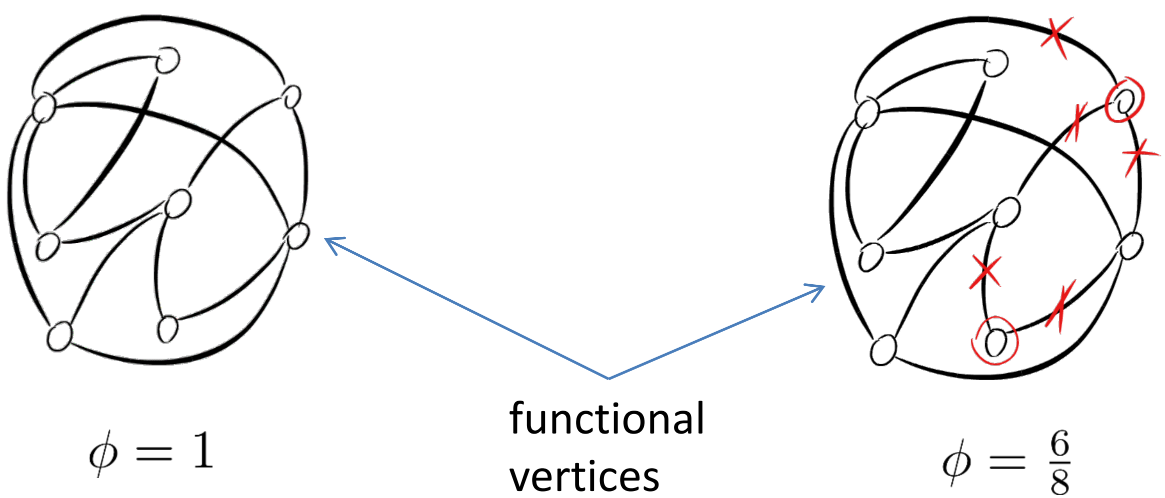
Percolation and Network Resilience

Percolation is a process of removing some fraction of network's nodes with adjacent edges. (more precisely node/link/cluster percolation)

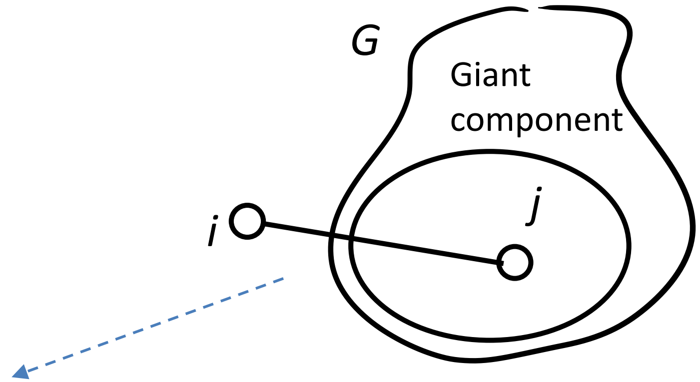
- models real-life phenomena such as router failure, immunization of people, and disasters
- the process is parameterized by **occupation probability** ϕ
- **Percolation transition**: when ϕ is large there is a giant component but as ϕ is decreased then giant component breaks into many small components or clusters (similar to phase transition in Poisson random graphs when giant component appears)
- Percolation can be defined on both nodes and edges. Here is an example of percolation on nodes:



$\phi = 1$ $\qquad$ $\phi = \frac{6}{8}$

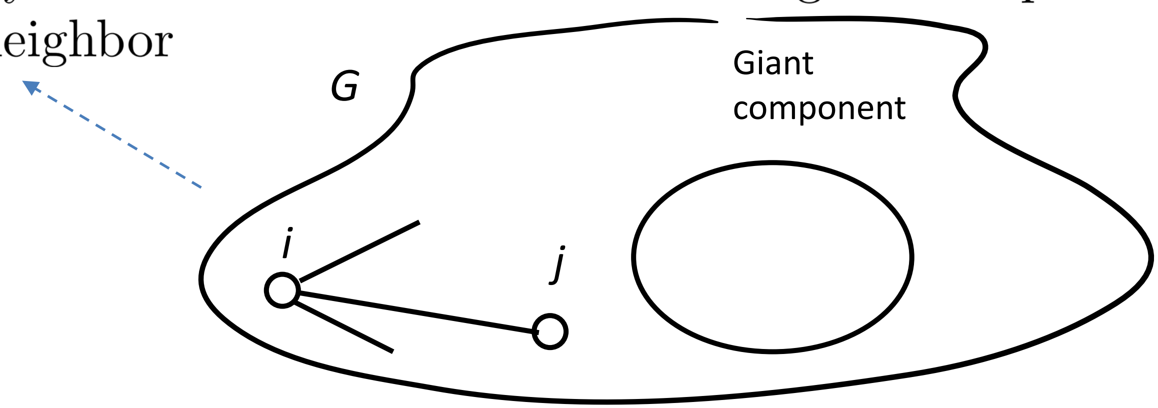
Percolation and Configuration Model

Consider a configuration model with
- degree distribution p_k
- occupation probability ϕ

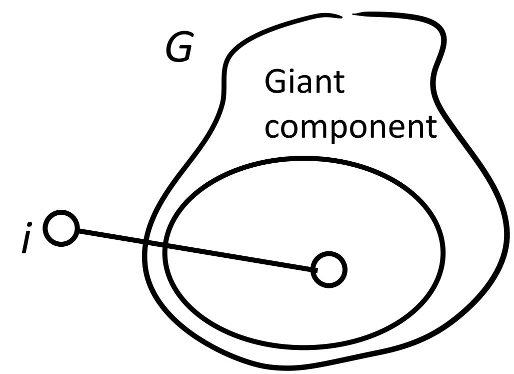


Consider node i which

- can belong to giant component, i.e., connected to it through some $j \in N(i)$
- is not in giant component (i.e., not connected to it via any of $N(i)$)
  - if $deg(i) = k$ then the total probability that it is not connected to a giant component is u^k, where
  - u is the probability that a vertex is not connected to giant component via a particular neighbor



