Models of Network Formation
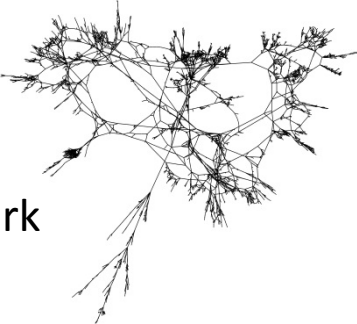
Happy families are all alike, every unhappy family is unhappy in its own way.
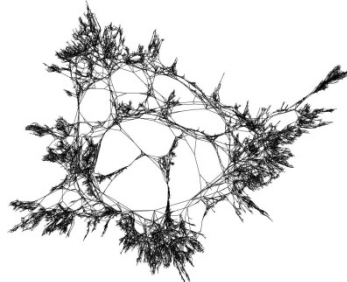
Leo Tolstoy

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
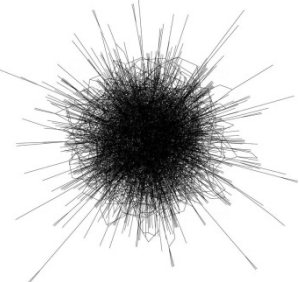
Artificial network

Artificial network

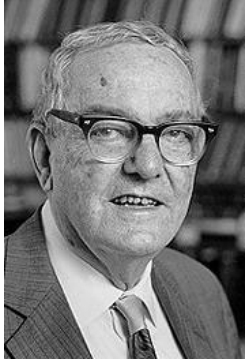**Artificial network**

Original 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"
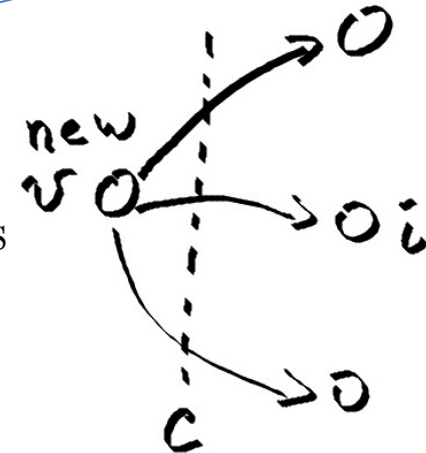
Price's model

c on the average

random

random

New

Consider adding a single vertex v in Price's model, where $p_q(n)$ is the fraction of vertices in the network with in-degree q.

Probability of $v \longrightarrow i$ citation is

$$\frac{q_i + a}{\sum_i (q_i + a)} = \frac{q_i + a}{n\langle q\rangle + na} = \frac{q_i + a}{n(c + a)}$$

Expected number of new citations to all nodes with degree q is

nodes with in-deg q

$$np_q(n) \cdot c \cdot \frac{q + a}{n(c + a)} = \frac{c(q + a)}{c + a}p_q(n)$$

Thus, the number of vertices with in-deg q after adding v is

$$(n + 1)p_q(n + 1) = np_q(n) + \frac{c(q - 1 + a)}{c + a}p_{q-1}(n) - \frac{c(q + a)}{c + a}p_q(n)$$

were previously
with in-deg q were with in-deg q-1   were with in-deg q and left that set

$$\implies p_q = \frac{q + a - 1}{q + a + 1 + a/c}p_{q-1} \implies p_q \, (q + a)^{-\alpha}$$

