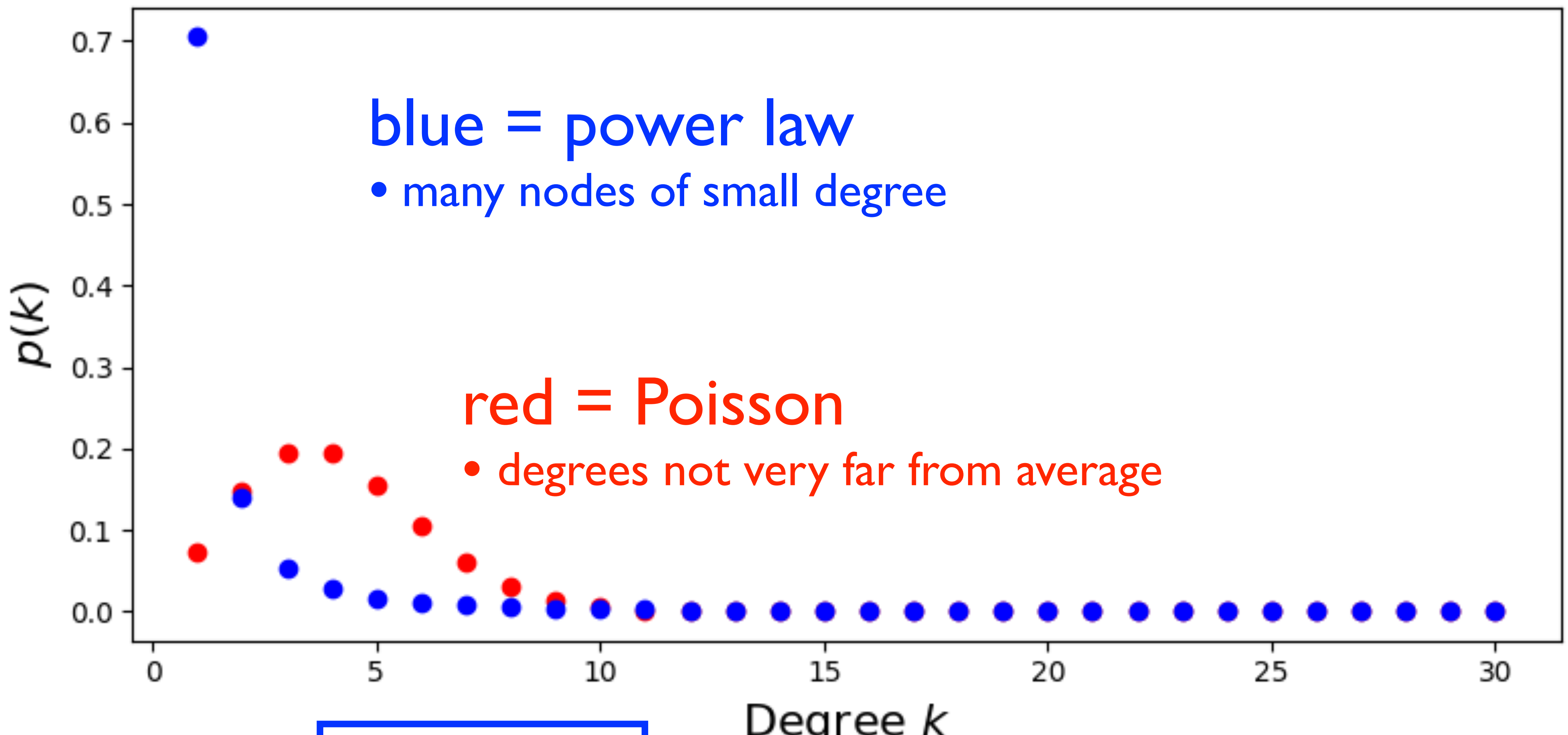
- The tail of the degree distribution can often be approximated with a power law

$$\boxed{P(k) \propto k^{-\gamma}}$$

$$\boxed{\log P(k) = -A \log k + B \Rightarrow P(k) = e^{B} k^{-A}}$$

networks with power-law distributed degrees are called **scale-free** networks

What does a power-law degree distribution mean?