Percolation and Configuration Model



Consider a configuration model with
- degree distribution p_k
- occupation probability ϕ
- u is the probability that a vertex is not connected to giant component via particular neighbor

For node i

- Avgerage probability of not being in giant component

$$\sum_k p_k u^k = g_0(u), \text{ where } g_0(z) = \sum_k^{\infty} p_k z^k$$

generating function for the degree distribution

or

$$\Pr[i \in \text{giant component}] = 1 - g_0(u)$$

- Total fraction of nodes in giant component when percolation is running

$$S = \phi(1 - g_0(u))$$

So, the question is how to calculate *u*?

Let us calculate u, the probability that i is not connected to giant component via a particular neighbor. There are two cases:

- i is connected to j which is removed with prob $1 - \phi$
- or j is not removed with prob ϕ but it is not in giant component

$$\Pr[i \notin gc \text{ via } j] = 1 - \phi + \phi u^k$$

j is on and its *k* neighbors are not in giant component

Node j is reached by following an edge, so average probability

$$u = \sum_{k=0}^{\infty} q_k (1 - \phi + \phi u^k) = 1 - \phi + \phi \sum_{k=0}^{\infty} q_k u^k = 1 - \phi + \phi g_1(u)$$

This equation is not easy to solve

Reminder: excess degree is the number of edges attached to a vertex other than the edge we arrived along.

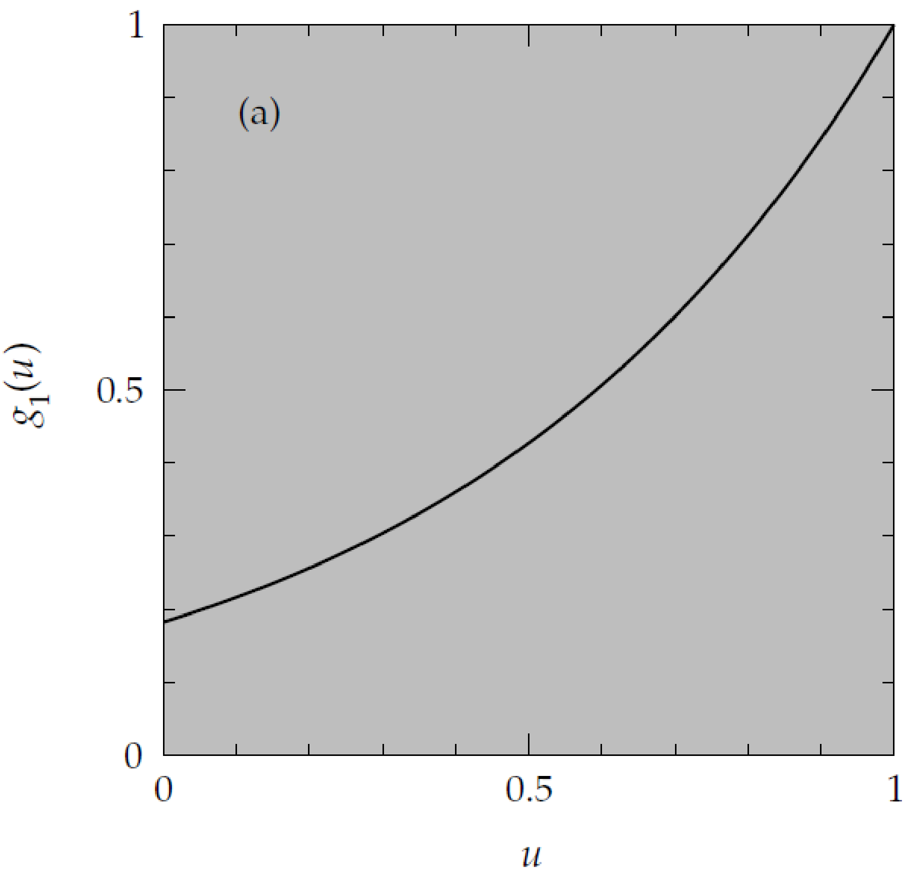
$$q_k = \frac{(k+1)\, p_{k+1}}{\langle k \rangle}, \quad \sum_k q_k = 1, \quad g_1(z) = \sum_{k=0}^{\infty} q_k z^k$$

Let us calculate u, the probability that i is not connected to giant component via a particular neighbor.

Node j is reached by following an edge, so average probability

$$u = \sum_{k=0}^{\infty} q_k(1 - \phi + \phi u^k) = 1 - \phi + \phi \sum_{k=0}^{\infty} q_k u^k = 1 - \phi + \phi g_1(u)$$

This equation is not easy to solve



The exact form of the curve will depend on the degree distribution, but we know the general shape: g_1 is a polynomial with all coefficients non-negative (because they are probabilities), so for $u>0$ it must have a non-negative value and all derivatives non-negative. Thus in general it is an increasing function of u and curves upward as shown in the figure.

To get the full right-hand side, we first multiply $g_1(u)$ by ϕ and then add $1 - \phi$. Graphically, this is equivalent to compressing the unit square along with the curve it contains, until it has height ϕ.
The shift it upward a distance $1 - \phi$. The point or points at which the resulting curve crosses the line $y = u$ are then the solutions.

