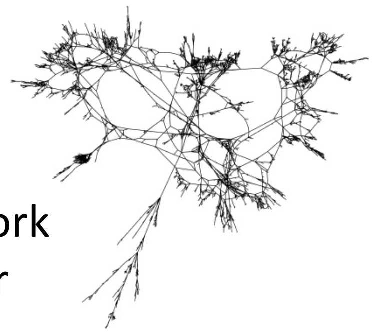
Models of Network Formation

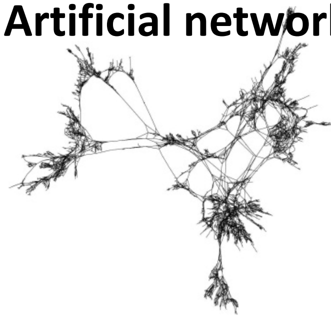
Fundamental theoretical and practical questions

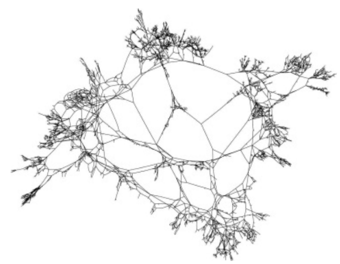
- What are the fundamental processes that form a network?
- How to predict its future structure?
- Why should networks have property X?
- Will my algorithm/heuristic work on networks created by similar processes?

Artificial network



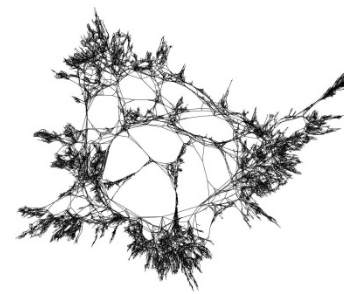
Artificial network





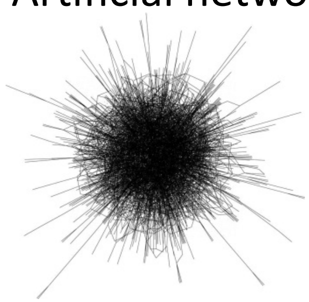
Original network

Artificial network



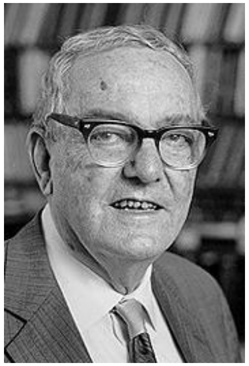
Artificial network

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

The crucial central assumption of Price's model is that a newly appearing paper cites previous ones chosen at random **with probability proportional to the number of citations those previous papers already have**.

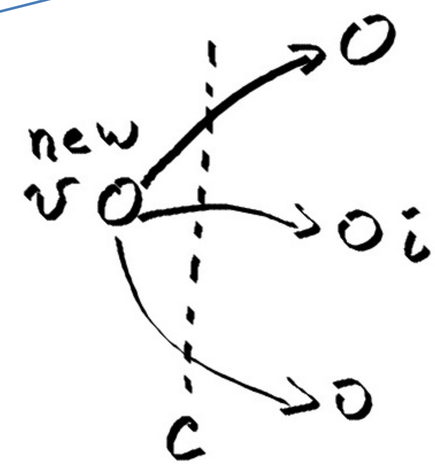
Consider adding a single vertex v in Price's model, where $p_q(n)$ is the fraction of vertices in the network with in-degree q.

Probability of $v \longrightarrow i$ citation is

constant to get some citations for "free"

degree of node *i*

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



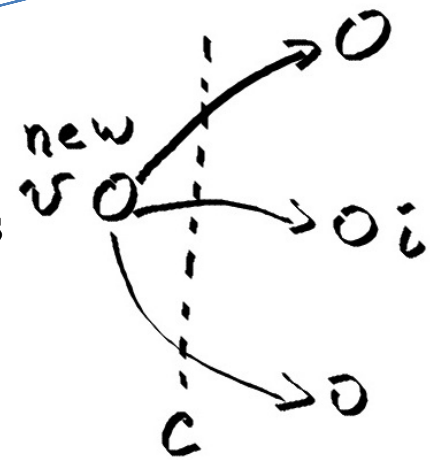
Consider adding a single vertex v in Price's model, where $p_q(n)$ is the fraction of vertices in the network with in-degree q.