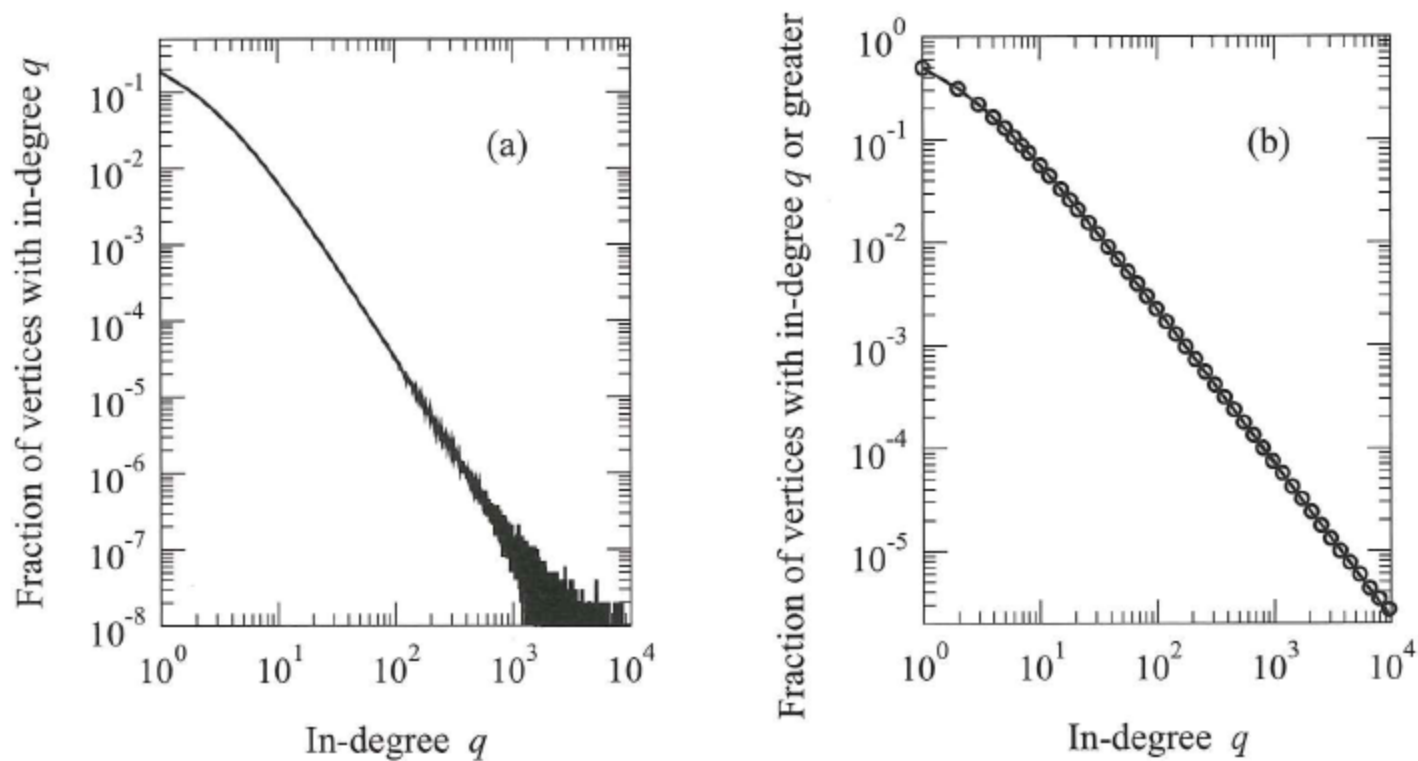Use properties of gamma
and beta functions