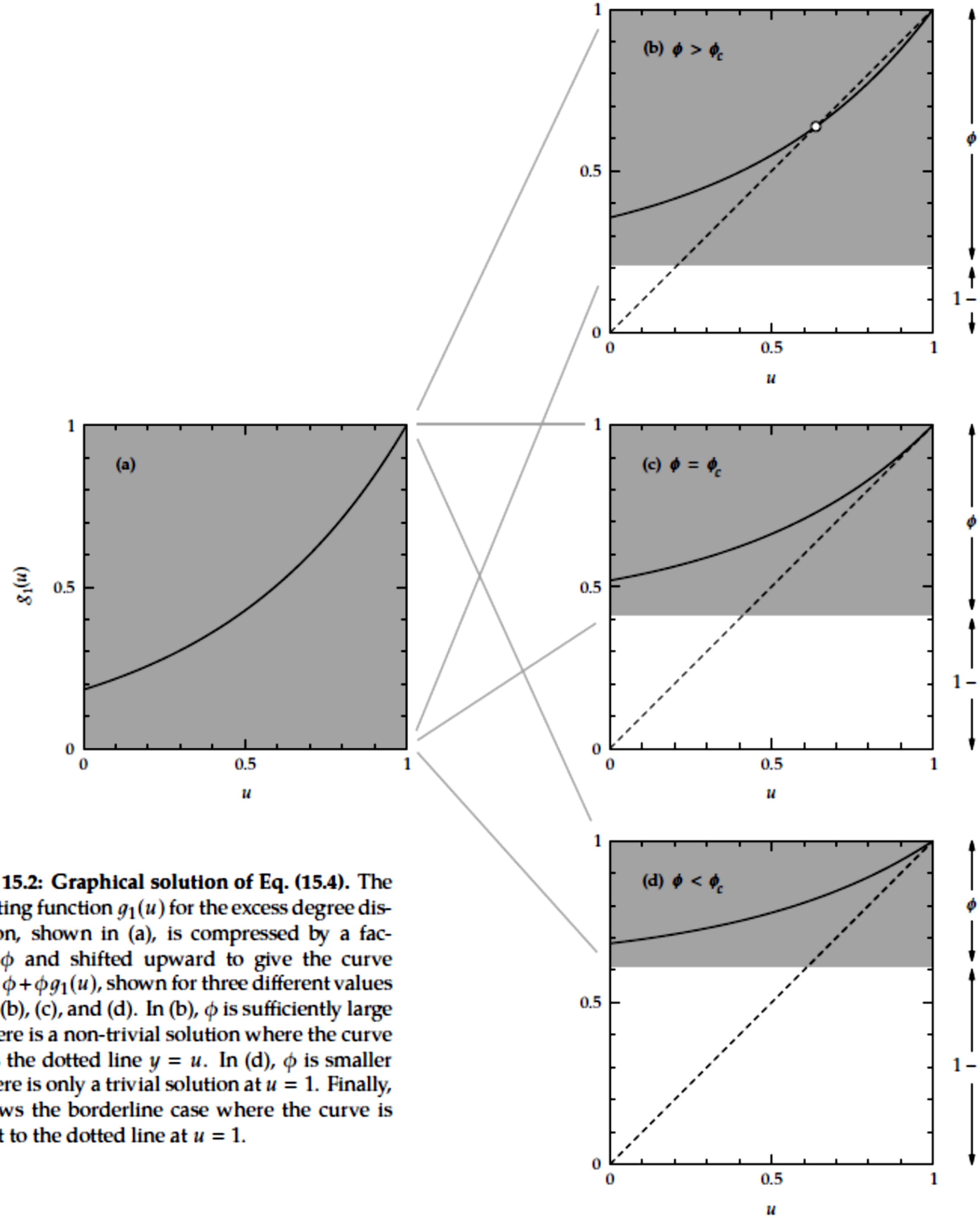


Figure 15.2: Graphical solution of Eq. (15.4). The generating function $g_1(u)$ for the excess degree distribution, shown in (a), is compressed by a factor of ϕ and shifted upward to give the curve $y = 1-\phi+\phi g_1(u)$, shown for three different values of ϕ in (b), (c), and (d). In (b), ϕ is sufficiently large that there is a non-trivial solution where the curve crosses the dotted line $y = u$. In (d), ϕ is smaller and there is only a trivial solution at $u = 1$. Finally, (c) shows the borderline case where the curve is tangent to the dotted line at $u = 1$.

Let us calculate u, the probability that i is not connected to gc via a particular neighbor. Two cases:

- i is connected to j which is removed with prob $1 - \phi$
- or j is not removed with prob ϕ but it is not in gc

$$\Pr[i \notin gc \text{ via } j] = 1 - \phi + \phi u^k$$

j is on and its *k* neighbors are not in gc

Node j is reached by following an edge, so average probability

$$u = \sum_{k=0}^{\infty} q_k (1 - \phi + \phi u^k) = 1 - \phi + \phi + \sum q_k u^k = 1 - \phi + \phi g_1(u)$$

[Cohen, Erez, Ben-Avraham, Halvin]:

$$\phi_c = \frac{1}{g_1'(1)} = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}$$

Minimum fraction of nodes that must be occupied in configuration model for a giant component to exist

[Cohen, Erez, Ben-Avraham, Halvin]: $\phi_c = \frac{1}{g_1'(1)} = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}$

- In configuration model fixing low ϕ_c leads to gc, for example $\langle k^2 \rangle > \langle k \rangle$

- Example: given Poisson degree distribution with c - mean degree

$$p_k = e^{-c} \frac{c^k}{k!} \Rightarrow \langle k \rangle = c, \ \langle k^2 \rangle = c(c+1) \Rightarrow \phi_c = \frac{1}{c}$$

  i.e., $c = 4$ means that $\frac{3}{4}$ vertices will fail before giant component disappears.