Probability of $v \longrightarrow i$ citation is

constant to get some citations for "free"

degree of node *i*

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



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

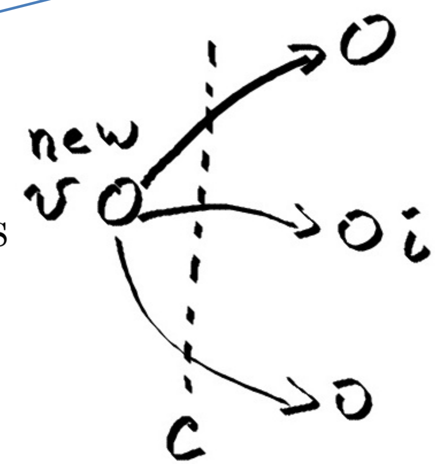
Consider adding a single vertex v in Price's model, where $p_q(n)$ is the fraction of vertices in the network with in-degree q.

Probability of $v \longrightarrow i$ citation is

constant to get some citations for "free"

degree of node *i*

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



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

nodes with in-deg *q*

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

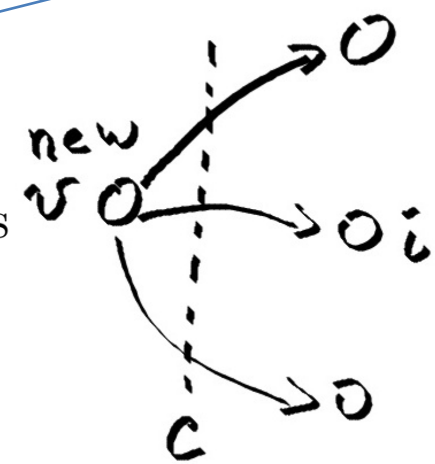
Thus, the number of vertices with in-deg q after adding v is

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

constant to get some citations for "free"

