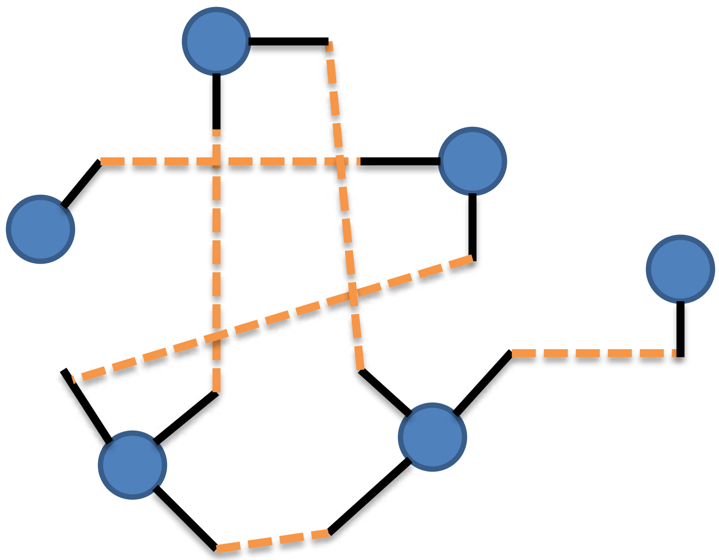
Random Graphs and Configuration Model

Degrees: 1, 1, 2, 2, 3, 3



1. Add *n* nodes
2. Add initial *d(i)* stubs to each *i*
3. Connect stubs iteratively

Problems? Total degree is even; Can create self-loops, multi-edges

Configuration Model

Multi-edges: Probability of adding an edge between i and j with degrees k_i, and k_j is

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

in the limit we can omit -1

Probability of second edge is $(k_i - 1)(k_j - 1)/2m$

Expected number of multiedges in conf model

$$\frac{1}{2(2m)^2}\sum_{ij} k_i k_j (k_i - 1)(k_j - 1) = \frac{1}{2\langle k\rangle^2 n^2}\sum_i k_i(k_i - 1)\sum_j k_j(k_j - 1) = \frac{1}{2}\left[\frac{\langle k^2\rangle - \langle k\rangle}{\langle k\rangle}\right]^2$$

Similar result for self-edges

$$\sum_i p_{ii} = \sum_i \frac{k_i(k_i - 1)}{4m} = \frac{\langle k^2\rangle - \langle k\rangle}{2\langle k\rangle}$$

Conclusion? Expected number of multi-edges remains constant as network grows.

Expected number of common neighbors

$$n_{ij} = \sum_l \frac{k_i k_l}{2m}\frac{k_j(k_l - 1)}{2m} = \frac{k_i k_j}{2m}\frac{\sum_l k_l(k_l - 1)}{n\langle k\rangle} = p_{ij}\frac{\langle k^2\rangle - \langle k\rangle}{\langle k\rangle}$$

i is connected to *l*

j is connected to *l* given *il*

Random graphs with given expected degree

$\forall i \in V$ define parameter c_i. Then edge probability

$$p_{ij} = \begin{cases} c_i c_j / 2m & i \neq j \\ c_i^2 / 4m & i = j \end{cases}, \text{ where } \sum_i c_i = 2m$$

average number of edges in network

$$\sum_{i \leq j} p_{ij} = \sum_{i<j} \frac{c_i c_j}{2m} + \sum_i \frac{c_i^2}{4m} = m$$

average degree

$$\langle k_i \rangle = 2p_{ii} + \sum_{j \neq i} p_{ij} = \frac{c_i^2}{2m} + \sum_{j \neq i} \frac{c_i c_j}{2m} = \sum_j \frac{c_i c_j}{2m} = c_i$$

More properties of random model

Excess degree distribution is the probability distribution, for a vertex reached by following an edge, of the number of other edges attached to that vertex.

$$q_k = \frac{(k+1)p_{k+1}}{\langle k \rangle}$$

Two academic collaboration networks, in which scientists are connected together by edges if they have coauthored scientific papers, and a snapshot of the structure of the Internet at the autonomous system level.

Network	n	Average degree	Average neighbor degree	$\frac{\langle k^2 \rangle}{\langle k \rangle}$
Biologists	1 520 252	15.5	68.4	130.2
Mathematicians	253 339	3.9	9.5	13.2
Internet	22 963	4.2	224.3	261.5

According to these results a biologist's collaborators have, on average, more than four times as many collaborators as they do themselves. On the Internet, a node's neighbors have more than 50 times the average degree! Note that in each of the cases in the table the configuration model value of $\langle k^2 \rangle / \langle k \rangle$ overestimates the real average neighbor degree.

