

Complex Networks percolation

2019.1.31(Thu)

Goal

metrics

algorithms

models

processes

contents of this chapter

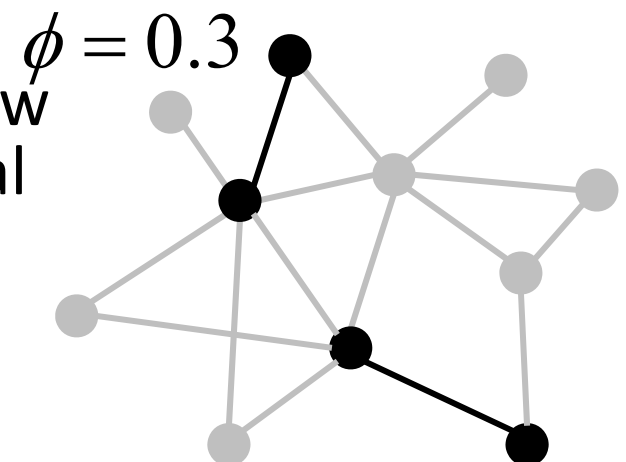
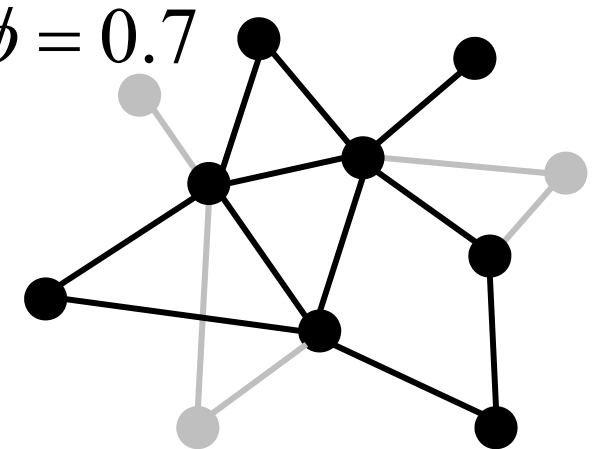
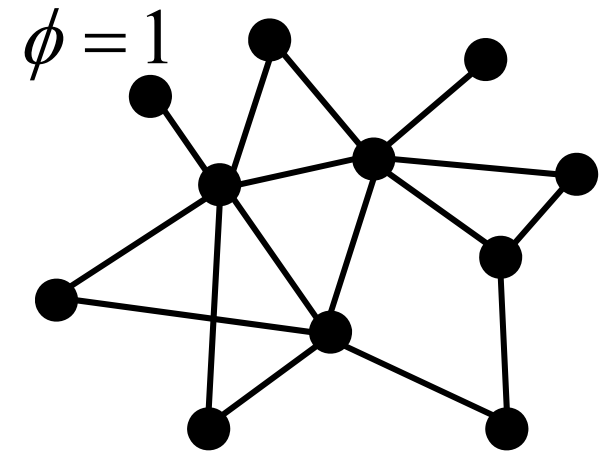
- Percolation : one of the simplest processes taking place on networks
 - When vertices are removed from a network, how long its giant component still remains?
 - Similar to the phase transition of a random graph (when a giant component appears?)

Processes on networks



- To make the connection between network structure and function
 - Network failure and resilience
 - Dynamic systems on networks
 - Epidemic and spreading processes

percolation

- Removing some vertices in a network
 - failure of routers on the Internet
 - vaccination / immunization against the spread of disease
- “knock-on” effect
 - vaccination of small fraction of the population can effectively prevent the spread of disease
 - “herd immunity”
- Percolation theory : to understand how the knock-on effects of vertex removal affect the network as a whole

