

Complex Networks

random graphs

2019.1.17(Thu)

Goal

metrics

algorithms

models

processes

contents of this chapter

- random graphs (12)
- models of network formation (14)
 - BA model
- other network models (15)
 - the small-world model
- percolation

network models

- “If I know a network as some particular property, such as a particular degree distribution, what effect will that have on the wider behavior of the system?”
- building mathematical models of networks
 - mimic the patterns of connections in real networks
 - understand the implications of the patterns

random graph

- a model network in which some specific set of parameters take fixed values, but the network is random in other respects
- simplest example: $G(n, m)$
 - take n vertices and place m edges at random
 - simple graph (no multiedges or self-edges)

⇒ a probability distribution $P(G)$ over possible networks

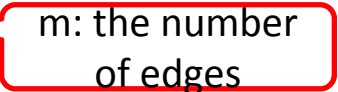
$$P(G) = \begin{cases} 1/\Omega & \text{if } G \text{ is a simple graph with } n \text{ vertices and } m \text{ edges} \\ 0 & \text{otherwise} \end{cases}$$

Ω : the total number of simple graphs with n vertices and m edges

random graph model = an ensemble of networks

- properties of random graphs = the average properties of the ensemble
- diameter of $G(n,m)$: $\langle l \rangle = \sum_G P(G) l(G) = \frac{1}{\Omega} \sum_G l(G)$
- this is a useful definition for several reasons:
 - many average properties can be calculated exactly
 - we are often interested in typical properties of the networks
 - distributions of values for many network measures is sharply peaked
- average degree : $\langle k \rangle = 2m/n$

another random graph model

- $G(n,p)$
 - n : the number of vertices
 - p : the probability of edges between vertices
- $G(n,p)$ is the ensemble of networks with n vertices in which each simple G appears with probability $P(G) = p^m (1-p)^{\binom{n}{2}-m}$
 - m: the number of edges
- often called as “Erdos-Renyi model”, “Poisson random graph”, “Bernoulli random graph”, or “the random graph”

mean number of edges and mean degree of $G(n,p)$

- total probability of drawing a graph with m edges from the ensemble is

$$P(m) = \binom{\binom{n}{2}}{m} p^m (1-p)^{\binom{n}{2}-m}$$

the expected value of binomially distributed variable (see the next page)

- the mean value of m is $\langle m \rangle = \sum_{m=0}^{\binom{n}{2}} m P(m) = \binom{n}{2} p$
- the mean degree is $\langle k \rangle = \sum_{m=0}^{\binom{n}{2}} \frac{2m}{n} P(m) = \frac{2}{n} \binom{n}{2} p = (n-1)p$

often denoted as c

$$c = (n-1)p$$

$$n(n-1)/2$$

the expected value of binomially distributed variable

$$\langle m \rangle = \sum_{m=0}^{\binom{n}{2}} m P(m) = \sum_{m=1}^{\binom{n}{2}} m P(m) = \sum_{m=1}^{\binom{n}{2}} m \binom{\binom{n}{2}}{m} p^m (1-p)^{\binom{n}{2}-m}$$

$0 \cdot P(0) = 0$

$$\binom{N}{k} = \frac{N!}{k!(N-k)!} = \frac{N \cdot (N-1)!}{k \cdot (k-1)!((N-1)-(k-1))!}$$

$$= \frac{N}{k} \cdot \frac{(N-1)!}{(k-1)!((N-1)-(k-1))!} = \frac{N}{k} \cdot \binom{N-1}{k-1}$$

$N = \binom{n}{2}, k = m$

$$\langle m \rangle = \sum_{m=1}^{\binom{n}{2}} m \binom{\binom{n}{2}}{m} p^m (1-p)^{\binom{n}{2}-m} = \binom{n}{2} p \sum_{m=1}^{\binom{n}{2}} \binom{\binom{n}{2}-1}{m-1} p^{m-1} (1-p)^{\binom{n}{2}-m}$$

$j = m-1$

$$\langle m \rangle = \binom{n}{2} p \sum_{j=0}^{\binom{n}{2}-1} \binom{\binom{n}{2}-1}{j} p^j (1-p)^{\binom{n}{2}-1-j} = \binom{n}{2} p (p + (1-p))^{\binom{n}{2}-1} = \binom{n}{2} p$$