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



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

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

Use properties of gamma and beta functions

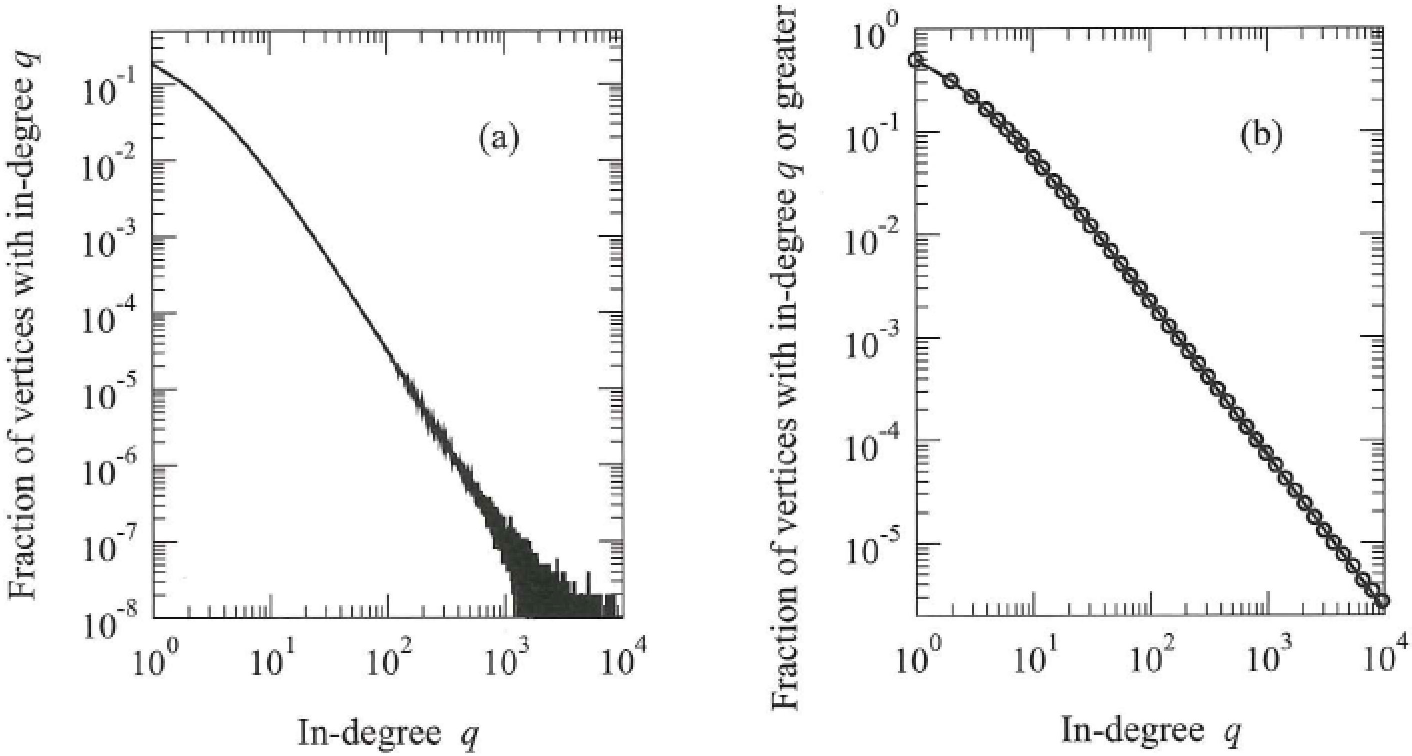


Figure 14.2: Degree distribution in Price's model of a growing network. (a) A histogram of the in-degree distribution for a computer-generated network with $c = 3$ and $a = 1.5$ which was grown until it had $n = 10^8$ vertices. The simulation took about 80 seconds on the author's computer using the fast 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"

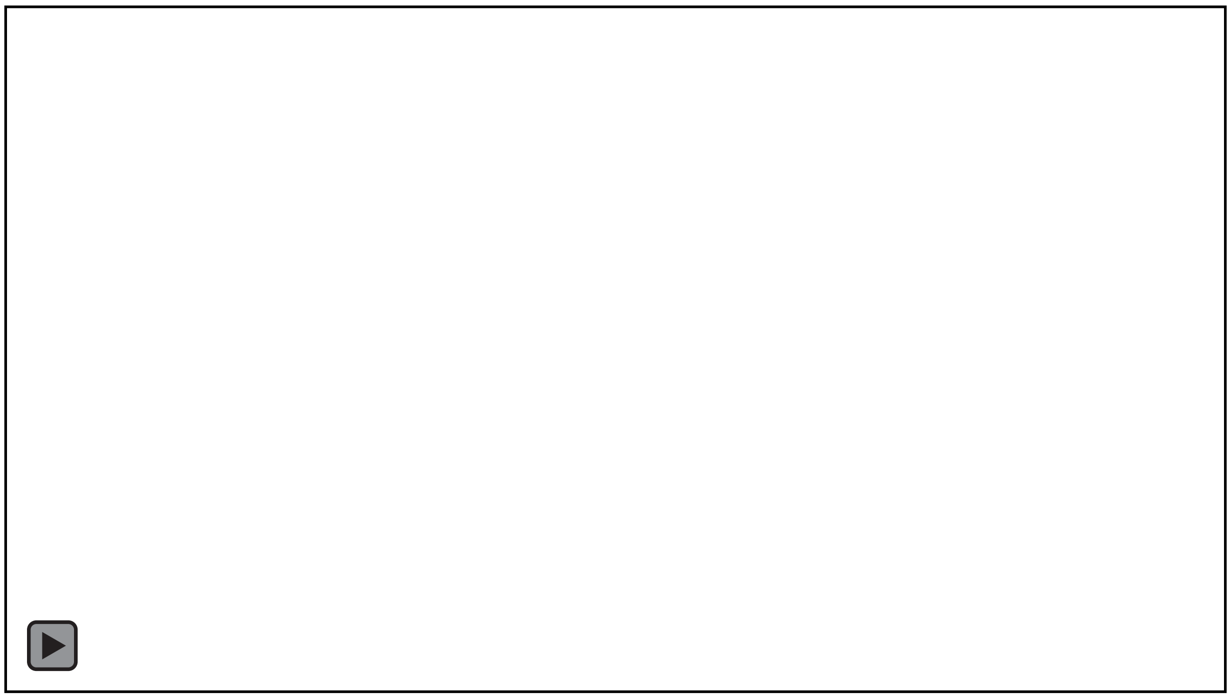
Preferential Attachment (Barabasi-Albert)

- 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

$$np_k(n)\,c\frac{a_k}{\sum_i a_{k_i}} = \frac{c}{\mu(n)} a_k p_k(n)$$

edges added at each step

(see Slide 3 for $a_k = k$)

