Random graph

Network models

Empirical network features:

- Power-law (heavy-tailed) degree distribution
- Small average distance (graph diameter)
- Large clustering coefficient (transitivity)
- Giant connected component, hierarchical structure,etc

Generative models:

- Random graph model (Erdos & Renyi, 1959)
- "Small world" model (Watts & Strogatz, 1998)
- Preferential attachement model (Barabasi & Albert, 1999)

Random graph model

Graph $G\{E, V\}$, nodes $n = |V|$, edges $m = |E|$
Erdos and Renyi, 1959.
Random graph models

- $G_{n,m}$, a randomly selected graph from the set of $C_N^m graphs, N = \frac{n(n-1)}{2}$, with n nodes and m edges
- $G_{n,p}$, each pair out of $N = \frac{n(n-1)}{2}$ pairs of nodes is connected with probability p, m - random number

$$\langle m \rangle = p\frac{n(n-1)}{2}$$

Average degree $$\langle k \rangle = \frac{1}{n}\sum_i k_i = \frac{2\langle m \rangle}{n} = p\ (n-1) \approx pn$$

Graph density $$\rho = \frac{\langle m \rangle}{n(n-1)/2} = p$$

Random graph model

- Probability that i-th node has a degree $k_i = k$

$$P(k_i = k) = P(k) = C_{n-1}^{k} p^k (1-p)^{n-1-k}$$

Binomial distribution

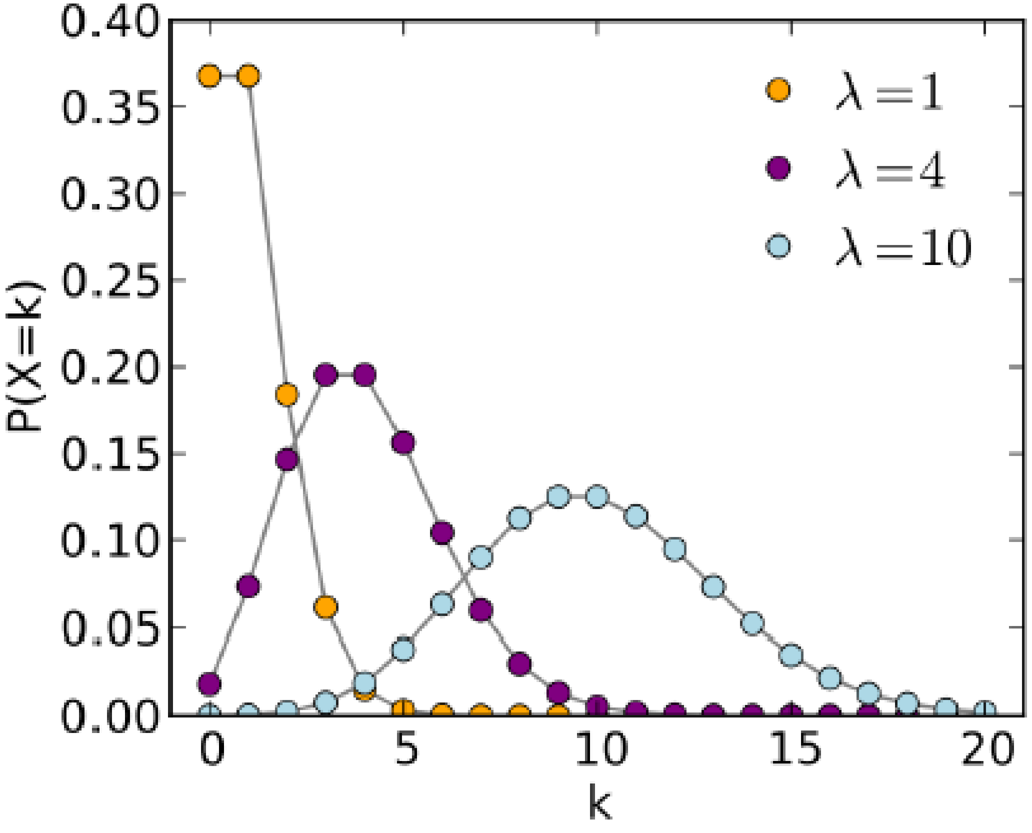
(Bernoulli distribution)
p^k - probability that connects to k nodes (has k-edges)
$(1-p)^{n-k-1}$ - probability that does not connect to any other node
C_{n-1}^{k} - number of ways to select k nodes out of all to connect to

- Limiting case of Bernoulli distribution, when $n \to \infty$ at fixed $\langle k \rangle = pn = \lambda$

$$P(k) = \frac{\langle k \rangle^k e^{-\langle k \rangle}}{k!} = \frac{\lambda^k e^{-\lambda}}{k!}$$

(Poisson distribution)

Poisson distribution



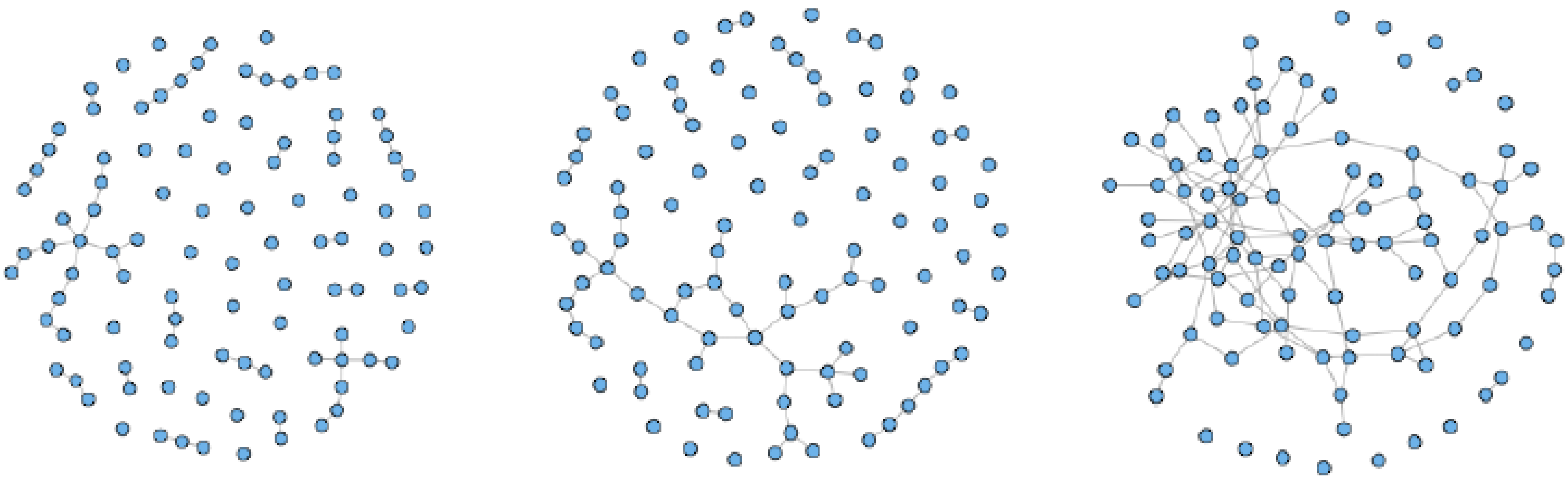
$$P(k_i = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad \lambda = pn$$