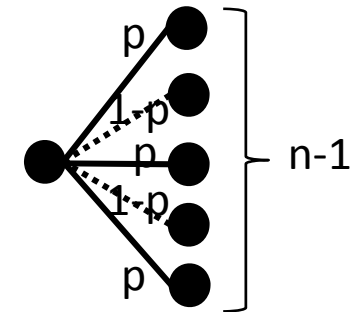
binomial theorem

1

degree distribution of $G(n,p)$ (1)

- a vertex is connected with probability p to each of the $n-1$ other vertices

$$p_k = \binom{n-1}{k} p^k (1-p)^{n-1-k}$$



- we are interested in large networks
 - a mean degree is approximately constant
 - $p = c/(n-1)$ becomes vanishingly small as $n \rightarrow \infty$

$$\ln[(1-p)^{n-1-k}] = (n-1-k) \ln\left(1 - \frac{c}{n-1}\right) \cong -(n-1-k) \frac{c}{n-1} \cong -c$$

$$(1-p)^{n-1-k} = e^{-c}$$

for large n

$$\begin{aligned} \ln(1+x) &= \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} x^n \quad \text{for all } |x| < 1 \\ &= x - \frac{1}{2} x^2 + \frac{1}{3} x^3 - \dots \end{aligned} \quad \text{Taylor expansion}$$

degree distribution of $G(n,p)$ (2)

- for large n

$$\binom{n-1}{k} = \frac{(n-1)!}{(n-1-k)!k!} \cong \frac{(n-1)^k}{k!}$$

$(n-1)(n-2)\dots(n-k)/k!$

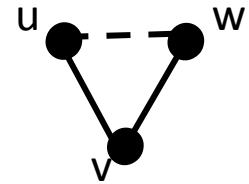
- p_k becomes as follows in the limit of large n

$$p_k = \binom{n-1}{k} p^k (1-p)^{n-1-k} = \frac{(n-1)^k}{k!} p^k e^{-c} = \frac{(n-1)^k}{k!} \left(\frac{c}{n-1}\right)^k e^{-c} = e^{-c} \frac{c^k}{k!}$$

Poisson distribution

clustering coefficient of $G(n,p)$

- clustering coefficient : the probability that two network neighbors of a vertex are also neighbors of each other



- in a random graph, the probability is $p = c/(n-1)$

$$C = \frac{c}{n-1}$$

- tends to zero in the limit $n \rightarrow \infty$
- differs sharply from most of the real-world networks (quite high clustering coefficient)

Giant component (1)

- the size of the largest component in a network

- $p=0 \rightarrow \text{size}=1$

independent of the size of the network

- $p=1 \rightarrow \text{size}=n$

it will grow with the network

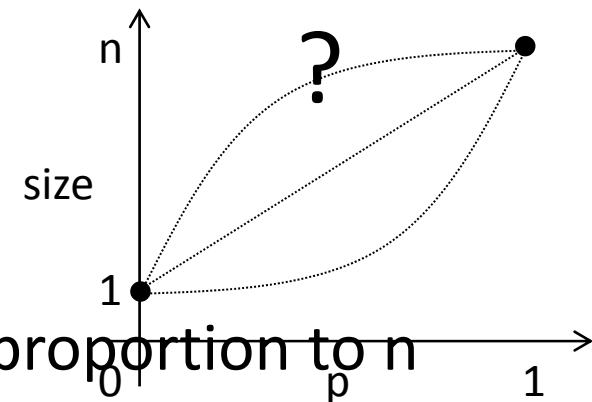
- giant component

- a component whose size grows in proportion to n

- u : average fraction of vertices that do not belong to the giant component

- $u = 1$ if there is no giant component

- u is the probability that a randomly chosen vertex does not belong to the giant component



Giant component (2)

- if vertex i does not belong to the giant component, for every other vertex j
 - i is not connected to j by an edge probability: $1-p$
 - i is connected to j , but j is not a member of the giant component probability: puprobability of not being connected to g.c. via j : $1-p+pu$
- total probability of not being connected to g.c. via any of $n-1$ vertices:

$$u = (1 - p + pu)^{n-1} = \left[1 - \frac{c}{n-1} (1-u) \right]^{n-1}$$

Giant component (3)

- taking logs of both sides

$$\ln u = (u - 1) \ln \left[1 - \frac{c}{n-1} (1-u) \right] \cong -(n-1) \frac{c}{n-1} (1-u) = -c(1-u)$$