- Example: power laws with $2 < \alpha < 3$ (Internet, etc)

$$\langle k \rangle \text{ is final, } \langle k^2 \rangle \text{ diverges } \Rightarrow \phi_c \to 0$$

  i.e., remove many vertices form the network $\Rightarrow$ giant component will be there

- Opposite Example: Epidemiological networks. Small ϕ_c are bad! The fewer individuals we need to vaccinate to destroy giant component the better.

[Cohen, Erez, Ben-Avraham, Halvin]: $\phi_c = \frac{1}{g_1'(1)} = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}$

- Example: Exponential degree distribution $p_k = (1 - e^{-\lambda})e^{-\lambda k}$, $\lambda > 0$

$$g_0(z) = \frac{e^\lambda - 1}{e^\lambda - z},\ g_1(z) = \left(\frac{e^\lambda - 1}{e^\lambda - z}\right)^2 \Rightarrow$$

$$u\left(e^\lambda - u\right)^2 - (1-\phi)\left(e^\lambda - u\right)^2 - \phi\left(e^\lambda - 1\right)^2 = 0 \Rightarrow$$

u=1 is always a solution and (u-1) is always factor

$$u = e^\lambda - \frac{1}{2}\phi - \sqrt{\frac{1}{4}\phi^2 + \phi(e^\lambda - 1)} \Rightarrow$$

size of giant component

$$S = \frac{3}{2}\phi - \sqrt{\frac{1}{4}\phi^2 + \phi(e^\lambda - 1)} \Rightarrow$$

percolation threshold

$$\phi_c = \frac{1}{2}(e^\lambda - 1)$$

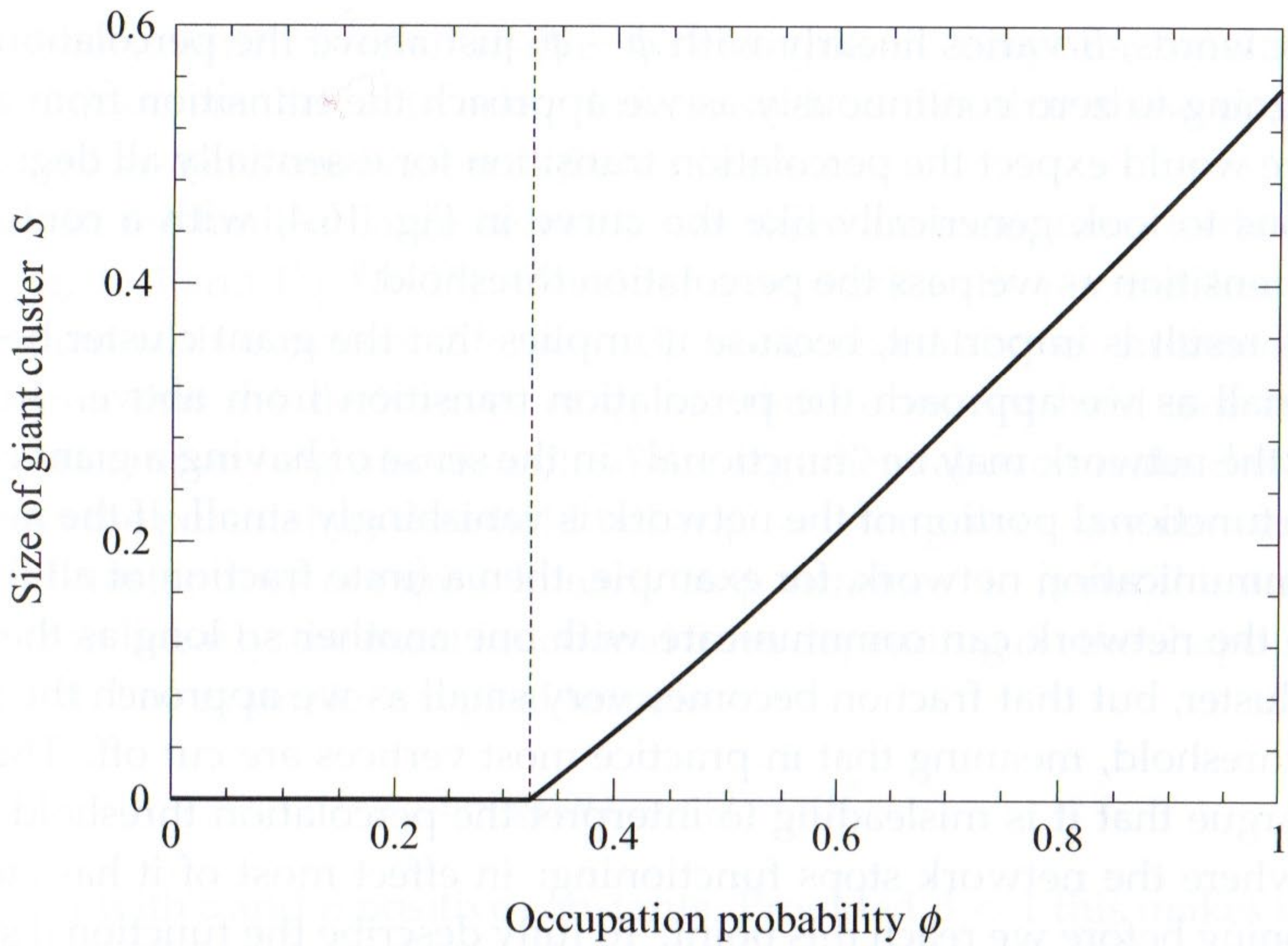


Figure 16.4: Size of the giant cluster for site percolation in the configuration model. The curve indicates the size of the giant cluster for a configuration model with an exponential degree distribution of the form (16.12) with $\lambda = \frac{1}{2}$, as given by Eq. (16.18). The dotted line indicates the position of the percolation transition, Eq. (16.20).