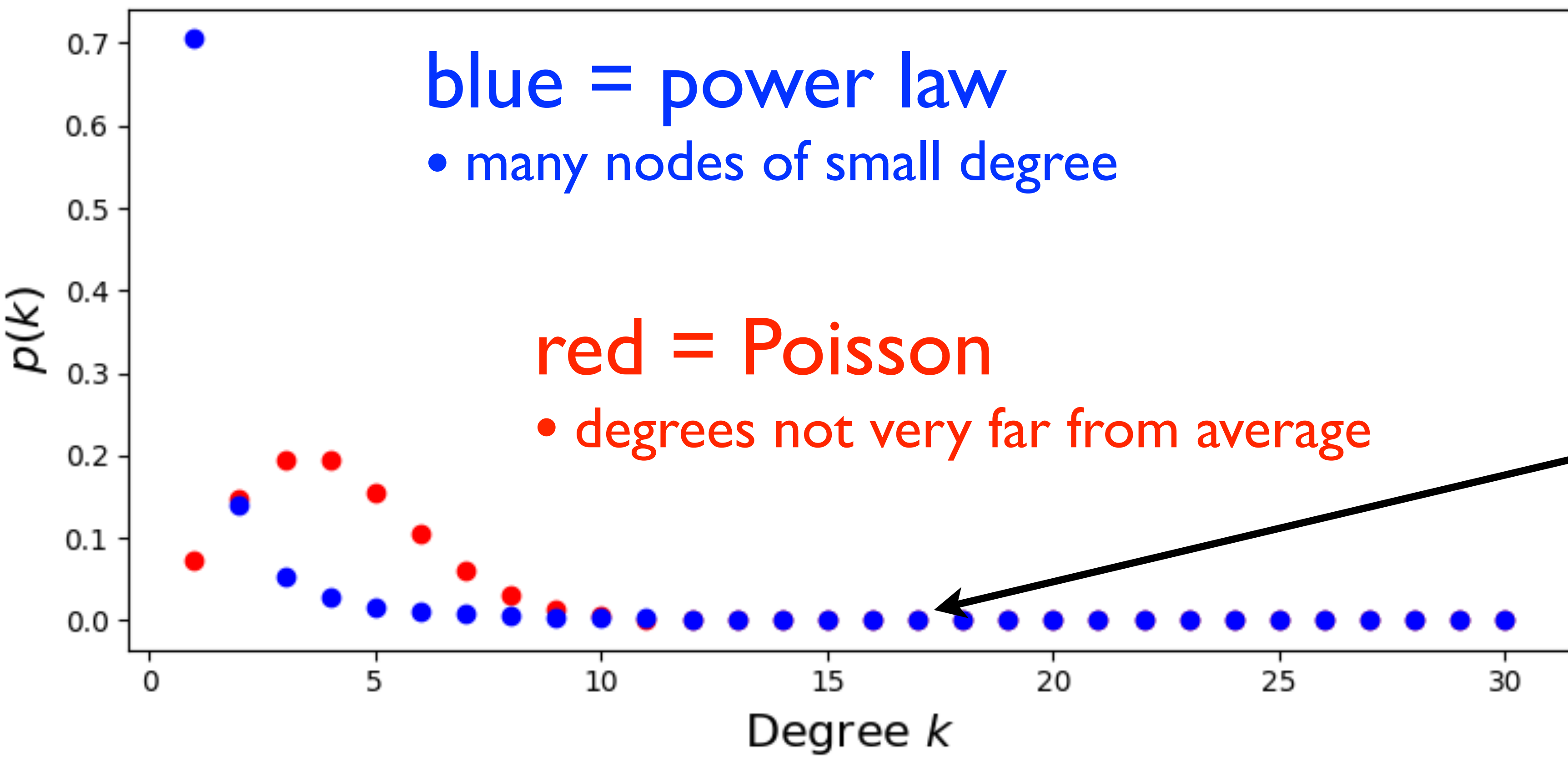


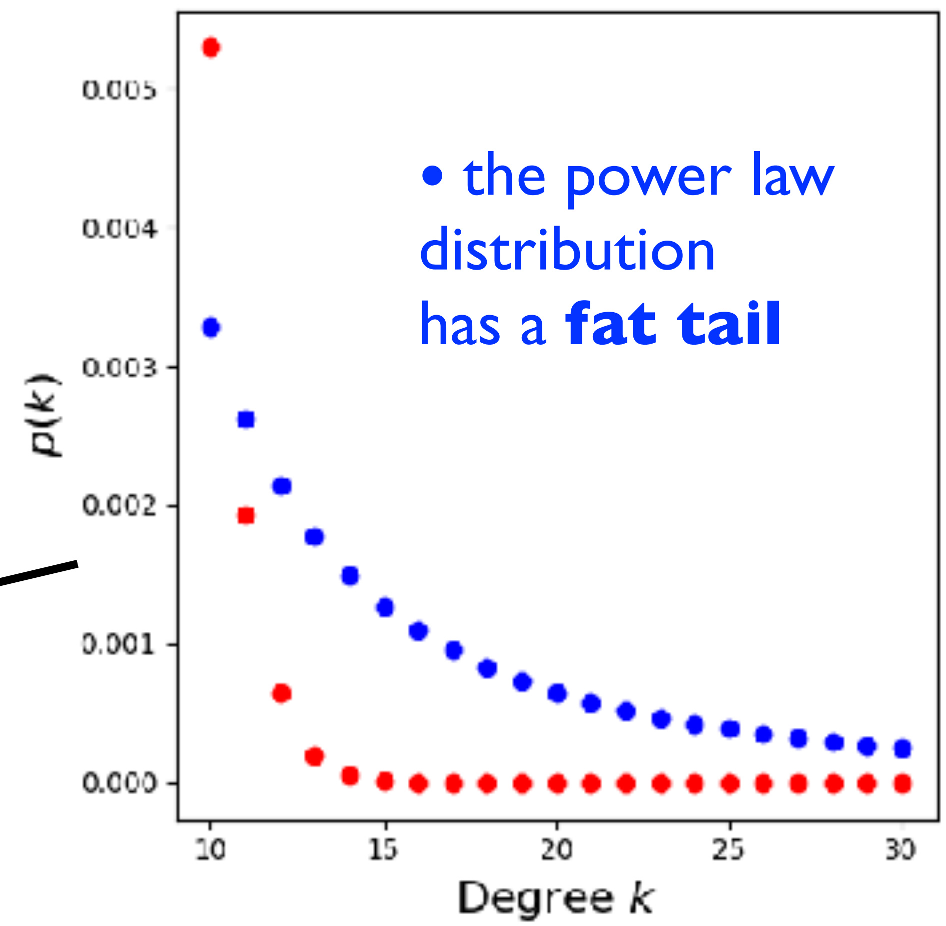
$$\boxed{P(k) \propto k^{-\gamma}}$$

$$\boxed{P(k) \sim e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}}$$

$\langle k \rangle = \langle k \rangle = 4$

What does a power-law degree distribution mean?