- taking exponentials of both sides

$$u = e^{-c(1-u)}$$

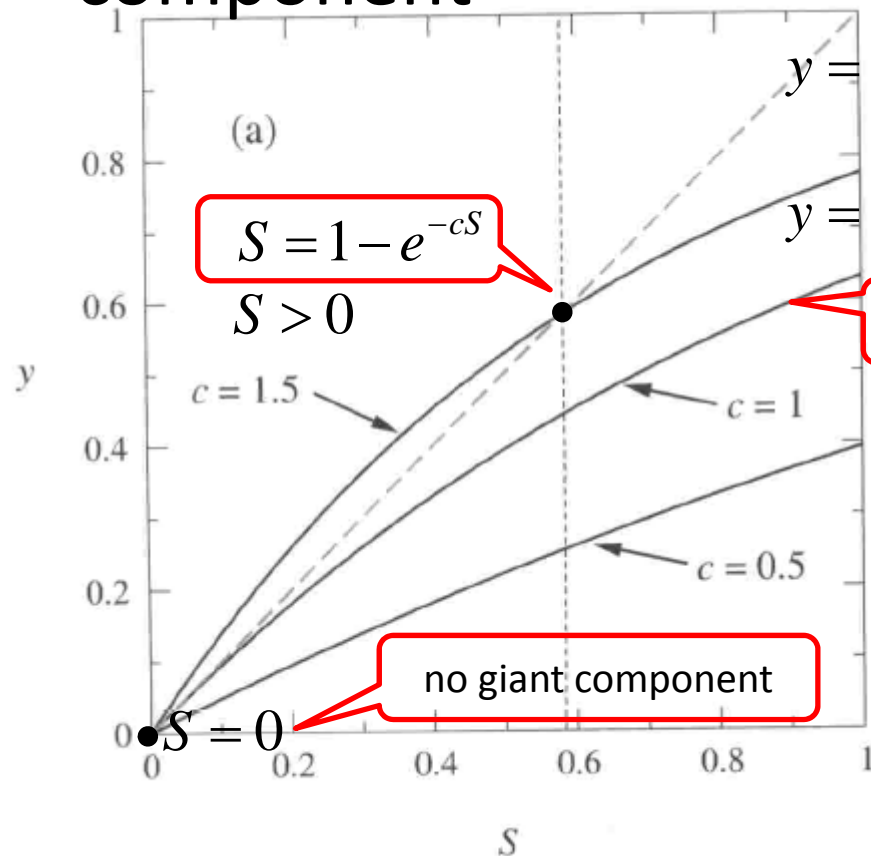
- u is the fraction of vertices not in the giant component
- the fraction of vertices that are in the giant component is $S = 1 - u$

$$S = 1 - e^{-cS}$$

it doesn't have a simple solution for S

Giant component (4)

- graphical solution for the size of the giant component



$$\frac{d}{dS}(1 - e^{-cS}) = 1$$

$$ce^{-cS} = 1$$

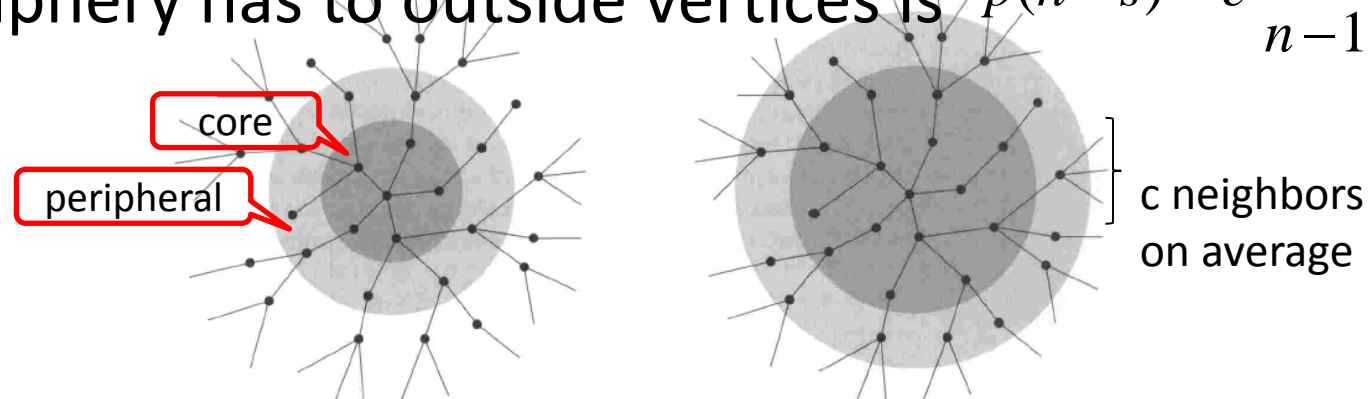
$$S = 0 \rightarrow c = 1$$

if $c \leq 1$, no giant component

if $c > 1$, two solutions for S ($S=0$ & $S>0$)

the value of c and the growth of a set (1)