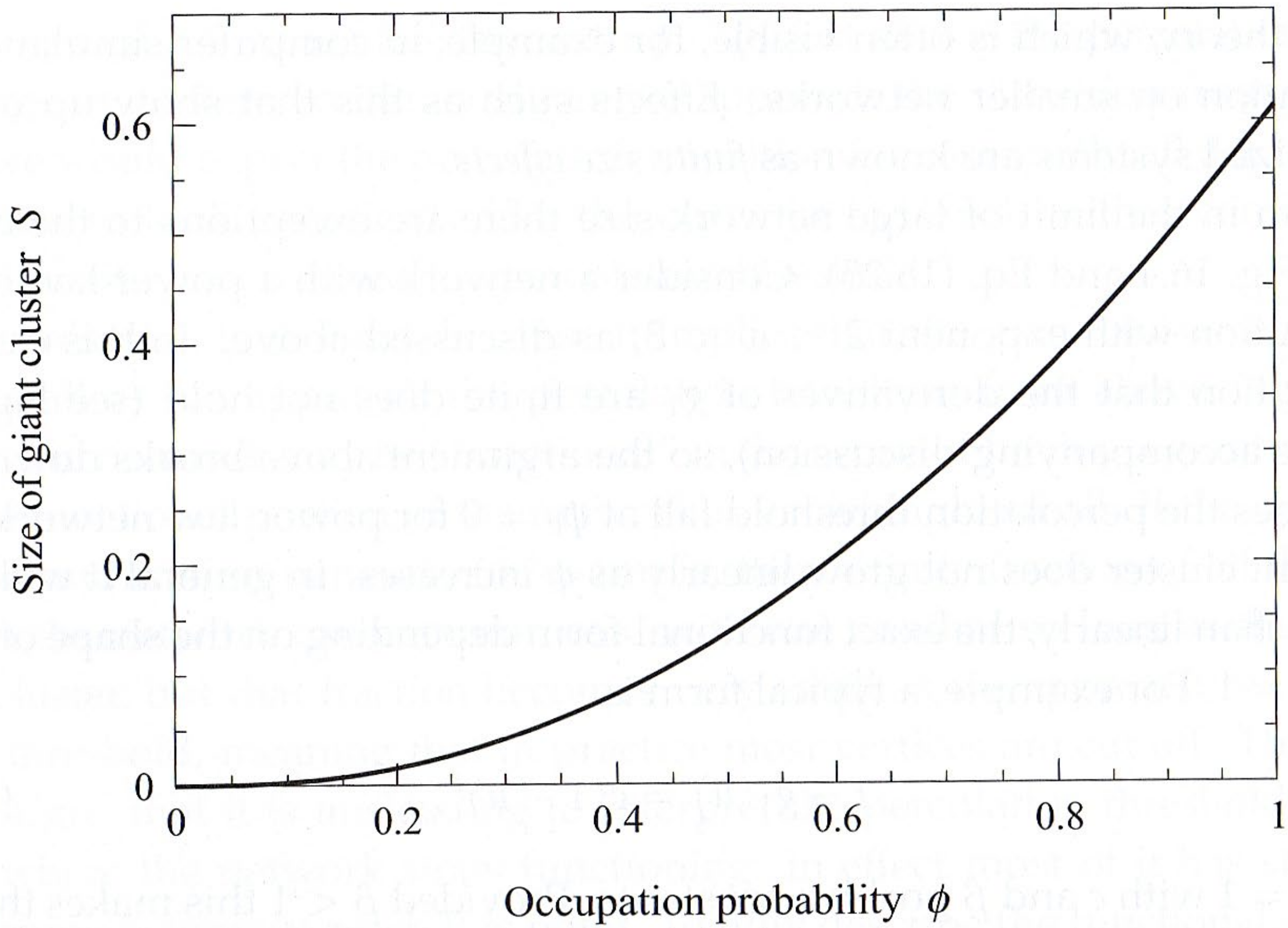


Figure 16.5: Size of the giant cluster for a network with power-law degree distribution. The size of the giant cluster for a scale-free configuration model network with exponent $\alpha = 2.5$, a typical value for real-world networks. Note the non-linear form of the curve near $\phi = 0$, which means that S, while technically non-zero, becomes very small in this regime. Contrast this figure with Fig. 16.4 for the giant cluster size in a network with an exponential degree distribution.

