Random Graphs, Generative Models

Lecture 3

Natalia Semenova

Higher School of Economics

nasemenova_2@edu.hse.ru

Structural Analysis and Visualization of Networks
30.01.2024

Lecture Overview

1. Random graph model (Erdos & Renyi, 1959)
   - Random Graph Growth
2. Preferential attachment model (Barabasi & Albert, 1999)
   - Non-linear preferential attachment
   - Link selection model
   - Copying model
3. Small world model (Watts and Strogatz, 1998)
   - Model comparison

Summary: Network models

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

Defining Random Networks

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

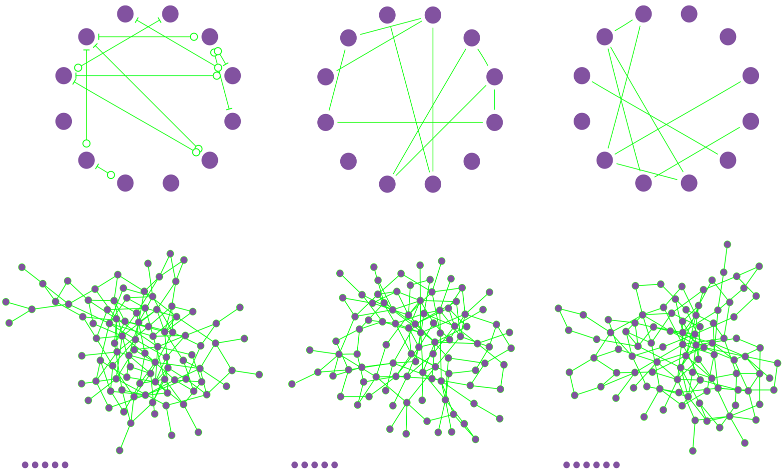
- $G(n, m)$ a randomly selected graph from the set of C_N^m graphs, $N = \frac{n(n-1)}{2}$, with n nodes and m edges
- $G(n, p)$ each pair out of $N = \frac{n(n-1)}{2}$ pairs of nodes is connected with probability p, m - random number

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

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

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

Random graph model $G(n, p)$



n = 12, p = 1/6   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

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

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

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

Degree distribution

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

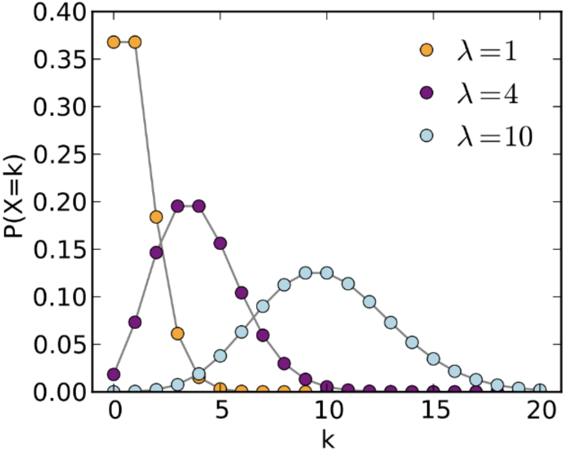
$$P(k_i = k) = P(k) = C_{n-1}^{k} p^k (1-p)^{n-1-k}$$

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:

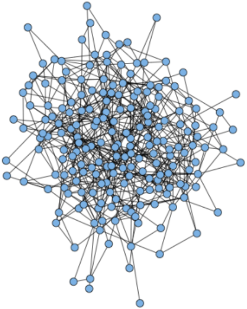
$$P(k) = \frac{\langle k \rangle^k e^{-\langle k \rangle}}{k!} = \frac{\lambda^k e^{-\lambda}}{k!}, \;\; \langle k \rangle = pn = \lambda$$