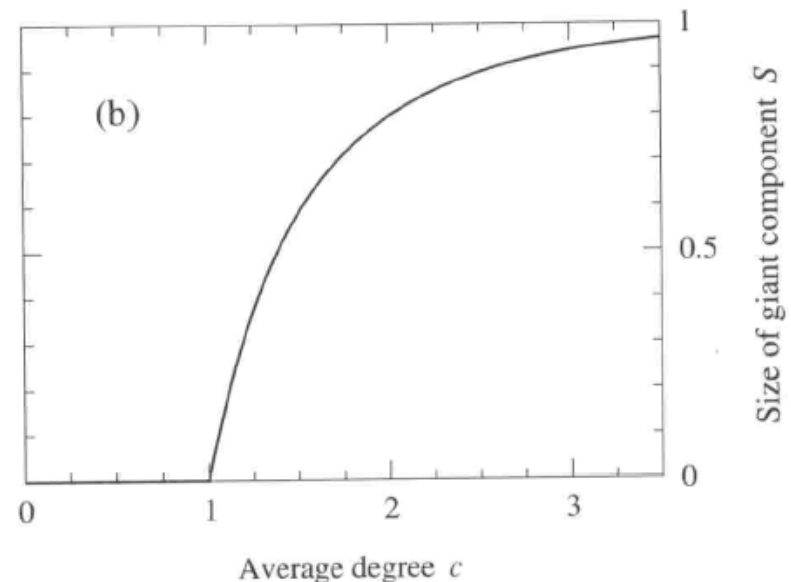
- core : all neighbor are inside the set
- peripheral : at least one neighbor is outside
- enlarging the set by adding immediate neighbor
 - s vertices in the set, $n-s$ vertices outside the set
 - the average number of connections a vertex in the periphery has to outside vertices is $p(n-s) = c \frac{n-s}{n-1} \cong c$



the value of c and the growth of a set

(2)

- each peripheral has c neighbors outside
 - (# of new peripheral) = $c \times$ (# of old peripheral)
- if $c > 1$, the average size of the periphery will grow exponentially \rightarrow giant component
- the size of the giant component is the larger solution of $S = 1 - e^{-cS}$



small components (1)

- when $c > 1$, there exist a giant component
- what is the structure of the remainder of the network?
 - it is made up of many small components whose average size is constant and doesn't increase with the size of the network
- there is only one giant component
 - suppose there are two giant components which have size $S_1 n$ and $S_2 n$
 - the number of distinct pairs of vertices between the two is $S_1 S_2 n^2$
 - the probability that there is no edge between the two component is
is
$$q = (1 - p)^{S_1 S_2 n^2} = \left(1 - \frac{c}{n-1}\right)^{S_1 S_2 n^2}$$

small components (2)

- taking logs of both sides and going to the limit

$$\begin{aligned}
 \ln q &= S_1 S_2 \lim_{n \rightarrow \infty} \left[n^2 \ln \left(1 - \frac{c}{n-1} \right) \right] = S_1 S_2 \left[-c(n+1) + \frac{1}{2} c^2 \right] \\
 &= c S_1 S_2 \left[-n + \left(\frac{1}{2} c - 1 \right) \right]
 \end{aligned}$$

Taylor expansion

$$\begin{aligned}
 \ln(1+x) &= \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} x^n \quad \text{for all } |x| < 1 \\
 &= x - \frac{1}{2} x^2 + \frac{1}{3} x^3 - \dots
 \end{aligned}$$

- taking the exponential again

$$q = q_0 e^{-c S_1 S_2 n}$$

$$q_0 = e^{c(c/2-1)S_1 S_2}$$

constant

- the probability that the two giant components are really components dwindles exponentially with increasing n

sizes of small components (1)

