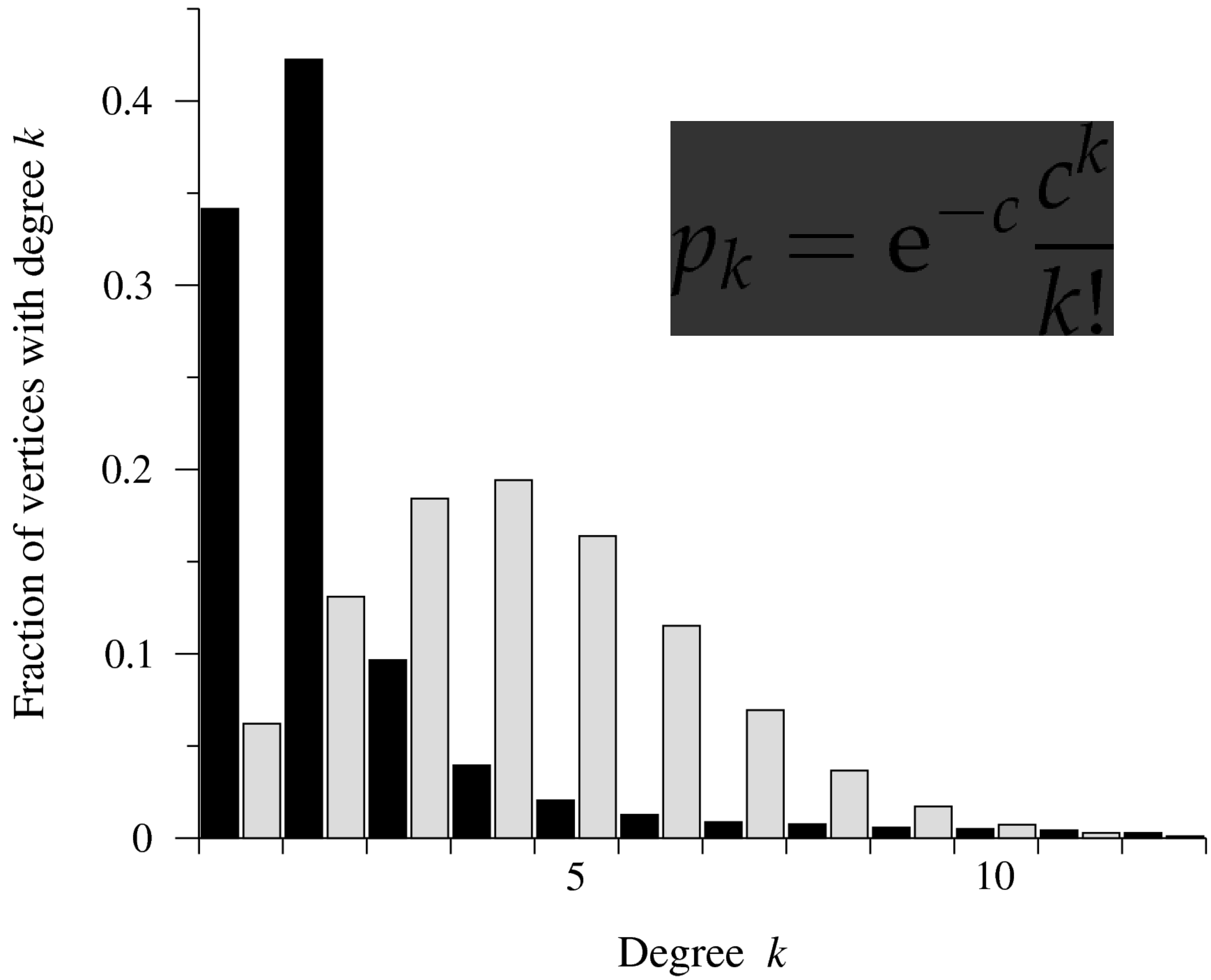
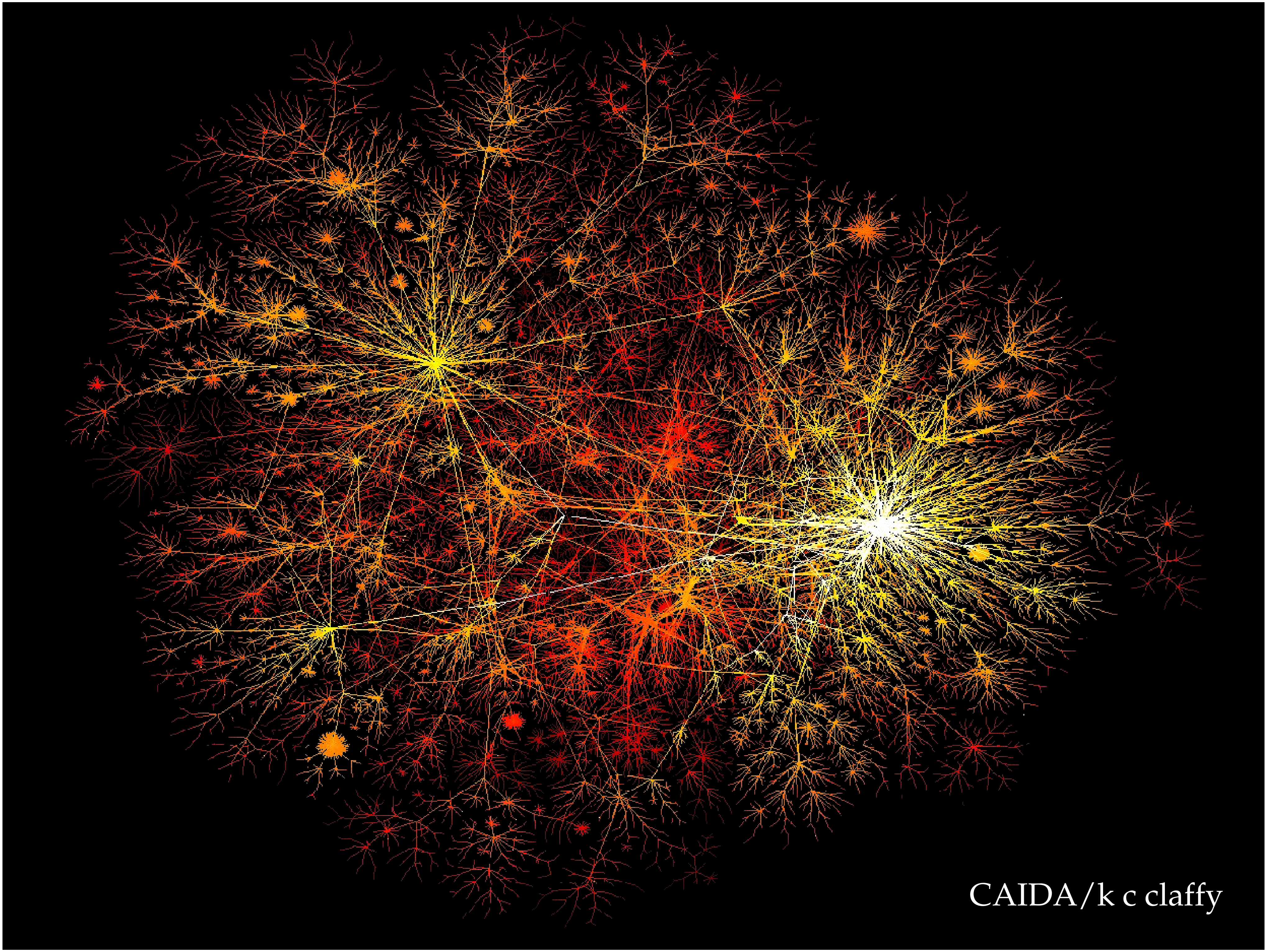
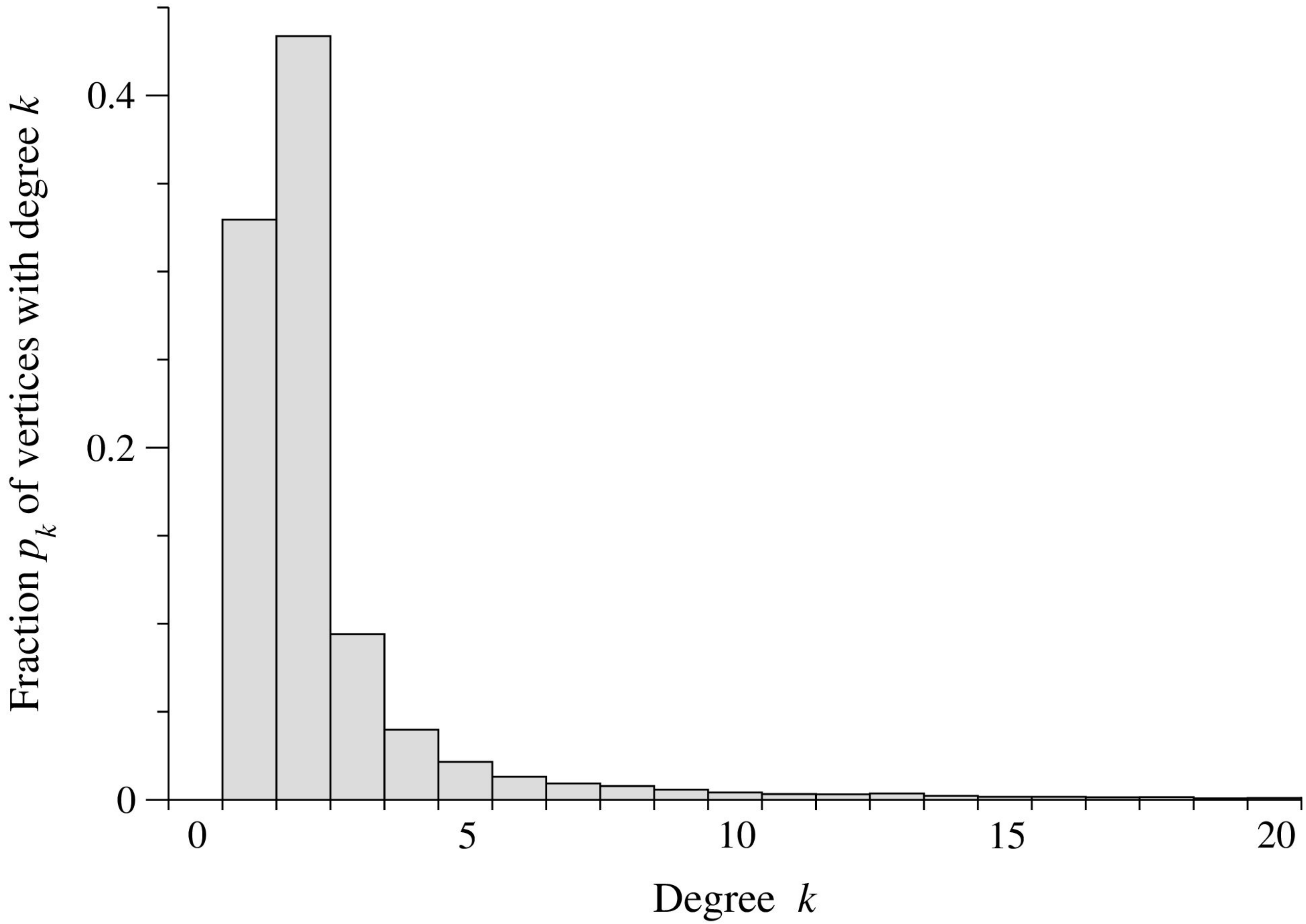
Degree distribution