Geometric graph: percolation example

Non-uniform Removal of Vertices

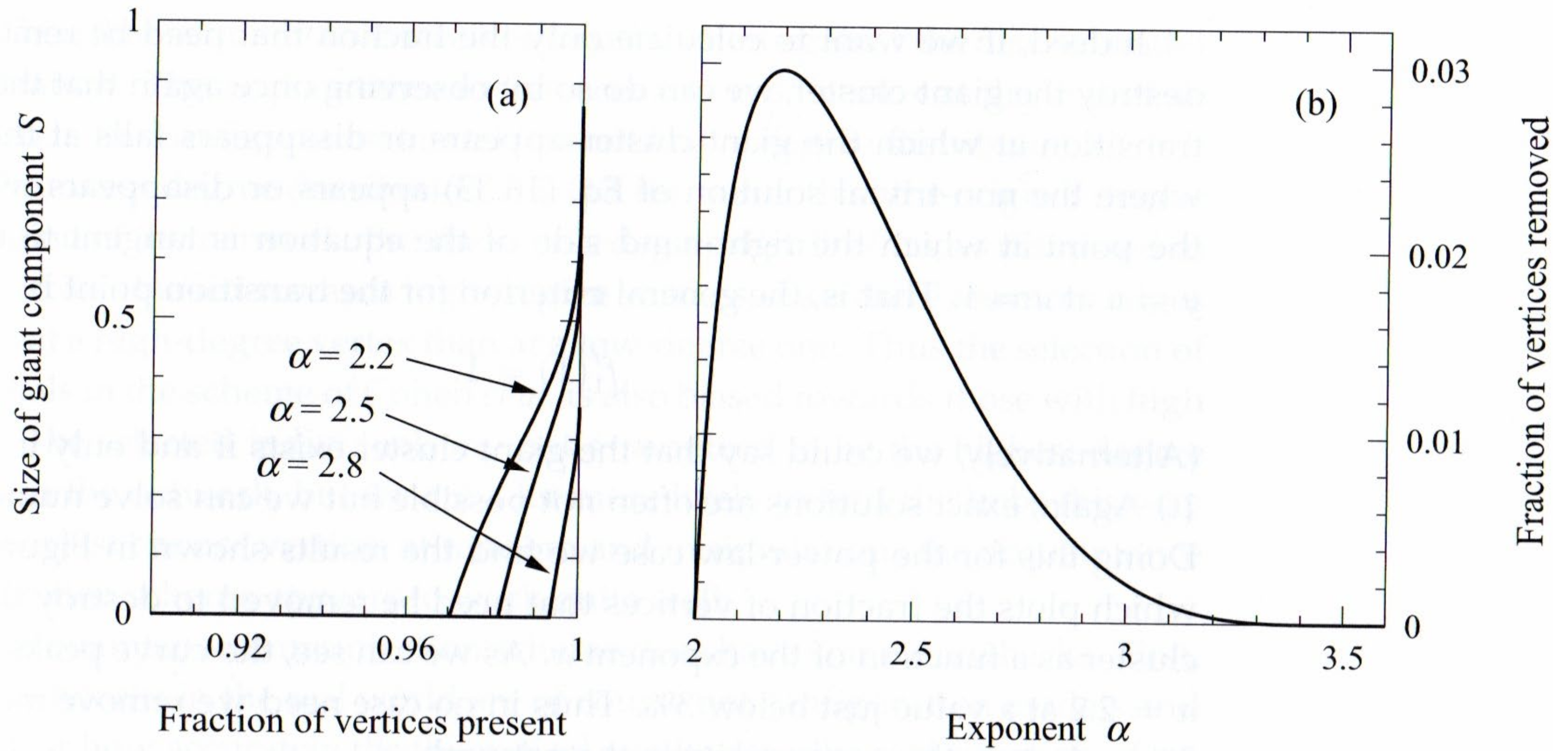


Figure 16.7: Removal of the highest-degree vertices in a scale-free network. (a) The size of the giant cluster in a configuration model network with a power-law degree distribution as vertices are removed in order of their degree, starting with the highest-degree vertices. Only a small fraction of the vertices need be removed to destroy the giant cluster completely. (b) The fraction of vertices that must be removed to destroy the giant cluster as a function of the exponent α of the power-law distribution. For no value of α does the fraction required exceed 3%.