- π_s : probability that a randomly chosen vertex belongs to a small component of size s

$$\sum_{s=0}^{\infty} \pi_s = 1 - S$$


- small components are trees

- a tree of s vertices contains $s-1$ edges

- the total number of places where we could add an extra edge to the tree: $\binom{s}{2} - (s-1) = \frac{1}{2}(s-1)(s-2)$

- the total number of extra edges :

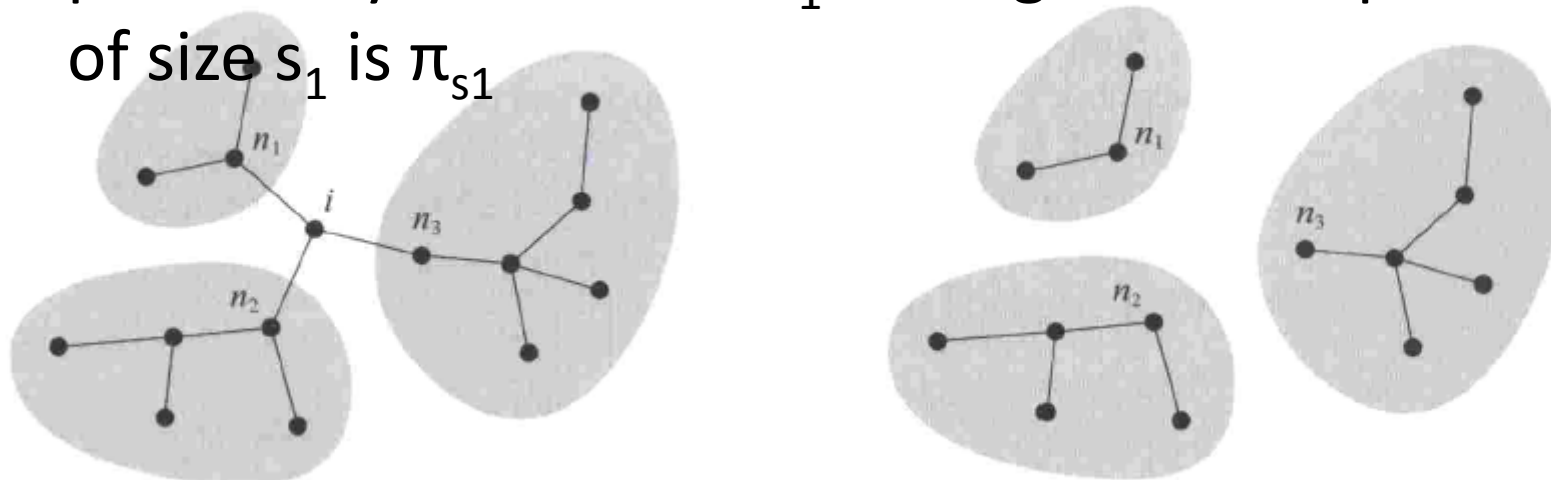
$$\frac{1}{2}(s-1)(s-2)p = \frac{1}{2}(s-1)(s-2)\frac{c}{n-1}$$

- s increases more slowly than \sqrt{n}  extra edges $\rightarrow 0$



sizes of small components (2)

- because the component is a tree,
 - (size of the component) = $\sum(\text{size of } n_i) + 1$
- if vertex i is removed, the subcomponents become components in their own right
 - probability that vertex n_1 belongs to a component of size s_1 is π_{s1}



size of small components (3)

- suppose that vertex i has degree k
- probability $P(s | k)$ that vertex i belongs to small component of size s is

$$P(s | k) = \sum_{s_1=1}^{\infty} \dots \sum_{s_k=1}^{\infty} \left[\prod_{j=1}^k \pi_{s_j} \right] \delta(s-1, \sum_j s_j)$$

- to get π_s , just average $P(s | k)$ over the distribution p_k of the degree

$$\begin{aligned} \pi_s &= \sum_{k=1}^{\infty} p_k P(s | k) = \sum_{k=0}^{\infty} p_k \sum_{s_1=1}^{\infty} \dots \sum_{s_k=1}^{\infty} \left[\prod_{j=1}^k \pi_{s_j} \right] \delta(s-1, \sum_j s_j) \\ &= e^{-c} \sum_{k=0}^{\infty} \frac{c^k}{k!} \sum_{s_1=1}^{\infty} \dots \sum_{s_k=1}^{\infty} \left[\prod_{j=1}^k \pi_{s_j} \right] \delta(s-1, \sum_j s_j) \end{aligned}$$