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

Master equation for $p_k(n)$

new vertices of deg k

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

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 $\gamma = 1/2$

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

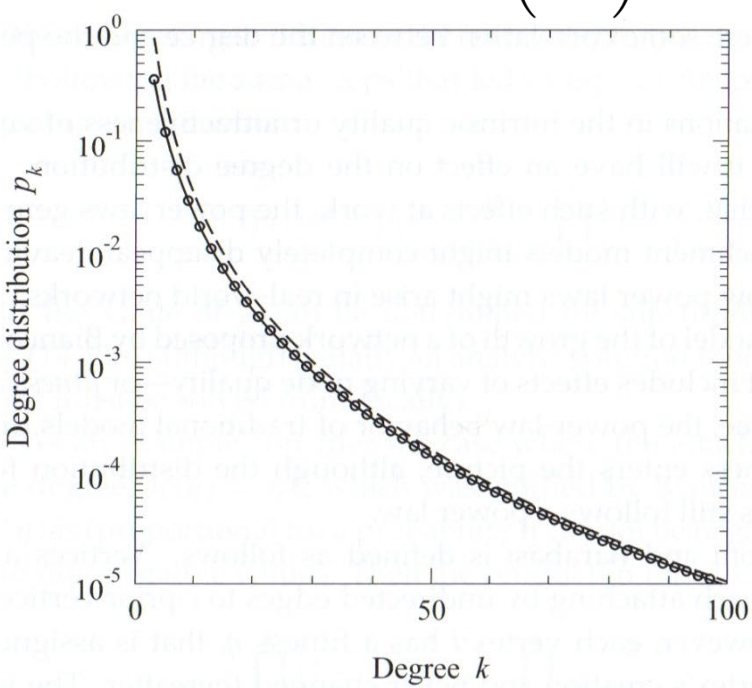
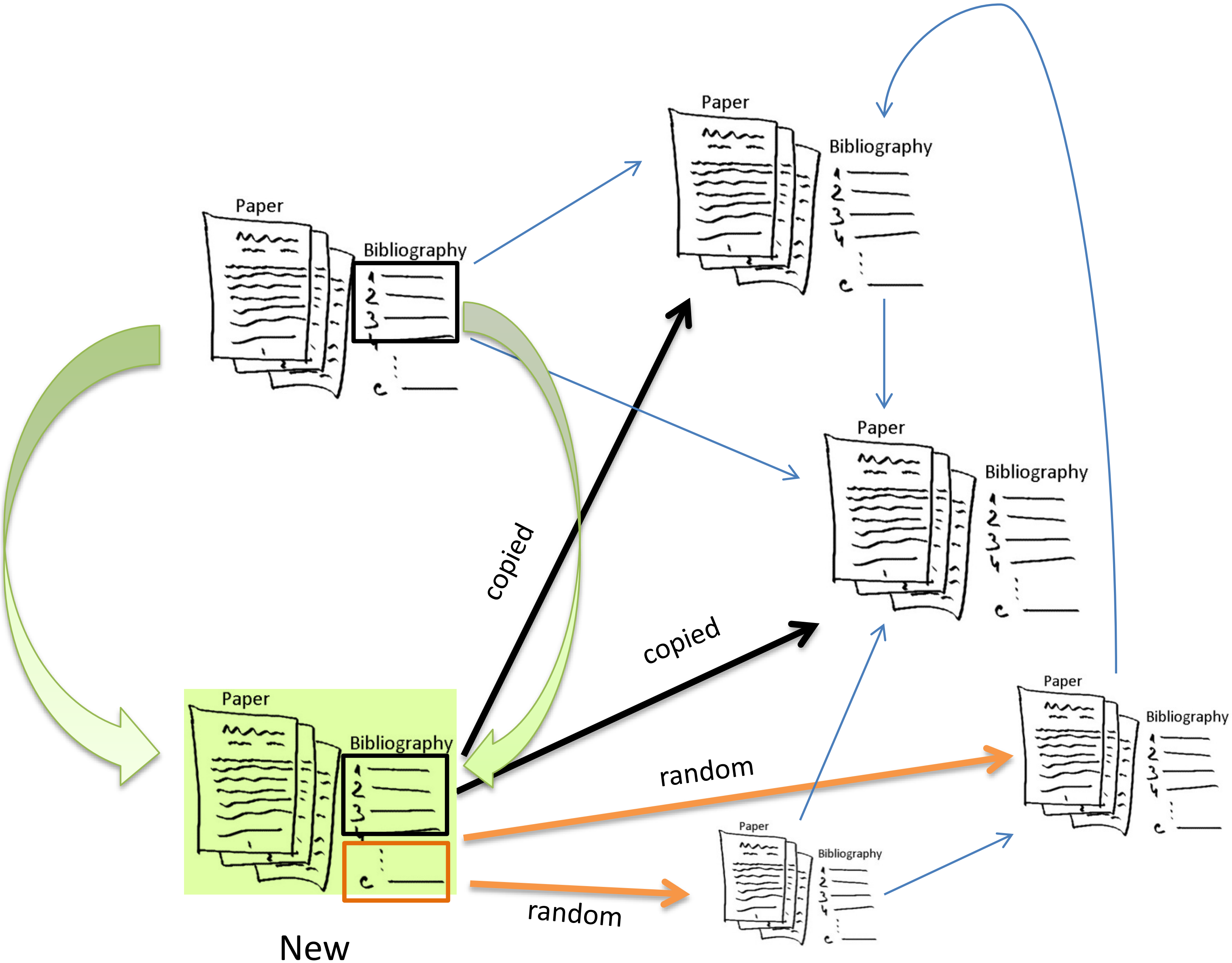