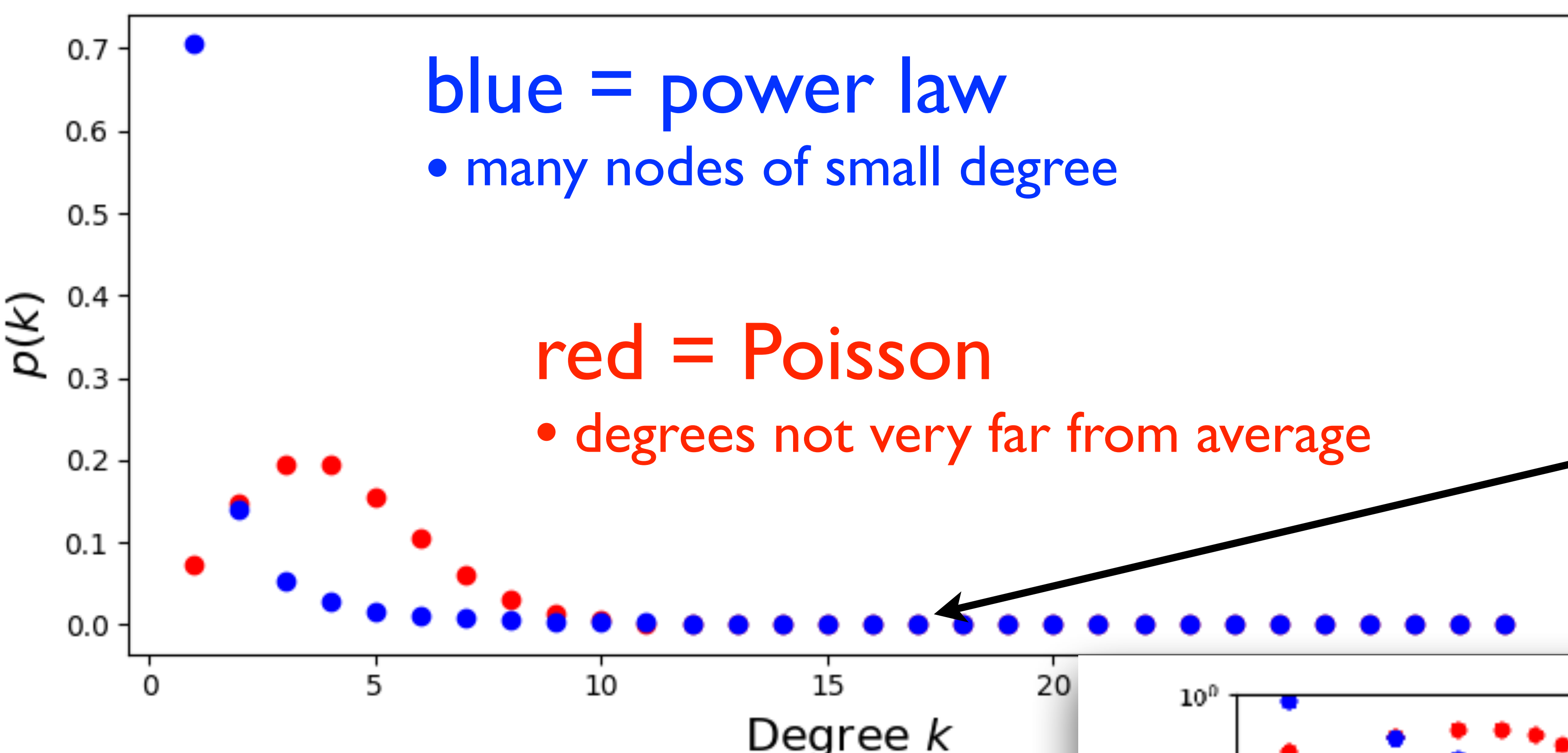


$$\boxed{P(k) \propto k^{-\gamma}}$$

$$\boxed{P(k) \sim e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}}$$

$<k> = <k> = 4$

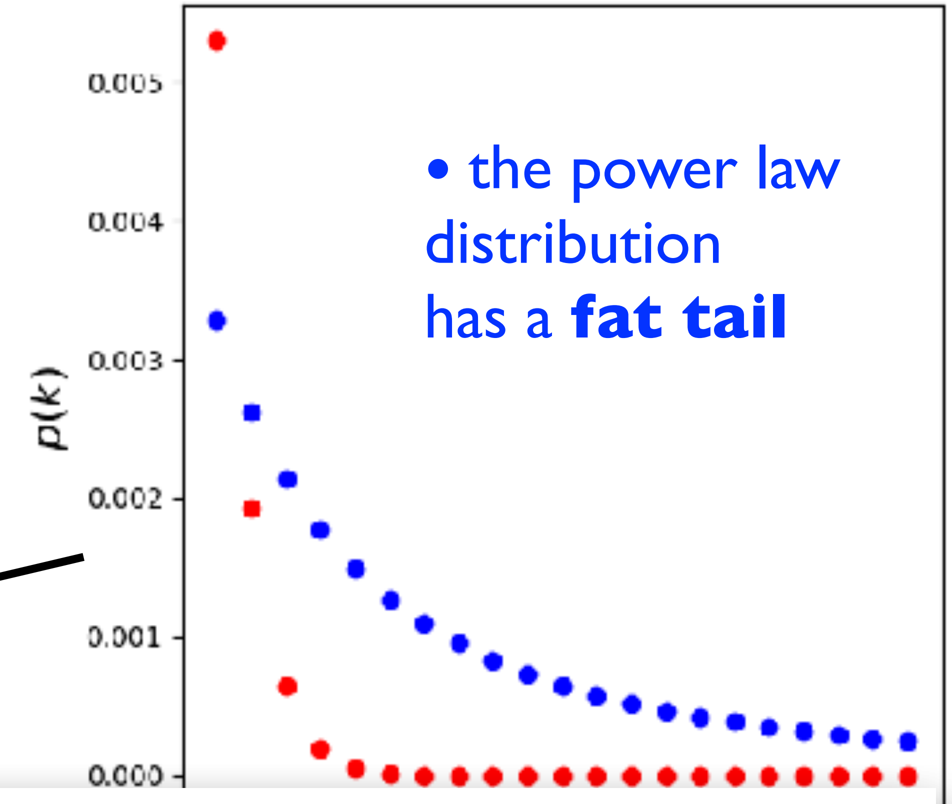
What does a power-law degree distribution mean?

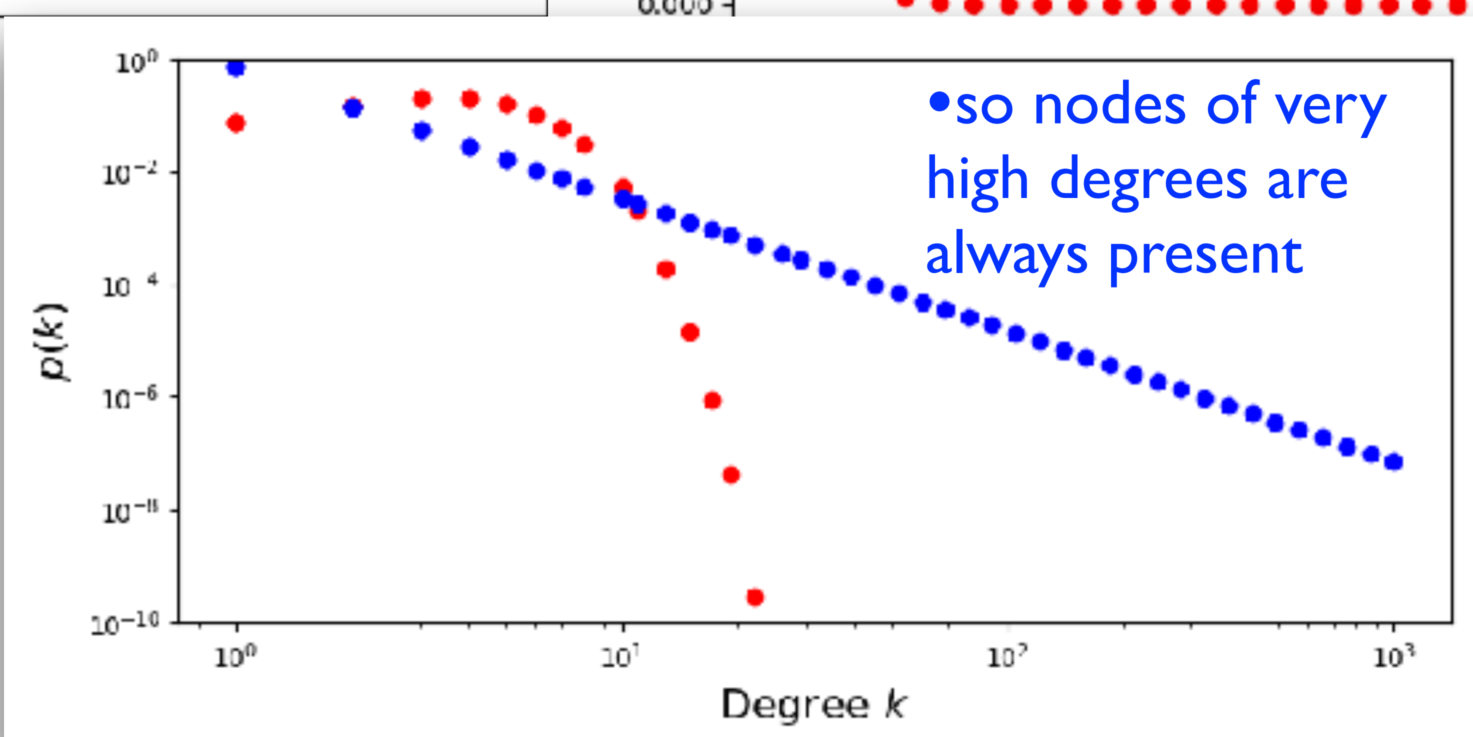


$$\boxed{P(k) \propto k^{-\gamma}}$$

$$\boxed{P(k) \sim e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}}$$

$\langle k \rangle = \langle k \rangle = 4$





Scale-Free Networks

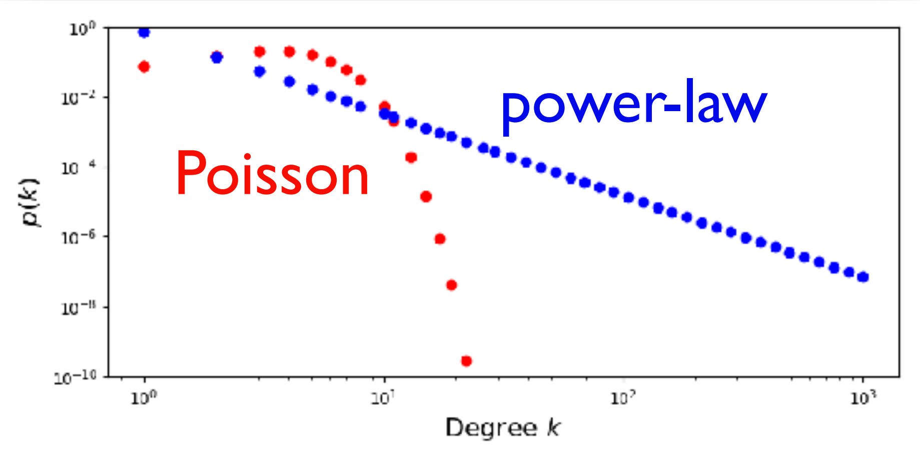
- Networks with power-law degree distributions are called **scale-free networks**
- This is because there is no characteristic scale in the distribution
- If degrees are rescaled, the form of the distribution does not change:

$$P(\alpha k) \propto (\alpha k)^{-\gamma} = \alpha^{-\gamma} P(k)$$

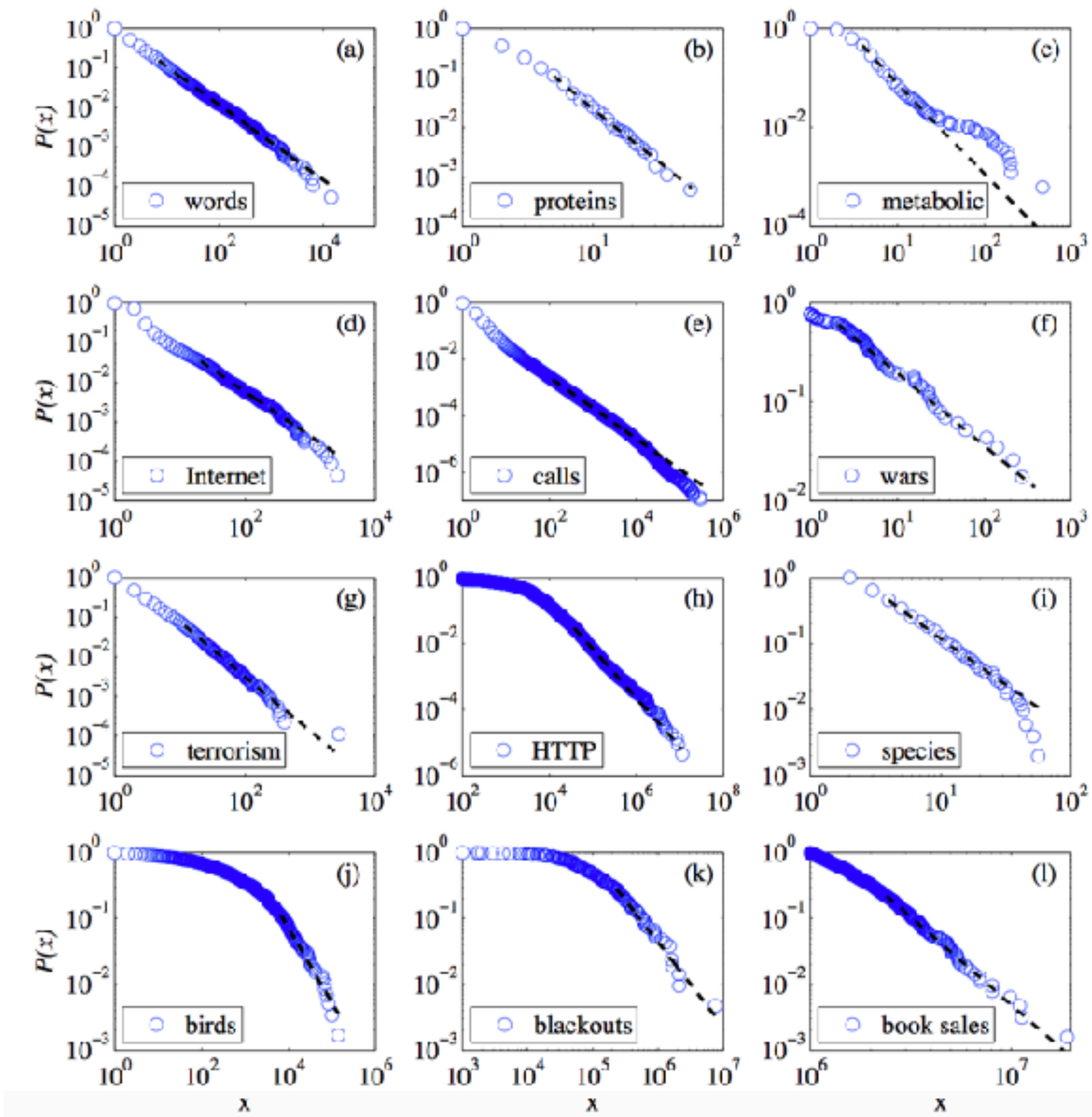
$$\Rightarrow \quad \frac{P(k=20)}{P(k=2)} = \frac{P(k=200)}{P(k=20)} = \frac{P\left(k=2\times 10^6\right)}{P\left(k=2\times 10^5\right)} = \ldots$$

- For comparison, the Poisson distribution behaves like this:

$$\begin{aligned} \frac{P(\alpha k)}{P(k)} &= \frac{\langle k\rangle^{\alpha k}\, k!}{(\alpha k)!\,\langle k\rangle^{k}} \\ &= \langle k\rangle^{(\alpha-1)} \frac{k!}{(\alpha k)!} \\ &\sim 0, \text{when } \alpha > 1, k \gg \langle k\rangle \end{aligned}$$

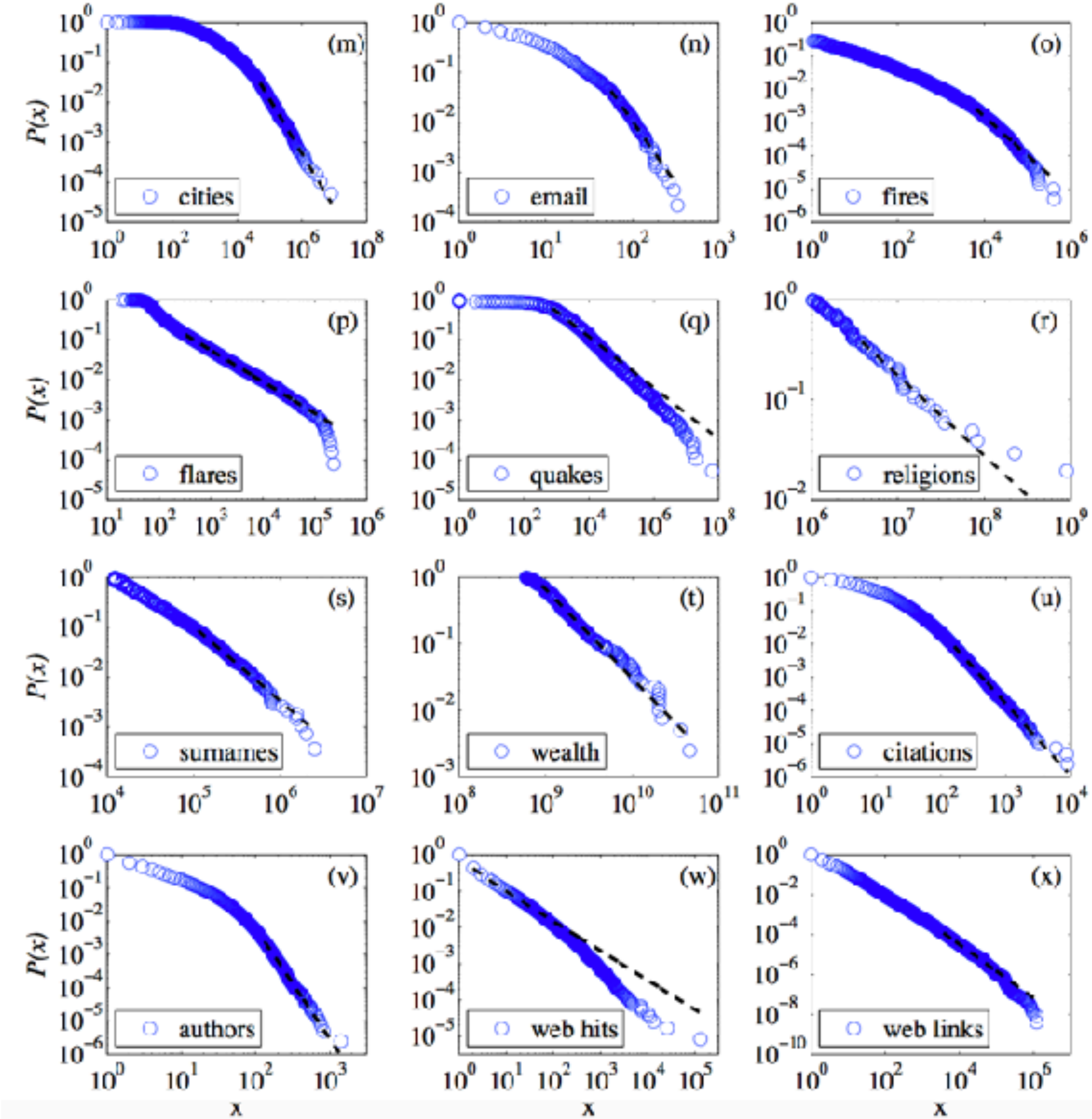


Power laws are common



See: A. Clauset et. al. "POWER-LAW DISTRIBUTIONS IN EMPIRICAL DATA", *SIAM Rev.*, 51(4), 661–703 (2009)