$$p_k = \mathrm{e}^{-c}\frac{c^k}{k!}$$

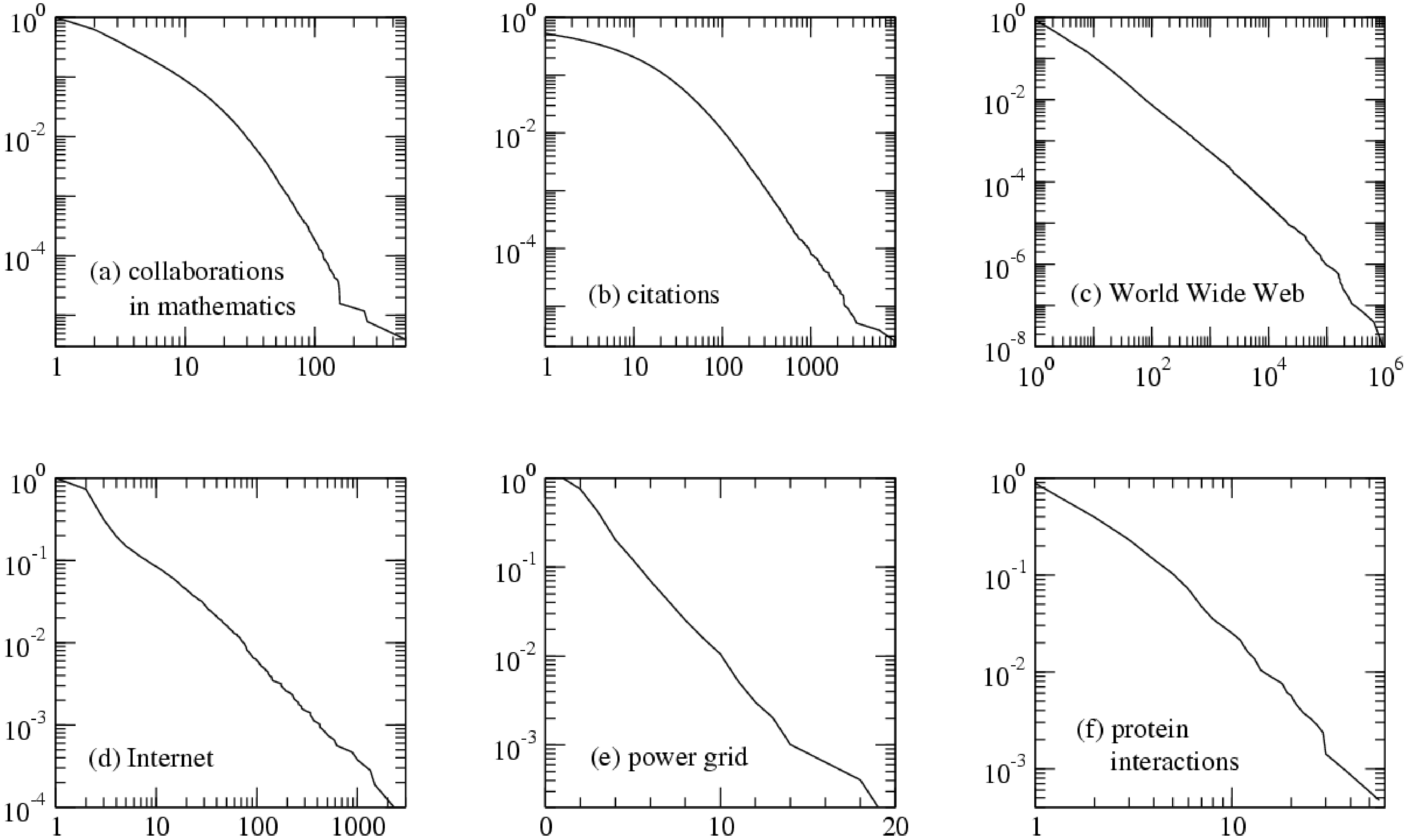
- Many other properties also don't match, including degree correlations and clustering coefficient


CAIDA/k c claffy



Fraction p_k of vertices with degree k
0.4
0.2
0
0
5
10
15
20
Degree k

Degree distributions



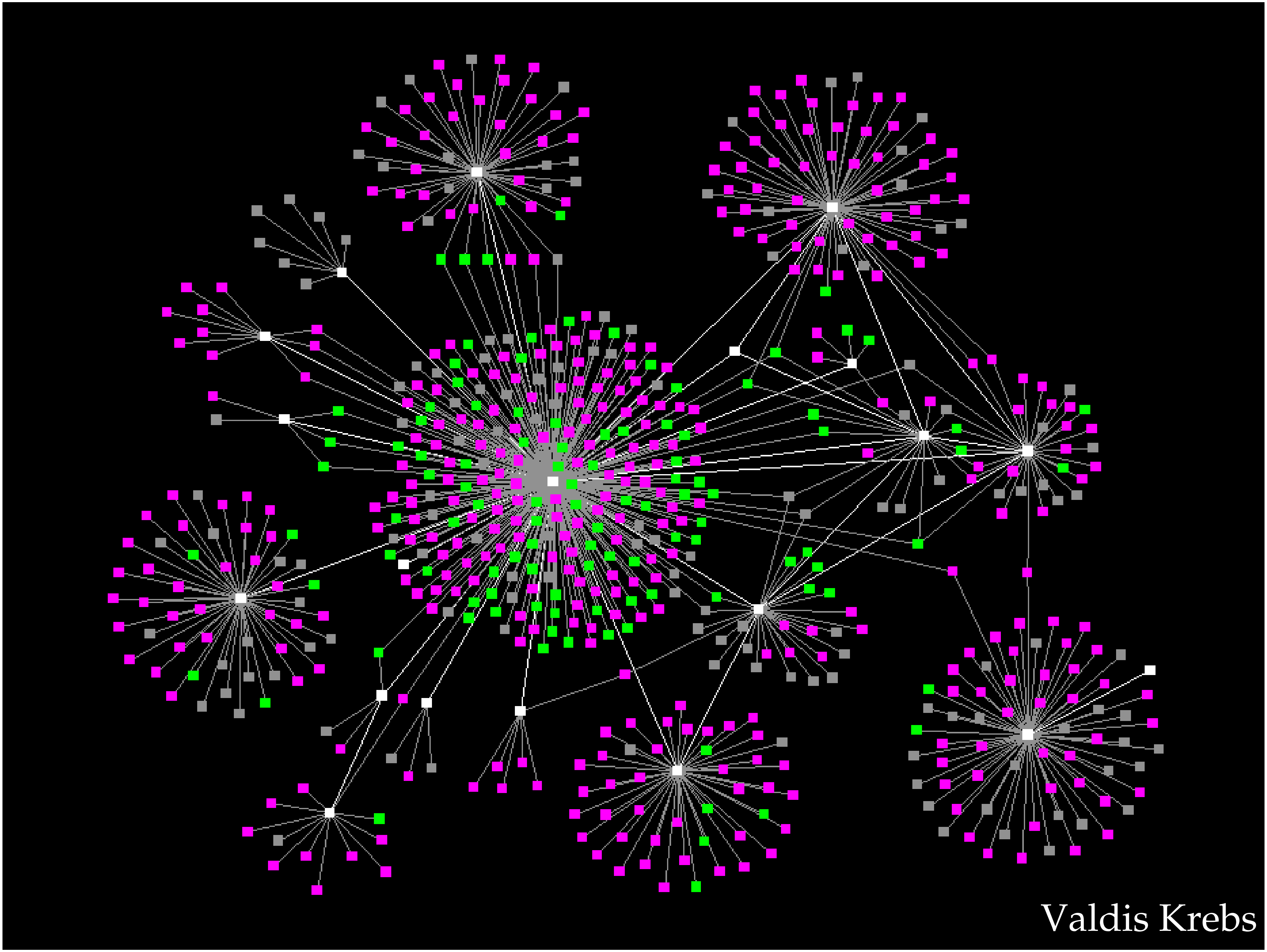
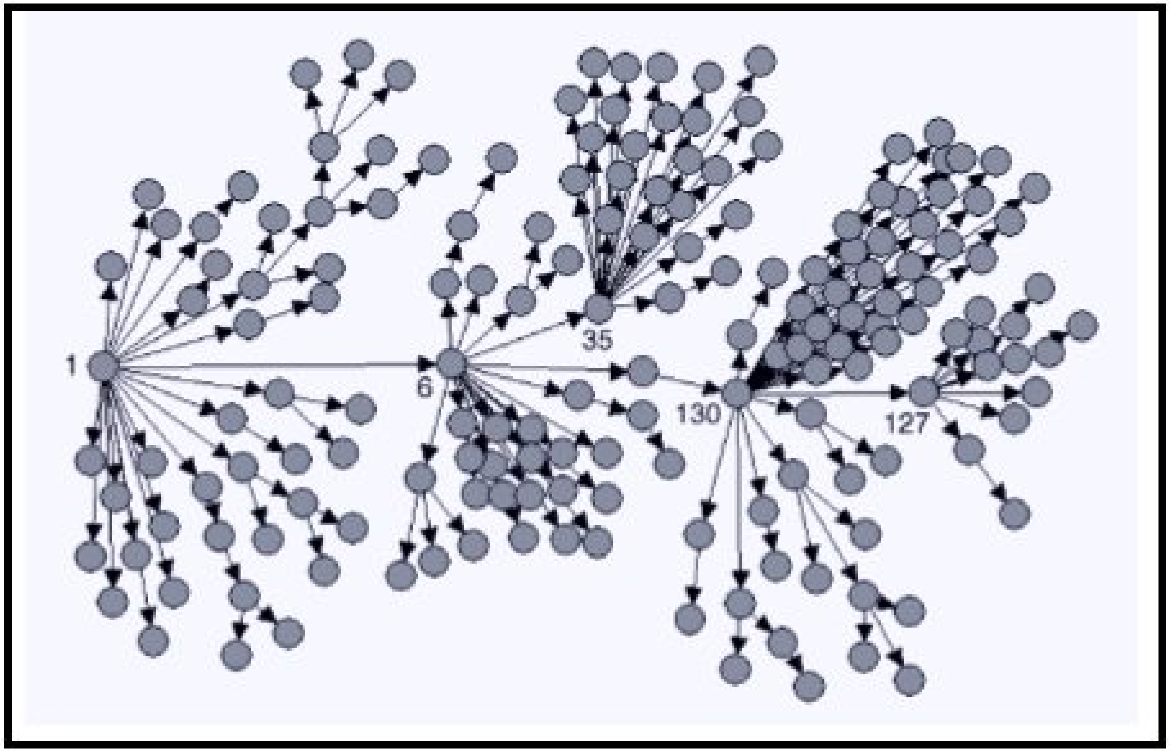

Valdis Krebs

FIGURE 2. Probable cases of severe acute respiratory syndrome, by reported source of infection* — Singapore, February 25–April 30, 2003

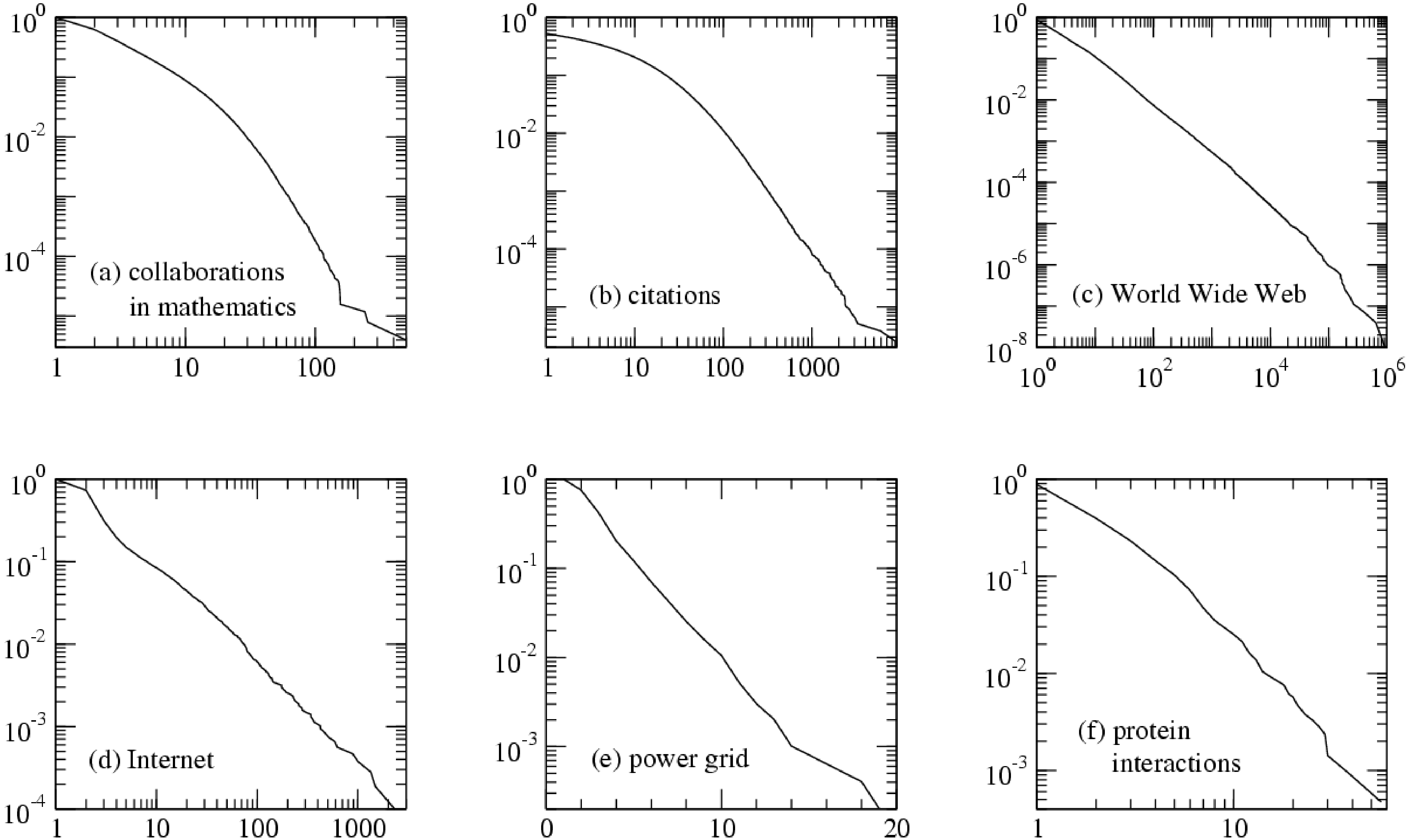


* Patient 1 represents Case 1; Patient 6, Case 2; Patient 35, Case 3; Patient 130, Case 4; and Patient 127, Case 5. Excludes 22 cases with either no or poorly defined direct contacts or who were cases translocated to Singapore and the seven contacts of one of these cases.