Figure 14.8: Degree distribution for sublinear preferential attachment. This plot shows the fraction p_k of vertices with degree k in a growing network with attachment kernel k^γ as described in the text. In this case $\gamma = 0.8$ and $c = 3$. The points are results from computer simulations, averaged 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.

Homework (10/27/20):Jeong, Néda, Barabási Measuring preferential attachment in evolving networks, Europhys. Lett.

Vertex Copying Models



Algorithm:

- Initialize network with $n_0 > c$ nodes ($d(\cdot) \overset{\text{random}}{=} c$)
- 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 $|V| = n$

When new node j is added ...

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

- it will have γc copied entries on the average.
- probability new edge is copied $\Pr_1[j \to i] = \gamma q_i/n$, where $q_i = d^-(i)$
- probability new edge is randomly created $\Pr_2[j \to i] = (1-\gamma)c/n$
- 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**

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

Define

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

then

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

Same as in Price's model!

Conclusion: Vertex copying behaves as the Price's model with $a = c(1/\gamma - 1)$.

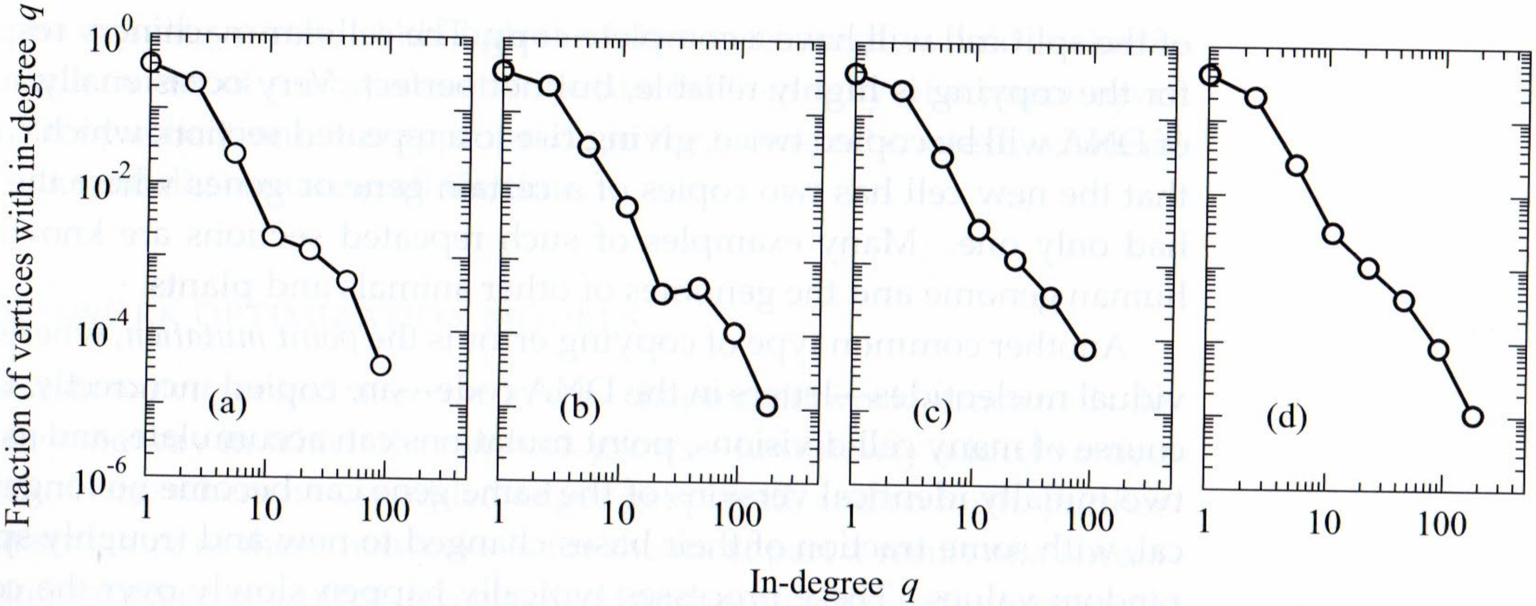
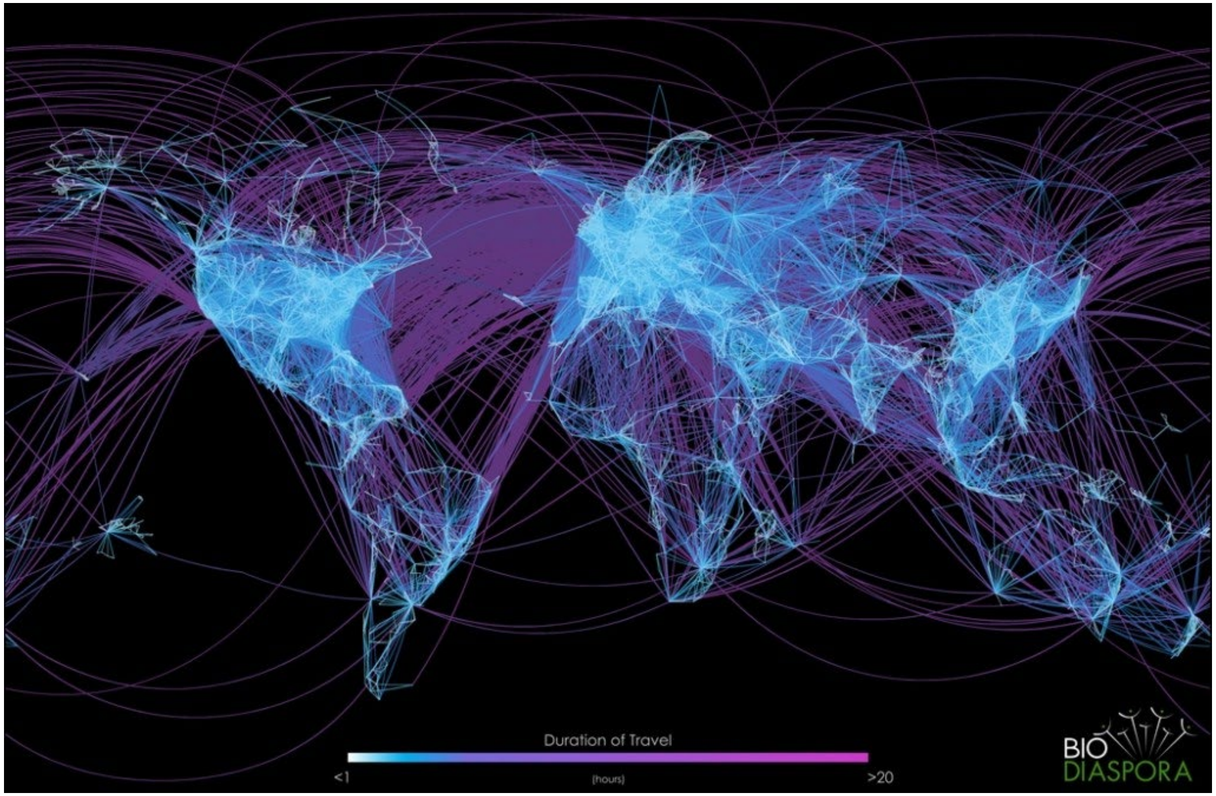
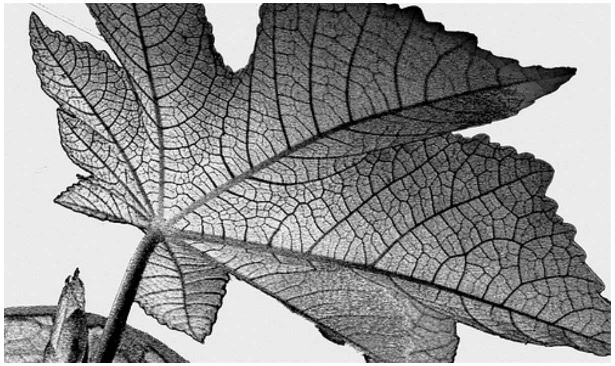


Figure 14.9: Distribution of in-degrees in the metabolic networks of various organisms. Jeong *et al.* [166] examined the degree distributions of the known portions of the metabolic networks of 43 organisms, finding some of them to follow power laws, at least approximately. Show here are the in-degree 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 $\lambda < 2/(n^2 + 2)$