M. Newman "Networks"

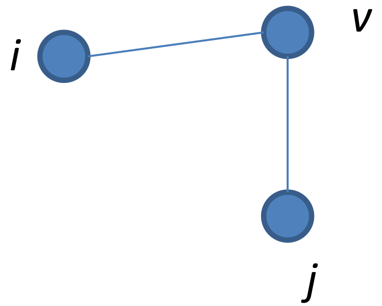
More properties of random model

Excess degree distribution is the probability distribution, for a vertex reached by following an edge, of the number of other edges attached to that vertex.

$$q_k = \frac{(k+1)p_{k+1}}{\langle k \rangle}$$

Clustering coefficient for configuration model

$$C = \sum_{k_i, k_j = 0}^{\infty} q_{k_i} q_{k_j} \frac{k_i k_j}{2m} = \frac{1}{2m} \Big(\sum_{k=0}^{\infty} k q_k \Big)^2 = \cdots = \frac{1}{n} \frac{(\langle k \rangle^2 - \langle k \rangle)^2}{\langle k \rangle^3}$$



Generating Functions and Degree Distributions

For degree and excess degree distributions we define generating functions

$$g_0(z) = \sum_{k=0}^{\infty} p_k z^k \text{ and } g_1(z) = \sum_{k=0}^{\infty} q_k z^k, \text{ respectively}$$

They are not independent

we add zero term because of infinity

$$g_1(z) = \frac{1}{\langle k \rangle} \sum_{k=0}^{\infty} (k+1) p_{k+1} z^k = \frac{1}{\langle k \rangle} \sum_{k=0}^{\infty} k p_k z^{k-1} = \frac{1}{\langle k \rangle} \frac{\mathrm{d} g_0}{\mathrm{d} z} = \frac{g_0'(z)}{g_0'(1)}$$

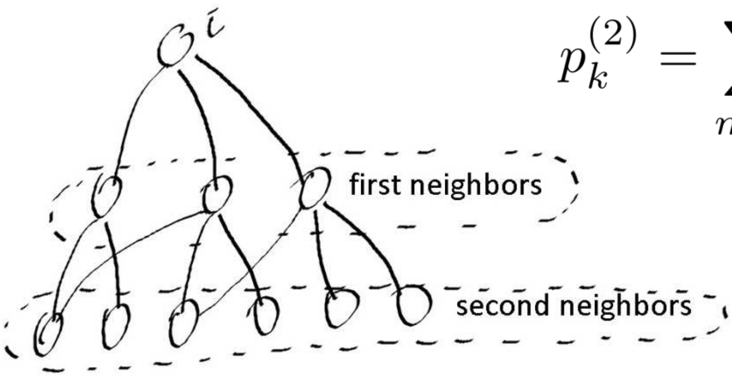
Example (Poisson): $p_k = e^{-c} \frac{c^k}{k!} \;\Rightarrow\; g_0(z) = e^{c(z-1)}$ and $g_1(z) = e^{c(z-1)}$

Example (power-law): $p_k = Ck^{-\alpha} \;\Rightarrow\; g_0(z) = \frac{Li_\alpha(z)}{\zeta(\alpha)}$. Thus,

$$g_1(z) = \frac{Li_{\alpha-1}(z)}{z Li_{\alpha-1}(1)} = \frac{Li_{\alpha-1}(z)}{z\zeta(\alpha-1)}$$

Number of second neighbors of a vertex

Probability that i has exactly k second neighbors



$$p_k^{(2)} = \sum_{m=0}^{\infty} p_m P^{(2)}(k|m)$$

degree distribution

Probability of having k second neighbors given m first neighbors

Prob excess degrees of m first neighbors take values $j_1, j_2, \ldots, j_m$

$$P^{(2)}(k|m) = \underbrace{\sum_{j_1=0}^{\infty} \cdots \sum_{j_m=0}^{\infty}}_{} \delta\left(k, \sum_{r=1}^{m} j_r\right) \prod_{r=1}^{\infty} q_{j_r}$$

all sets of values $j_1, j_2, \ldots, j_m$ that sum to k

$$g^{(2)}(z) = \sum_{k=0}^{\infty} p_k^{(2)} z^k = \sum_{k=0}^{\infty} z^k \cdot \sum_{m=0}^{\infty} p_m \sum_{j_1=0}^{\infty} \cdots \sum_{j_m=0}^{\infty} \delta\left(k, \sum_{r=1}^{m} j_r\right) \prod_{r=1}^{m} q_{j_r} = \cdots =$$

$$= \sum_{m=0}^{\infty} p_m \cdot \left(\sum_{j=0}^{\infty} q_j z^j\right)^m = g_0(g_1(z))$$