Reference: Bogatti SP. Netdraw 1.0 Network Visualization Software. Harvard, Massachusetts: Analytic Technologies, 2002.

Steve Borgatti

Degree distributions



Power laws

- Let x be the quantity we're interested in
- Let $p(x)\,\mathrm{d}x$ be the probability that it lies between x and $x + \mathrm{d}x$
- A straight line on the log-log plot implies that

$$\ln p(x) = -\alpha \ln x + c$$

- Take the exponential of both sides:

$$p(x) = C\,x^{-\alpha}$$

- Normally $2 < \alpha < 3$ and always greater than 1

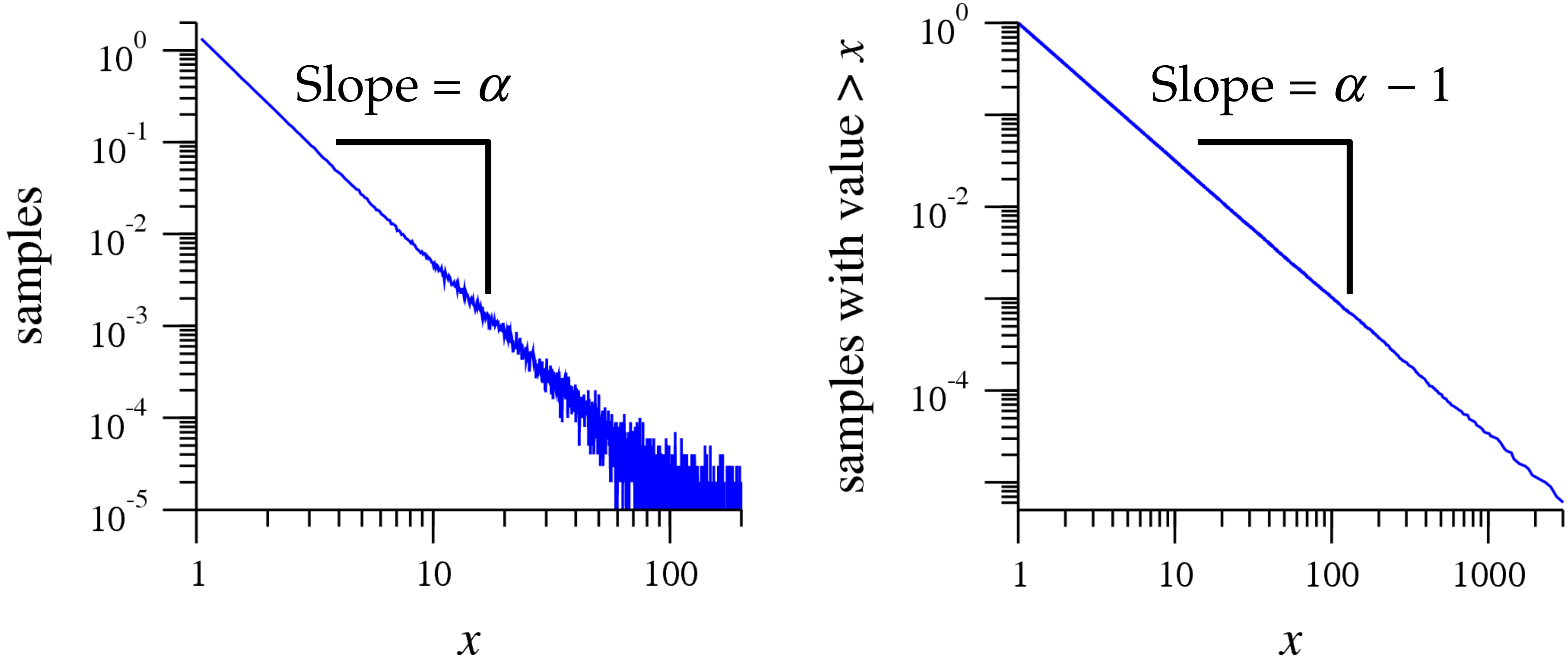
Cumulative distribution

- We define the *cumulative distribution function* $P(x)$ as the probability that the quantity of interest is larger than x:

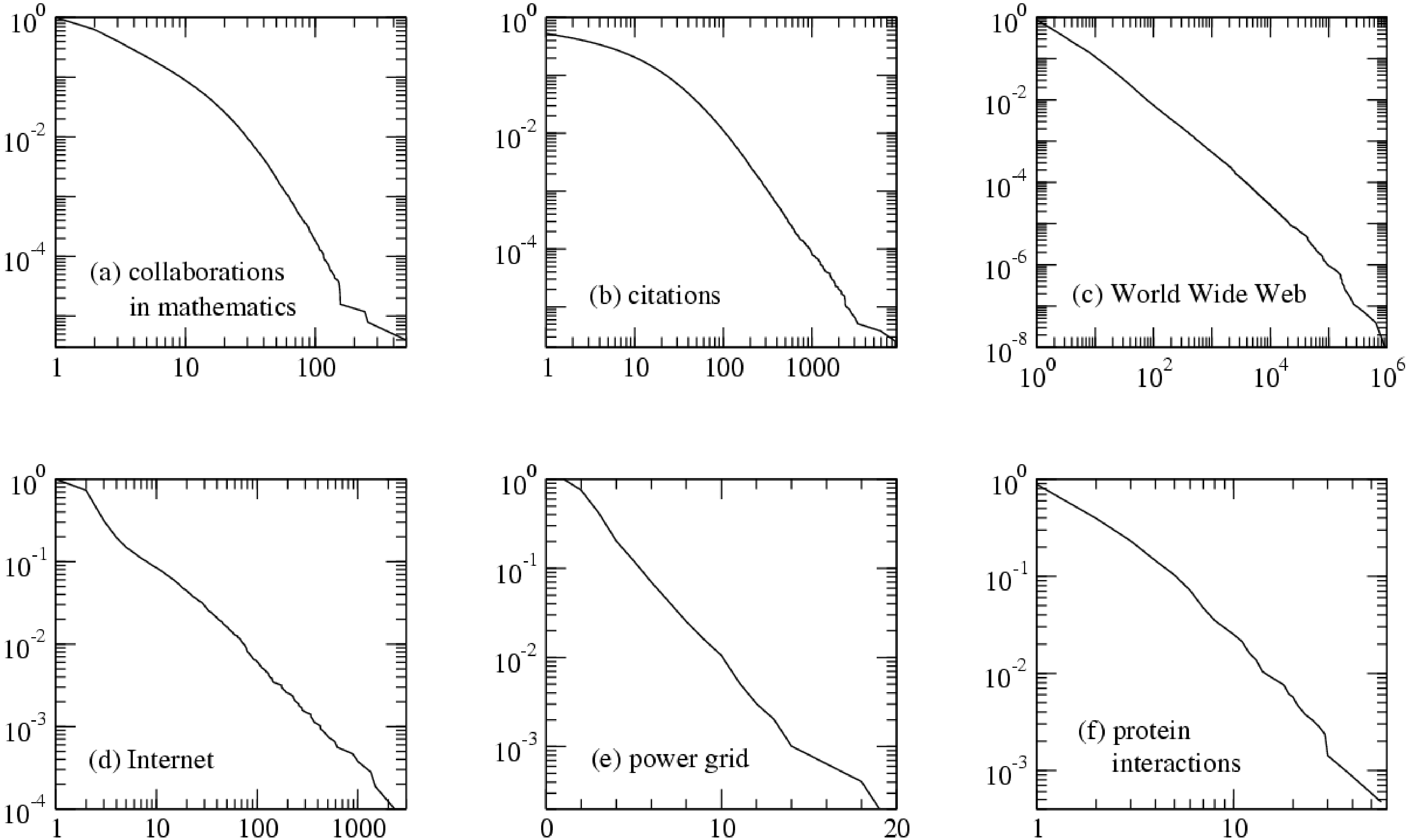
$$
\begin{aligned}
P(x) &= \int_x^\infty p(x)\,\mathrm{d}x = C\int_x^\infty x^{-\alpha}\,\mathrm{d}x \\
&= \frac{C}{\alpha - 1}\, x^{-(\alpha-1)}
\end{aligned}
$$

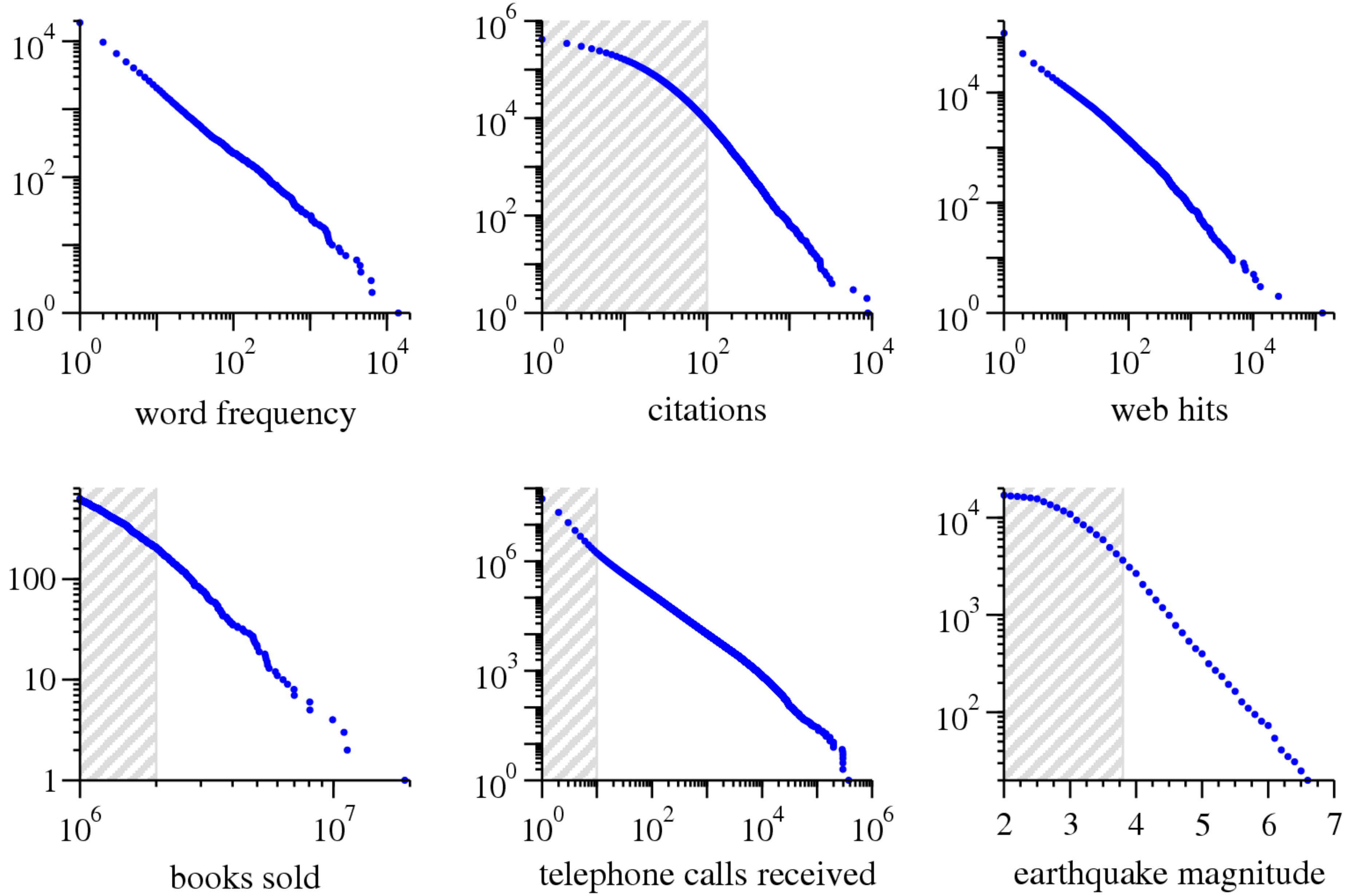
- This is called *Pareto's law* – the cumulative distribution goes as a power law also, but with exponent $\alpha - 1$

- Much better to plot the cumulative distribution $P(x)$ than to plot $p(x)$:

