Network models: random graphs

Leonid E. Zhukov

School of Data Analysis and Artificial Intelligence
Department of Computer Science
National Research University Higher School of Economics

Structural Analysis and Visualization of Networks
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 attachment 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 $C^m_N$ 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

- Probability that the $i$-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

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

\[ p < p_c \]

\[ p = p_c \]

\[ p > p_c \]
Random graph model

\[ p \gg p_c \]
Let $u$ – fraction of nodes that do not belong to GCC. The probability that a node does not belong to GCC

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

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

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

$$s = 1 - u$$

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

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

\[ s = 1 - e^{-\lambda s} \]

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

\[ \lambda e^{-\lambda s} > 1 \]

critical value:

\[ \lambda_c = 1 \]

\[ \lambda_c = \langle k \rangle = p_c n = 1, \quad p_c = \frac{1}{n} \]
Numerical simulations

\[ \langle k \rangle = pn \]
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
Phase transition

Clauset, 2014
Threshold probabilities

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

When $p > p_c$:
- $p_c \sim n^{-k/(k-1)}$, having a tree with $k$ nodes
- $p_c \sim n^{-1}$, having a cycle with $k$ nodes
- $p_c \sim n^{-2/(k-1)}$, complete subgraph with $k$ nodes

Barabasi, 2002
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 \( n \to \infty, \ C \to 0 \)
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 $p_c$, $\langle k \rangle^d \sim n$,

$$d \sim \frac{\ln n}{\ln \langle k \rangle}$$
Random graph with \( n \) nodes with a given degree sequence:
\[ D = \{ k_1, k_2, k_3..k_n \} \] and \( m = \frac{1}{2} \sum_i k_i \) 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

Clauset, 2014
References