$\because p_k = e^{-c} \frac{c^k}{k!}$
 in the limit of large k

size of small components (4)

- generating function encapsulates all of the information about the degree distribution in a single function

$$h(z) = \pi_1 z + \pi_2 z^2 + \pi_3 z^3 + \dots = \sum_{s=1}^{\infty} \pi_s z^s$$

- we can recover the probabilities by differentiating $\pi_s = \frac{1}{s!} \left. \frac{d^s h}{dz^s} \right|_{z=0}$
- substituting π_s into the equation of $h(z)$

$$h(z) = \sum_{s=1}^{\infty} z^s e^{-c} \sum_{k=0}^{\infty} \frac{c^k}{k!} \sum_{s_1=1}^{\infty} \dots \sum_{s_k=1}^{\infty} \left[\prod_{j=1}^k \pi_{s_j} \right] \delta(s-1, \sum_j s_j)$$

size of small components (5)

$$\begin{aligned}
 h(z) &= \sum_{s=1}^{\infty} z^s e^{-c} \sum_{k=0}^{\infty} \frac{c^k}{k!} \sum_{s_1=1}^{\infty} \dots \sum_{s_k=1}^{\infty} \left[\prod_{j=1}^k \pi_{s_j} \right] \delta(s-1, \sum_j s_j) \\
 &= e^{-c} \sum_{k=0}^{\infty} \frac{c^k}{k!} \sum_{s_1=1}^{\infty} \dots \sum_{s_k=1}^{\infty} \left[\prod_{j=1}^k \pi_{s_j} \right] z^{1+\sum_j s_j} \\
 &= z e^{-c} \sum_{k=0}^{\infty} \frac{c^k}{k!} \sum_{s_1=1}^{\infty} \dots \sum_{s_k=1}^{\infty} \left[\prod_{j=1}^k \pi_{s_j} z^{s_j} \right] \\
 &= z e^{-c} \sum_{k=0}^{\infty} \frac{c^k}{k!} \left[\sum_{s=1}^{\infty} \pi_s z^s \right]^k = z e^{-c} \sum_{k=0}^{\infty} \frac{c^k}{k!} [h(z)]^k = z \exp[c(h(z)-1)]
 \end{aligned}$$

$\delta=1$ only when $s-1=\sum_j s_j$

$\because \exp x = \sum_{n=0}^{\infty} \frac{1}{n!} x^n$

- it doesn't have a known closed-form solution for $h(z)$, but we can calculate many useful things from it without solving for $h(z)$ explicitly

size of small components (6)

- mean size of the component to which a randomly chosen vertex belongs

$$\langle s \rangle = \frac{\sum_s s \pi_s}{\sum_s \pi_s} = \frac{h'(1)}{1-S}$$

$h'(z)$: the first derivative of $h(z)$

- from the equation of $h(z)$

$$h'(z) = \exp[c(h(z)-1)] + cz h'(z) \exp[c(h(z)-1)] = \frac{h(z)}{z} + ch(z)h'(z)$$