Phase transition

Consider $G_{n,p}$ as a function of p

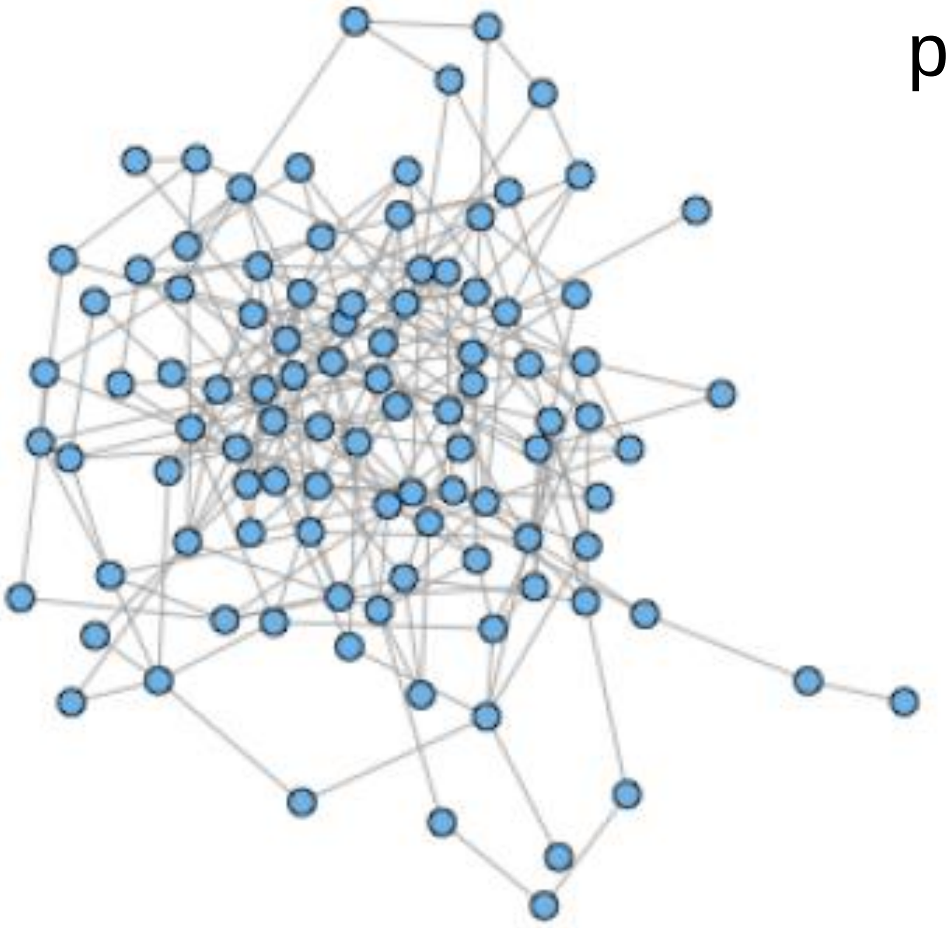
- $p = 0$, empty graph
- $p = 1$, complete (full) graph
- There are exist critical p_c, structural changes from $p < p_c$ to $p > p_c$
- Gigantic connected component appears at $p > p_c$

Random graph model



$p < p_c$ $\qquad$ $p = p_c$ $\qquad$ $p > p_c$

Random graph model



p>>pc

Phase transition

Let u – fraction of nodes that do not belong to GCC. The probability that a node does not belong to GCC

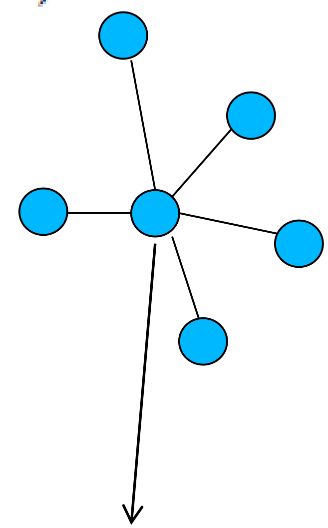
$$u = P(k=1)\cdot u + P(k=2)\cdot u^2 + P(k=3)\cdot u^3 \ldots =$$

$$= \sum_{k=0}^{\infty} P(k) u^k = \sum_{k=0} \frac{\lambda^k e^{-\lambda}}{k!} u^k = e^{-\lambda} e^{\lambda u} = e^{\lambda(u-1)}$$