Figure 14.2: Degree distribution in Price's model of a growing network. (a) A histogram of the in-degree distribution for a computer-generated network with $c = 3$ and $a = 1.5$ which was grown until it had $n = 10^8$ vertices. The simulation took about 80 seconds on the author's computer using the fast algorithm described in the text. (b) The cumulative distribution function for the same network. The points are the results from the simulation and the solid line is the analytic solution, Eq. (14.34).
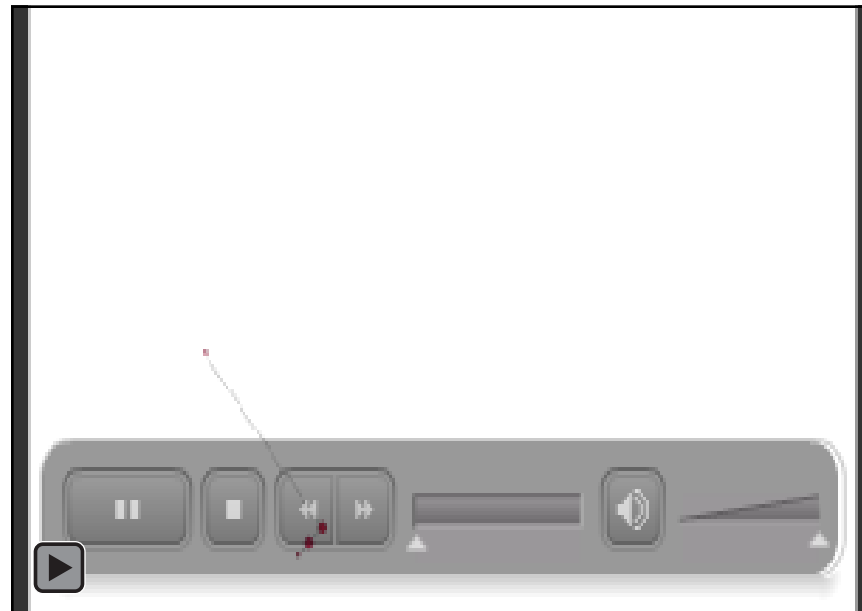
Newman "Networks, An Introduction"

Preferential Attachment (Barabasi-Albert)

- Initialize network with m_0 nodes ($m_0 \geq 2$, $d(i) \geq 1$)

- Add node i, connect it to exactly c out of m existing nodes with probability

$$\Pr[i - j] = \frac{k_j}{\sum_l k_l}$$

- Repeat previous step or stop if $|V| = n$

$$p_k \sim k^{-3}$$

http://probaperception.blogspot.com

Non-linear Preferential Attachment

Q: What if the probability of attachment is not linear in the degree of node

a_k - *attachment kernel*, i.e., functional form of the attachment probability

In B-A model $a_k = k$

In non-linear model $a_k = k^\gamma \leftarrow$ not a probability! normalized form $a_k / \sum_i a_{k_i}$

$p_k(n)$ - fraction of vertices with degree k when $|V| = n$

Expected number of k-deg nodes with a new connection when one node is added

$$np_k(n)\, c \frac{a_k}{\sum_i a_{k_i}} = \frac{c}{\mu(n)} a_k p_k(n) \qquad \text{(see Slide 3 for } a_k = k\text{)}$$

edges added at each step

$$\mu(n) = \frac{1}{n}\sum_{i=1}^{n} a_{k_i} = \sum_k a_k p_k(n)$$

Master equation for $p_k(n)$

$$\Rightarrow \quad (n+1)\,p_k(n+1) = np_k(n) + \frac{c}{\mu(n)}\left(a_{k-1}p_{k-1}(n) - a_k p_k(n)\right)$$

were with deg k-1   were with deg k and left

in ∞

$$\Rightarrow \quad p_c = \frac{\mu/c}{a_c + \mu/c} \qquad p_k = \frac{a_{k-1}}{a_k + \mu/c} p_{k-1}$$

Krapivsky, P. L., Redner, S., and Leyvraz, F., Connectivity of growing random networks, *Phys. Rev. Lett.*
Jeong, H., Néda, Z., and Barabási, A.-L., Measuring preferential attachment in evolving networks, *Europhys. Lett.*

If $a_k = k^\gamma$

$$\text{If } \gamma < 1 \quad p_k = \frac{\mu}{ck^\gamma} \prod_{r=c}^{k} \left(1 + \frac{\mu}{cr^\gamma}\right)^{-1}$$

No power-law tail!
(see handout with Taylor ser exp)