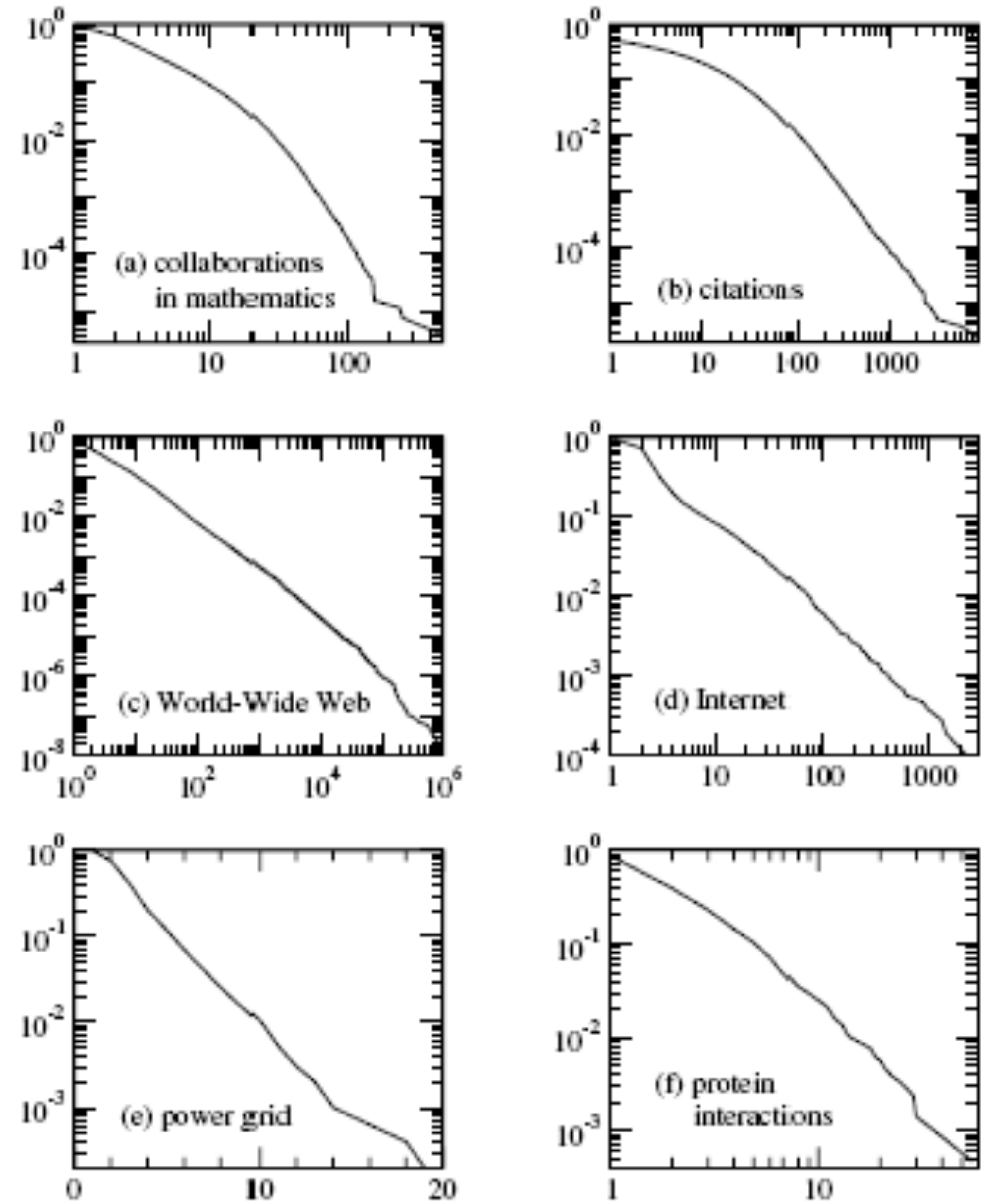
Power laws are common



See: A. Clauset et. al. "POWER-LAW DISTRIBUTIONS IN EMPIRICAL DATA", *SIAM Rev.*, 51(4), 661–703 (2009)

The cumulative degree distribution is also a power law

$$P\left(k > k'\right) \sim \sum_{k'=k}^{\infty} k'^{-\gamma} \sim k'^{-(\gamma-1)}$$



Moments of the power-law distribution

- The mth moment is

$$\langle k^m \rangle \propto \sum_{k=k_0}^{\infty} k^m k^{-\gamma}$$
$$\sim \int_{k_0}^{\infty} k^{-\gamma+m} dk$$
$$= \begin{cases} \infty, & \text{if } \gamma \leq m+1 \\ \text{const.}, & \text{if } \gamma > m+1 \end{cases}$$

γ	$\langle k \rangle$	$\langle k^2 \rangle$	$\langle k^3 \rangle$
$(1, 2]$	∞	∞	∞
$(2, 3]$	const.	∞	∞
$(3, 4]$	const.	const.	∞
$(4, 5]$	const.	const.	const.
$\ldots$			

- Many real world networks have $\gamma \approx$ 2–3 $\Rightarrow \langle k \rangle < \infty$ and var $= \infty$.
- However, given a network we can of course always calculate $\langle k^m \rangle = \frac{1}{n} \sum_{i=1}^{n} k_i^m$

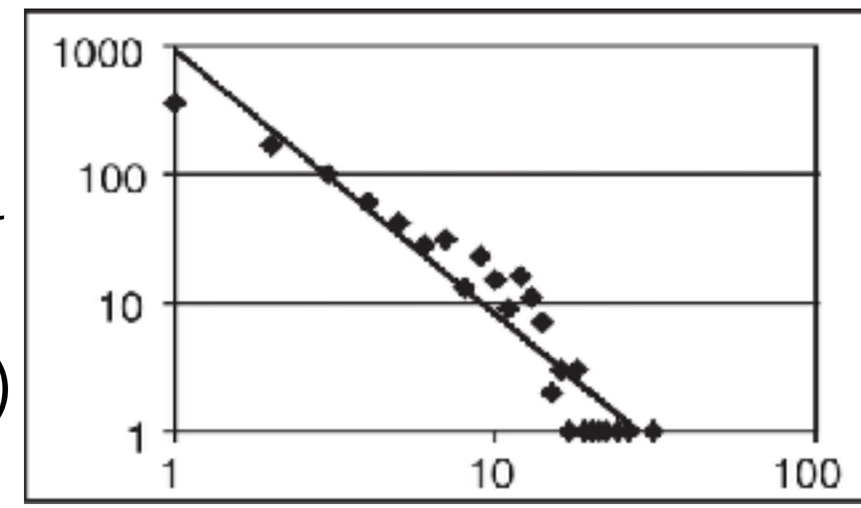
(but if the moments of the "underlying" network are infinite, they will change if we get a bigger sample)

Disclaimer: reality check

Note! there are other distributions that look like power laws (e.g. stretched exponential, lognormal) and those often better describe real data... **not all networks are scale-free, but typically their degree distributions are still broad**