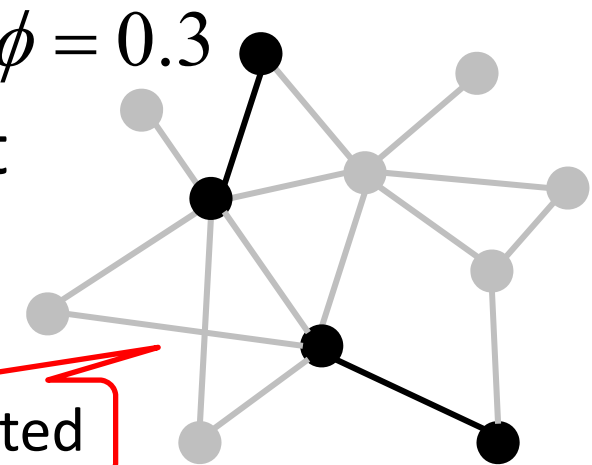
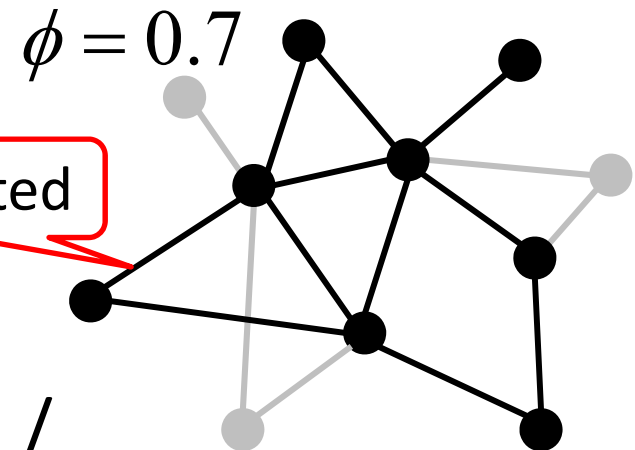
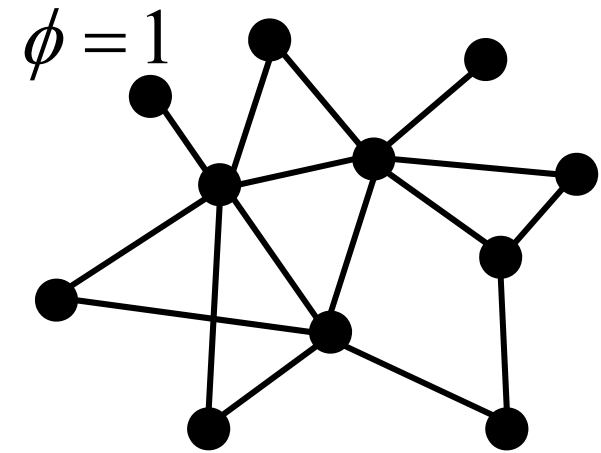


vertex / edge removal

- site percolation : vertex removal 
- bond percolation : edge removal
- random removal 
- removal from highest degrees
- removal of highest betweenness

Uniform random removal of vertices

- ϕ : probability that a vertex is present (occupation probability)
 - $\phi = 1$: no vertices have been removed
 - $\phi = 0$: all vertices have been removed
- Percolation transition : formation / dissolution of a giant component
- Percolation threshold : the point at which the transition occurs

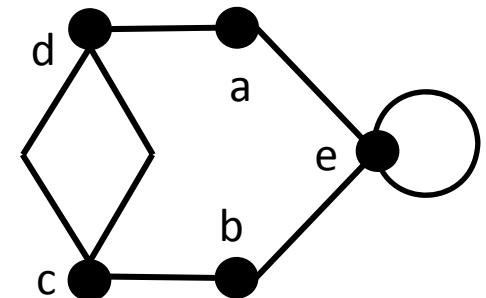
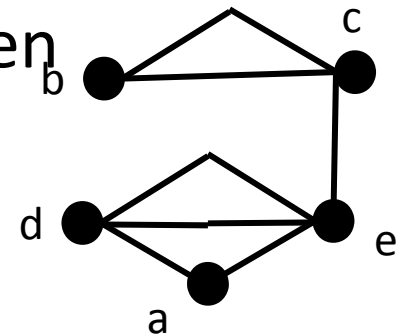
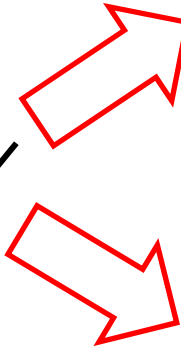
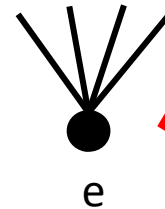
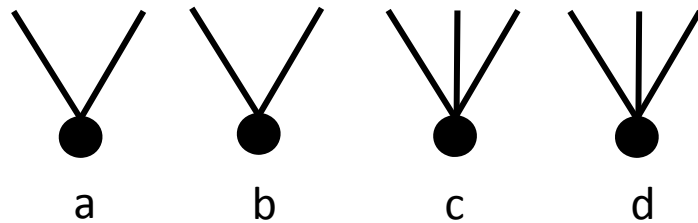


configuration model

- Degree distribution of a naïve random graph is Poisson distribution
- Configuration model: a method for generating networks of arbitrary degree distribution