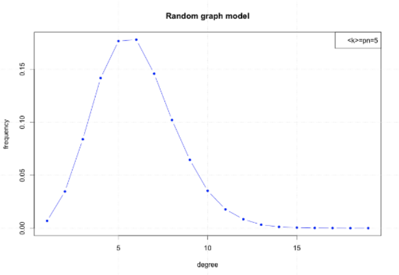
Poisson Distribution



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

Random graph



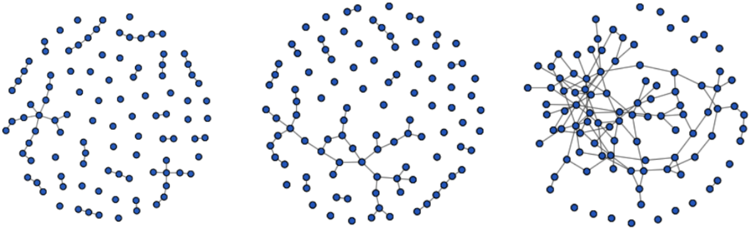


$$\langle k \rangle = pn = 5$$

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

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)

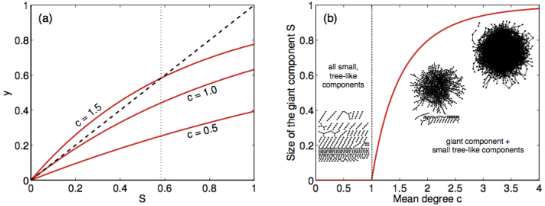
$$u = e^{\lambda(u-1)}$$

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

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

Phase transition/Expansion/Evolution

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



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

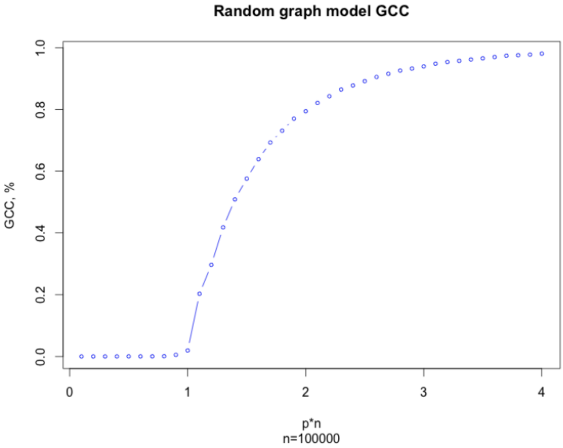
$$\lambda e^{-\lambda s} > 1$$

critical value:

$$\lambda_c = 1$$

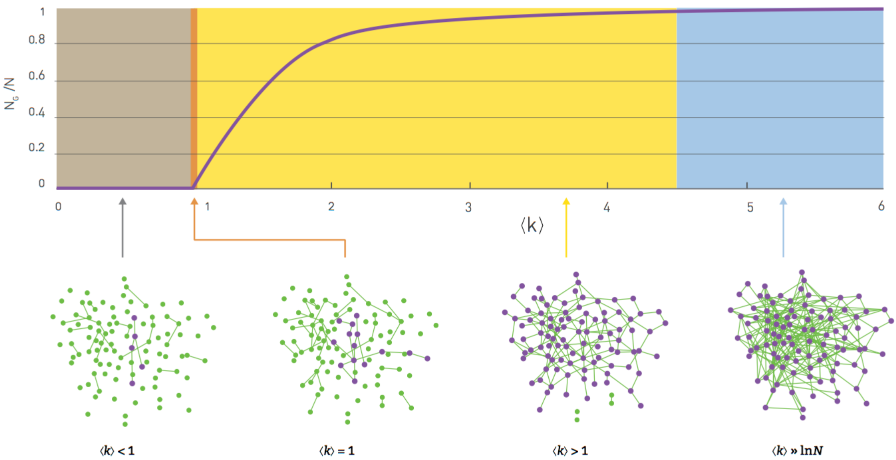
$$\lambda_c = \langle k \rangle = p_c n = 1, \quad p_c = \frac{1}{n}$$