Let s -fraction of nodes belonging to GCC (size of GCC)

$$s = 1 - u$$

$$1 - s = e^{-\lambda s}$$

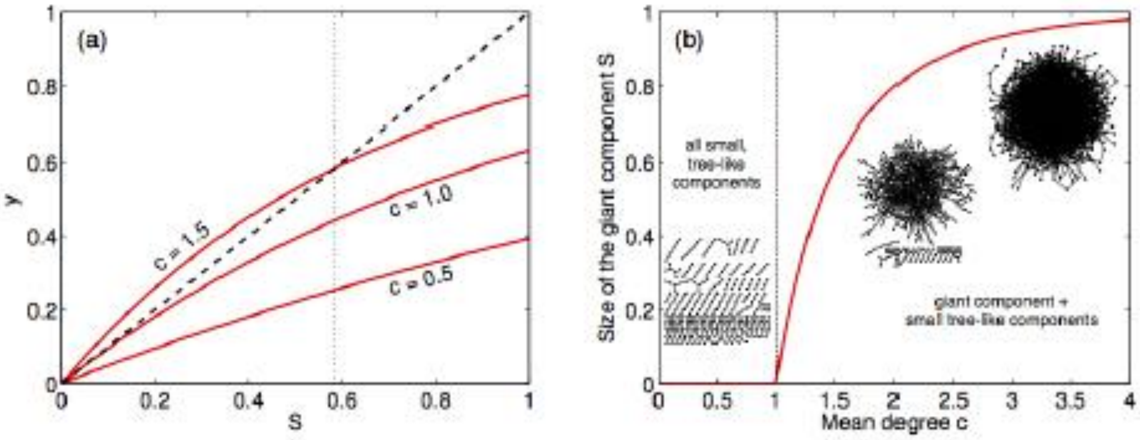


when $\lambda \to \infty, \quad s \to 1$
when $\lambda \to 0, \quad s \to 0$
$(\lambda = pn)$

This node does not belong
to GCC , if k neighbors do not
participate

Phase transition

$$s = 1 - e^{-\lambda s}$$



non-zero solution exists when (at $s = 0$):

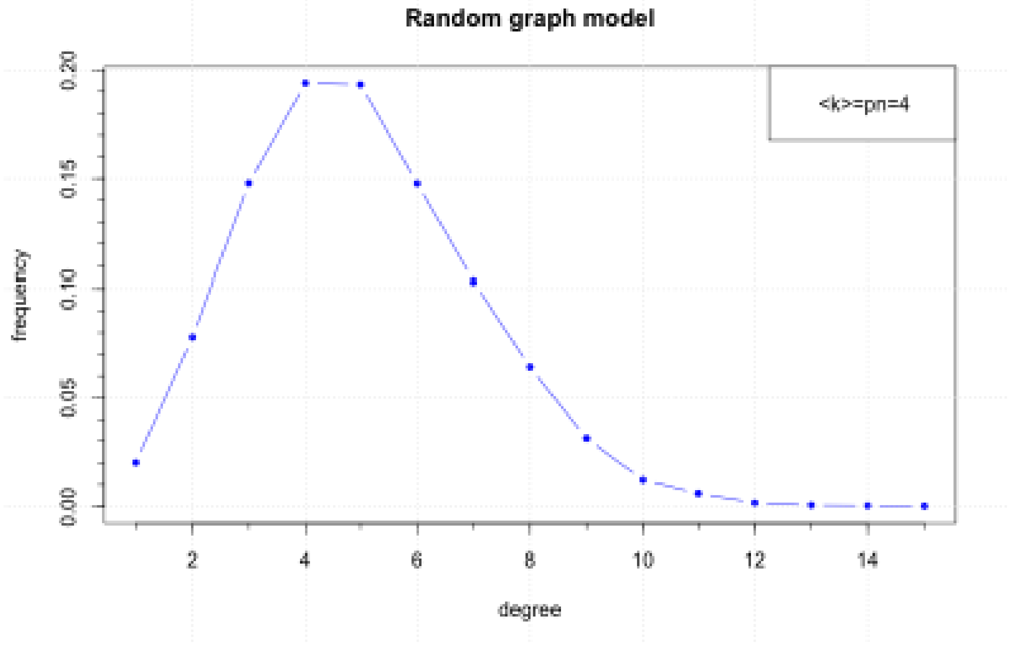
$$\lambda e^{-\lambda s} > 1$$

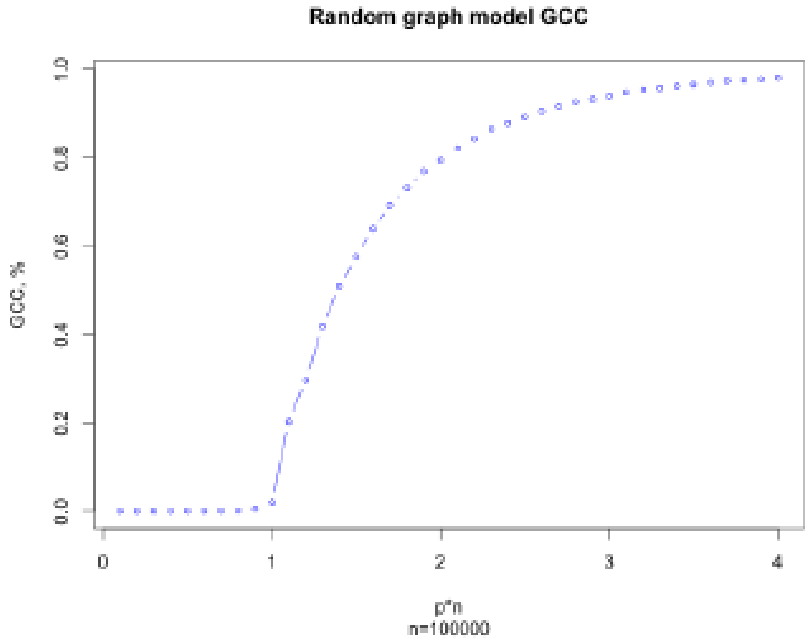
critical value:

$$\lambda_c = 1$$

$$\lambda_c = \langle k \rangle = p_c n = 1, \quad p_c = \frac{1}{n}$$

Simulations





$$\langle k \rangle = pn$$

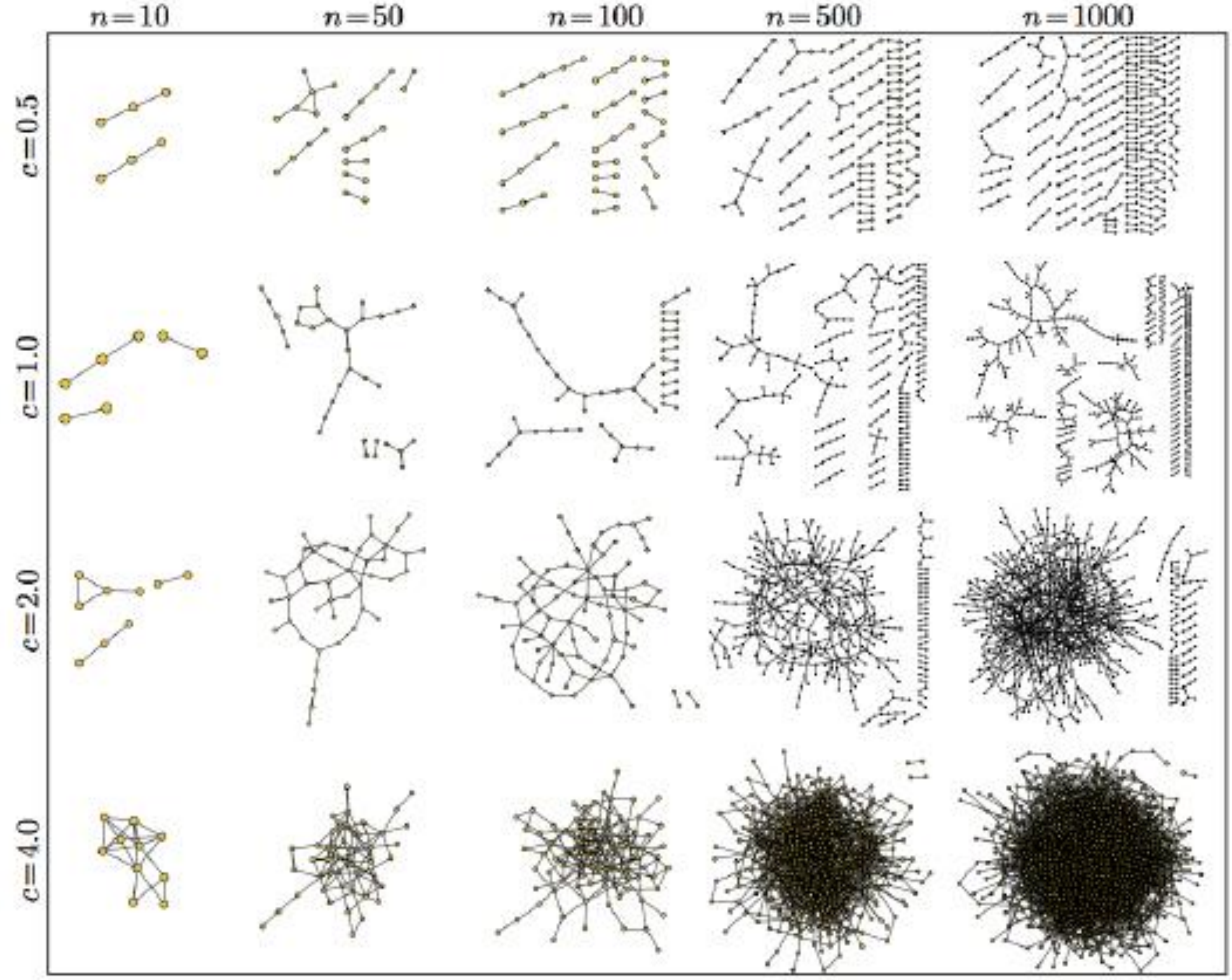
Phase transitions

Graph $G(n, p)$, for $n \to \infty$, critical value $p_c = 1/n$

- when $p < p_c$, $(\langle k \rangle < 1)$ there is no components with more than $O(\ln n)$ nodes, largest component is a tree
- when $p = p_c$, $(\langle k \rangle = 1)$ the largest component has $O(n^{2/3})$ nodes
- when $p > p_c$, $(\langle k \rangle > 1)$ gigantic component has all $O(n)$ nodes

Critical value: $\langle k \rangle = p_c n = 1$- on average one neighbor for a node

Phase transitions



Clustering coefficient

- Clustering coefficient

$$C(k) = \frac{\#\text{of links between NN}}{\#\text{max number of links NN}} = \frac{pk(k-1)/2}{k(k-1)/2} = p$$

$$C = p = \frac{\langle k \rangle}{n}$$

- when $n \to \infty, \quad C \to 0$

Graph diameter

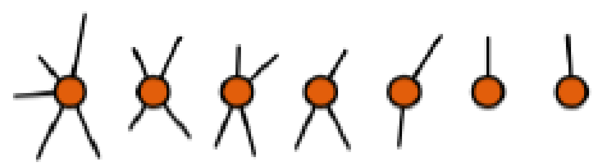
- $G(n,p)$ is locally tree-like (GCC) (no loops; low clustering coefficient)

- on average, the number of nodes d steps away from a node $\langle k \rangle^d$
- in GCC, around p_c , $\langle k \rangle^d \sim n$,

$$d \sim \frac{\ln n}{\ln \langle k \rangle}$$

General degree distribution Configuration model

- Random graph with n nodes with a given degree sequence: $D = \{k_1, k_2, k_3..k_n\}$ and $m = 1/2\sum_i k_i$ edges.
- Construct by randomly matching two stubs and connecting them by an edge.



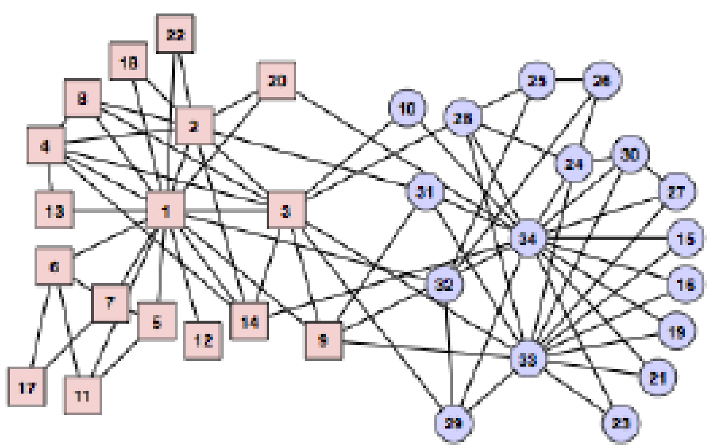
- Can contain self loops and multiple edges
- Probability that two nodes i and j are connected