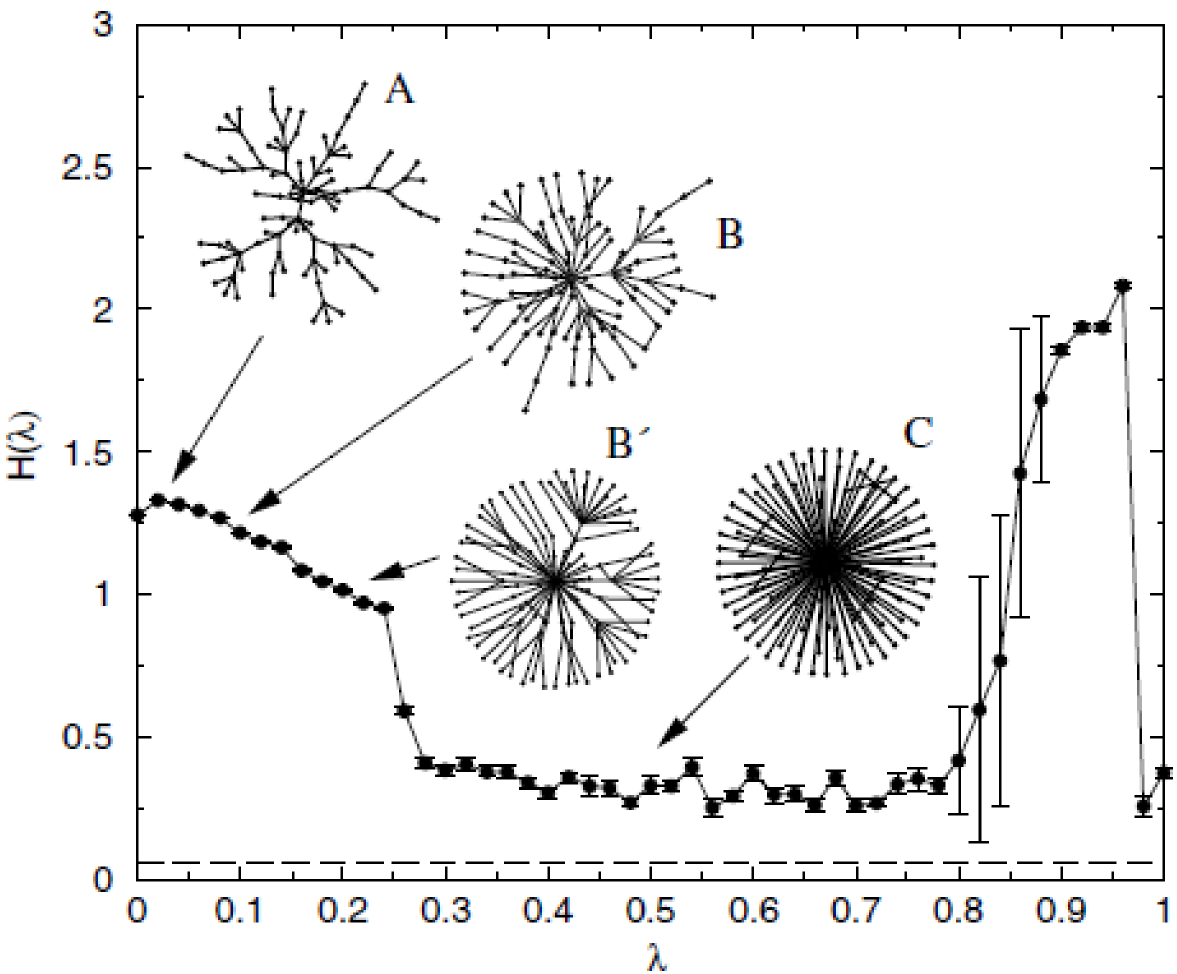


Fig. 7.4. Average (over 50 replicas) degree entropy as a function of λ with $n = 100$, $T = \binom{n}{2}$, $\nu = 2/\binom{n}{2}$ and $\rho(0) = 0.2$. Optimal networks for selected values of λ are plotted. The entropy of a star network, $H_{star} = \log n - [(n-1)/n]\log(n-1) = 0.056$ is provided as 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