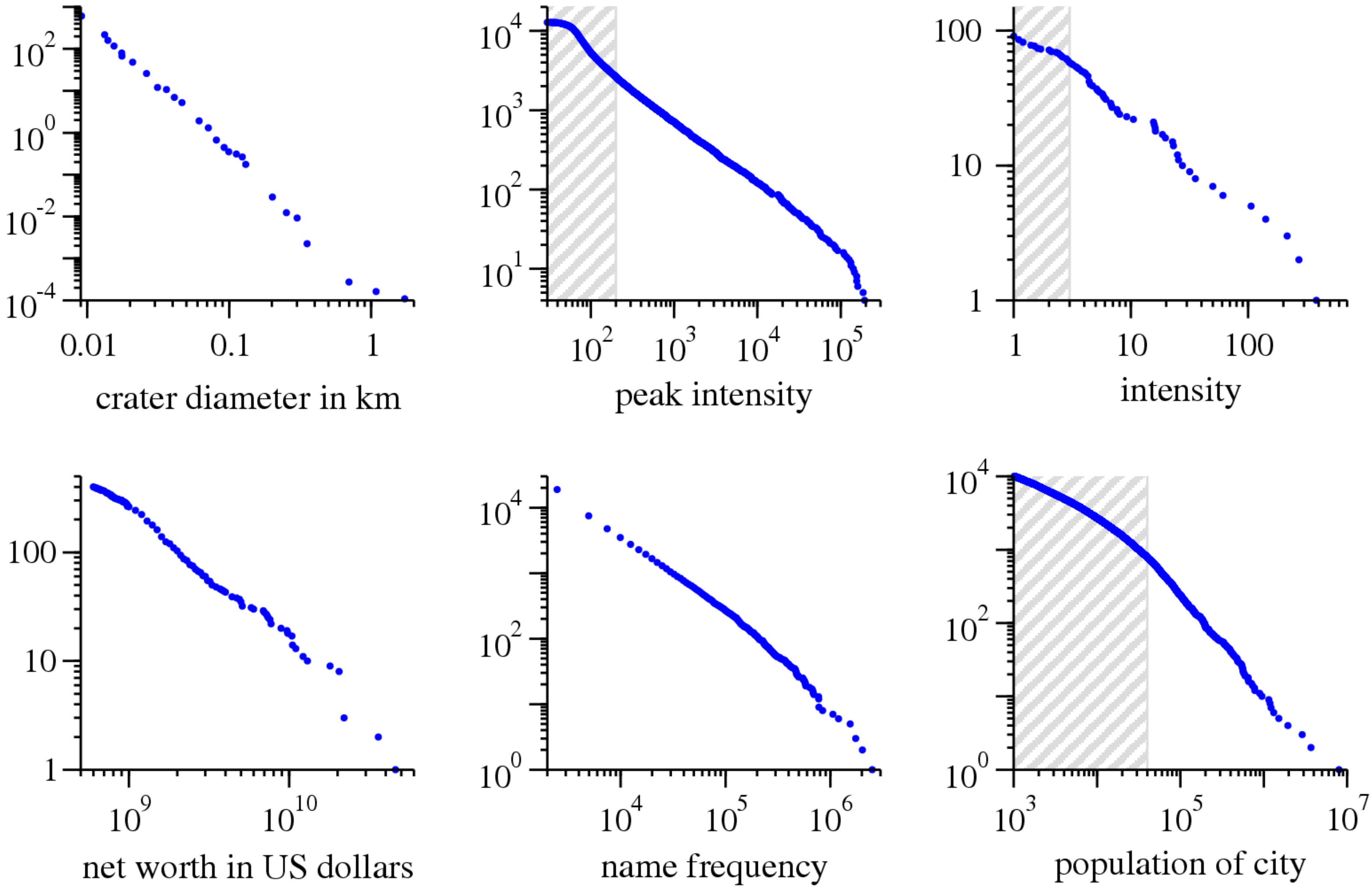


Degree distributions





word frequency
citations
web hits
books sold
telephone calls received
earthquake magnitude

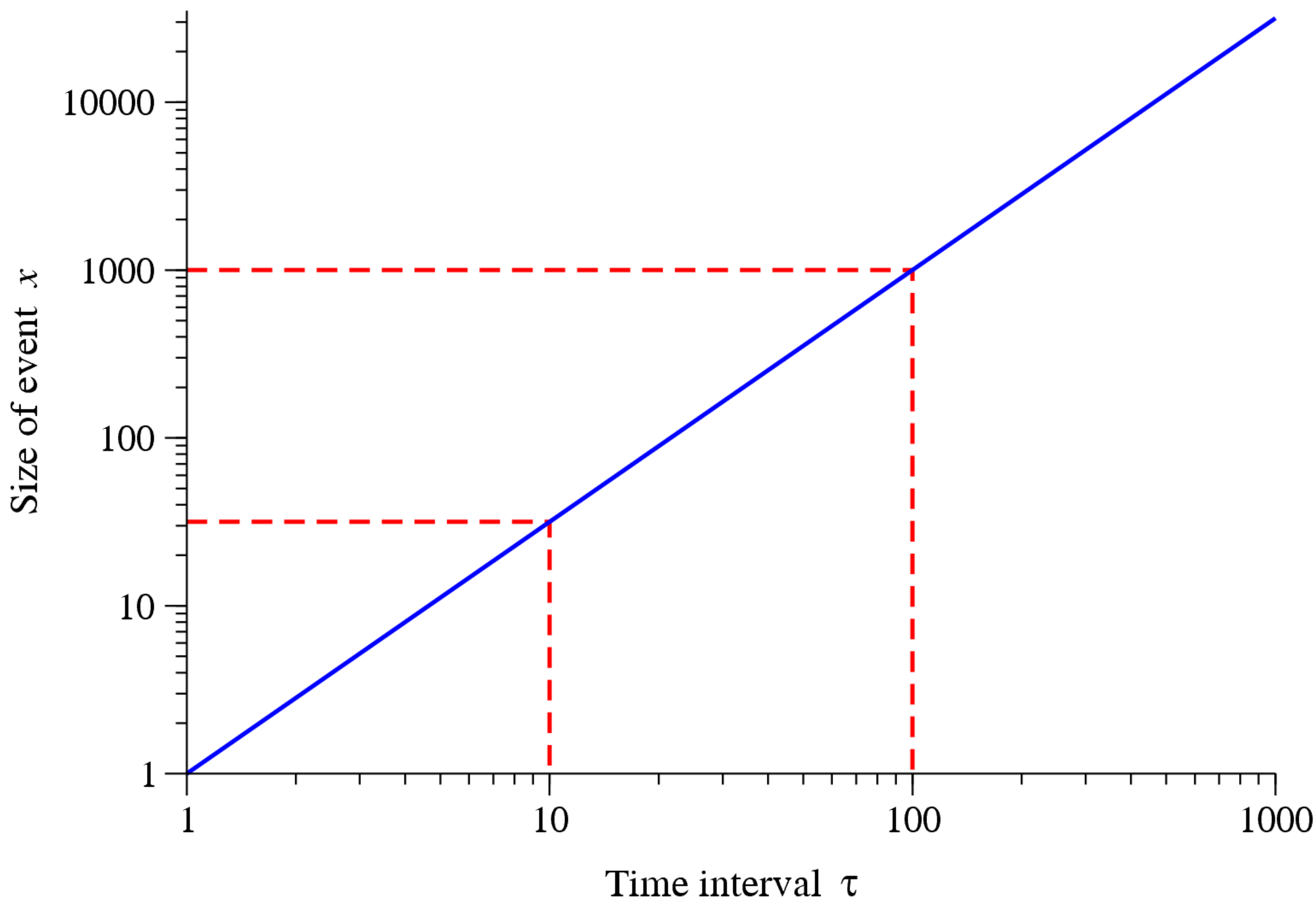


crater diameter in km
peak intensity
intensity
net worth in US dollars
name frequency
population of city

Interesting things about power laws

- The cumulative distribution is interesting in its own right. It tells you how often an event of a given size (or larger) occurs
- Suppose we measure n events in time t
- Then $nP(x)$ are size x or greater and the mean interval between events of size x or greater is

$$\tau = \frac{t}{nP(x)} = \frac{t}{n}\left(\frac{x}{x_{\min}}\right)^{\alpha-1}$$



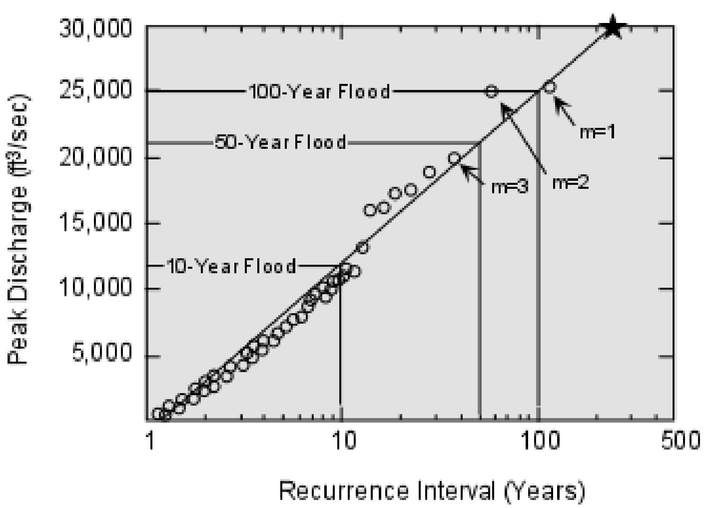
10000
1000
100
10
1
Size of event x
1
10
100
1000
Time interval τ

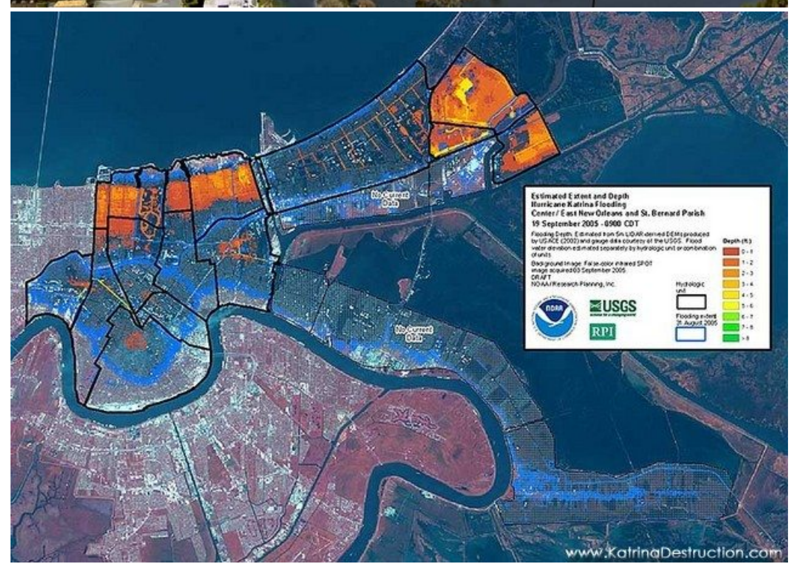
Interesting things about power laws

- The cumulative distribution can tell you:
  - How long will it be until the next large earthquake?
  - How often do floods of a certain size occur?
  - How long will it be before an asteroid wipes out life on Earth
- These questions belong to the field of study called *large deviation theory* or *rare event dynamics*

The "100-Year Flood"

- From the cumulative distribution we can calculate a flood curve of magnitude against average waiting time





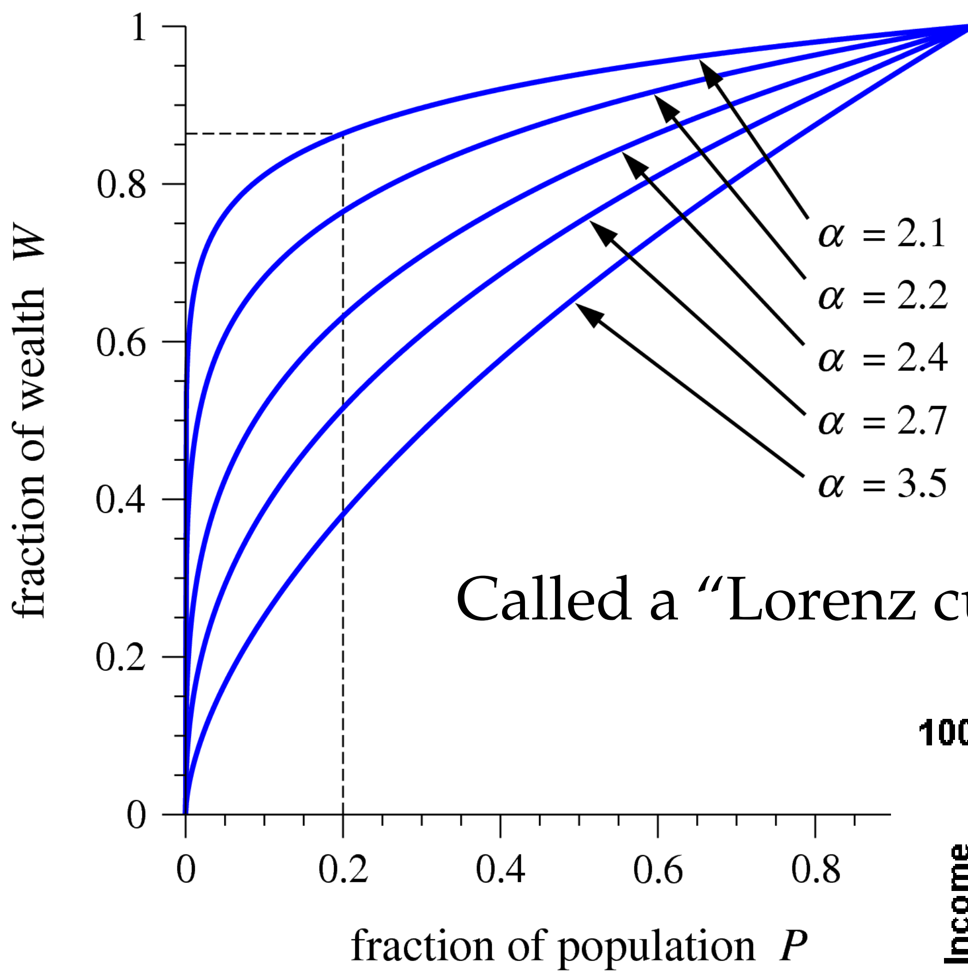
The "80-20 Rule"