Power-law fitting

Power-laws give straight lines in log-log plots → Draw points to log-log plot, fit a line (e.g. with least squares fit)



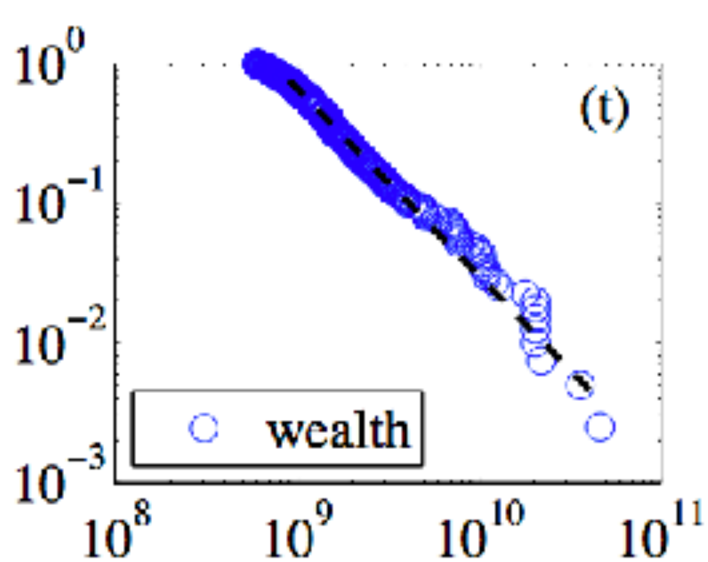
Tong et al. Science (2004)

Find maximum likelihood estimate for parameters of the PL distribution → Goodness-of-fit tests (p-value for data being produced with PL) → Test for alternative hypothesis (log-normal etc.)

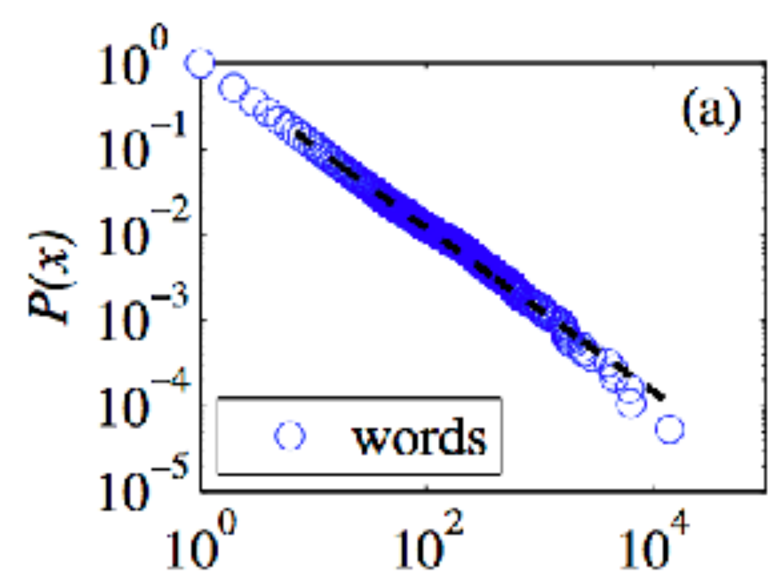
See: A. Clauset et. al. “POWER-LAW DISTRIBUTIONS IN EMPIRICAL DATA”, *SIAM Rev.*, 51(4), 661–703 (2009)

Why are there power laws?

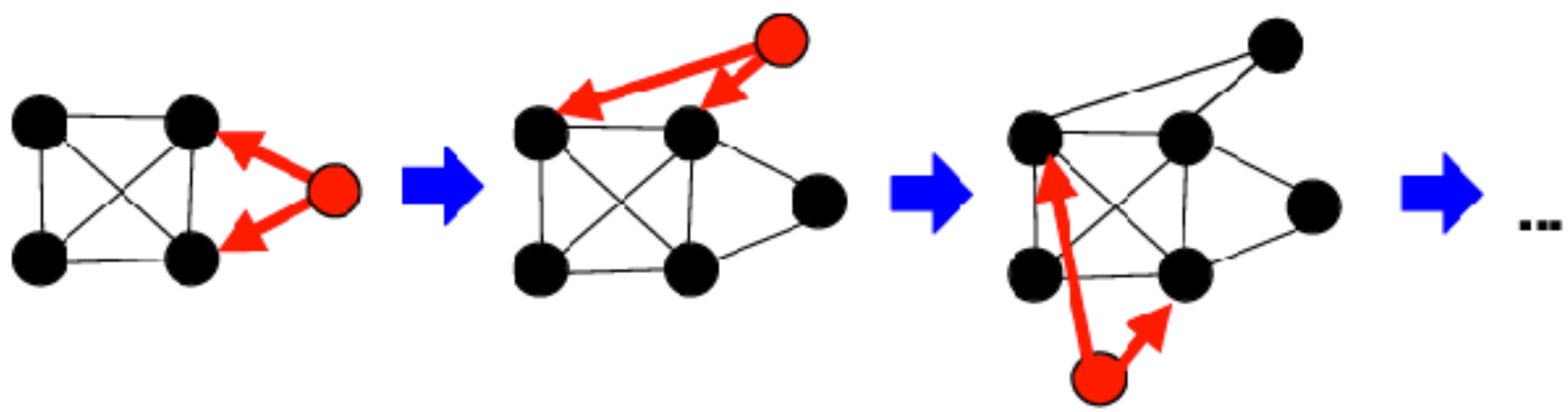
- Distribution of wealth
  - Pareto distribution, 80/20
  - "Rich get richer"
- Word appearance in texts
  - Zipf's law
- …



- Any dynamic system where elements get "larger" with probability that is proportional to their current "size" (Simon 1955)
- Networks: preferential attachment



The Barabási-Albert scale-free model

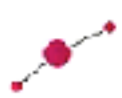


Model definition

- Take a small seed network (a few connected nodes)
- Repeat until you have N nodes:
  1. Add a new node with m stubs (unconnected links)
  2. Connect each stub to an existing node i, chosen with probability $p_i = k_i / \sum k_i$.

step 2. is called
preferential attachment

Animation: B-A network growth (m=1)



Properties of the BA model

- Results in a power-law degree distribution:

$$P(k) = \frac{2m^2}{k^3}$$

- Average degree $\langle k \rangle \approx 2m$ (m new edges added with each node).
- Ultra-small world:

$$\ell \propto \frac{\ln N}{\ln(\ln N)}$$

- Clustering coefficient is

$$C(k, N) \propto \frac{(\ln N)^2}{N}$$

  - No k-dependence!
  - Way too small for large k
  - $C(k, N) \to 0$ when $N \to \infty$

Real-wold networks: between randomness and order

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

* (depending on layout & clustering measure)

Q: which mechanism leads to preferential attachment in networks?

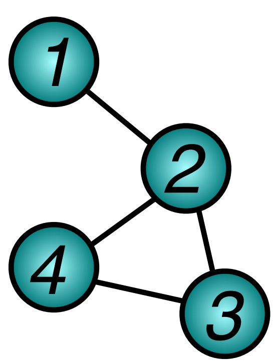
Q: which mechanism leads to preferential attachment in networks?