generating function of $p_k^{(2)}$

Conclusion: Once we know generating functions of g_0 and g_1 the generating function of second neighbor distribution is straightforward to calculate. Moreover, this can be extended to

$$g^{(3)}(z) = \sum_{k=0}^{\infty} \sum_{m=0}^{\infty} p_m^{(2)} P^{(3)}(k|m) z^k = \sum_{n=0}^{\infty} p_m^{(2)} (g_1(z))^m = g^{(2)}(g_1(z)) = g_0(g_1(g_1(z)))$$

$$\Longrightarrow g^{(d)}(z) = g^{(d-1)}(g_1(z)) = g_0(g_1(\ldots g_1(z)\ldots))$$

Problem: Sometimes it is difficult to extract explicit probabilities for numbers of second neighbors and it is hard to evaluate *n* derivatives (in order to recover the probabilities).
Solution: calculate the average number of neighbors at distance *d.* At *z=1* of the first derivative we can evaluate the average of a distribution (see Slide 16).

$$\frac{\mathrm{d}g^{(2)}}{\mathrm{d}z} = g_0'(g_1(z))g_1'(z) \xRightarrow{z=1, g_1(1)=1} \underbrace{c_2}_{\text{mean number of second neighbors}} = g_0'(1)g_1'(1) \xRightarrow{g_0'(1)=\langle k\rangle} g_1'(k) = \sum_{k=0}^{\infty} kq_k =$$

$$\frac{1}{\langle k\rangle}\sum_{k=0}^{\infty} k(k+1)p_{k+1} = \frac{1}{\langle k\rangle}(\langle k^2\rangle - \langle k\rangle)$$

Conclusion: $c_2 = \langle k^2\rangle - \langle k\rangle$ and more general

$$c_d = \left(\frac{c_2}{c_1}\right)^{d-1} c_1 \Longrightarrow \begin{array}{l}\text{Condition of giant component's existance} \\ \text{in configuration model is } \langle k^2\rangle - 2\langle k\rangle > 0\end{array}$$

[MR] A critical point for random graphs with given degree sequence

Let's use theory for more practical results …

Given a network with **power-law distribution** $p_k = Ck^{-\alpha},\ \alpha > 0,\ k > 0$
Reminder: C is calculated from normalization condition, i.e., $C = 1/\zeta(\alpha)$

$$p_k = \begin{cases} 0 & k = 0 \\ k^{-\alpha}/\zeta(\alpha) & k > 0 \end{cases}$$

This network will have a giant component iff $\langle k^2 \rangle - 2\langle k \rangle > 0$

$$\langle k \rangle = \sum_{k=0}^{\infty} k p_k = \frac{1}{\zeta(\alpha)} \sum_{k=1}^{\infty} k^{-\alpha+1} = \frac{\zeta(\alpha-1)}{\zeta(\alpha)}$$

$$\langle k^2 \rangle = \sum_{k=0}^{\infty} k^2 p_k = \frac{1}{\zeta(\alpha)} \sum_{k=1}^{\infty} k^{-\alpha+2} = \frac{\zeta(\alpha-2)}{\zeta(\alpha)}$$