$$p_{ij} = \frac{k_i k_j}{2m-1}$$

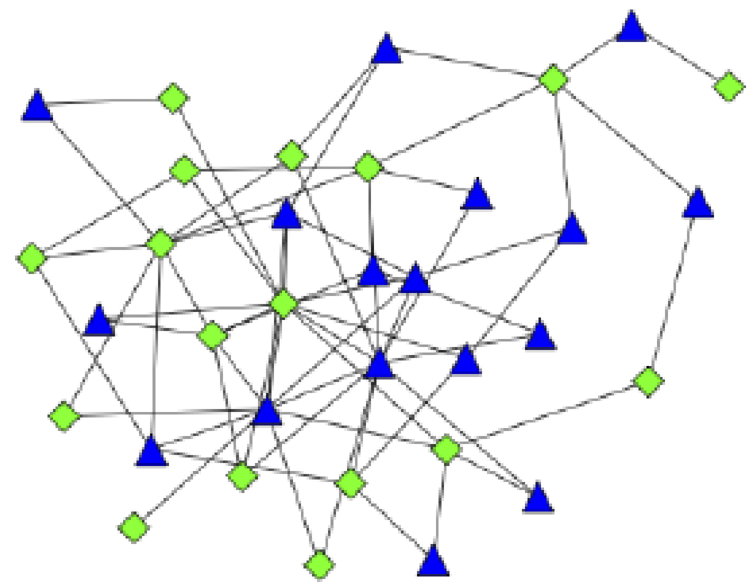
- Will be a simple graph for special "graphical degree sequence"

Configuration model

Can be used as a "null model" for comparative network analysis



karate club



configuration model

Generating function

A. Generating functions

Our approach is based on generating functions [42], the most fundamental of which, for our purposes, is the generating function $G_0(x)$ for the probability distribution of vertex degrees k. Suppose that we have a unipartite undirected graph—an acquaintance network, for example—of N vertices, with N large. We define

$$G_0(x)=\sum_{k=0}^{\infty} p_k x^k, \tag{2}$$

where p_k is the probability that a randomly chosen vertex on the graph has degree k. The distribution p_k is assumed correctly normalized, so that

$$G_0(1)=1. \tag{3}$$

in a community of 1000 people, each person knows between zero and five of the others, the exact numbers of people in each category being, from zero to five: $\{86,150,363,238,109,54\}$. This distribution will then be generated by the polynomial

$$G_0(x) = \frac{86 + 150x + 363x^2 + 238x^3 + 109x^4 + 54x^5}{1000}. \tag{24}$$

Generating function

Derivatives. The probability p_k is given by the kth derivative of G_0 according to

$$p_k = \frac{1}{k!} \frac{d^k G_0}{dx^k}\bigg|_{x=0}. \tag{4}$$

Thus the one function $G_0(x)$ encapsulates all the information contained in the discrete probability distribution p_k. We say that the function $G_0(x)$ "generates" the probability distribution p_k.

Moments. The average over the probability distribution generated by a generating function—for instance, the average degree z of a vertex in the case of $G_0(x)$—is given by

$$z = \langle k \rangle = \sum_k k p_k = G_0'(1).$$

$$\langle k^n \rangle = \sum_k k^n p_k = \left[\left(x \frac{d}{dx} \right)^n G_0(x) \right]_{x=1}$$

$$\langle k^2 \rangle = \sum k^2 p_k$$

$$= 1^2 p_1 + 2^2 p_2 + 3^2 p_3 + 4^2 p_4 + \cdots$$

$$G_0(x) = \sum p_k x^k = p_0 + p_1 x + p_2 x^2 + p_3 x^3 + p_4 x^4 + \cdots$$

$$\Rightarrow G_0'(x) = p_1 + 2p_2 x + 3p_3 x^2 + 4p_4 x^3 + \cdots$$

$$\Rightarrow x G_0'(x) = x p_1 + 2p_2 x^2 + 3p_3 x^3 + 4p_4 x^4 + \cdots$$

$$\Rightarrow \frac{d}{dx}\left(x G_0'(x)\right) = p_1 + 2^2 p_2 x + 3^2 p_3 x^2 + 4^2 p_4 x^3 + \cdot$$

$$\Rightarrow \left(x G_0'(x)\right)'_{x=1} = p_1 + 2^2 p_2 + 3^2 p_3 + 4^2 p_4 \cdots = \langle k^2 \rangle$$

Powers

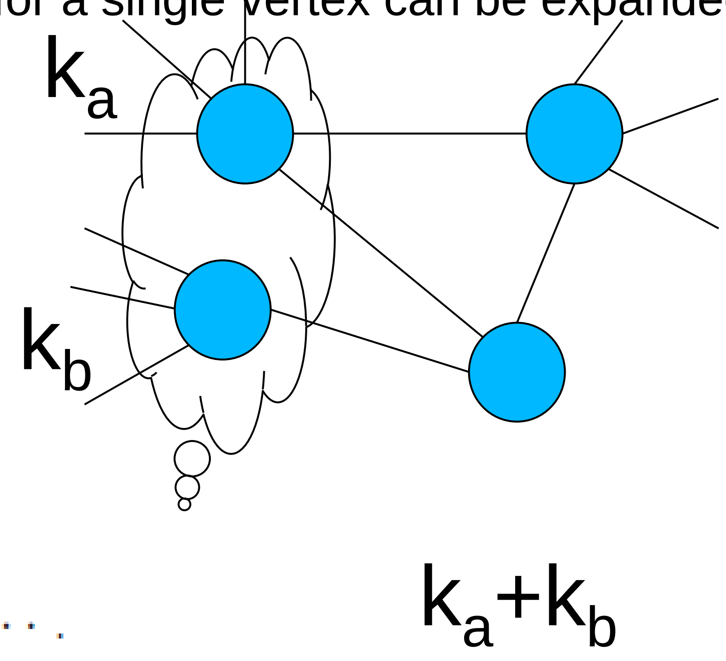
If the distribution of a property (degree) *k* of an object is generated by a given generating function, then the distribution of the total of *k* summed over *m* independent realizations of the object is generated by the *m*th power of that generating function.

