Network models: random graphs

Leonid E. Zhukov

School of Applied Mathematics and Information Science
National Research University Higher School of Economics

27.01.2013

Network models

Empirical network features:

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

Generative models:

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

Random Graph models

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 $C^m_{n(n-1)/2}$ graphs with n nodes and m edges
- $G_{n,p}$, each pair out of $n(n-1)/2$ pairs of nodes is connected with probability p, m - random number

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

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

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

Random Graph models

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

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

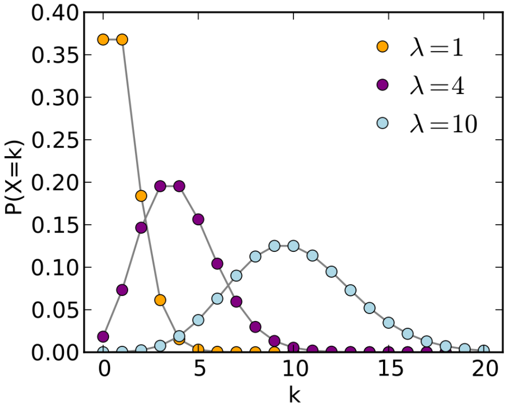
(Bernoulli distribution)
p^k - probability that connects to k nodes (has k-edges)
$(1-p)^{n-k-1}$ - probability that does not connect to any other node
C_{n-1}^{k} - number of ways to select k nodes out of all to connect to

- 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



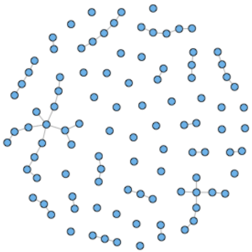
$$P(k_i = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad \lambda = pn$$

Phase transition

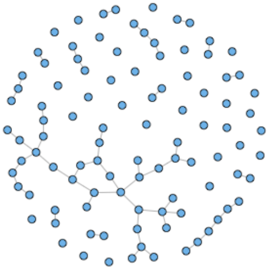
Consider $G_{n,p}$ as a function of p

- $p = 0$, 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

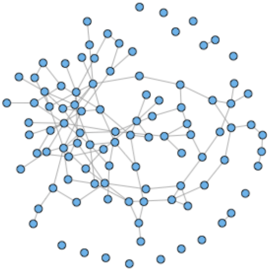


$p < p_c$

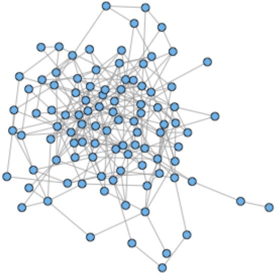


$p = p_c$

Random graph model



$p > p_c$



$p >> p_c$

Phase transition

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

$$u = P(k=1)\cdot u + P(k=2)\cdot u^2 + P(k=3)\cdot u^3 ... =$$
$$= \sum_{k=0}^{\infty} P(k)u^k = \sum_{k=0} \frac{\lambda^k e^{-\lambda}}{k!} u^k = e^{-\lambda} e^{\lambda u} = e^{\lambda(u-1)}$$