$$h'(z) = \frac{h(z)}{z[1-ch(z)]}$$

$$h(1) = \sum_s \pi_s = 1-S$$

$$h'(1) = \frac{h(1)}{1-ch(1)} = \frac{1-S}{1-c+cS}$$

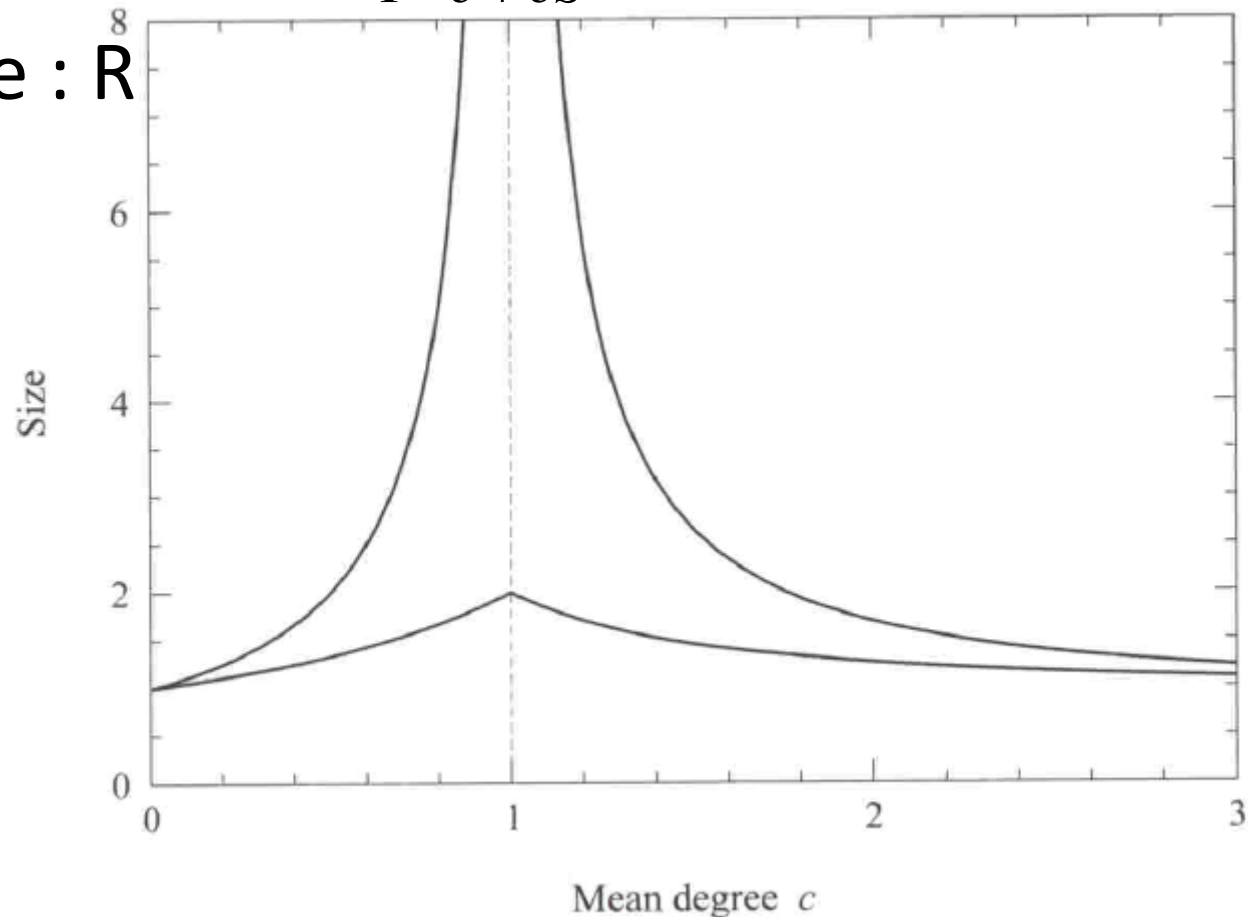
$$\langle s \rangle = \frac{1}{1-c+cS}$$

it doesn't grow with the number of vertices n

when $c < 1$ and there is no giant component,
 $\langle s \rangle = 1/(1-c)$

divergence of the average size $\langle s \rangle$

- upper curve : $\langle s \rangle = \frac{1}{1-c+cS}$ diverges when $c=1$
- lower curve : R



average size of a small component

- $\langle s \rangle$: the average size of the component to which a randomly chosen vertex belongs
 \neq average size of a component
- n_s : the actual number of components of size s
- sn_s : the number of vertices that belong to components of size s
- the probability that a randomly chosen vertex belongs to a component of size s is $\pi_s = \frac{sn_s}{n}$

average size of a small component

- R: average size of a component

$$R = \frac{\sum_s s n_s}{\sum_s n_s} = \frac{n \sum_s \pi_s}{n \sum_s \pi_s / s} = \frac{1 - S}{\sum_s \pi_s / s}$$

$$\int_0^1 \frac{h(z)}{z} dz = \sum_{s=1}^{\infty} \pi_s \int_0^1 z^{s-1} dz = \sum_{s=1}^{\infty} \frac{\pi_s}{s}$$

$$h(z) = \sum_{s=1}^{\infty} \pi_s z^s$$

$$\frac{h(z)}{z} = [1 - ch(z)] \frac{dh}{dz} \quad \because h'(z) = \frac{h(z)}{z[1 - ch(z)]}$$

$$\sum_{s=1}^{\infty} \frac{\pi_s}{s} = \int_0^1 [1 - ch(z)] \frac{dh}{dz} dz = \int_0^{1-S} (1 - ch) dh = 1 - S - \frac{1}{2} c(1 - S)^2$$

$$\because h(1) = \sum_s \pi_s = 1 - S$$

$$\therefore R = \frac{2}{2 - c + cS}$$

it does not
diverge at c=1

the complete distribution of component sizes

- p.416

path lengths (1)