Numerical simulations



$$\langle k \rangle = pn$$

Evolution of random network



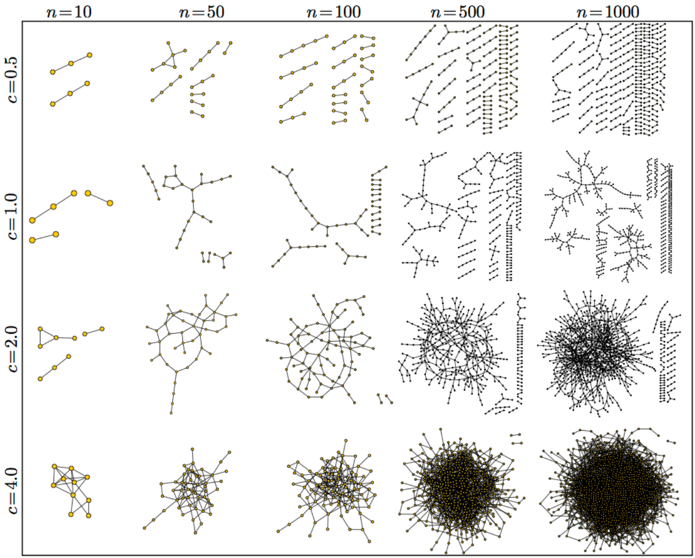
Phase transition/Expansion/Evolution

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

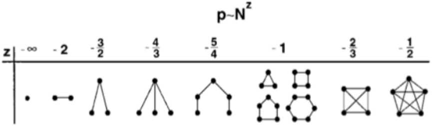
Critical value: $\langle k \rangle = p_c n = 1$- on average one neighbor for a node

Numerical simulation



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

Graph diameter

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

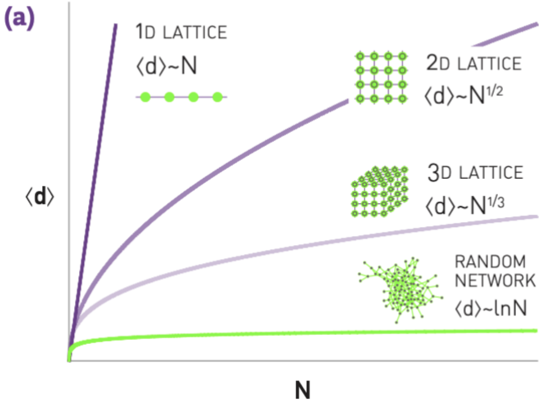
- on average, the number of nodes d steps away from a node

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

- in GCC, 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 - Poisson:

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

- Average path length:

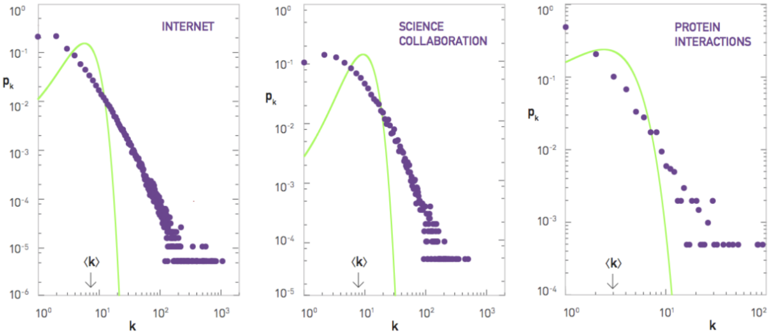
$$\langle L \rangle = \frac{\ln n}{\ln \langle k \rangle}$$

- Clustering coefficient:

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

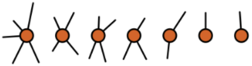
Real networks

Degree distribution in real networks



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.



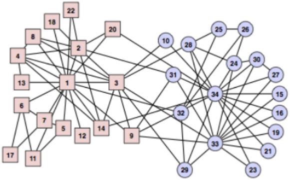
- Can contain self loops and multiple edges
- Probability that two nodes i and j are connected

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

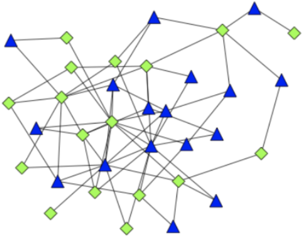
- Will be a simple graph for special "graphical degree sequence"

Configuration model

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



karate club



configuration model

Clauset, 2014

Motivation

Most of the networks we study are evolving over time, they expand by adding new nodes:

- Citation networks
- Collaboration networks
- Web
- Social networks

Preferential attachment model

Continues approximation: continues time, real variable node degree $\langle k_i(t) \rangle$- expected value over multiple realizations
Time-dependent degree of a single node:

$$k_i(t + \delta t) = k_i(t) + m\Pi(k_i)\delta t$$

$$\frac{dk_i(t)}{dt} = m\Pi(k_i) = m\frac{k_i}{\sum_i k_i} = \frac{mk_i}{2mt} = \frac{k_i(t)}{2t}$$

node i is added at time t_i: $k_i(t_i) = m$

$$\int_m^{k_i(t)} \frac{dk_i}{k_i} = \int_{t_i}^{t} \frac{dt}{2t}$$

Solution:

$$k_i(t) = m\left(\frac{t}{t_i}\right)^{1/2}$$

Preferential attachment model

Barabasi and Albert, 1999

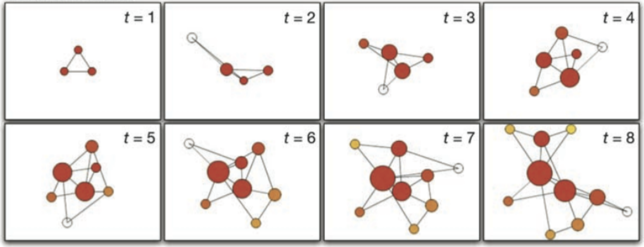
Dynamic growth: start at $t = 0$ with n_0 nodes and $m_0 \geq n_0$ edges

1. **Growth**
   At each time step add a new node with m edges ($m \leq n_0$), connecting to m nodes already in netwrok $k_i(i) = m$
2. **Preferential attachment**
   The probability of linking to existing node i is proportional to the node degree k_i
   $$\Pi(k_i) = \frac{k_i}{\sum_i k_i}$$

after t timesteps: $t + n_0$ nodes, $mt + m_0$ edges

Preferential attachment model

Scale-Free Model



Barabasi, 1999

Preferential attachment

Continues approximation: continues time, real variable node degree $\langle k_i(t) \rangle$- expected value over multiple realizations
Time-dependent degree of a single node:

$$k_i(t+\delta t) = k_i(t) + m\Pi(k_i)\delta t$$

$$\frac{dk_i(t)}{dt} = m\Pi(k_i) = m\frac{k_i}{\sum_i k_i} = \frac{mk_i}{2mt} = \frac{k_i(t)}{2t}$$

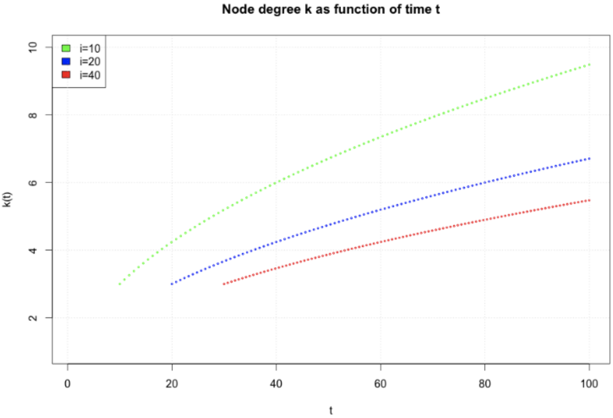
node i is added at time t_i: $k_i(t_i) = m$

$$\int_m^{k_i(t)} \frac{dk_i}{k_i} = \int_{t_i}^{t} \frac{dt}{2t}$$

Solution:

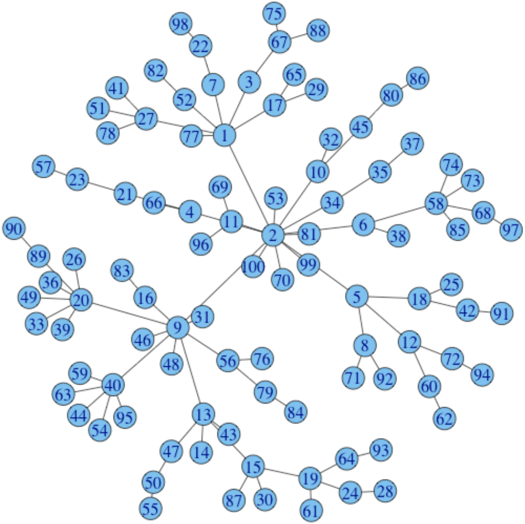
$$k_i(t) = m\left(\frac{t}{t_i}\right)^{1/2}$$

Preferential attachment



$$k_i(t) = m\left(\frac{t}{t_i}\right)^{1/2}; \qquad \frac{dk_i(t)}{dt} = \frac{m}{2}\frac{1}{\sqrt{tt_i}}$$