- The fraction of the population with wealth over x is $P(x)$, and the fraction of the total wealth in the hands of those people is

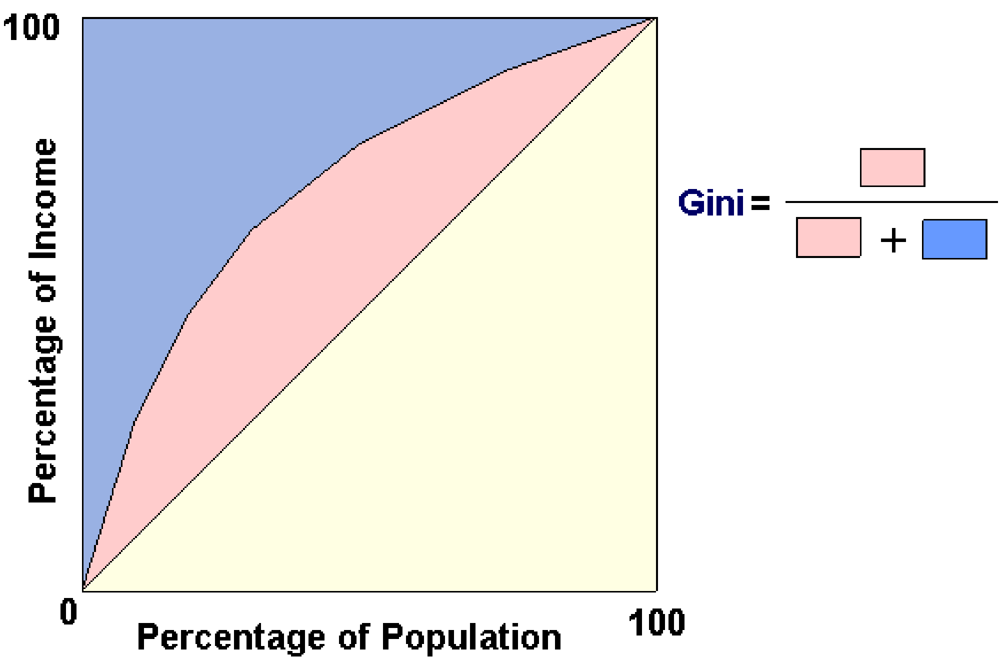
$$W(x) = \frac{\int_x^\infty t p(t)\, dt}{\int_{x_{\min}}^\infty t p(t)\, dt} = \left(\frac{x}{x_{\min}} \right)^{-\alpha+2}$$

- Eliminating $x/x_{\min}$ we get

$$W = P^{(\alpha-2)/(\alpha-1)}$$



Called a "Lorenz curve"



Gini coefficients

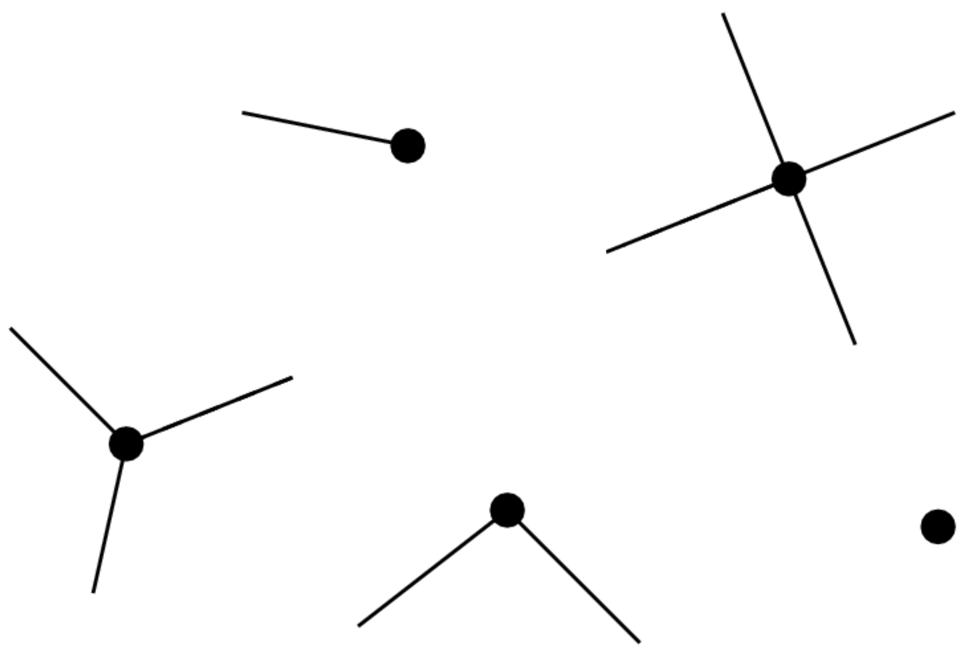
- Denmark 24.7
- Japan 24.9
- Sweden 25
- Czech Republic 25.4
- Norway 25.8
- Slovakia 25.8
- Bosnia 26.2
- Uzbekistan 26.8
- Finland 26.9

- Haiti 59.2
- Bolivia 60.1
- Swaziland 60.9
- Zimbabwe 61
- Central African Republic 61.3
- Sierra Leone 62.9
- Botswana 63
- Lesotho 63.2
- Namibia 74.3

- United States is 74th out of 125
- United Kingdom is 51st
- India and Canada are 27th and 28th respectively
- China is 90th
- Germany is 13th
- France is 30th

The configuration model

- A random graph with given degrees
  - You choose the degrees of all nodes
  - Create nodes with the right numbers of spokes
  - Join them up at random:



Degree of a neighbor

- Let p_k be the fraction of nodes with degree k
- If we follow an edge then we reach nodes of high degree with probability proportional to k
- The distribution is:

$$\frac{kp_k}{\sum_k kp_k} = \frac{kp_k}{\langle k \rangle}$$

- For example, average degree of a neighbor is:

$$\sum_k k \frac{kp_k}{\langle k \rangle} = \frac{\sum_k k^2 p_k}{\langle k \rangle} = \frac{\langle k^2 \rangle}{\langle k \rangle}$$

Degree of a neighbor

- The difference between my neighbor's degree and mine is:

$$\frac{\langle k^2 \rangle}{\langle k \rangle} - \langle k \rangle$$

- So:

$$\frac{\langle k^2 \rangle}{\langle k \rangle} > \langle k \rangle$$

Your friends have more friends than you do

Degree of a neighbor

Network	n	Average degree	Average neighbor degree	$\frac{\langle k^2 \rangle}{\langle k \rangle}$
Biologists	1 520 252	15.5	68.4	130.2
Mathematicians	253 339	3.9	9.5	13.2
Internet	22 963	4.2	224.3	261.5

- Obviously the model is not perfect, but it gives results of the right order of magnitude

Excess degree distribution

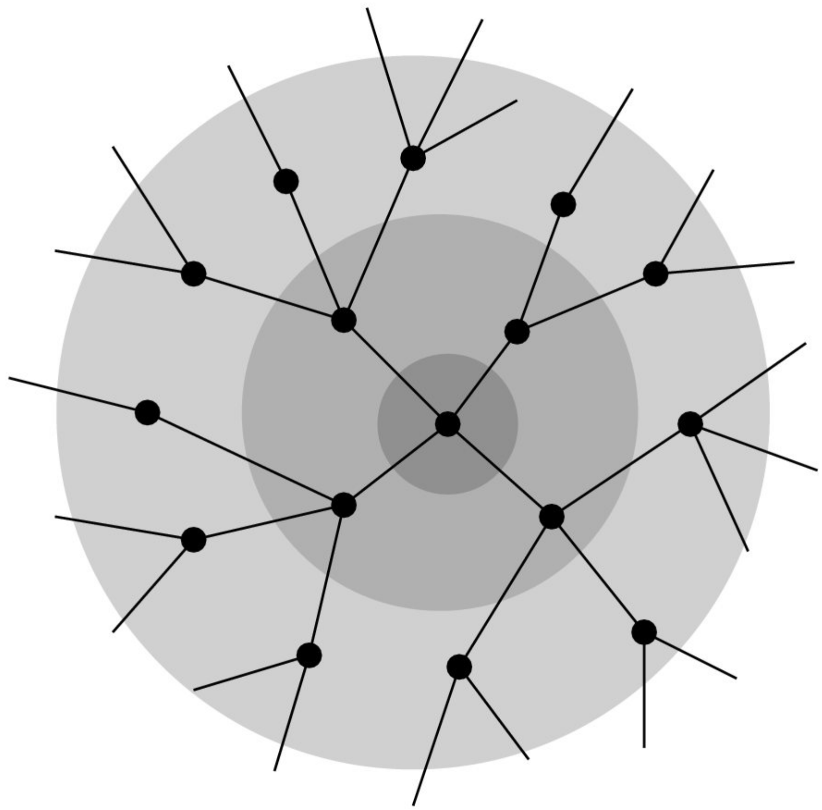
- Usually we are interested not in the number of edges connected to our neighbor, but in the number of edges *other* than the one we arrived along
- If this *excess degree* is k, then the total degree is $k + 1$, and the excess degree thus has distribution

$$q_k = \frac{(k+1)\, p_{k+1}}{\langle k \rangle}$$

- This *excess degree distribution* plays a big role in the theory of networks

Giant component

- For instance you can work out whether there will be a giant component in the network as follows:
  - Start at a single vertex and grow outward:



Giant component

- The expected number of nodes in the ring d steps from the center is equal to the number in the $d-1$ ring, times the average excess degree
- Thus the number reached grows or shrinks exponentially depending on the average excess degree
- Average excess degree:

$$\sum_k k\, q_k = \sum_k \frac{k\,(k+1)\, p_{k+1}}{\langle k \rangle} = \sum_k \frac{k(k-1)\, p_k}{\langle k \rangle} = \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle}$$

Giant component

- Thus we conclude there is a giant component if

$$\frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} > 1$$

- Or equivalently if

$$\langle k^2 \rangle - 2\langle k \rangle > 0.$$

- For instance, in a network with a power-law degree distribution the second moment diverges and hence there is *always* a giant component