A: any process where links are followed

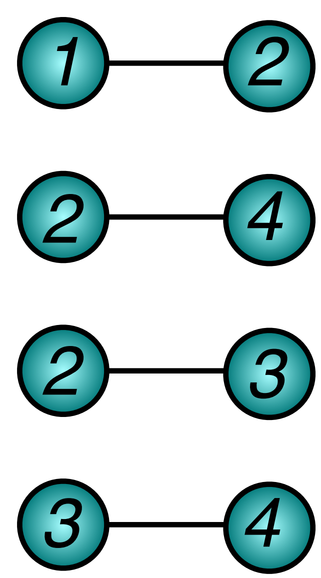
Following a link leads to high degree nodes

- *Follow a random link, what is the probability that node i with degree k_i is picked?*
- There are in total $2m=\sum k_i$ endpoints for links, a node with degree k_i has k_i endpoints leading to it

Example network:



Each link picked with probability 1/4



After that, each node is picked with probability 1/2

$$p_1 = \frac{1}{4} \cdot \frac{1}{2} = \frac{1}{8}$$

$$p_2 = 3 \cdot \frac{1}{4} \cdot \frac{1}{2} = \frac{3}{8}$$

$$p_3 = 2 \cdot \frac{1}{4} \cdot \frac{1}{2} = \frac{2}{8}$$

$$p_4 = 2 \cdot \frac{1}{4} \cdot \frac{1}{2} = \frac{2}{8}$$

$$p(\text{'follow random link, reach node i'}) = \frac{k_i}{\sum_j k_j}$$

Following a link leads to high degree nodes

- *Follow a random link, what is the expected degree of the node at the end of the link?*
- Nodes with degree k have k opportunities to be picked: $p(k_{nn} = k) \propto kp(k)$
- Expected value depends on the 2nd moment:

$$\langle k_{nn} \rangle = \sum_k kp(k_{nn} = k) = \sum_k k \frac{kp(k)}{\langle k \rangle} = \frac{\langle k^2 \rangle}{\langle k \rangle}$$

Friendship paradox

- On average, *a person has less friends than a friend has friends*
  - Direct consequence of following a link (and variation in degree distributions in social networks)

$$\langle k_{nn} \rangle = \frac{\langle k^2 \rangle}{\langle k \rangle}$$

- Applies to any type of network:

Network	$\langle k \rangle$	$\langle k_{nn} \rangle$	$p(k_{nn} > k)$
Short messages	2.2	146	0.62
Airports & flights	11	65	0.93
Protein interaction	3.0	20	0.85
Internet AS	13	1445	0.96

Note that expected neighbour degree $\langle k_{nn} \rangle$ is different than expected average neighbour degree

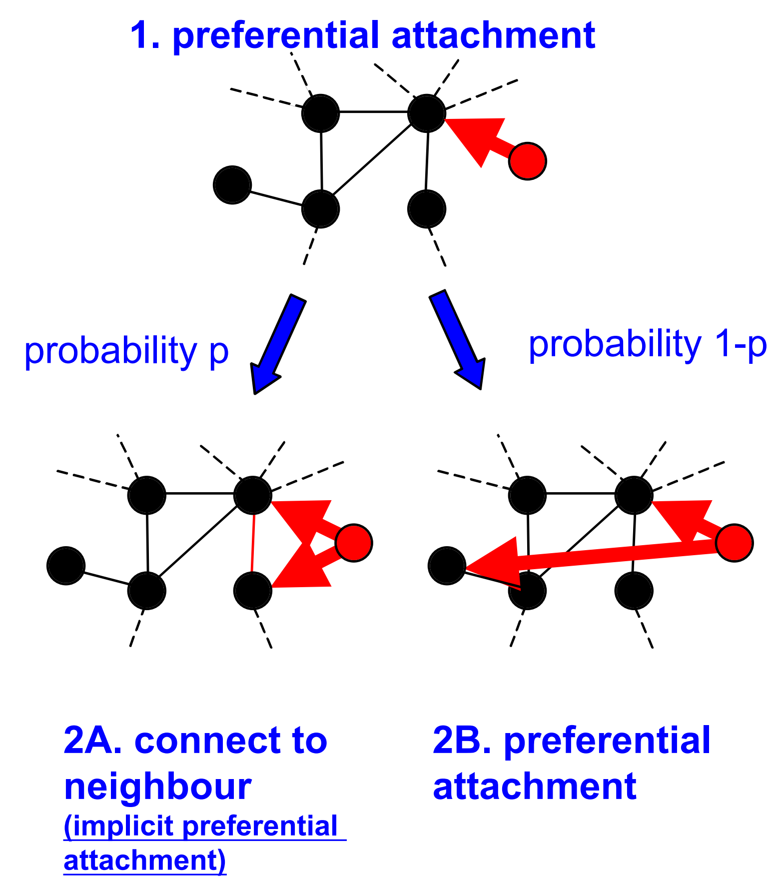
Following link as implicit preferential attachment

- Links are often followed in networks even when it is not explicit
  - Triadic closure (social networks)
  - Vertex copying (protein interaction networks, citation networks)
  - Random walks (WWW)
- Models of network growth with these processes as part of them lead to preferential attachment and power-laws

Other scale-free network models: Holme-Kim

- **the Holme-Kim Model**
  - **motivation: to get realistic clustering**

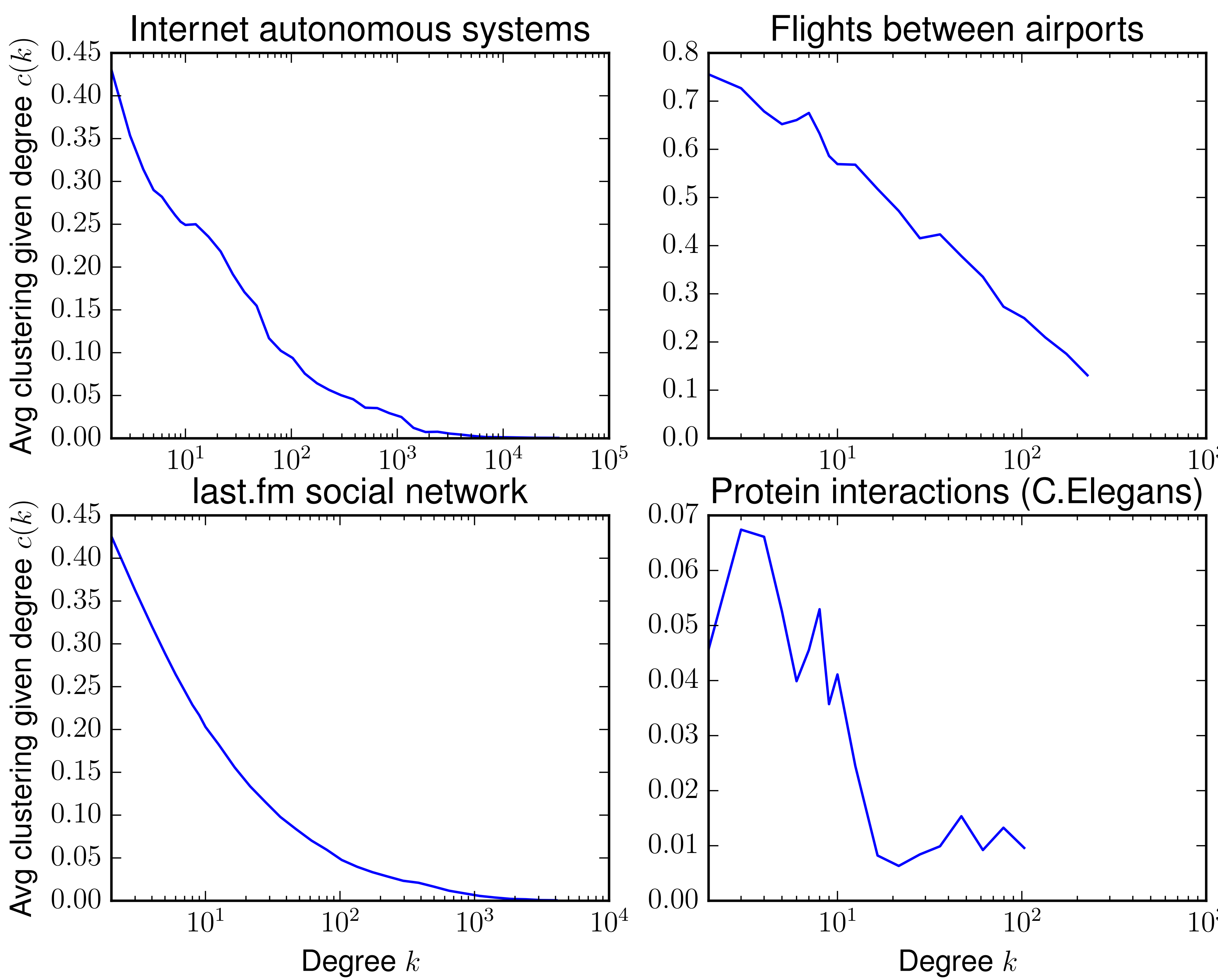
1. Take a small seed network
2. Create a new vertex with m edges
3. Connect the first of the m edges to existing vertices with a probability proportional to their degree k (just like BA)
4. With probability p, connect the next edge to a random neighbour of the vertex of step 3., otherwise do 3. again
5. Repeat 2.-4. until the network has grown to desired size of N vertices