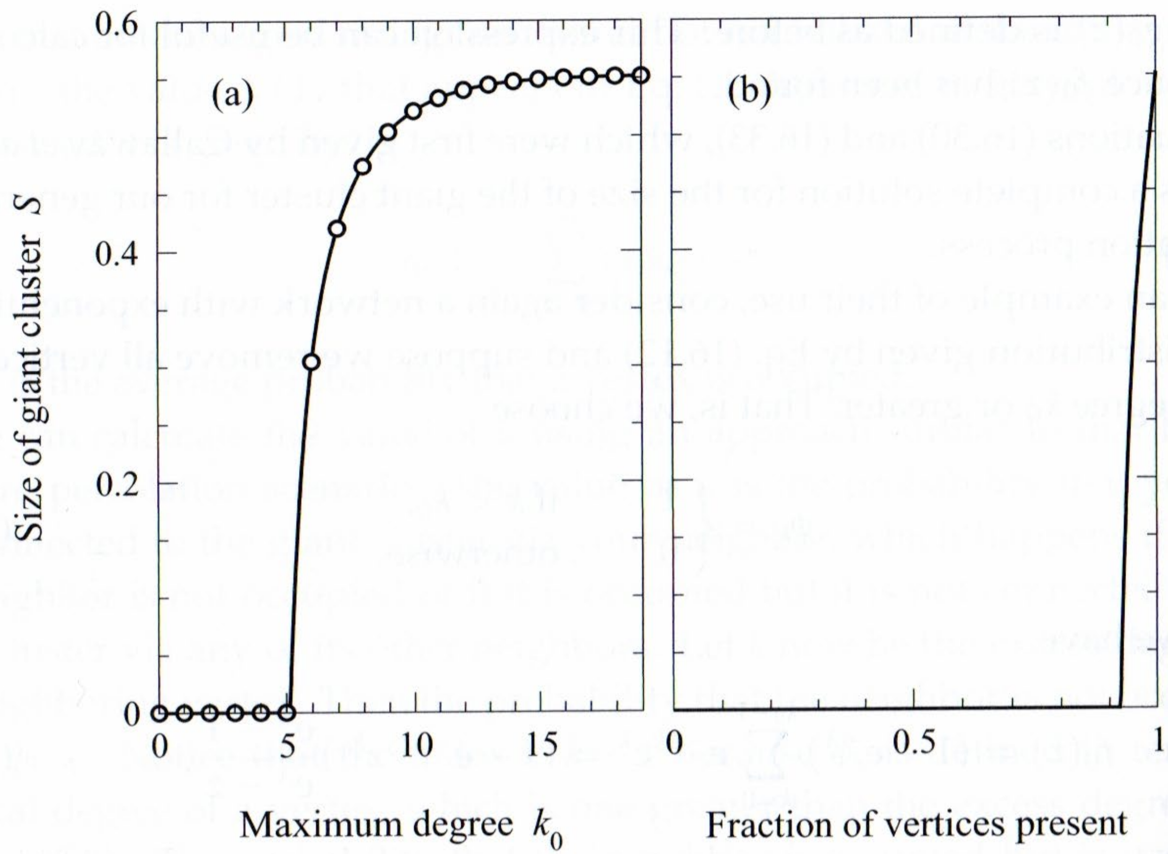
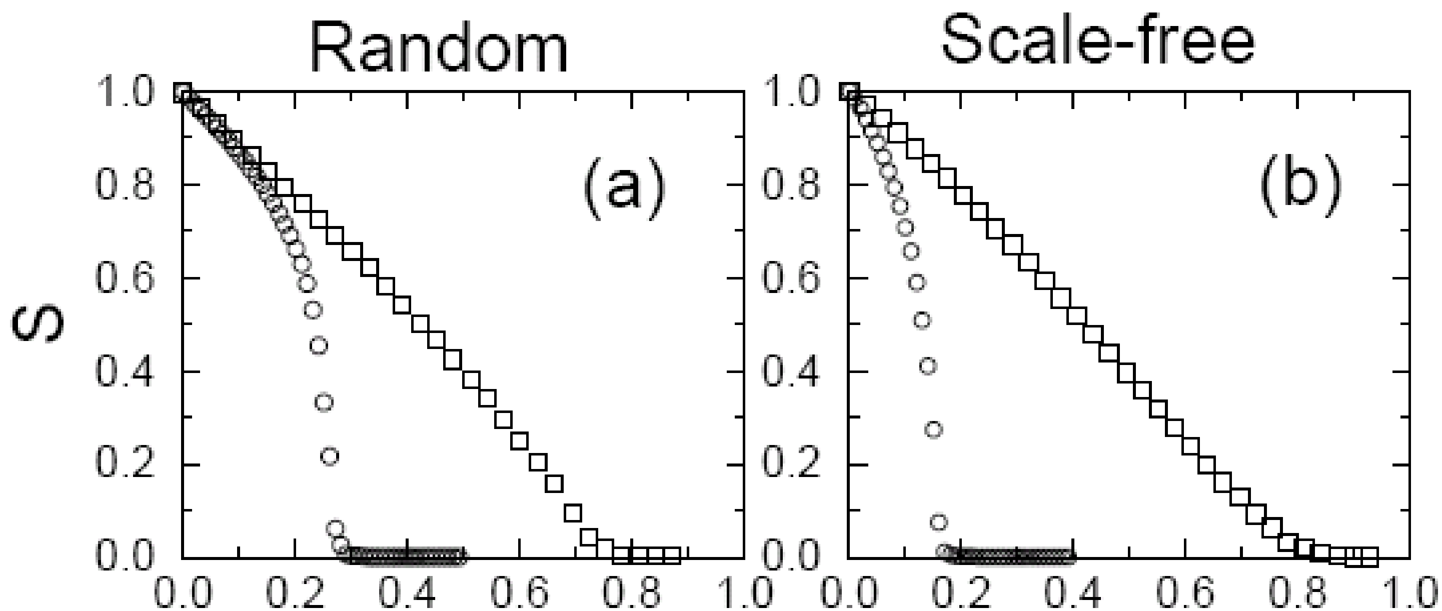


Figure 16.6: Size of the giant percolation cluster as the highest degree vertices in a network are removed. (a) The size of the giant cluster in a network with an exponential degree distribution $p_k \sim e^{-\lambda k}$ with $\lambda = \frac{1}{2}$ as vertices are removed in order of degree, starting from those with the highest degree. The curve is shown as a function of the degree k_0 of the highest-degree vertex remaining in the network. Technically, since k_0 must be an integer, the plot is only valid at the integer points marked by the circles; the curves are just an aid to the eye. (b) The same data plotted now as a function of the fraction $\bar{\phi}$ of vertices remaining in the network.

Network resilience to targeted attacks

Scale-free graphs are resilient to random attacks, but sensitive to targeted attacks. For random networks there is smaller difference between the two