$$\Longrightarrow \zeta(\alpha - 2) > 2\zeta(\alpha - 1)$$

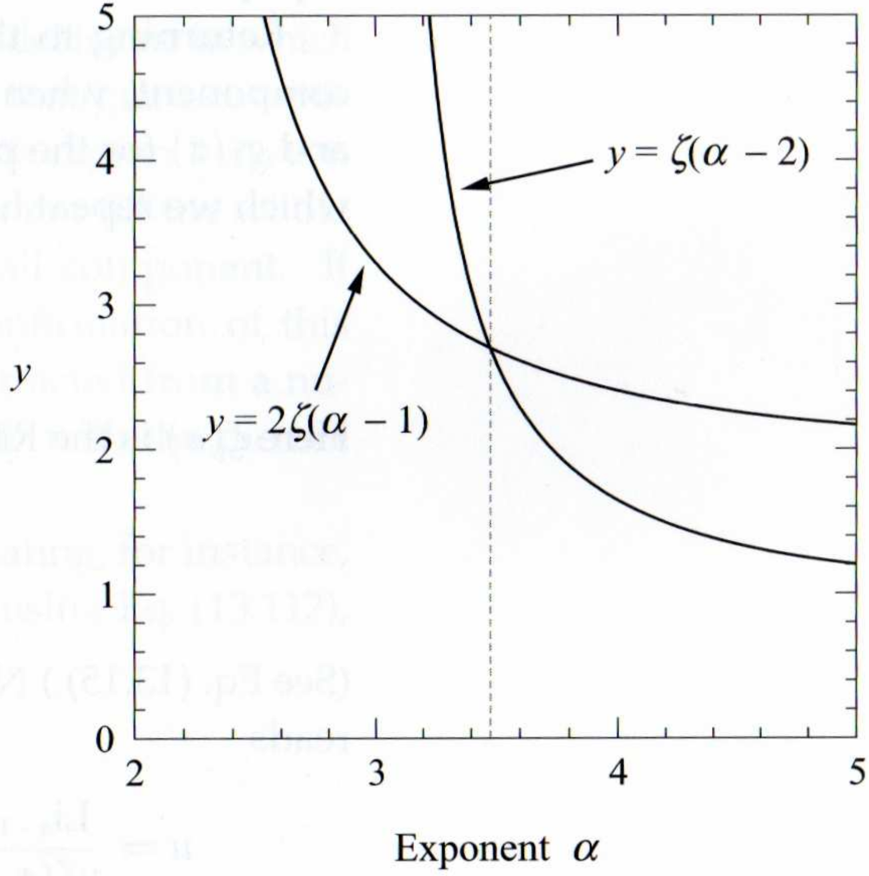


Figure 13.8: Graphical solution of Eq. (13.138). The configuration model with a pure power-law degree distribution (Eq. (13.133)) has a giant component if $\zeta(\alpha-2) > 2\zeta(\alpha-1)$. This happens for values of α below the crossing point of the two curves.

Newman "Networks: An Introduction"

Models of Network Formation

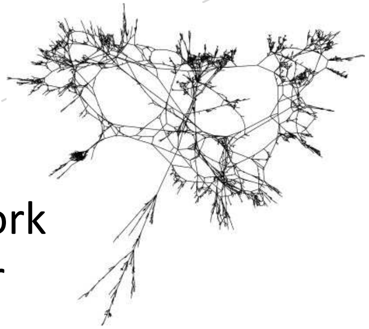
Happy families are all alike, every unhappy family is unhappy in its own way.

Leo Tolstoy

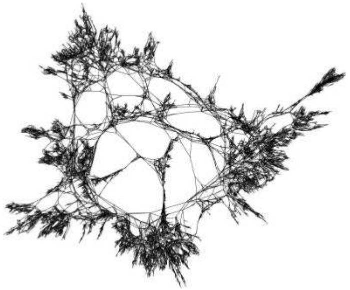
Fundamental theoretical and practical questions

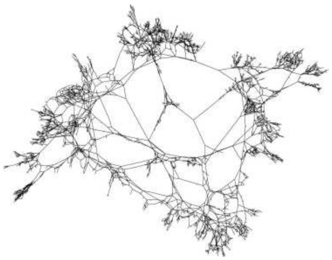
- What are the fundamental processes that form a network?
- How to predict its future structure?
- Why should network have property X?
- Will my algorithm/heuristic work on networks created by similar processes?

Artificial network



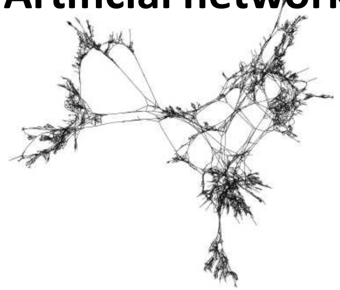
Artificial network





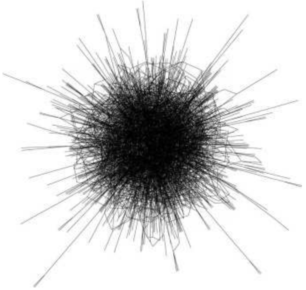
Original network

Artificial network



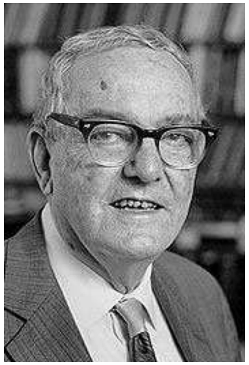
Artificial network

Artificial network



Is it similar to the original network?

Rich-get-richer effect



Herbert Simon
1916-2001

Analyzed the power laws in economic data, suggested explanation of wealth distribution: return of investment is proportional to the amount invested, i.e., wealthy people will get more and more.

Simon (1976). “On a class of skew distribution functions"

Derek Price
1922-1983

Studied information science; in particular, citation networks; his main assumption was about newly appearing papers that cite old papers with probability proportional to the number of citations those old papers have → the model is similar to Simon’s model.

Price (1976). "A general theory of bibliometric and other cumulative advantage processes"

Price's model