For example, if we choose *m* vertices at random from a large graph, then the distribution of the sum of the degrees of those vertices is generated by [*G*_0(*x*)]^*m*.

To see why this is so, consider the simple case of just two vertices

The square [*G*0(*x*)]^2 of the generating function for a single vertex can be expanded as

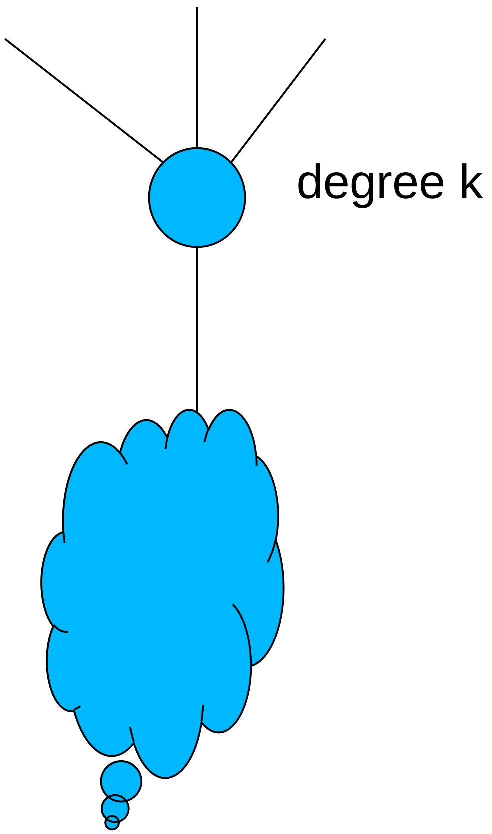
$$
\begin{aligned}
[G_0(x)]^2 &= \left[\sum_k p_k x^k\right]^2 \\
&= \sum_{jk} p_j p_k x^{j+k} \\
&= p_0 p_0 x^0 + (p_0 p_1 + p_1 p_0) x^1 \\
&\quad + (p_0 p_2 + p_1 p_1 + p_2 p_0) x^2 \\
&\quad + (p_0 p_3 + p_1 p_2 + p_2 p_1 + p_3 p_0) x^3 + \cdots .
\end{aligned}
$$



Powers

It is clear that the coefficient of the power of x^n in this expression is precisely the sum of all products $p_j p_k$ such that $j+k=n$, and hence correctly gives the probability that the sum of the degrees of the two vertices will be n. It is straightforward to convince oneself that this property extends also to all higher powers of the generating function.

Degree distribution through randomly chosen edge



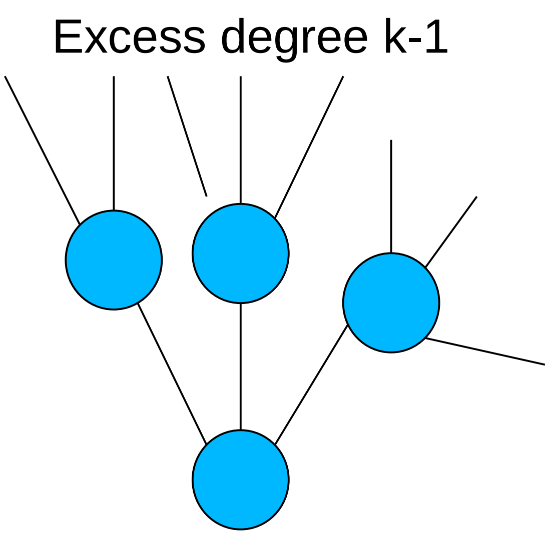
Distribution of the degree of the vertices that we arrive at by following a randomly chosen edge.

Such an edge arrives at a vertex with probability proportional to the degree of that vertex,

the vertex therefore has a probability distribution of degree proportional to ***kpk***

$$\frac{\sum_k k p_k x^k}{\sum_k k p_k} = x \frac{G_0'(x)}{G_0'(1)}.$$

Distribution of outgoing edges of neighbors

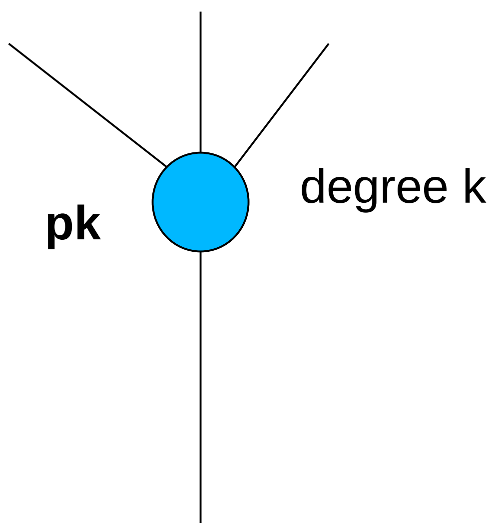


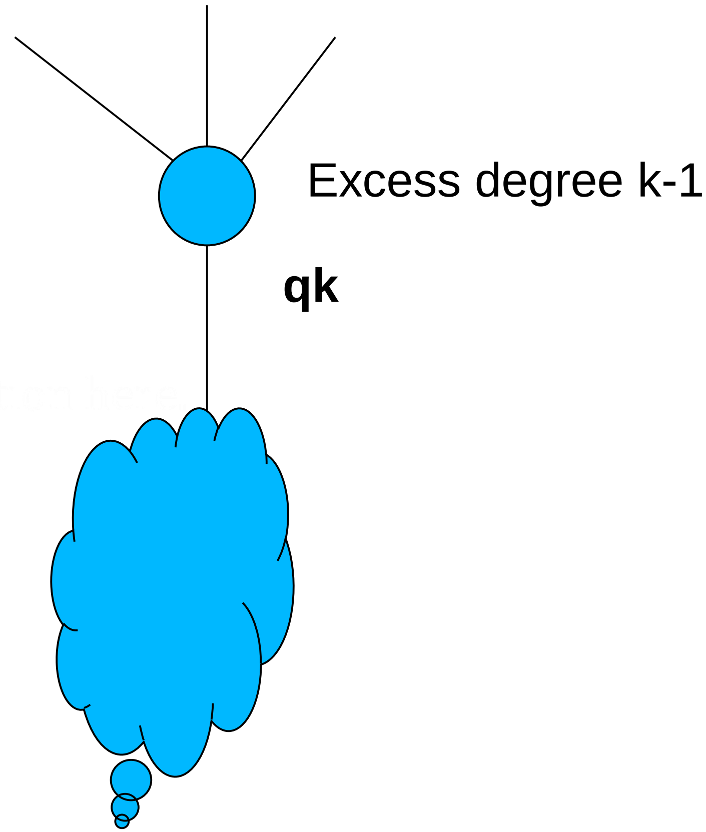
If we start at a randomly chosen vertex and follow each of the edges at that vertex to reach the k nearest neighbors, then the vertices arrived at each have the distribution of remaining outgoing edges generated by this function, less one power of x, to allow for the edge that we arrived along. Thus the distribution of outgoing edges is generated by the function