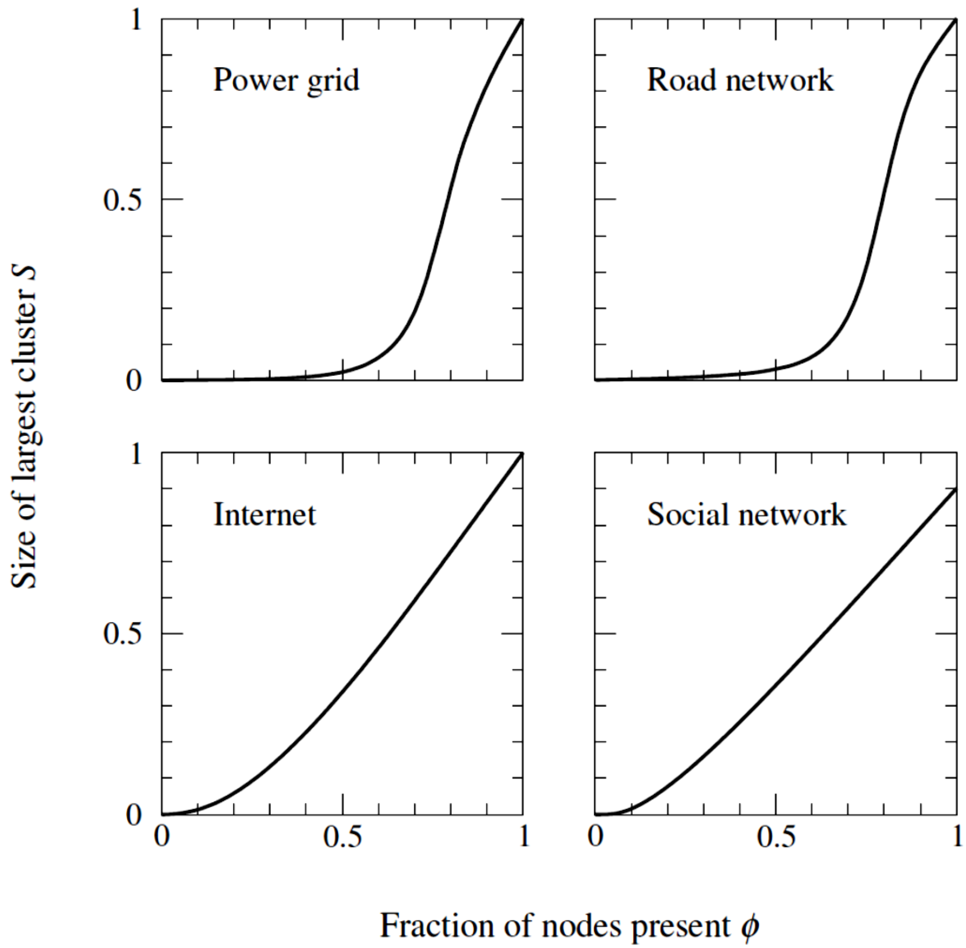


Figure 15.10: Size of the largest cluster as a function of occupation probability for percolation on four networks. The four frames of this figure show the size of the largest percolation cluster, measured as a fraction of network size, for random removal of nodes from four real-world networks: the western United States power grid, the network formed by the US interstate highways, the Internet at the level of autonomous systems, and a social network of professional collaborations between physicists. Each curve is averaged over 1000 random repetitions of the calculation, which is why the curves appear smooth.

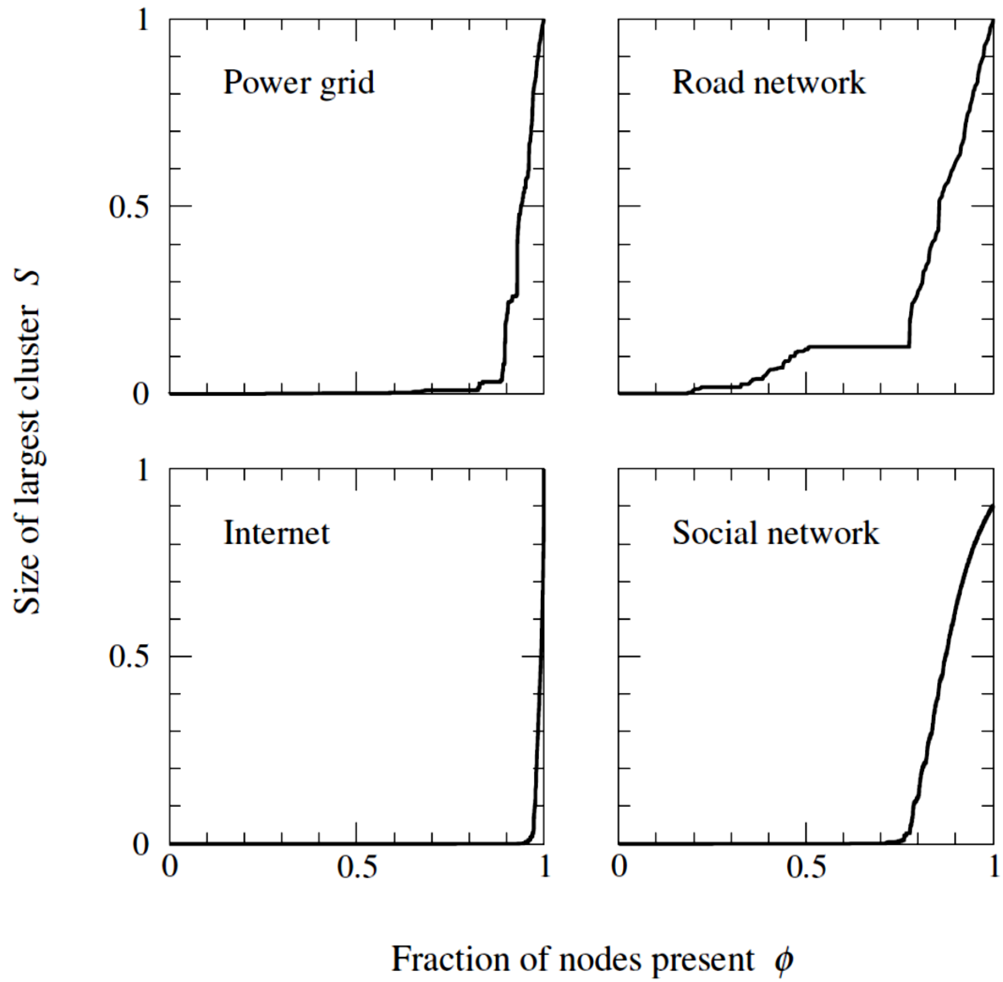


Figure 15.11: Size of the largest percolation cluster as a function of occupation probability for targeted attacks on four networks. The four frames in this figure show the size of the largest cluster, as a fraction of network size, when nodes are removed in degree order, highest degrees first, from the same four networks as Fig. 15.10. Since this is mostly a deterministic process and not a random one (except for random choices between nodes of the same degree) the curves cannot be averaged as in Fig. 15.10 and so are relatively jagged.