– “stumps” are prepared according to given degree distribution and they are connected randomly

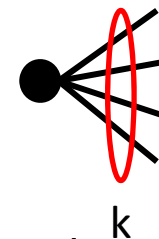
v	deg
a	2
b	2
c	3
d	3
e	4



Uniform removal in the configuration model (1)

- Site percolation process on networks generated using the configuration model
- Configuration model network with degree distribution : p_k
- Occupation probability : ϕ
- Size of the giant cluster S
 - k : degree
 - u : average probability that the vertex is not connected to the giant cluster via a particular neighbor
 - u^k : probability of its not belonging to the giant cluster
 - $\sum_k p_k u^k (= g_0(u))$: average probability of not being in the giant cluster
 - $1 - g_0(u)$: average probability of being in the giant cluster
 - $S = \phi[1 - g_0(u)]$: total fraction of the giant cluster (because ϕ is the probability of remaining nodes)

removed ($1 - \phi$) or present but not connected to GC via its neighbors



$$g_0(z) = \sum_{k=0}^n p_k z^k \quad \text{:generating function for degree distribution}$$

Uniform removal of the configuration model (2)

All neighbors of vertex A are not connected to GC

- Total probability of not connecting to the giant component via vertex A is $1 - \phi + \phi u^k$
- Excess degree distribution
 - Probability of an edge reaches to a vertex of degree k (other than the edge we arrived along)

vertex A is removed

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

- Average probability that a vertex is not connected to the giant component :

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

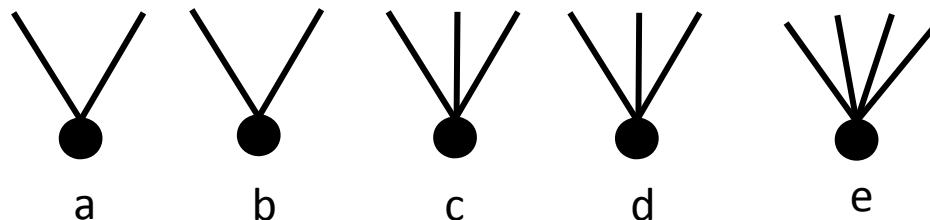
- $= 1 - \phi + \phi g_1(u)$

$$g_1(z) = \sum_{k=0}^{\infty} q_k z^k \quad \text{:generating function for the excess degree distribution}$$

Excess degree distribution

- Configuration model with degree distribution p_k when vertices are chosen randomly
- If we take a vertex and follow one of its edges to the vertex at the other end, what is the probability that this vertex will have degree k ?
 - It is not p_k (because a vertex of degree zero will not be reached)
 - Probability of selecting an edge ending at any particular vertex of degree k : $k/(2m - 1) \approx k/2m$
 - There are np_k such vertices \rightarrow probability of an edge attaching to any vertex with degree k is $\frac{k}{2m} \times np_k = \frac{kp_k}{\langle k \rangle}$
 - The probability is proportional not to p_k but to kp_k

high degree is more likely



$2m$ edges

$$\langle k \rangle = \frac{2m}{n}$$

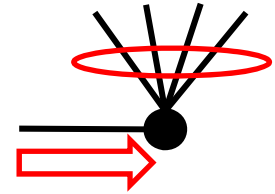
Your friends have more friends than you do

- Average degree of a neighbor : $\sum_k k \frac{k p_k}{\langle k \rangle} = \frac{\langle k^2 \rangle}{\langle k \rangle}$
- Average degree : $\langle k \rangle = \frac{2m}{n}$
- $\frac{\langle k^2 \rangle}{\langle k \rangle} - \langle k \rangle = \frac{1}{\langle k \rangle} (\langle k^2 \rangle - \langle k \rangle^2) = \frac{\sigma_k^2}{\langle k \rangle} > 0$