For $1/2 < \gamma < 1$

$$p_k \sim k^{-\gamma} \exp\left(-\frac{\mu k^{1-\gamma}}{c(1-\gamma)}\right)$$

For $\gamma = 1/2$

$$p_k \sim \left(\sqrt{k}\right)^{\mu^2/c^2 - 1} \exp\left(-\frac{2\mu}{c}\sqrt{k}\right)$$
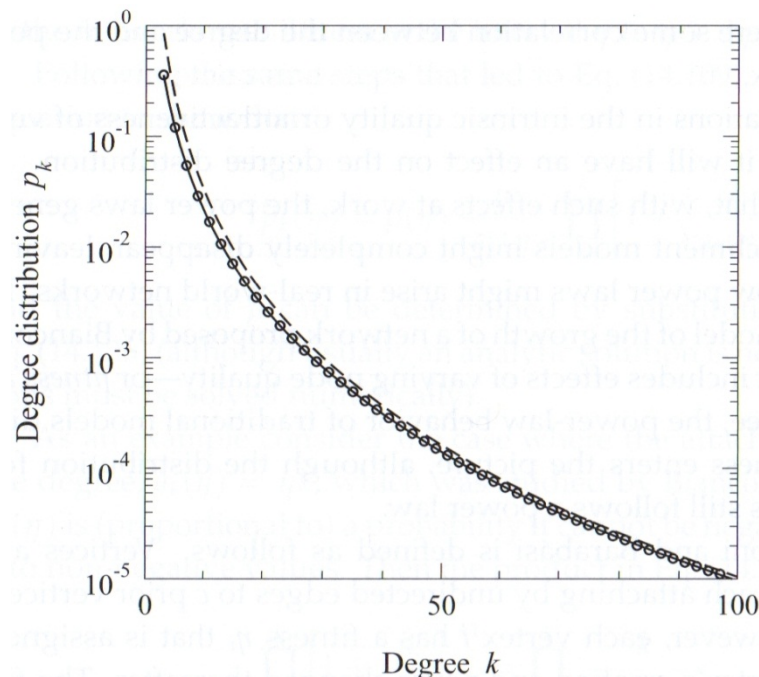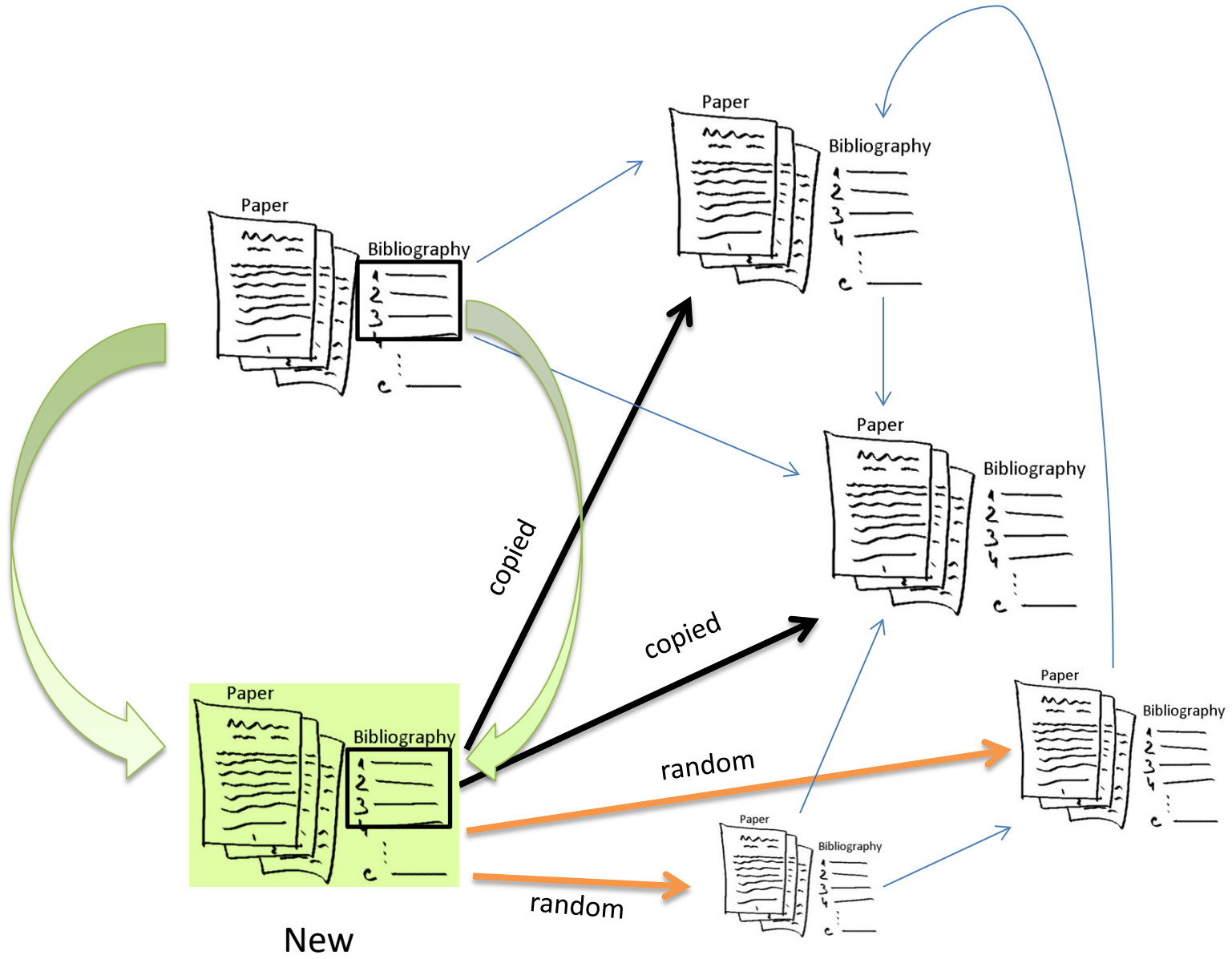