Preferential attachment



Preferential attachment

Time evolution of a node degree

$$k_i(t) = m\left(\frac{t}{t_i}\right)^{1/2}$$

Find probability $P(k' \le k)$ of a randomly selected node to have $k' \le k$ at time t (fraction of nodes with $k' \le k$). Nodes with $k_i(t) \le k$:

$$m\left(\frac{t}{t_i}\right)^{1/2} \le k \quad \Rightarrow \quad t_i \ge \frac{m^2}{k^2}t$$

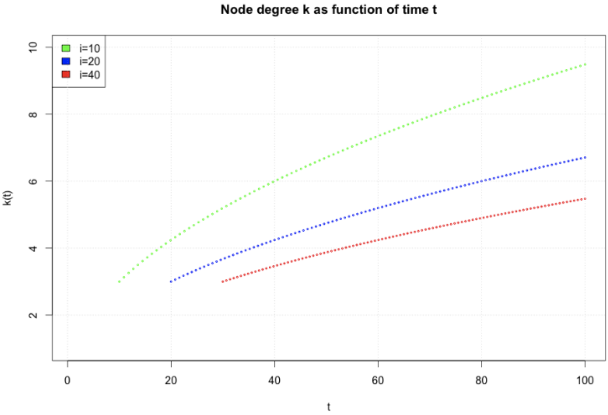
Cumulative function:

$$F(k) = P(k' \le k) = \frac{n_0 + t - m^2 t/k^2}{n_0 + t} \approx 1 - \frac{m^2}{k^2}$$

Distribution function:

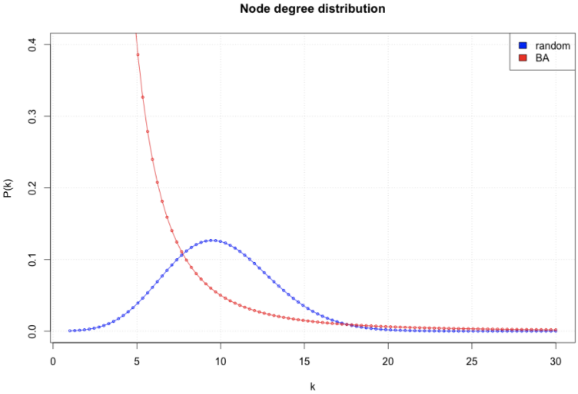
$$P(k) = \frac{d}{dk}F(k) = \frac{2m^2}{k^3}$$

Preferential attachment



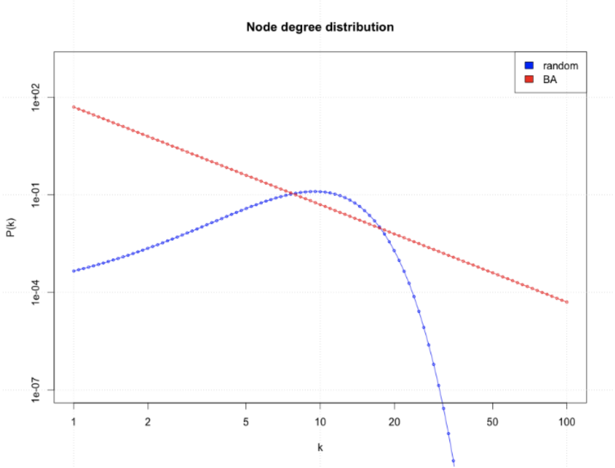
$$m\left(\frac{t}{t_i}\right)^{1/2} \leq k \quad \Rightarrow \quad t_i \geq \frac{m^2}{k^2}t$$

Preferential attachment vs random graph



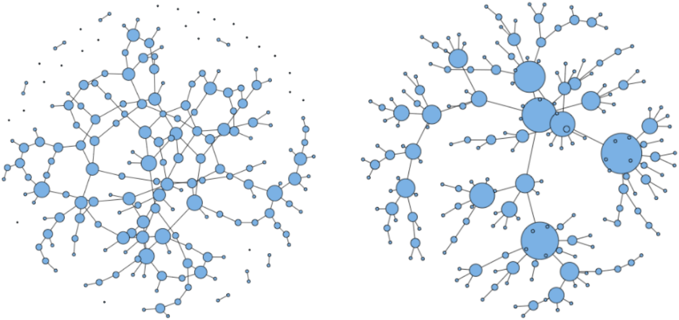
$$BA: P(k) = \frac{2m^2}{k^3}, \quad ER: P(k) = \frac{\langle k\rangle^k e^{-\langle k\rangle}}{k!}, \quad \langle k\rangle = pn$$