Network	n	Average degree	Average neighbor degree	$\frac{\langle k^2 \rangle}{\langle k \rangle}$
Biologists	1520252	15.5	68.4	130.2
Mathematicians	253339	3.9	9.5	13.2
Internet	22963	4.2	224.3	261.5

a vertex of degree k appears as one of the neighbors of exactly k other vertices -> overrepresented

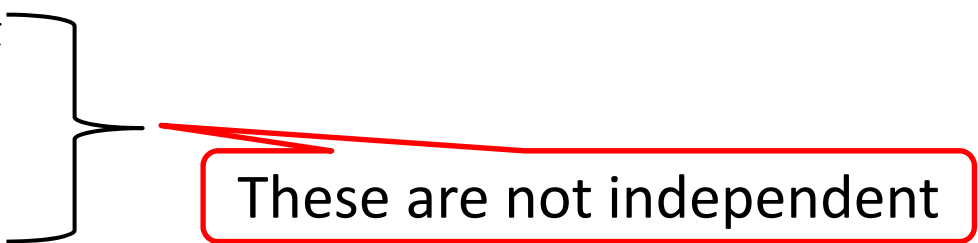
Excess degree



- The number of edges attached to a vertex other than the edge we arrived along
- One less than the total degree
- $q_k = \frac{(k+1)p_{k+1}}{\langle k \rangle}$: excess degree distribution

denominator is still just $\langle k \rangle$ ($\sum_{k=0}^{\infty} q_k = 1$)

Generating functions for degree distributions

- $g_0(z) = \sum_{k=0}^{\infty} p_k z^k$
 - $g_1(z) = \sum_{k=0}^{\infty} q_k z^k$
- 
- These are not independent
- $q_k = \frac{(k+1)p_{k+1}}{\langle k \rangle}$
 - $g_1(z) = \frac{1}{\langle k \rangle} \sum_{k=0}^{\infty} (k+1)p_{k+1} z^k = \frac{1}{\langle k \rangle} \sum_{k=0}^{\infty} k p_k z^{k-1} = \frac{1}{\langle k \rangle} \frac{d g_0}{d z}$
 - $\langle k \rangle = g'_0(1)$
 - $g_1(z) = \frac{g'_0(z)}{g'_0(1)}$