Giant component

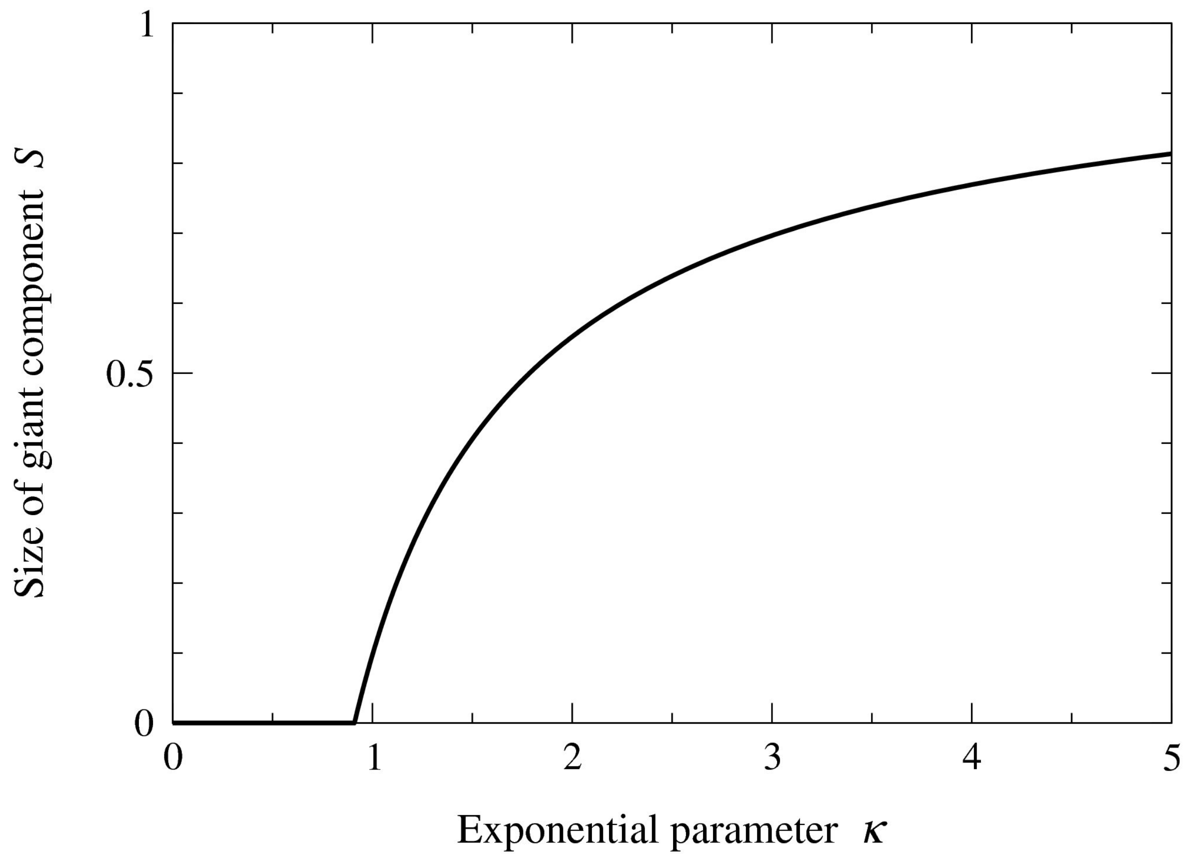
- We can also calculate the size of the giant component
  - Let u be the probability that the vertex at the end of an edge is not in the giant component
  - If excess degree is k then the probability is u^k
  - Averaging over the distribution q_k we get

$$u = \sum_{k=0}^{\infty} q_k u^k = g_1(u)$$

  - Where $g_1(u)$ is the *generating function* for the excess degree distribution

Example

- Exponential distribution: $p_k = Ce^{-k/\kappa}$
- Giant component size: $S = \frac{3}{2} - \sqrt{e^{1/\kappa} - \frac{3}{4}}$



Distribution of components

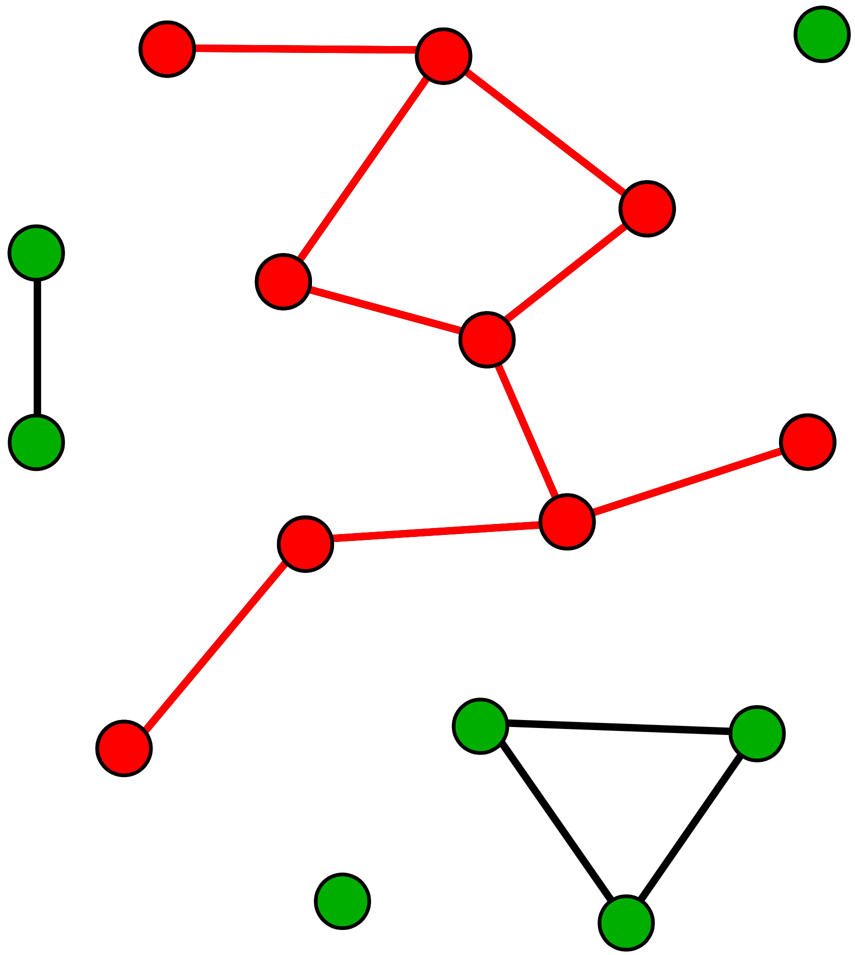
$$h_0(z) = \sum_{s=1}^{\infty} \pi_s z^s = z g_0(h_1(z)), \qquad h_1(z) = z g_1(h_1(z)),$$

$$\begin{aligned}
\pi_s &= \frac{1}{(s-1)!}\left[\frac{\mathrm{d}^{s-1}}{\mathrm{d}z^{s-1}}\left(\frac{h_0(z)}{z}\right)\right]_{z=0} = \frac{1}{(s-1)!}\left[\frac{\mathrm{d}^{s-1}}{\mathrm{d}z^{s-1}} g_0(h_1(z))\right]_{z=0} \\
&= \frac{1}{(s-1)!}\left[\frac{\mathrm{d}^{s-2}}{\mathrm{d}z^{s-2}}\left[g_0'(h_1(z))h_1'(z)\right]\right]_{z=0} = \frac{1}{2\pi\mathrm{i}(s-1)} \oint \frac{g_0'(h_1(z))}{z^{s-1}} \frac{\mathrm{d}h_1}{\mathrm{d}z}\,\mathrm{d}z \\
&= \frac{\langle k\rangle}{2\pi\mathrm{i}(s-1)} \oint \frac{g_1(h_1)}{z^{s-1}}\,\mathrm{d}h_1 = \frac{\langle k\rangle}{2\pi\mathrm{i}(s-1)} \oint \frac{\left[g_1(h_1)\right]^s}{h_1^{s-1}}\,\mathrm{d}h_1
\end{aligned}$$

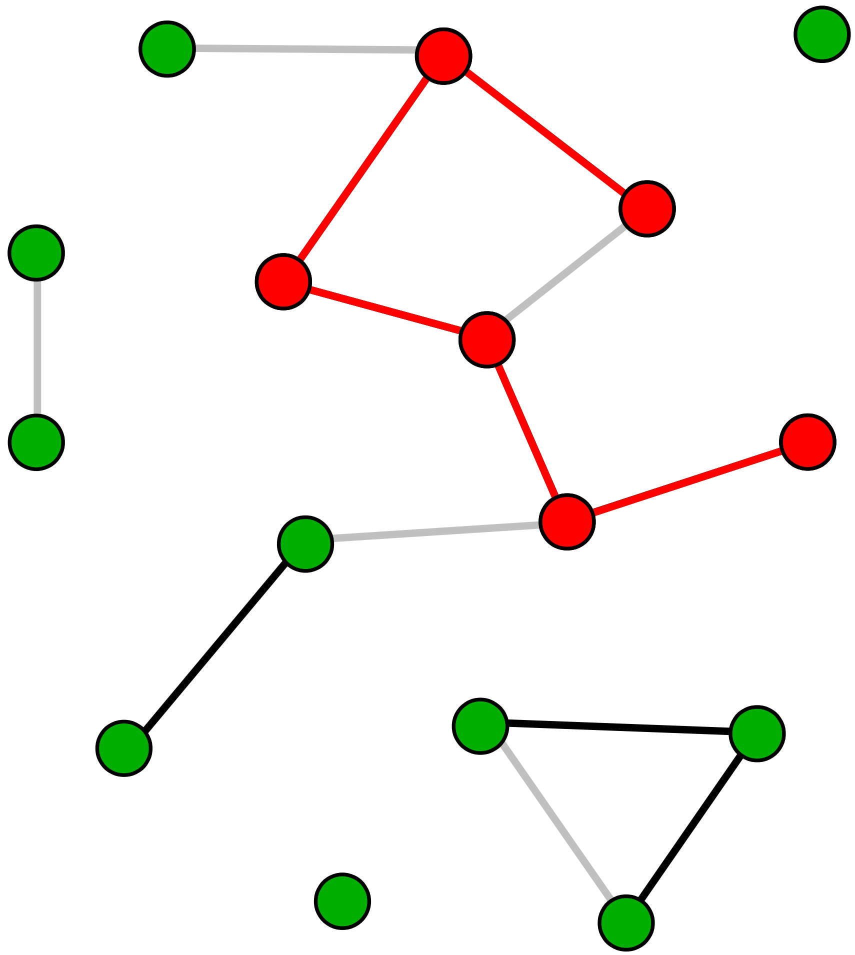
And hence

$$\boxed{\pi_s = \frac{\langle k\rangle}{(s-1)!}\left[\frac{\mathrm{d}^{s-2}}{\mathrm{d}z^{s-2}}\left[g_1(z)\right]^s\right]_{z=0}}$$