Figure 14.8: Degree distribution for sublinear preferential attachment. This plot shows the fraction p_k of vertices with degree k in a growing network with attachment kernel k^γ as described in the text. In this case $\gamma = 0.8$ and $c = 3$. The points are results from computer simulations, averaged over 100 networks of (final) size 10^7 vertices each. The solid line is the exact solution, Eq. (14.112), evaluated numerically. The dashed line is the asymptotic form, Eq. (14.119), with the overall constant of proportionality chosen to coincide with the exact solution for large values of k.

Vertex Copying Models

New

Algorithm:

- Initialize network with $n_0 > c$ nodes $(d(\cdot) \stackrel{\mathrm{random}}{=} c)$

- Choose uniformly at random existing vertex i with prob $\frac{1}{n}$

- Add new node j with out-degree c

- Go through all bibliographic entries of i and either (a) copy it to j with prob γ or (b) add new random entry to j with prob $1 - \gamma$

- Repeat previous step or stop if $|V| = n$

1/n is a probability to choose a node with connections to *i*

When new node j is added ...
- it will have γc copied entries on the average.
- probability new edge is copied $\mathrm{Pr}_1[j \to i] = \gamma q_i / n$, where $q_i = d^-(i)$
- probability new edge is randomly created $\mathrm{Pr}_2[j \to i] = (1 - \gamma)c/n$
- if $p_q(n)$ - fraction of nodes with in-deg q then **total expected number of nodes of in-deg q receiving new edge**

i gets a new link Pr$_1$ + Pr$_2$

$$np_q(n) \times \frac{\gamma q + (1 - \gamma)c}{n} = (\gamma q + (1 - \gamma)c)\,p_q(n)$$

...

• if $p_q(n)$ - fraction of nodes with in-deg q then **total expected number of nodes of in-deg q receiving new edge**

$$np_q\left(n\right) \times \frac{\gamma q + \left(1 - \gamma\right)c}{n} = \left(\gamma q + \left(1 - \gamma\right)c\right)p_q\left(n\right)$$

Define

$$a = c\left(\frac{1}{\gamma} - 1\right) \implies \gamma = \frac{c}{c + a}$$

then

$$\left(\gamma q + \left(1 - \gamma\right)c\right)p_q\left(n\right) = \frac{c\left(q + a\right)}{c + a}p_q\left(n\right)$$

Same as in Price's model!