$$C(k) \propto \frac{1}{k}$$

for large N, ie clustering more realistic! This type of clustering is found in many real-world networks.

Clustering coefficient as a function of degree



Real-wold networks: between randomness and order

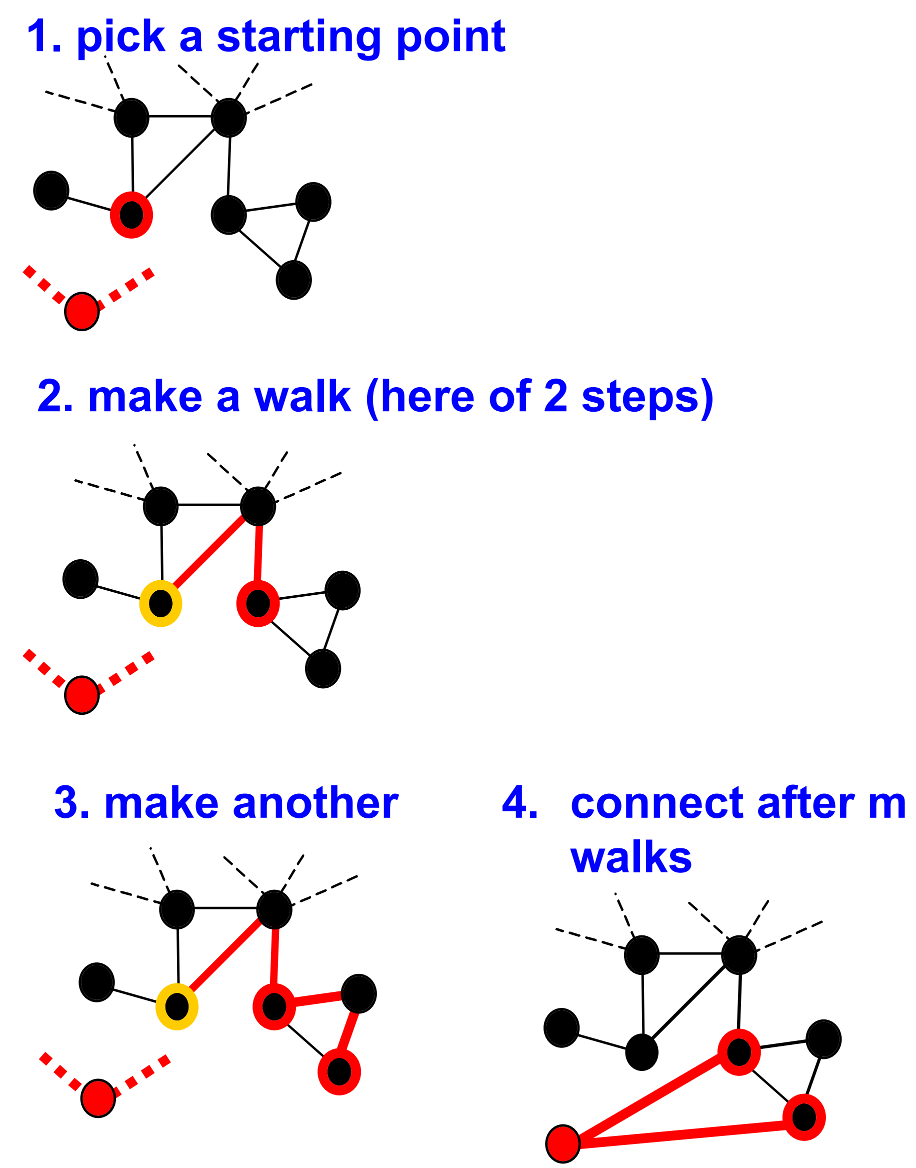
Network	Degrees	Paths	Clustering
Real-world	Fat-tailed	short	high
Erdős-Renyi	Poissonian	short	low
Regular lattices	Fixed	long	high*
Configuration	Free to choose	short	low
Watts-Strogatz	Fixed to Poisson	short	high
Barabási-Albert	Power-law	short	low
Holme-Kim	Power-law	short	high

* (depending on layout & clustering measure)

Other scale-free network models: random walks

- **Random walks**
  - **e.g. people learn to know people through other people, which leads to popular people without looking for them**

1. Take a small seed network
2. Create a new vertex with m edges
3. Pick a random vertex
4. Make a l-step random walk starting from this vertex
5. Connect one of the edges of the new vertex to wherever you are
6. Repeat 3.-5. or 4.-5. m times
7. Repeat 2.-6. until N vertices

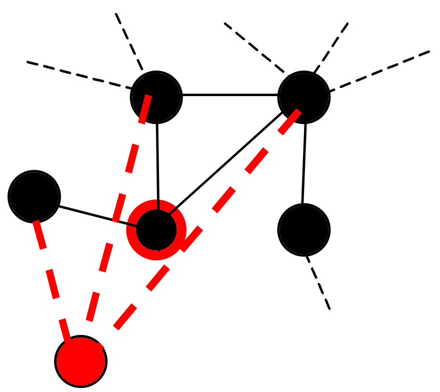


Other scale-free network models: vertex copying

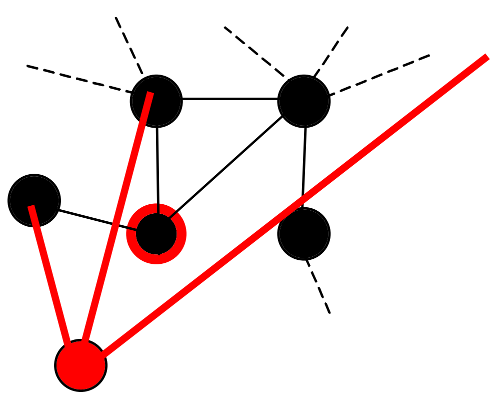
- **the vertex copying model**
  - **motivation: citations or WWW link lists are often copied**
  - **a "local" explanation to preferential attachment**
  - **asymptotically SF with $\gamma \geq 3$**

1. Take a small seed network
2. Pick a random vertex
3. Make a copy of it
4. With probability p, move each edge of the copy to point to a random vertex
5. Repeat 2.-4. until the network has grown to desired size of N vertices

1. copy a vertex



2. rewire edges with *p*



Summary

- Real-wold networks have fat-tailed degree distributions
- Following a link -> preferential attachment
  - Many realistic processes have this component
- Any model of network growth where one follows a link leads to scale-free networks
- Following a link twice from a node, or following link twice can be used to create clustering

Degree distributions matter

- Scale-free networks are resistant to random failures
  - but vulnerable to attacks
  - and good at spreading epidemics
- More on this next week…