- small world effect : typical length of paths between vertices in network tend to be short
- the diameter of a random graph varies with the number n of vertices as $\ln n$
 - the average number of vertices s steps away from a randomly chosen vertex in a random graph is c^s
 - it grows exponentially with s $c^s \cong n$
 - diameter of the network is approximately $s \cong \ln n / \ln c$
- this argument is true when c^s is much less than n

path lengths (2)

- two different starting vertices (i and j)
- if there is a dashed line between the surfaces, the shortest path between i and j is $s+t+1$
- the absence of an edge between the surfaces is a necessary and sufficient condition for $d_{ij} > s+t+1$

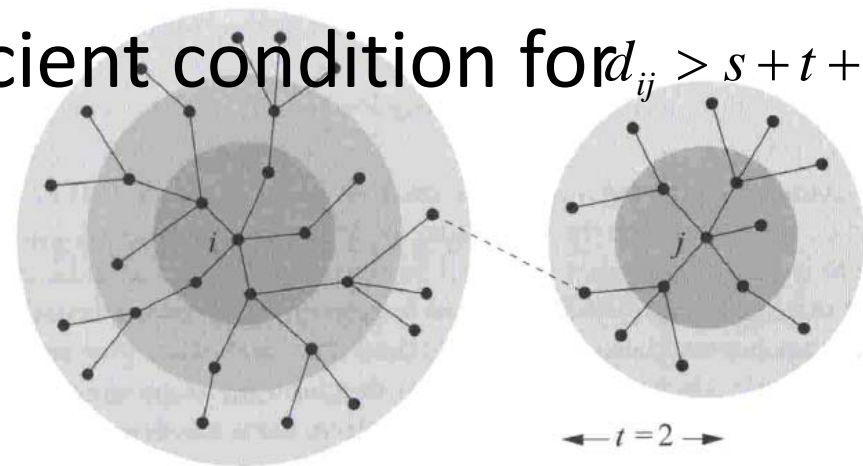
- $c^s \times c^t$ pairs of vertices

$$P(d_{ij} > s+t+1) = (1-p)^{c^{s+t}}$$

$$l = s+t+1$$

$$P(d_{ij} > l) = (1-p)^{c^{l-1}} = \left(1 - \frac{c}{n}\right)^{c^{l-1}}$$

$$\ln P(d_{ij} > l) = c^{l-1} \ln \left(1 - \frac{c}{n}\right) \cong -\frac{c^l}{n}$$



$$\ln(1+x) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} x^n \quad \text{for all } |x| < 1$$

path length (3)

$$P(d_{ij} > l) = \exp\left(-\frac{c^l}{n}\right)$$

tend to zero only if c^l grows faster than n

$c^l = an^{1+\varepsilon}$
 $\varepsilon \rightarrow 0$

- diameter : the smallest value of l s.t. $P(d_{ij} > l) = 0$

$$l = \frac{\ln a}{\ln c} + \lim_{\varepsilon \rightarrow 0} \frac{(1 + \varepsilon) \ln n}{\ln c} = A + \frac{\ln n}{\ln c}$$

diameter increases slowly with n

- logarithmic dependence of the diameter on n
 - acquaintance network of the entire world (7 billion people)

$$l = \frac{\ln n}{\ln c} = \frac{\ln(7 \times 10^9)}{\ln 1000} = 3.3..$$

small enough to account for the results of the small-world experiments of Milgram

problems with the random graph (1)

- no transitivity or clustering
 - $C = \frac{c}{n-1}$ tends to zero in the limit of large n
random graph rarely contains triangles
 - the acquaintance network of the human population in the world
 - $n \cong 7,000,000,000$
 - $C \cong \frac{1000}{7,000,000,000} \cong 10^{-7}$
- no correlation between the degrees of adjacent vertices (no communities)

clustering coefficient of
real acquaintance
network is much bigger
(0.01 or 0.5)

problems with the random graph (2)

- the shape of degree distribution is different
 - real network : right-skewed