Preferential attachment vs random graph

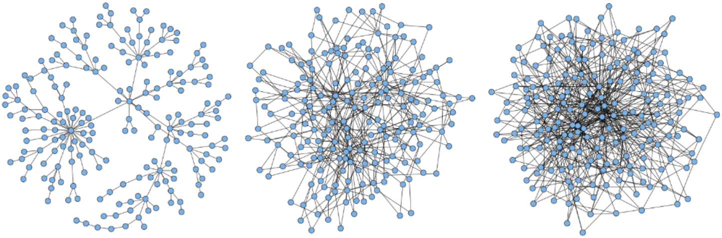


$$BA : P(k) = \frac{2m^2}{k^3}, \quad ER : P(k) = \frac{\langle k \rangle^k e^{-\langle k \rangle}}{k!}, \quad \langle k \rangle = pn$$

Preferential attachment vs random graph



Preferential attachment model



$m = 1$ $\qquad$ $m = 2$ $\qquad$ $m = 3$

Growing random graph

1. **Growth**
   At each time step add a new node with m edges ($m \leq n_0$), connecting to m nodes already in network $k_i(i) = m$
2. **Attachment uniformly at random**
   The probability of linking to existing node i is

$$\Pi(k_i) = \frac{1}{n_0 + t - 1}$$

Node degree growth:

$$k_i(t) = m\left(1 + \log\left(\frac{t}{i}\right)\right)$$

Node degree distribution function:

$$P(k) = \frac{e}{m}\exp\left(-\frac{k}{m}\right)$$

Preferential attachment

- Power law distribution function:

$$P(k) = \frac{2m^2}{k^3}$$

- Average path length (analytical result) :

$$\langle L \rangle \sim \log(N)/\log(\log(N))$$

- Clustering coefficient (numerical result):

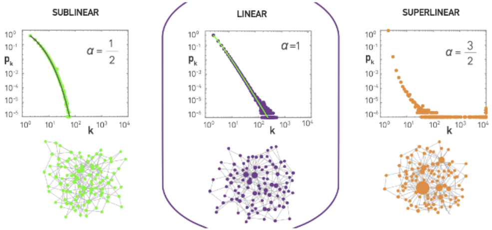
$$C \sim N^{-0.75}$$

Non-linear preferential attachment

- Non-linear preferential attachment models:

$$\Pi(k) \sim k^{\alpha}$$

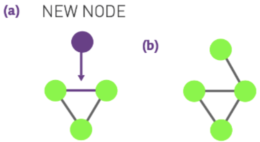
- $\alpha = 0$, no hubs, exponential dsitribution
- $0 < \alpha < 1$, sublinear, smaller hubs, stretched exponential
- $\alpha = 1$, scale-free, hubs, power law
- $\alpha > 1$, superlinear, super hubs, hubs-and-spoke



Link selection model

Local growth mechanism:

- Growth: at each time step add a new node
- Link selection: select link at random and connect to one of two nodes at the ends



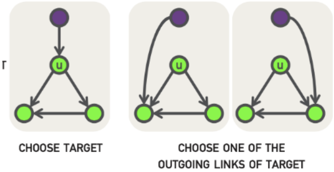
Probability to connect to a node with degree k:

$$\Pi(k) = \frac{kp_k}{\langle k \rangle}$$

Copying model

Local growth mechanism:

- Random connection: with probability p connect to a random node u
- Copying: with probability $1 - p$ randomly choose an outgoing link from u and connect to its target



Probability to connect to a node with degree k:

$$\Pi(k) = \frac{p}{n} + (1-p)\frac{k}{2m}$$

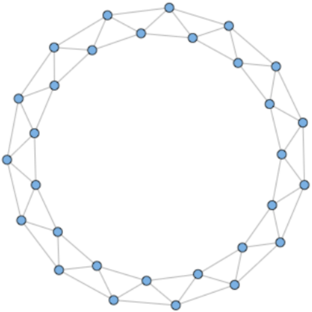
Historical note

- Polya urn model, George Polya, 1923
- Yule process, Udny Yule, 1925
- Distribution of wealth, Herbert Simon,1955
- Evolution of citation networks, cumulative advantage, Derek de Solla Price, 1976
- Preferential attachment network model, Barabasi and Albert, 1999