Graphical solution of function u

size of giant component

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

generating function for
degree distribution

- $g_0(z) = \sum_{k=0}^{\infty} p_k z^k$

- $u = 1 - \phi + \phi g_1(u)$

total probability of not
connecting to the giant
component via vertex A

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

generating function for
excess degree distribution

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

excess degree distribution

Graphical solution of function u

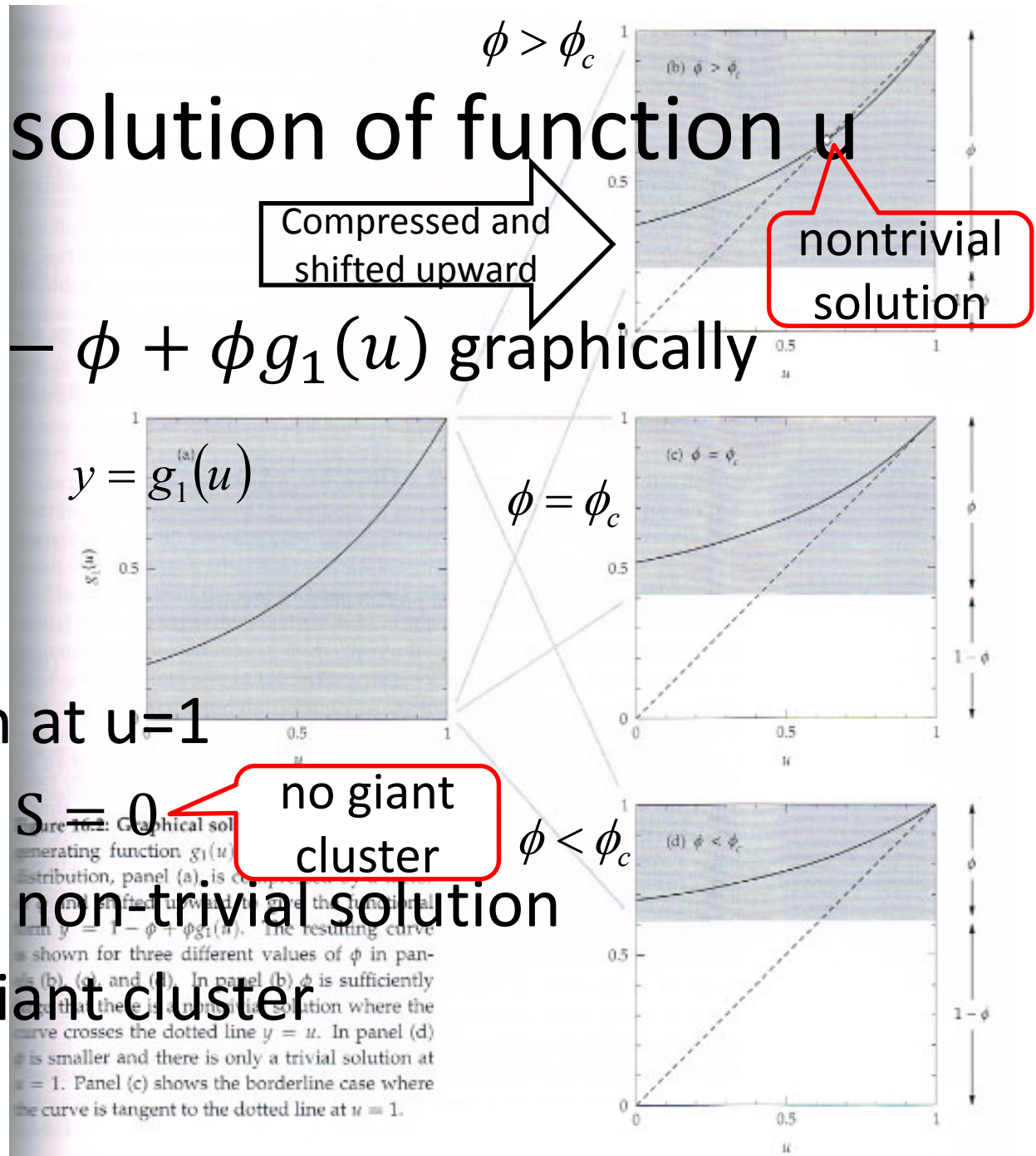
- Solving $u = 1 - \phi + \phi g_1(u)$ graphically

$$\begin{cases} y = 1 - \phi + \phi g_1(u) \\ y = u \end{cases}$$

- Trivial solution at $u=1$

$$-g_1(1) = 1 \rightarrow S = 0$$

- When there is non-trivial solution there can be a giant cluster



Critical occupation probability

curve is tangent to the dotted line at $u=1$

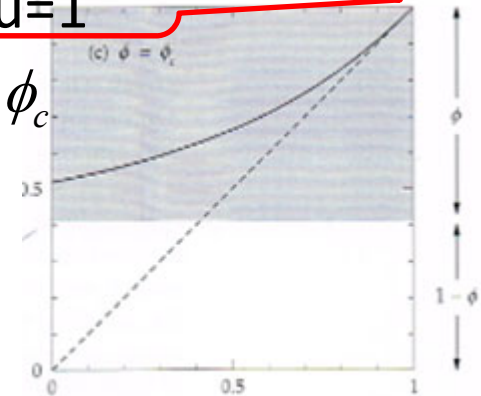
- Percolation threshold

- $\left[\frac{d}{du} (1 - \phi + \phi g_1(u)) \right]_{u=1} = 1$

- $\phi_c = \frac{1}{g'_1(1)}$

- $g'_1(1) = \frac{1}{\langle k \rangle} \sum_{k=0}^{\infty} k(k+1)p_{k+1} = \frac{1}{\langle k \rangle} \sum_{k=0}^{\infty} k(k-1)p_k$
 $= \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle}$

$$\phi = \phi_c$$



- $\phi_c = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}$: minimum fraction of vertices that must be present in the configuration model network for a giant component to exist

if $\langle k^2 \rangle \gg \langle k \rangle$, then ϕ_c is low and the network will have a giant cluster