Conclusion: Vertex copying behaves as the Price's model with $a = c(1/\gamma - 1)$.
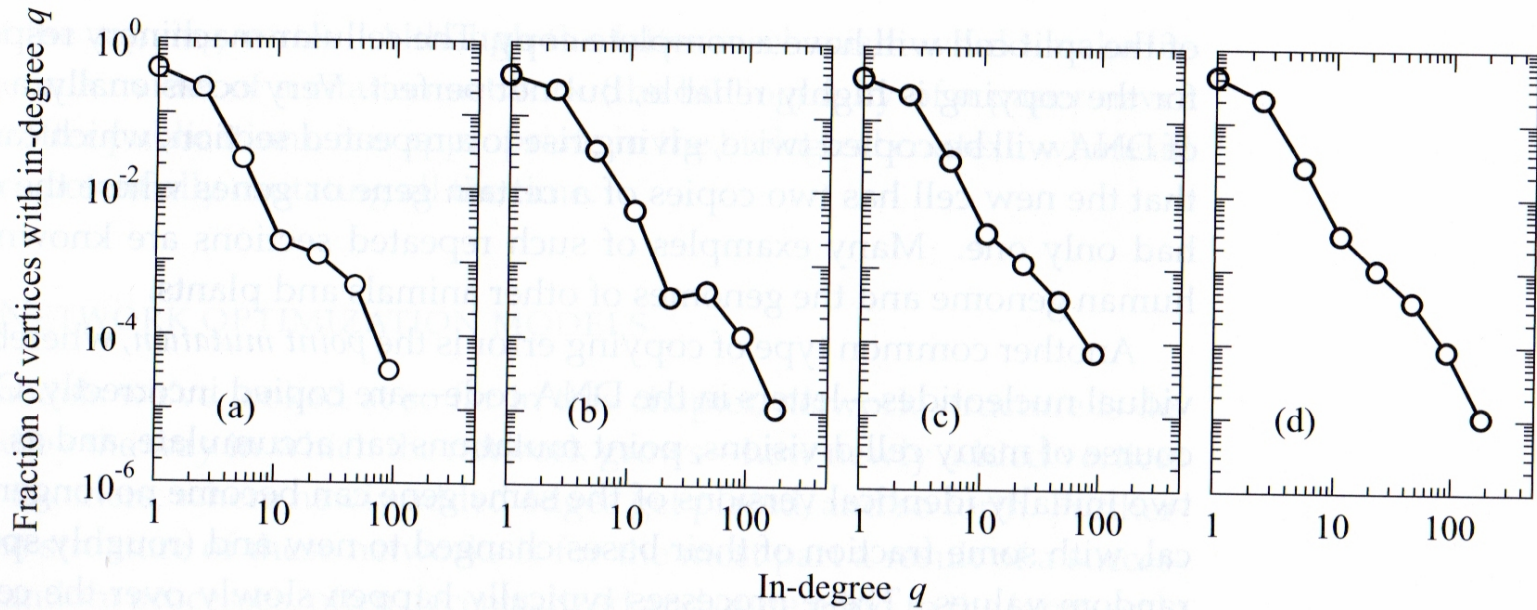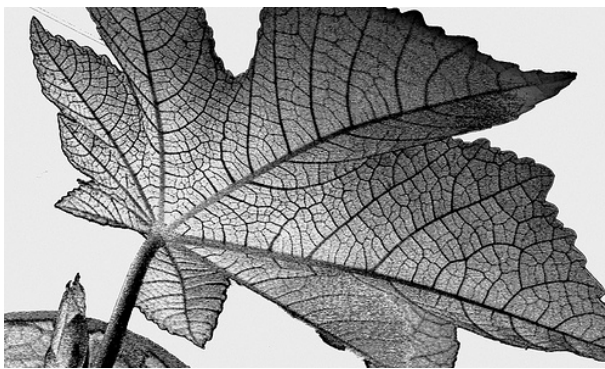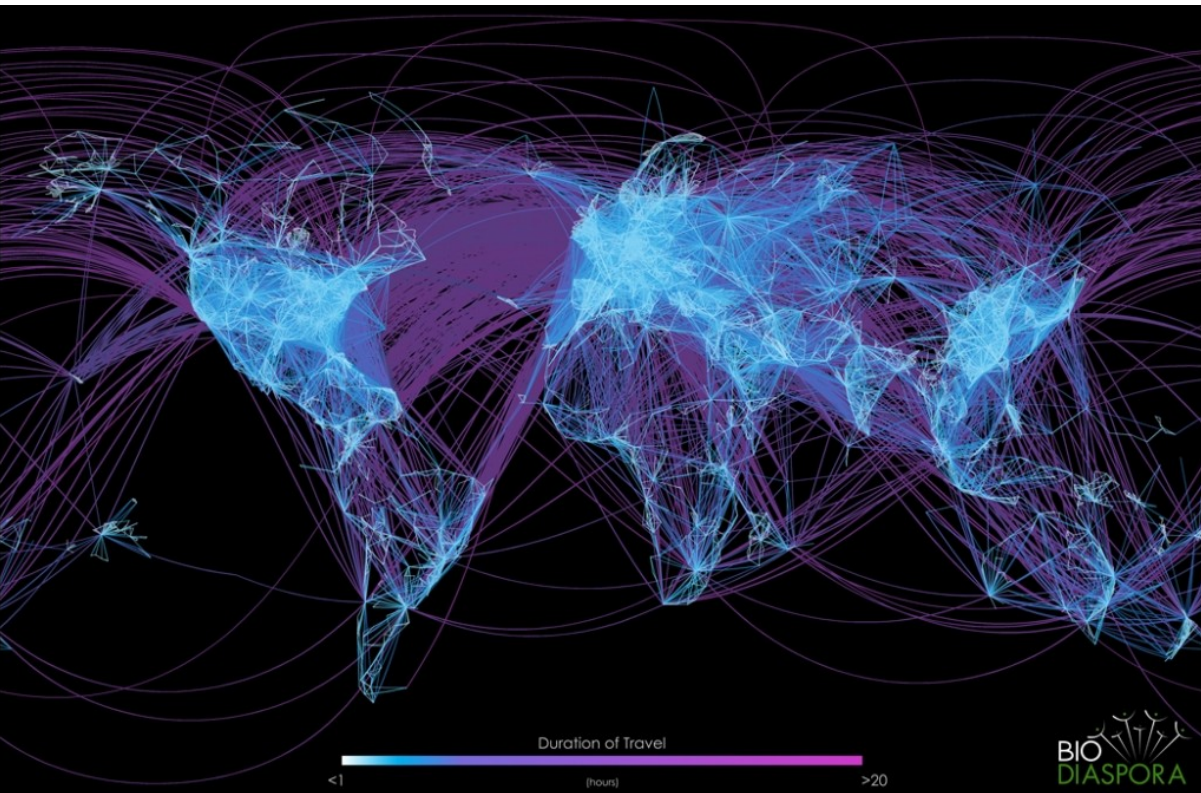
Figure 14.9: Distribution of in-degrees in the metabolic networks of various organisms. Jeong *et al.* [166] examined the degree distributions of the known portions of the metabolic networks of 43 organisms, finding some of them to follow power laws, at least approximately. Show here are the in-degree distributions for (a) the archaeon *A. fulgidus*, (b) the bacterium *E. coli*, (c) the worm *C. elegans* (a eukaryote), and (d) the aggregated in-degree distribution for all 43 organisms. After Jeong *et al.* [166].

