NATIONAL RESEARCH UNIVERSITY

Random graphs

Network Science

Lecture 3

Leonid Zhukov

lzhukov@hse.ru

www.leonidzhukov.net/hse/2021/networks

National Research University Higher School of Economics

School of Data Analysis and Artificial Intelligence, Department of Computer Science

Network models

Empirical network features:

- Power-law (heavy-tailed) degree distribution
- Small average distance (graph diameter)
- Large clustering coefficient (transitivity)

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

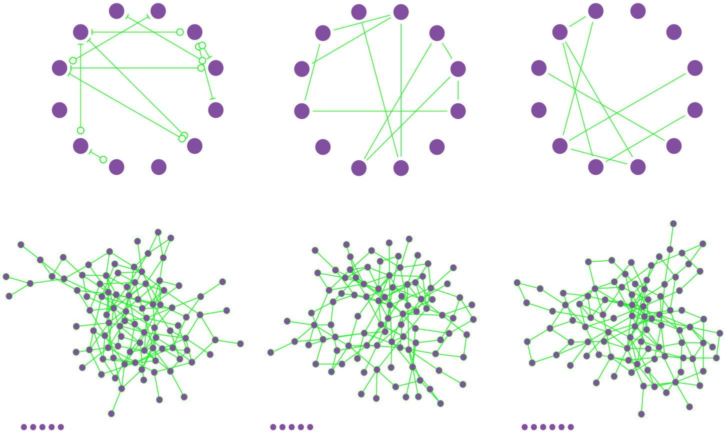
- $G(n, m)$ a randomly selected graph from the set of C_N^m graphs, $N = \frac{n(n-1)}{2}$, with n nodes and m edges
- $G(n, p)$ each pair out of $N = \frac{n(n-1)}{2}$ pairs of nodes is connected with probability p, m - random number

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

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

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

Random graph model $G(n, p)$



$n = 12, p = 1/6$ $\qquad$ $n = 100, p = 0.03$

Random graph model $G(n, p)$

- In $G(n, p)$ model, probability for a network to have m links is given by binomial distribution:

$$P(m) = C_N^m p^m (1-p)^{N-m}$$

  where $N = \frac{n(n-1)}{2}$
- p^m - probability that m links are present
  $(1-p)^{N-m}$ - probability that other links are not
  C_N^m - number of ways to select m links out of all N,
  $C_N^m = \frac{N!}{m!(N-m)!}$
- expected number of links

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

Degree distribution

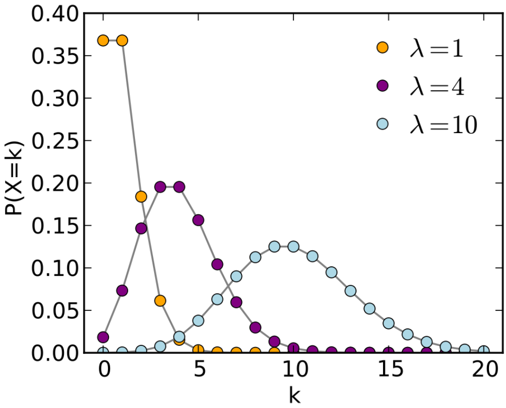
- Probability that i-th node has a degree $k_i = k$ is given by Binomial distribution:

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

  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, $C_{n-1}^{k} = \frac{(n-1)!}{k!(n-k-1)!}$
- Binomial distribution, when $\langle k \rangle << N$ or $n \to \infty$ and $p \to 0$ at fixed $\langle k \rangle$, is well approximated by Poisson distribution:

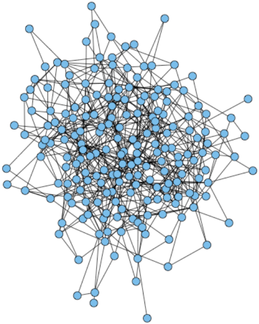
$$P(k) = \frac{\langle k \rangle^k e^{-\langle k \rangle}}{k!} = \frac{\lambda^k e^{-\lambda}}{k!}, \; \langle k \rangle = pn = \lambda$$