$$G_1(x) = \frac{G_0'(x)}{G_0'(1)} = \frac{1}{z} G_0'(x), \tag{9}$$

where z is the average vertex degree, as before. The probability that any of these outgoing edges connects to the origi-

Excess degree-revisit

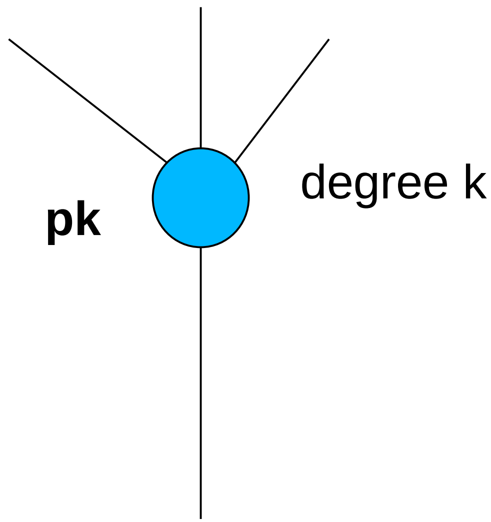


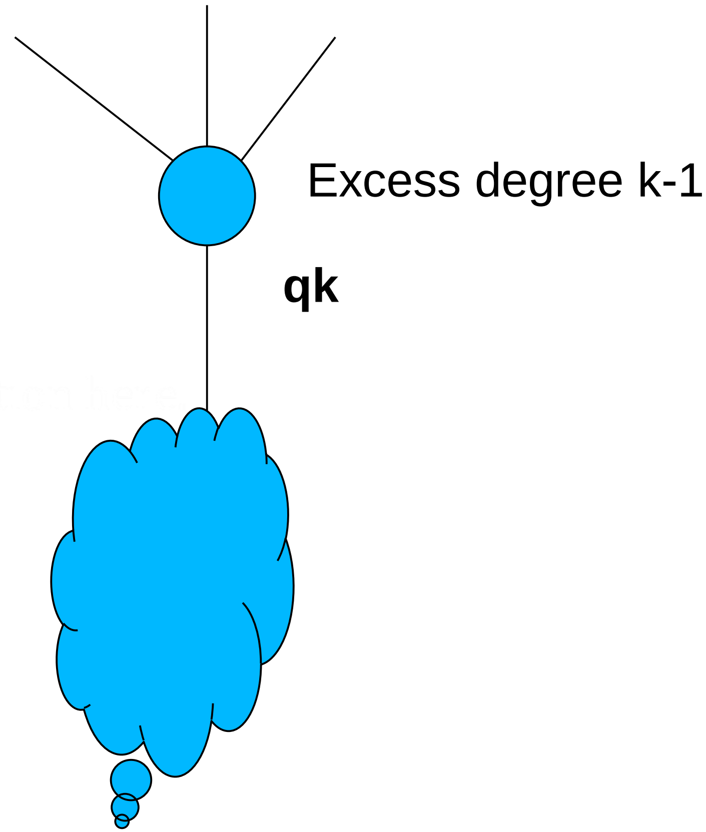


$$G_0(x) = \sum_{k=0}^{\infty} p_k x^k, \qquad G_1(x) = \sum_{k=0}^{\infty} q_k x^k.$$

$$q_k = \frac{(k+1)p_{k+1}}{\sum_k k p_k} = \frac{(k+1)p_{k+1}}{z},$$

Excess degree-revisit





$$G_0(x) = \sum_{k=0}^{\infty} p_k x^k, \qquad G_1(x) = \sum_{k=0}^{\infty} q_k x^k.$$

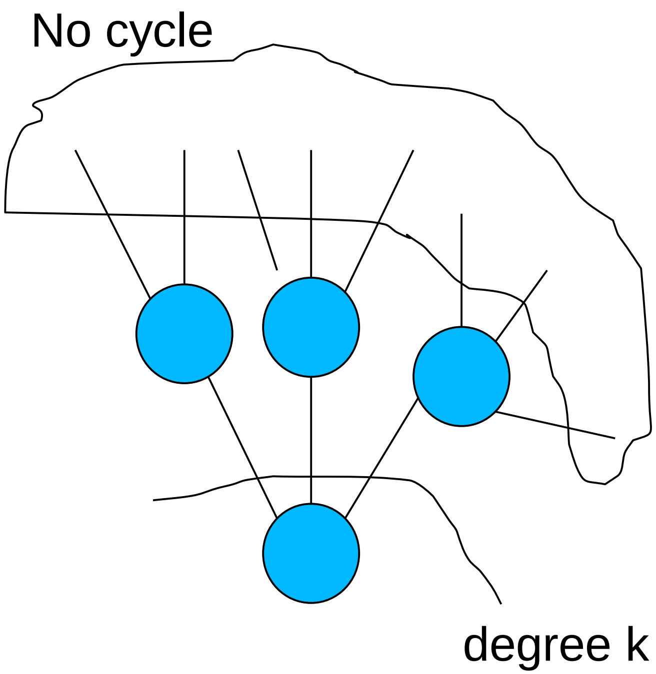
$$G_1(x) = \frac{G_0'(x)}{G_0'(1)}$$

$$G_1(x) = \frac{\sum k p_k x^{k-1}}{z}$$

q_{k-1}

$$q_k = \frac{(k+1)p_{k+1}}{\sum_k k p_k} = \frac{(k+1)p_{k+1}}{z},$$

Probability distribution of the number of second neighbors



The generating function for the probability distribution of the number of *second* neighbors of the original vertex can be written as

$$\sum_k p_k[G_1(x)]^k = G_0(G_1(x)).$$

Average number of second neighbors

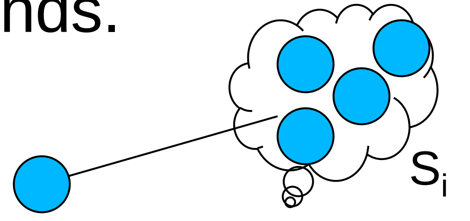
$$z_2 = \left[\frac{d}{dx} G_0(G_1(x))\right]_{x=1} = G_0'(1)\,G_1'(1) = G_0''(1),$$