Local random models vs global optimization models

Small world

Motivation: keep high clustering, get small diameter



Clustering coefficient $C = 1/2$
Graph diameter $d = 8$

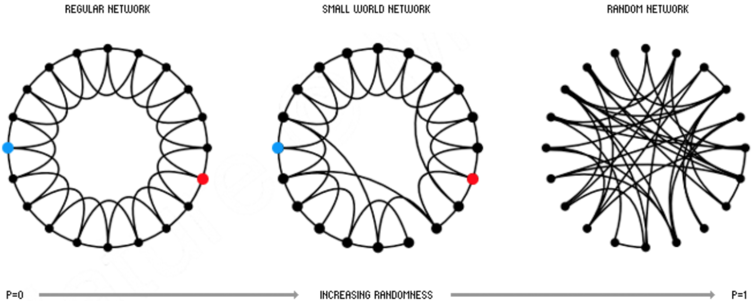
Small world

Watts and Strogatz, 1998

Single parameter model, interpolation between regular lattice and random graph

- start with regular lattice with n nodes, k edges per vertex (node degree), $k << n$
- randomly connect with other nodes with probability p, forms $pnk/2$ "long distance" connections from total of $nk/2$ edges
- $p = 0$ regular lattice, $p = 1$ random graph

Small world



High clustering
High diameter

$$h = \frac{N}{2\bar{k}} \qquad C = \frac{1}{2}$$

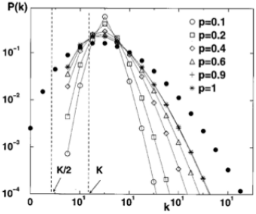
High clustering
Low diameter

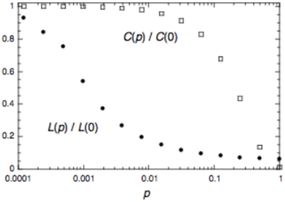
Low clustering
Low diameter

$$h = \frac{\log N}{\log \alpha} \qquad C = \frac{\bar{k}}{N}$$

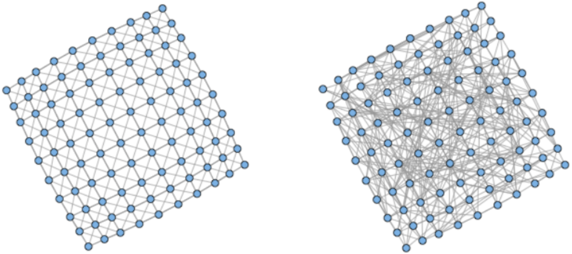
Small world model

- Node degree distribution:
  Poisson like
- Ave. path length $\langle L(p) \rangle$:
  $p \to 0$, ring lattice, $\langle L(0) \rangle = n/2k$
  $p \to 1$, random graph, $\langle L(1) \rangle = \log(n)/\log(k)$
- Clustering coefficient $C(p)$:
  $p \to 0$, ring lattice, $C(0) = 3/4 = const$
  $p \to 1$, random graph, $C(1) = k/n$





Small world model



20% rewiring:

ave. path length = 3.58 $\rightarrow$ ave. path length = 2.32

clust. coeff = 0.49 $\rightarrow$ clust. coeff = 0.19

Model comparison

	Random	BA model	WS model	Empirical networks
P(k)	$\frac{\lambda^k e^{-\lambda}}{k!}$	k^{-3}	poisson like	power law
C	$\langle k \rangle / N$	$N^{-0.75}$	const	large
$\langle L \rangle$	$\frac{\log(N)}{\log(\langle k \rangle)}$	$\frac{\log(N)}{\log\log(N)}$	$\log(N)$	small

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)
- Emergence of Scaling in Random Networks, A.L. Barabasi and R. Albert, Science 286, 509-512, 1999
- Collective dynamics of small-world networks. Duncan J. Watts and Steven H. Strogatz. Nature 393 (6684): 440-442, 1998