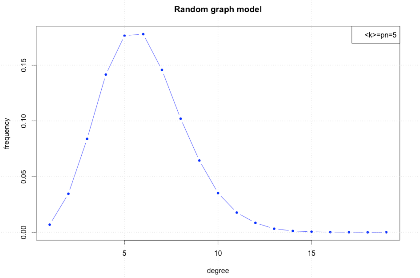
Poisson Distribution



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

Random graph



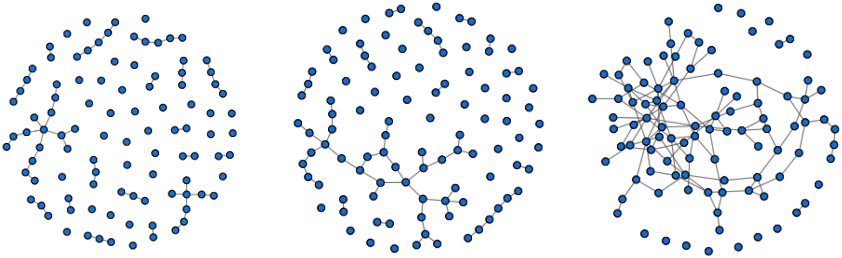


$$\langle k \rangle = pn = 5$$

Random graph model

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

- $p = 0$, empty graph - $\langle k \rangle = 0$
- $p = 1$, complete (full) graph - $\langle k \rangle = n - 1$
- n_G -largest connected component, $s = \frac{n_G}{n}$



p

Phase transition

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

$$u = \frac{n - n_G}{n} = P(k=0) + P(k=1) \cdot u + P(k=2) \cdot u^2 + P(k=3) \cdot u^3 \ldots =$$

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

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

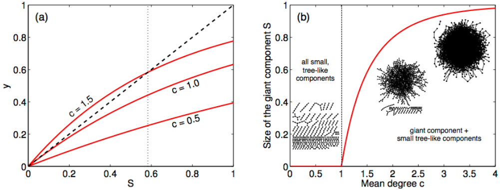
$$u = e^{\lambda(u-1)}$$

$$s = 1 - u, \; 1 - s = e^{-\lambda s}$$

$\lambda = pn = \langle k \rangle$
when $\lambda \to \infty, \; s \to 1$
when $\lambda \to 0, \; s \to 0$

Phase transition

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



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

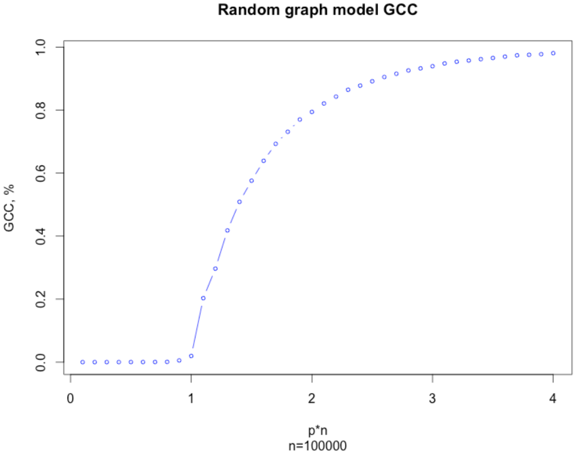
$$\lambda e^{-\lambda s} > 1$$

critical value:

$$\lambda_c = 1$$

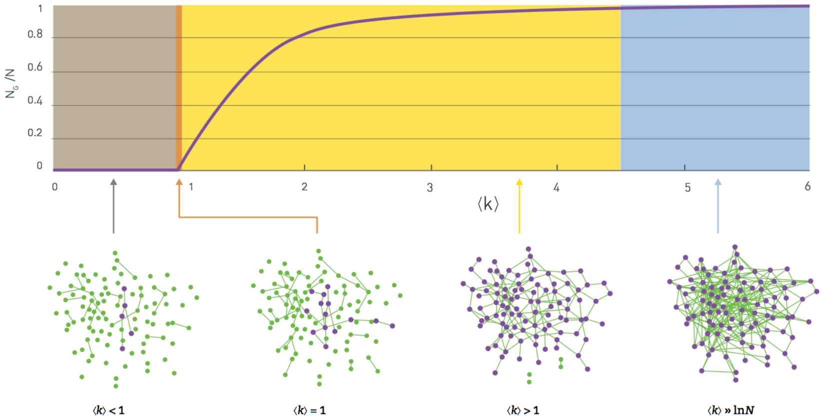
$$\lambda_c = \langle k \rangle = p_c n = 1, \;\; p_c = \frac{1}{n}$$

Numerical simulations



$$\langle k \rangle = pn$$

Evolution of random network



from A-L. Barabasi, 2016

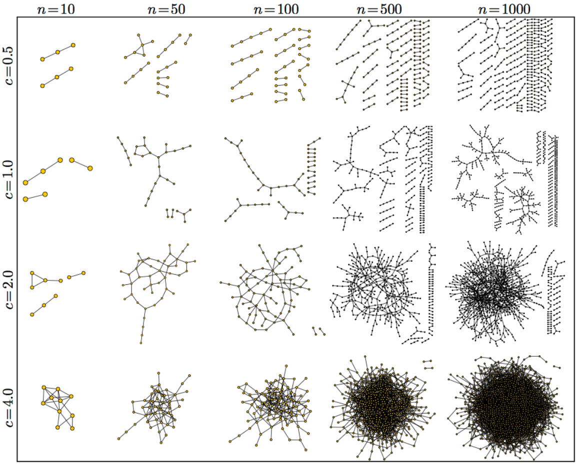
Phase transition

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

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

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

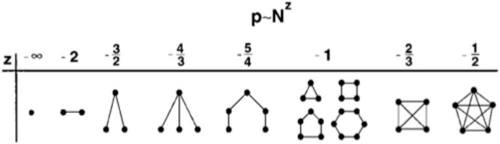
Numerical simulation



Clauset, 2014

Threshold probabilities

Graph $G(n, p)$
Threshold probabilities when different subgraphs of k-nodes and l-edges appear in a random graph $p_s \sim n^{-k/l}$