$G_0'(G_1(x))G_1'(x)$

z

Component size

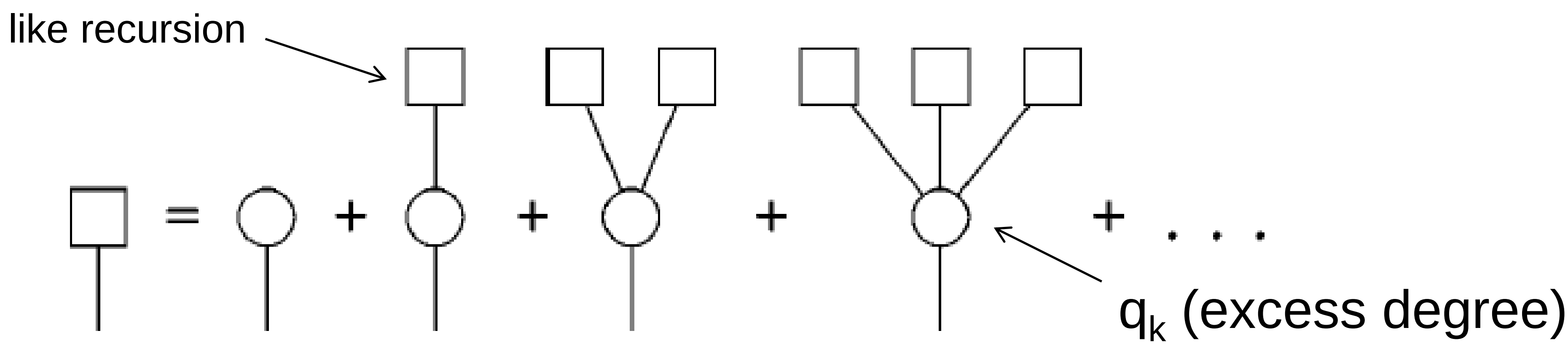
- First let us consider the distribution of the sizes (s_i) of connected components in the graph
- Let *H*1(*x*) be the generating function for the distribution of the sizes of components that are reached by choosing a random edge and following it to one of its ends.



- We explicitly exclude from *H*1(*x*) the giant component, if there is one; the giant component is dealt with separately.

Component size

This means that the distribution of components generated by $H1(x)$ can be represented graphically as



Each component is treelike in structure, consisting of the single node we reach by following our initial edge, plus any number of other treelike clusters,

Recursively follow the same size distribution,

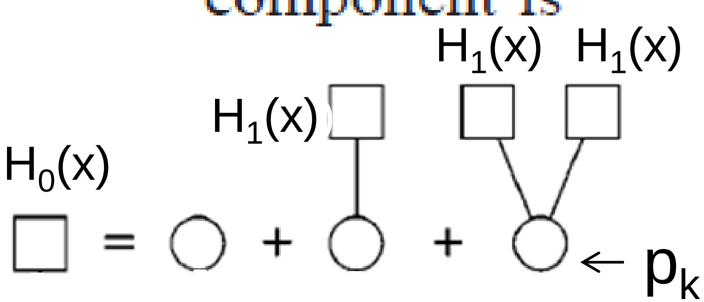
If we denote by qk the probability that the initial node has k edges coming out of it other than the edge we came in along the "powers" property of Sec. II A, $H_1(x)$ must satisfy a self-consistency condition of the form

$$H_1(x) = xq_0 + xq_1 H_1(x) + xq_2[H_1(x)]^2 + \cdots . \quad (25)$$

However, q_k is nothing other than the coefficient of x^k in the generating function $G_1(x)$, Eq. (9), and hence Eq. (25) can also be written

$$H_1(x) = xG_1(H_1(x)). \quad (26)$$

If we start at a randomly chosen vertex, then we have one such component at the end of each edge leaving that vertex, and hence the generating function for the size of the whole component is



$$H_0(x) = xG_0(H_1(x)). \quad (27)$$

$H_o(x)$ generates the probability s_k that a randomly chosen node belongs to a component of size k the average size of the component to which a randomly chosen vertex belongs, for the case where there is no giant component in the graph, is given in the normal fashion by

$$\langle s\rangle = H_0'(1) = 1 + G_0'(1)H_1'(1). \tag{29}$$

From Eq. (26) we have

$$H_1'(1) = 1 + G_1'(1)H_1'(1), \tag{30}$$

and hence

Avg component size

$$\langle s\rangle = 1 + \frac{G_0'(1)}{1 - G_1'(1)} = 1 + \frac{z_1^2}{z_1 - z_2}, \tag{31}$$

where $z_1 = z$ is the average number of neighbors of a vertex and z_2 is the average number of second neighbors. We see

Phase transition-Giant component formation

that this expression diverges when

$$G_1'(1) = 1. \qquad (32)$$

This point marks the phase transition at which a giant component first appears. Substituting Eqs. (2) and (9) into Eq. (32), we can also write the condition for the phase transition as

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

$$\sum_k k(k-2)p_k = 0. \qquad (33)$$

Indeed, since this sum increases monotonically as edges are added to the graph, it follows that the giant component exists if and only if this sum is positive.

ER graph – Phase transition

$$P(k) = \frac{\langle k \rangle^k e^{-\langle k \rangle}}{k!} = \frac{\lambda^k e^{-\lambda}}{k!}$$

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

$<k>= \lambda$

$<k^2>= \lambda^2 + \lambda$

$\lambda=1$