Newman "Networks, An Introduction"

Network Optimization Models

Simplified model of operating the network

m - number of edges.

l - mean shortest path between all pairs of nodes.

Assumption:

- (m) cost of running the network is proportional to the number of routes it operates;

- (l) customer dissatisfaction measure.

We are interested in minimizing both m and l but minimizing l maximizes m.

Consider a model with

$$E(m, l) = \lambda m + (1 - \lambda)l$$

Given $|V| = n$ what if we minimize $E(m, l)$?

large $\lambda \Rightarrow$ tree, $m = n - 1 \Rightarrow$ search over all possible trees to minimize l

small $\lambda \Rightarrow$ non-star-graph solutions appear when $\lambda < 2/(n^2 + 2)$
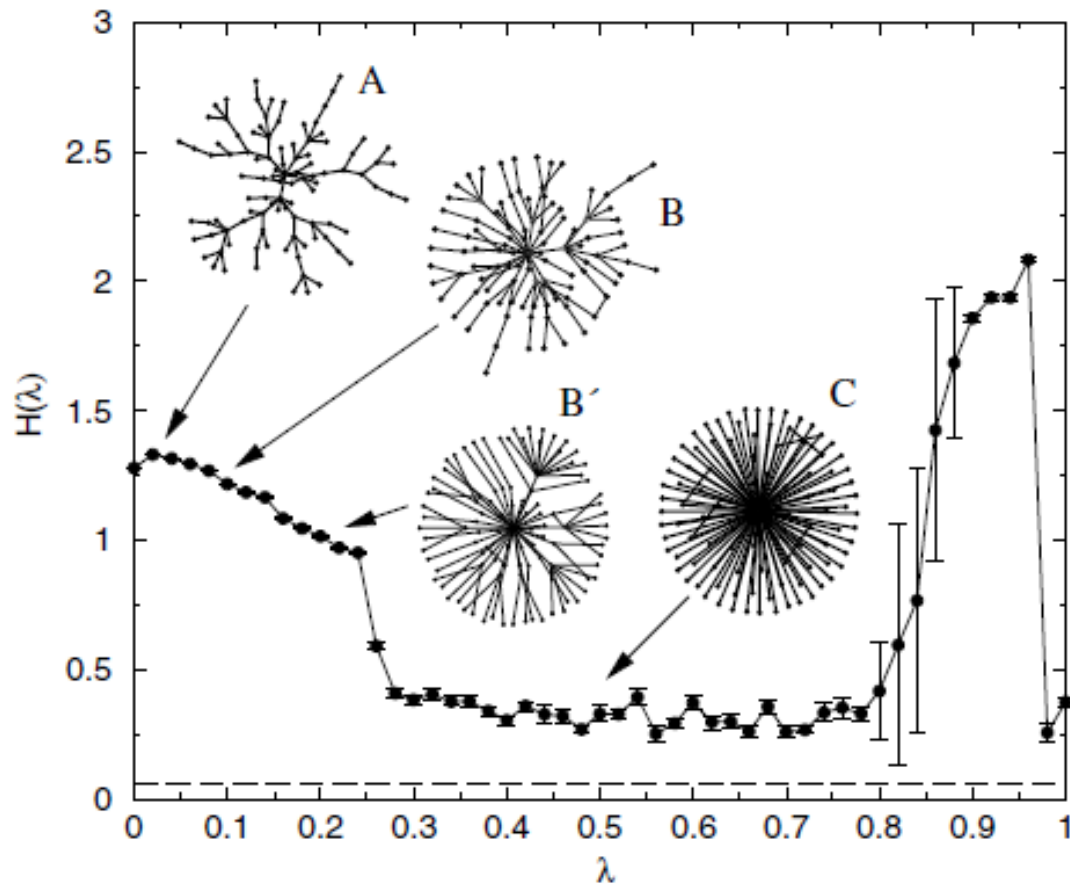
Fig. 7.4. Average (over 50 replicas) degree entropy as a function of λ with $n = 100$, $T = \binom{n}{2}$, $\nu = 2/\binom{n}{2}$ and $\rho(0) = 0.2$. Optimal networks for selected values of λ are plotted. The entropy of a star network, $H_{star} = \log n - [(n-1)/n]\log(n-1) = 0.056$ is provided as reference (dashed line). **A**: an exponential-like network with $\lambda = 0.01$. **B**: A scale-free network with $\lambda = 0.08$. Hubs involving multiple connections and a dominance of nodes with one connection can be seen. **C**: a star network with $\lambda = 0.5$. **B'**: a intermediate graph between **B** and **C** in which many hubs can be identified

Ferrer i Cancho, Sole "Optimization in Complex Networks"