When $p > p_s$:

- $p_s \sim n^{-k/(k-1)}$, having a tree with k nodes
- $p_s \sim n^{-1}$, having a cycle with k nodes
- $p_s \sim n^{-2/(k-1)}$, complete subgraph with k nodes

Barabasi, 2002

Clustering coefficient

- Clustering coefficient (probability that two neighbors link to each other):

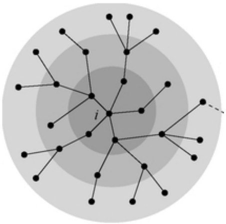
$$C_i(k) = \frac{\text{\#of links between NN}}{\text{\#max number of links NN}} = \frac{pk(k-1)/2}{k(k-1)/2} = p$$

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

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

Graph diameter

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



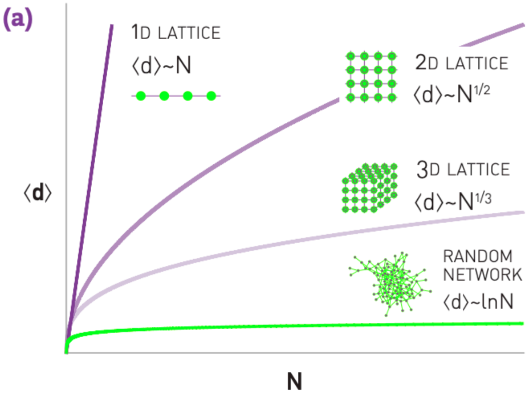
- on average, the number of nodes d steps away from a node

$$n = 1 + \langle k \rangle + \langle k \rangle^2 + \ldots \langle k \rangle^D = \frac{\langle k \rangle^{D+1} - 1}{\langle k \rangle - 1} \approx \langle k \rangle^D$$

- around p_c , $\langle k \rangle^D \sim n$,

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

Graph diameter



Random graph model

- Node degree distribution function - Binomial/Poisson:

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

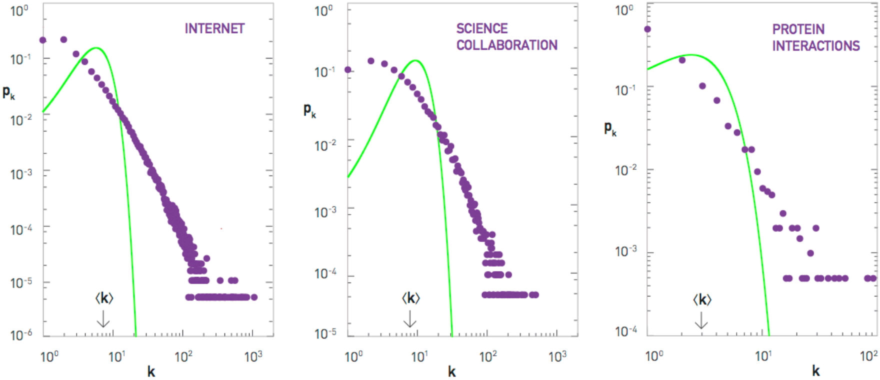
- Average path length:

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

- Clustering coefficient:

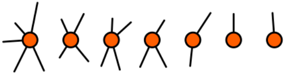
$$C = \frac{\langle k \rangle}{n}$$

Degree distribution in real networks



Configuration model

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



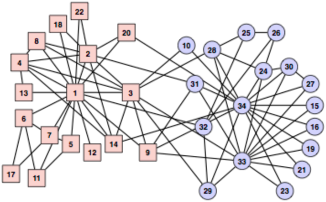
- Can contain self loops and multiple edges
- Probability that two nodes i and j are connected

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

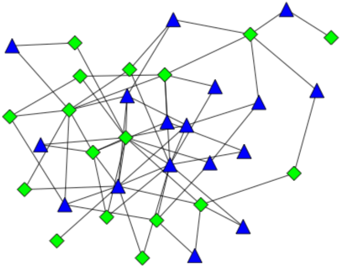
- Will be a simple graph for special "graphical degree sequence"

Configuration model

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



karate club



configuration model

Clauset, 2014

References

- 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)