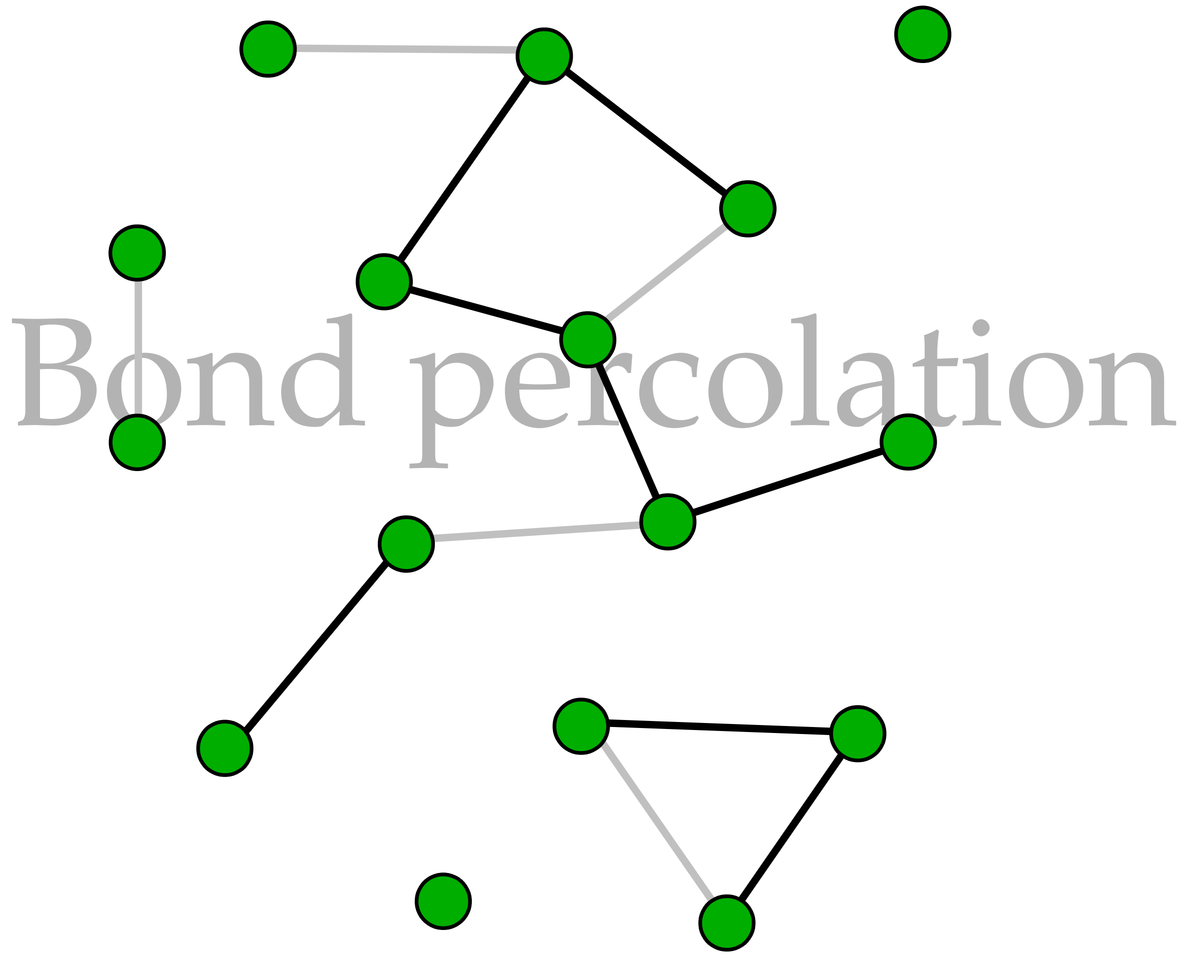
Epidemiology



Epidemiology



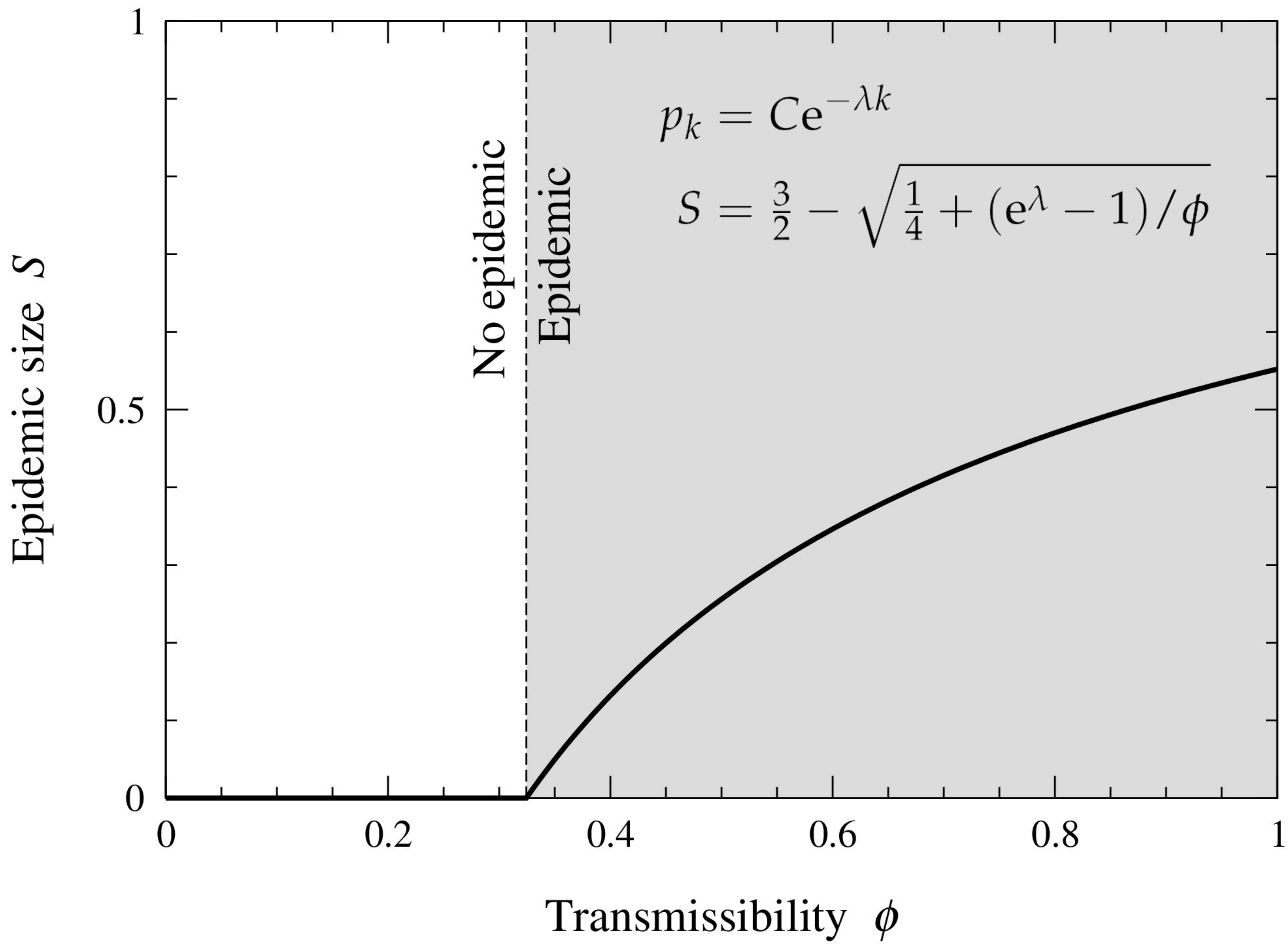
Epidemiology



Bond percolation

- Let u be the probability that an edge does not connect to the giant component
- The probability that a *particular* edge does not lead to the giant component is $1 - p + pu^k$, where k is the excess degree of the vertex at the end of the edge
- Then the average probability is

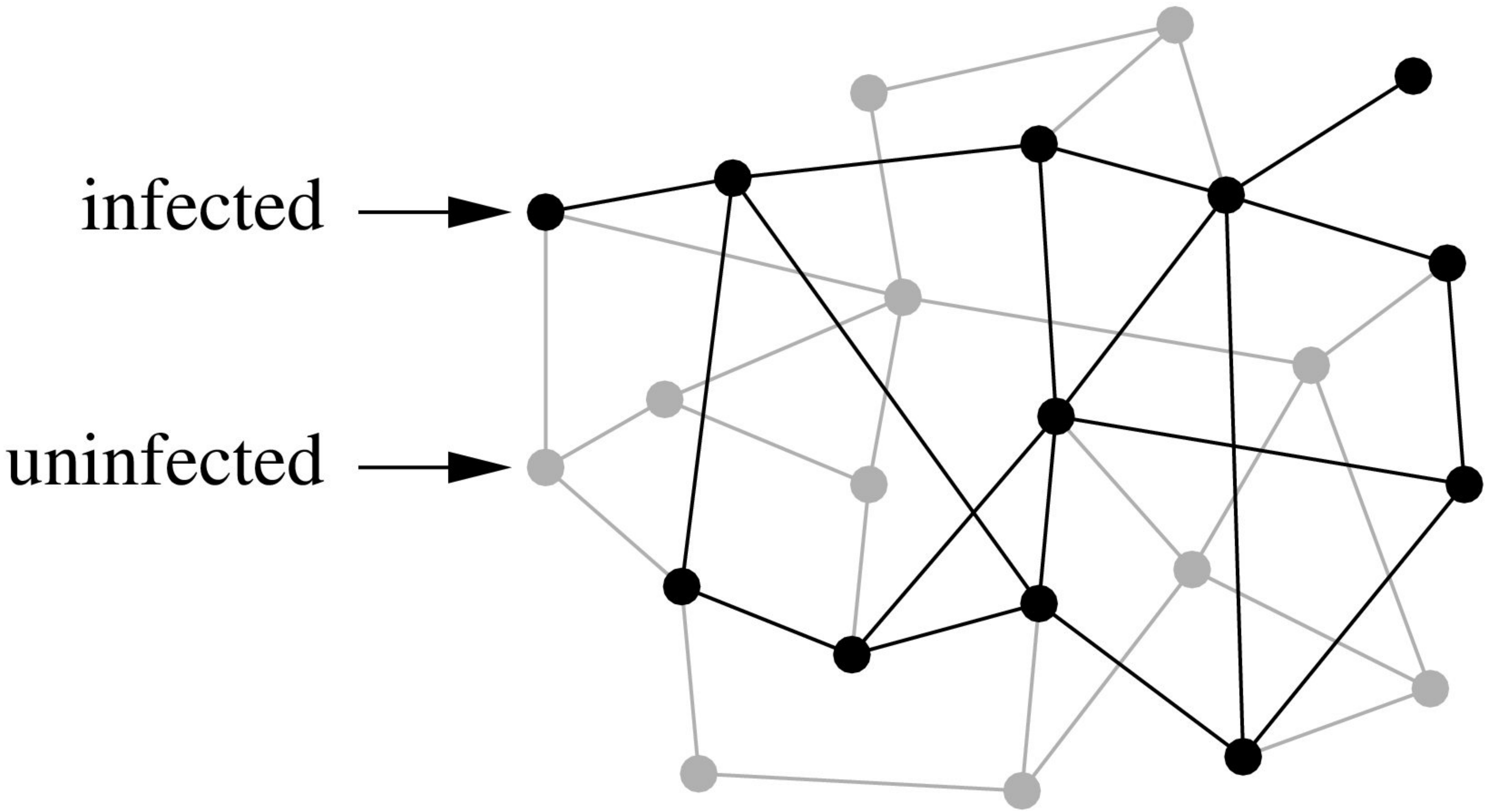
$$\sum_k q_k(1 - p + pu^k)$$

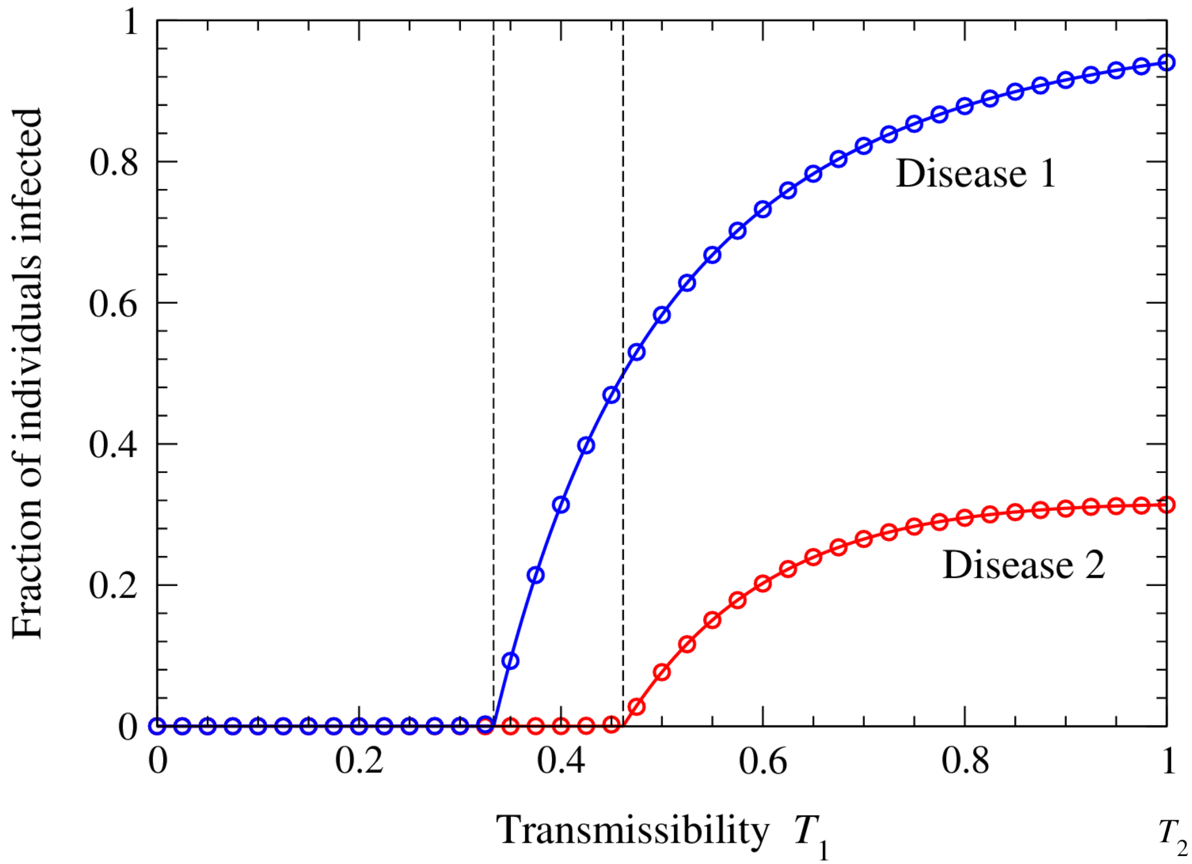


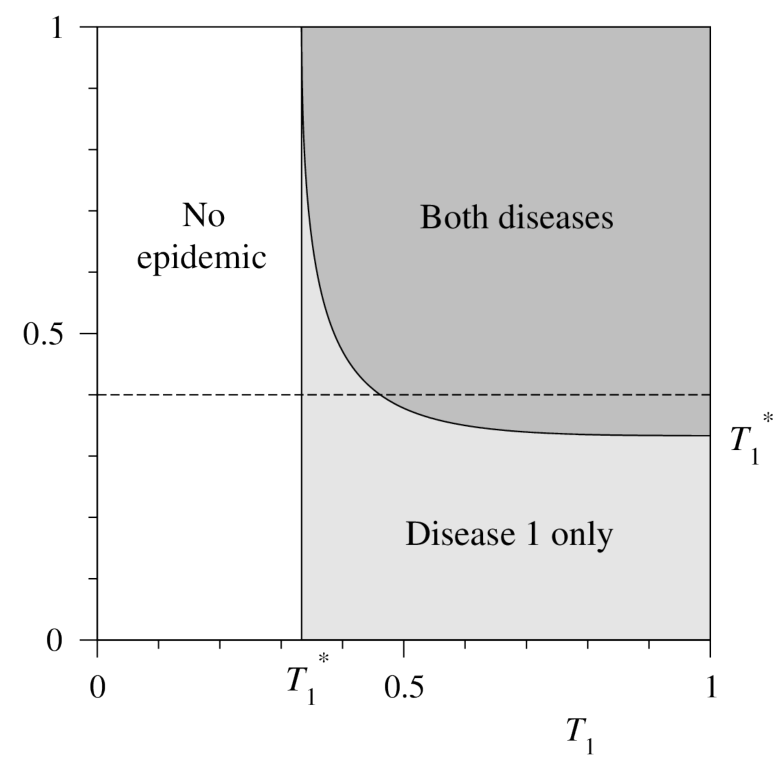
1
0.5
0
0
0.2
0.4
0.6
0.8
1
Epidemic size S
Transmissibility ϕ
No epidemic
Epidemic
$p_k = Ce^{-\lambda k}$
$S = \frac{3}{2} - \sqrt{\frac{1}{4} + (e^{\lambda} - 1)/\phi}$