Giant cluster of Poisson/power-law distribution

- $\phi_c = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}$
- Poisson distribution : $p_k = e^{-c} \frac{c^k}{k!} \rightarrow \phi_c = \frac{1}{c}$
 - $\frac{3}{4}$ of the vertices have to be removed before the giant cluster is destroyed
- Power-law distribution : $\langle k^2 \rangle$ diverges for large networks $\rightarrow \phi_c = 0$
 - No matter how many vertices are removed, there will always be a giant cluster



robust against
random failure

disease spread
out of control



Size of the giant cluster

- As well as percolation threshold, the size of the giant cluster also plays a role for assessing robustness of a network
- We can solve for u for some special cases
- A network with an exponential degree distribution
 - $p_k = (1 - e^{-\lambda})e^{-\lambda k}, \lambda > 0$
 - $g_0(z) = \frac{e^\lambda - 1}{e^\lambda - z}, g_1(z) = \left(\frac{e^\lambda - 1}{e^\lambda - z}\right)^2$ (p.468)
 - $u = 1 - \phi + \phi g_1(u)$
 - $u = 1 - \phi + \phi \left(\frac{e^\lambda - 1}{e^\lambda - u}\right)^2$
 - $u(e^\lambda - u)^2 - (1 - \phi)(e^\lambda - u)^2 - \phi(e^\lambda - 1)^2 = 0$

Size of giant cluster (exponential degree distribution)

- $u(e^\lambda - u)^2 - (1 - \phi)(e^\lambda - u)^2 - \phi(e^\lambda - 1)^2 = 0$
- $u = 1$ is always a solution \rightarrow it must contain a factor of $u - 1$
- $(u - 1)[u^2 + (\phi - 2e^\lambda)u + \phi - 2\phi e^\lambda + e^{2\lambda}] = 0$
- $u = e^\lambda - \frac{1}{2}\phi - \sqrt{\frac{1}{4}\phi^2 + \phi(e^\lambda - 1)}$
 - the other solution is greater than one so it cannot be probability u
- $S = \phi[1 - g_0(u)] = \phi \left[1 - \frac{2(e^\lambda - 1)}{\phi + \sqrt{\phi^2 + 4\phi(e^\lambda - 1)}} \right]$

$$= \phi \left[1 - 2(e^\lambda - 1) \frac{\phi - \sqrt{\phi^2 + 4\phi(e^\lambda - 1)}}{\phi^2 - (\phi^2 + 4\phi(e^\lambda - 1))} \right] =$$

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

This can become greater than 1 for small ϕ

Size of giant cluster (exponential degree distribution)

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

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

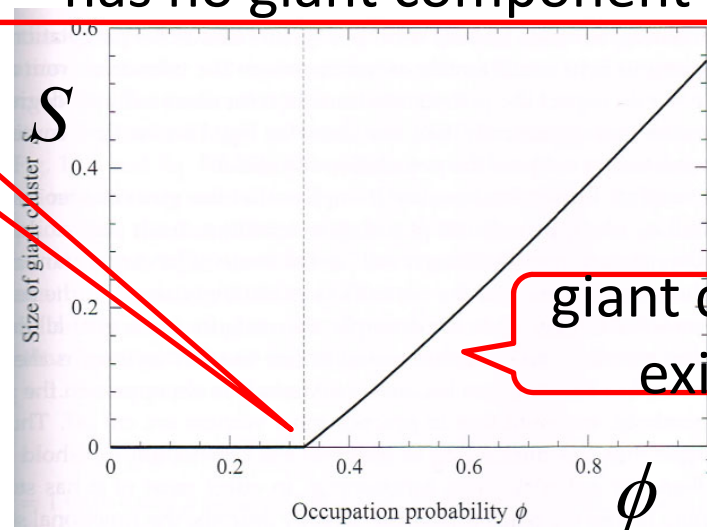
- $\phi_c = \frac{1}{2}(e^\lambda - 1)$ percolation threshold

- If λ becomes large, ϕ_c can become greater than one

- $-\frac{1}{2}(e^\lambda - 1) = 1 \rightarrow \lambda = \ln 3$

When $\lambda > \ln 3$, the network has no giant component

- When $\lambda = 1/2$, $\phi_c = 0.324...$
 - Phase transition
 - Sharp transition is true in an infinite networks



Non uniform removal of vertices

- Vertices are removed randomly (previous discussion)
- High-degree vertices are removed → percolation will be quite different