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

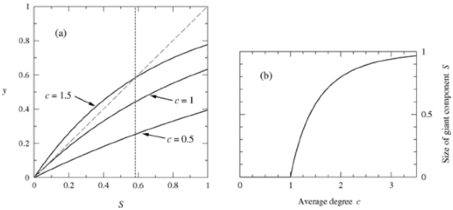
$$s = 1 - u$$

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

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

Phase transition

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



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

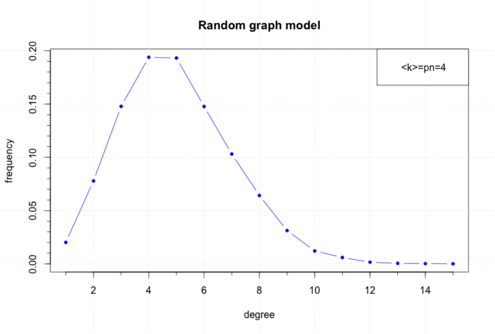
$$\lambda e^{-\lambda s} > 1$$

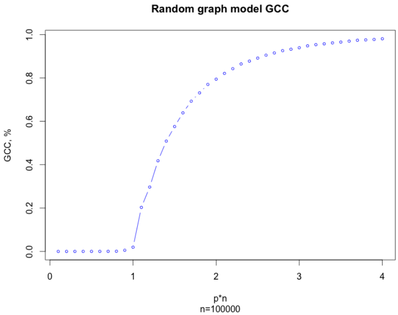
critical value:

$$\lambda_c = 1$$

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

Phase transition





Phase transition

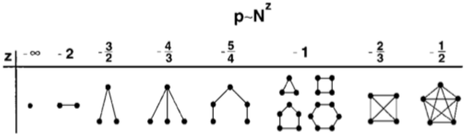
Graph $G(n, p)$, for $n \to \infty$, critical value $p_c = 1/n$

- 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

Threshold probabilities

Graph $G(n, p)$
Threshold probabilities when different subgraphs of g-nodes appear in a random graph

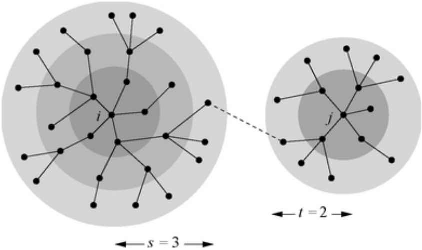


- $p_c \sim n^{-g/(g-1)}$, having a tree of order g
- $p_c \sim n^{-1}$, having a cycle of order g
- $p_c \sim n^{-2/(g-1)}$, complete subgraph of order g

Barabasi, 2002

Graph diameter

- On average, the number of nodes s steps away from a node $\langle k \rangle^s = \lambda^s$



- If graph is a tree (GCC, around p_c), $\lambda^d \sim n$, $d \sim \frac{\ln n}{\ln \lambda}$
- $P(d_{ij} > s + t + 1)$ - probability, that there is no edge between the surfaces
- $P(d_{ij} > s + t + 1) = (1 - p)^{\lambda^{s+t}}$,
  where $\lambda^s \lambda^t$ total number of possible pairs from different groups

Graph diameter

- define $l = s + t + 1$
- $P(d_{ij} > l) = (1-p)^{\lambda^{l-1}} = (1 - \frac{\lambda}{n})^{\lambda^{l-1}}$
  $\ln P(d_{ij} > l) = \lambda^{l-1} \ln(1 - \frac{\lambda}{n}) = -\frac{\lambda^l}{n}$
  $P(d_{ij} > l) = \exp\left(-\frac{\lambda^l}{n}\right)$
- Graph diameter is the smallest value l such that $P(d_{ij} > l) = 0$, i.e no matter which pair of nodes we pick, there is zero chance to be separated by greater distance, $\lambda^l = an$, should grow faster than n
- $d = \min(l) = \frac{lna}{\ln \lambda} + \frac{\ln n}{\ln \lambda} = A + \frac{\ln n}{\ln \lambda}$
- Graph diameter when $p \geq p_c$ ($\lambda = \langle k \rangle = pn$):

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

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

Configuration model

Select a sequence of nodes with degreees $D = \{k_1, k_2, k_3..k_n\}: \ \sum_i k_i = 2m$ to follow given distribution $P(k)$. For example: 1 1 1 1 1 2 2 2 3 3 3...

$$P(k) = \frac{\#(k_i = k)}{2m}$$

Randomly select two nodes from the sequence and form an edge between them

Refernces

- On random graphs I, P. Erdos and A. Renyi, Publicationes Mathematicae 6, 290–297 (1959).
- On the evolution of random graphs, P. Erdos and A. Renyi, Publicaton of the Mathematical Institute of the Hungarian Academy of Sciences, 17-61 (1960)