Coinfection

- Now suppose we have *two* diseases
- And suppose that one disease *depends* on the other:
  - Infection with the first disease is necessary for infection with the second
  - Or makes the second more likely
  - Example: HIV's immunosuppressant effects increase the chances of getting certain types of infections


infected
uninfected

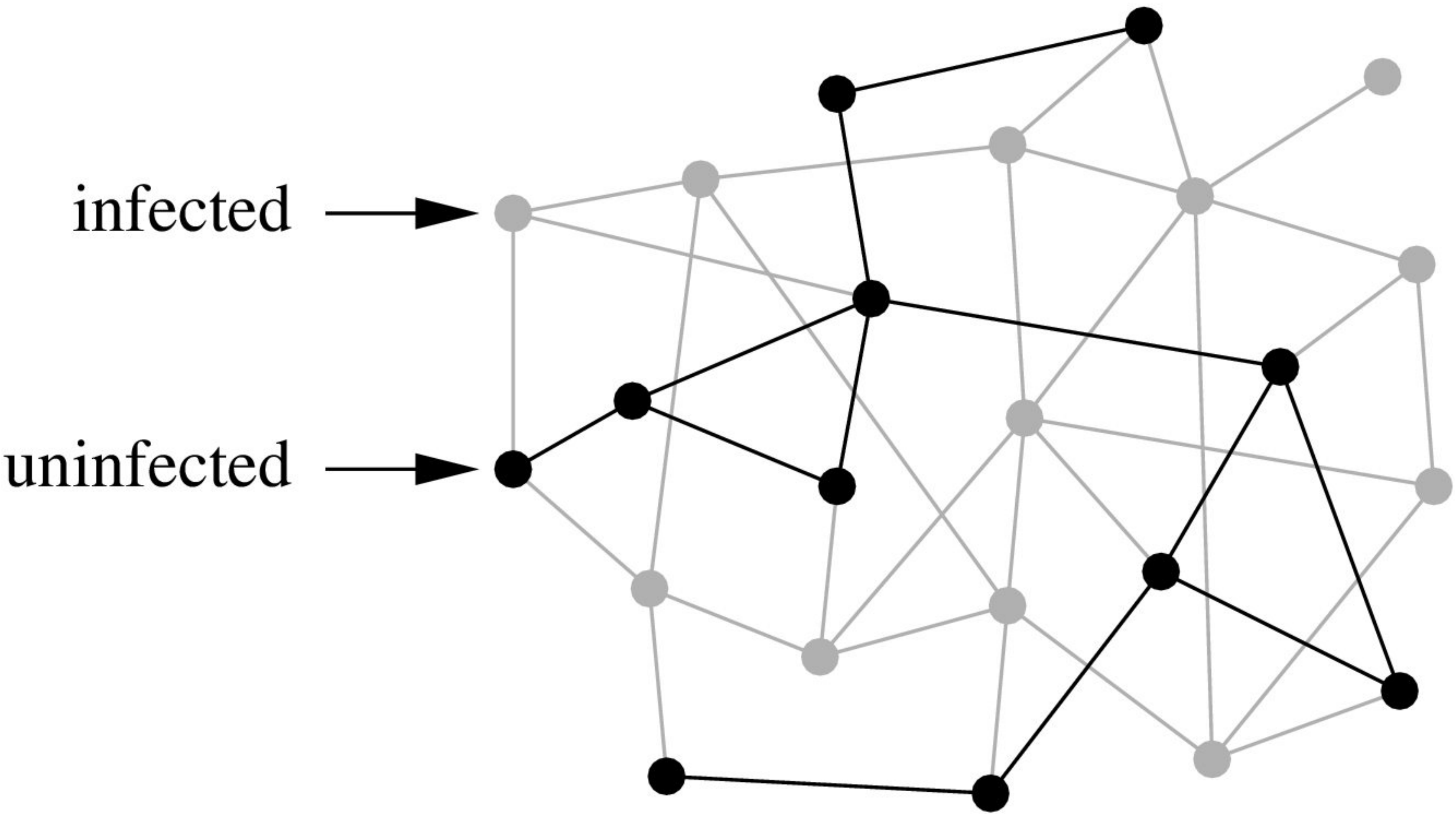




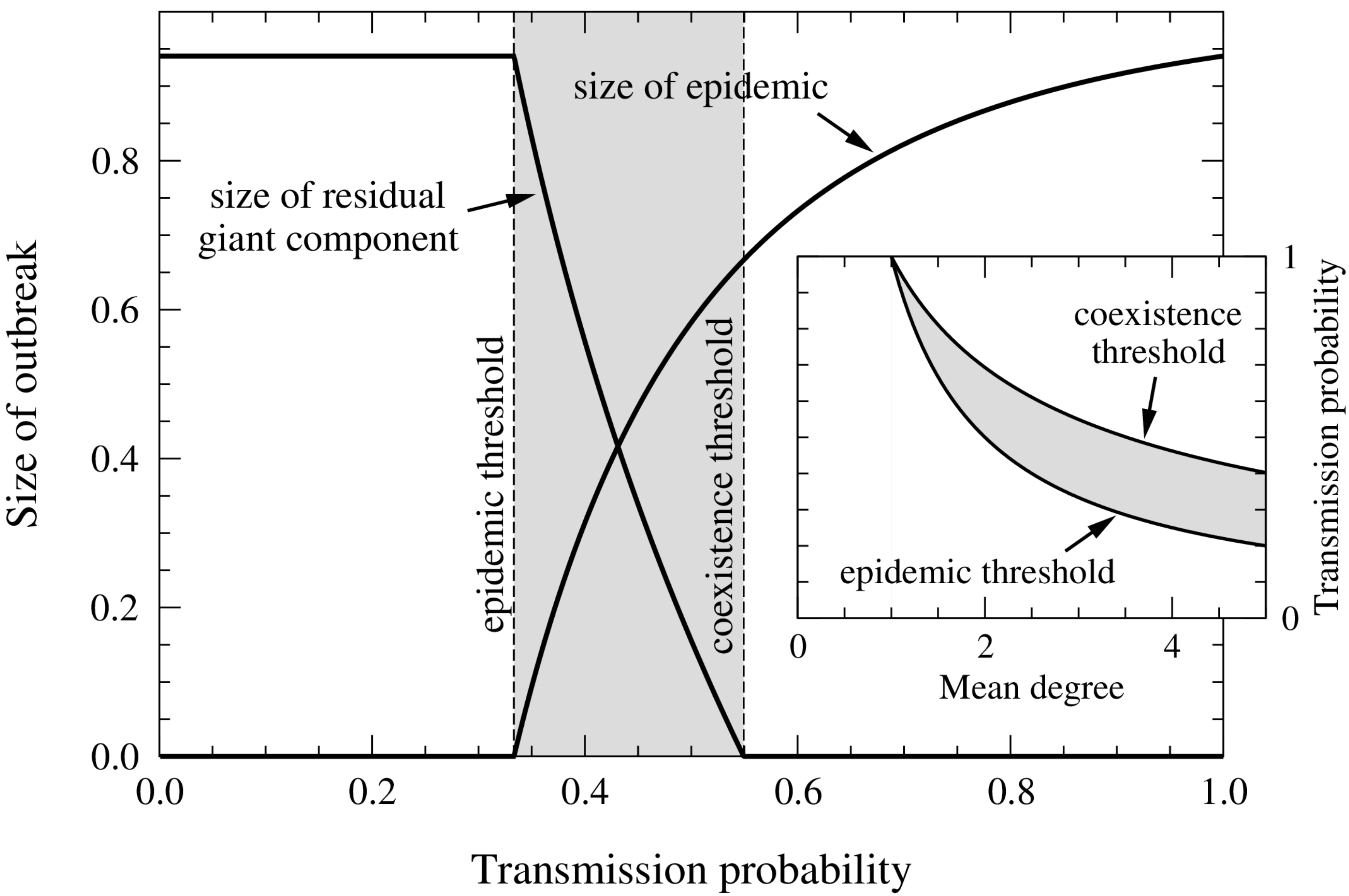
Newman and Ferrario, *PLOS One* 2013

Competing pathogens

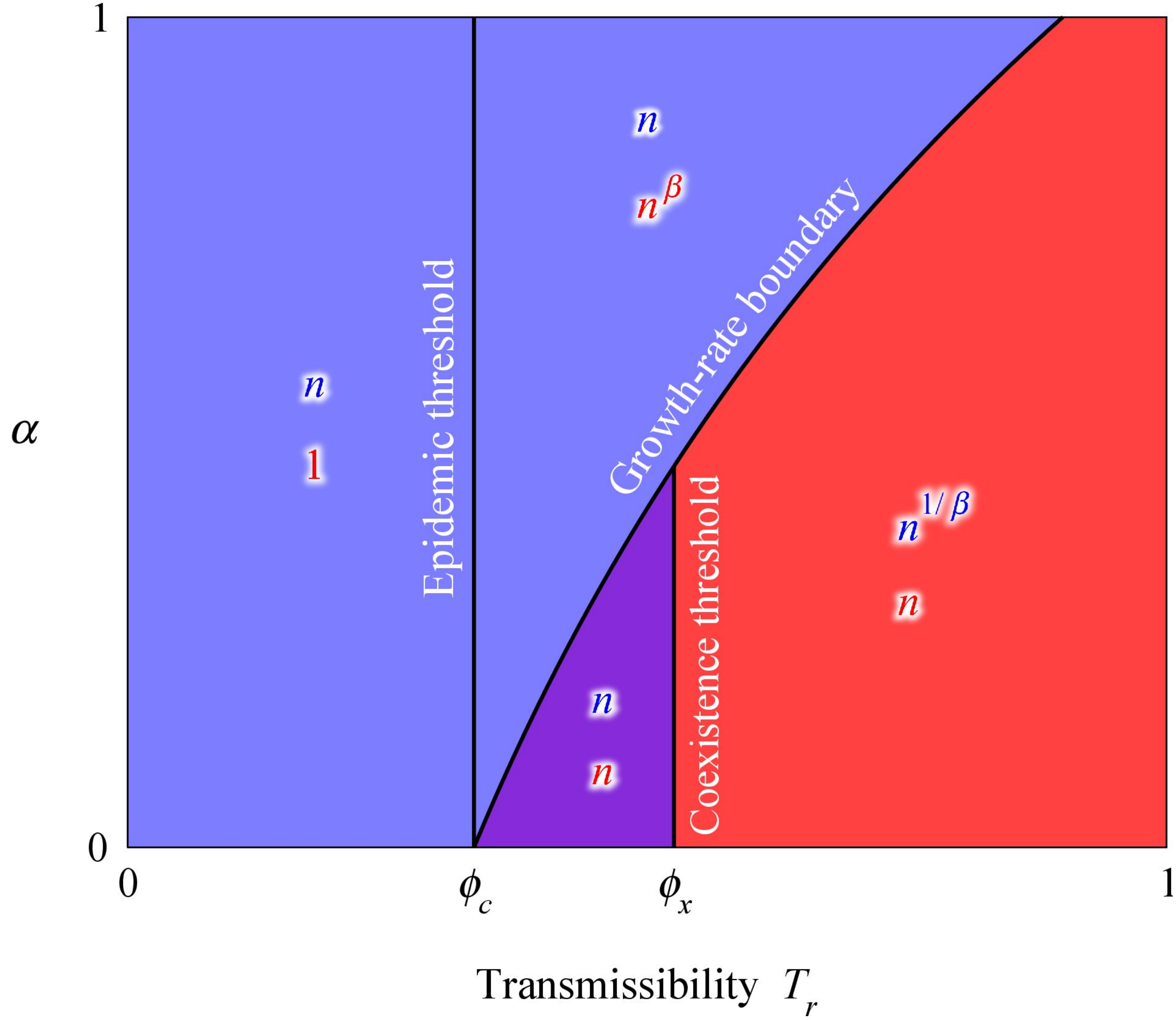
- Alternatively, the two diseases can compete:
  - One possibility is *cross-immunity*
  - Example: Different strains of the same disease, like the flu
  - Immunity to one strain gives you full or partial immunity to the other
  - The second disease can only infect those who didn't already catch the first



infected
uninfected



Newman, *PRL* 2005



Karrer and Newman *PRE* 2011