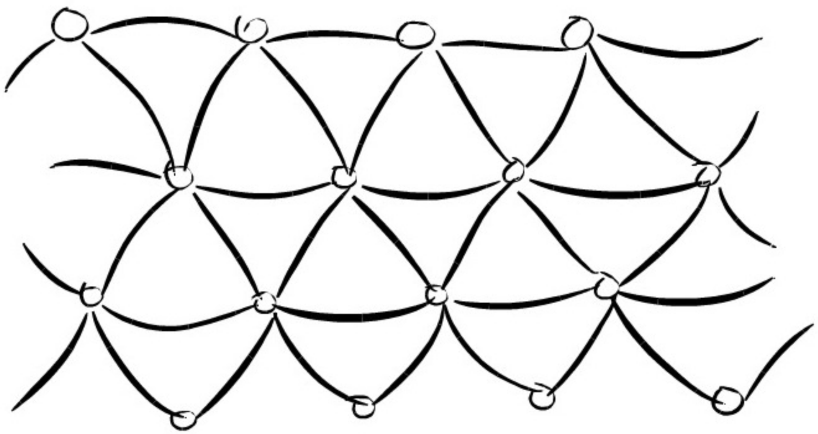
Ferrer i Cancho, Sole "Optimization in Complex Networks"

Small Worlds and High Clustering Coefficient

Network	n	z	C measured	C for random graph
Internet [153]	6,374	3.8	0.24	0.00060
World Wide Web (sites) [2]	153,127	35.2	0.11	0.00023
power grid [192]	4,941	2.7	0.080	0.00054
biology collaborations [140]	1,520,251	15.5	0.081	0.000010
mathematics collaborations [141]	253,339	3.9	0.15	0.000015
film actor collaborations [149]	449,913	113.4	0.20	0.00025
company directors [149]	7,673	14.4	0.59	0.0019
word co-occurrence [90]	460,902	70.1	0.44	0.00015
neural network [192]	282	14.0	0.28	0.049
metabolic network [69]	315	28.3	0.59	0.090
food web [138]	134	8.7	0.22	0.065

N. Przulj. Graph theory analysis of protein-protein interactions

Simple Models with High Clustering Coefficient



Triangular lattice

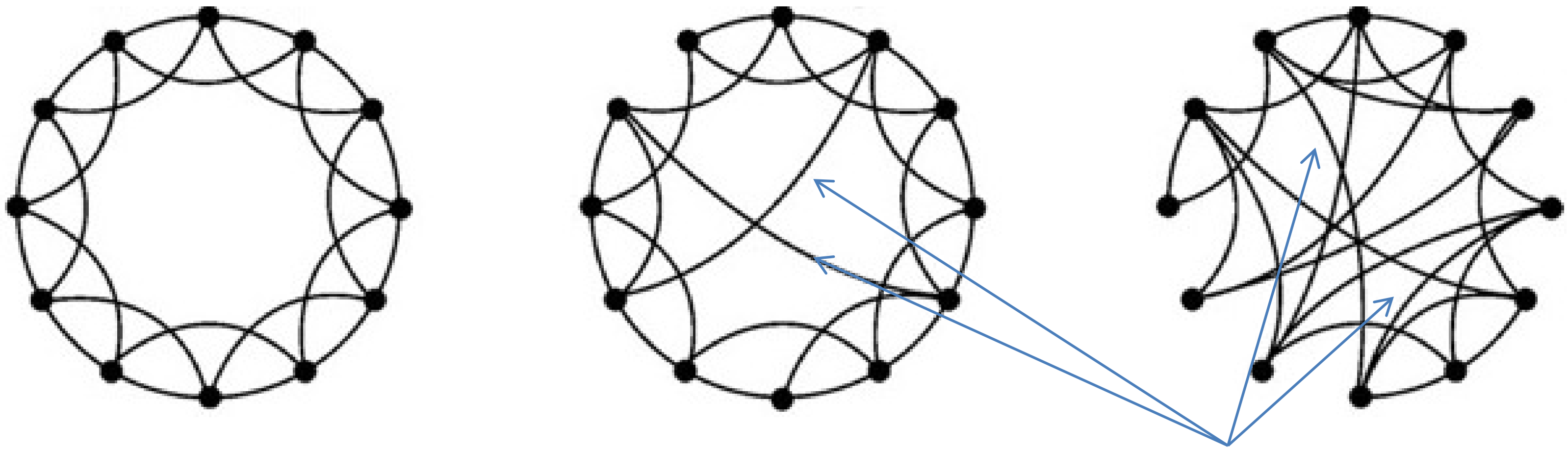
$c_i = triangles/triples = 6/15 = 0.4$
$C \to 0.4$

One-dimensional line

Clustering coefficient depends on the number of connected 1d line neighbors

Small-World Model

High C, High Diam | High C, Low Diam | Low C, Low Diam



Randomly rerouted (or added) edges
with probability *p* (for each of the edges in circle)

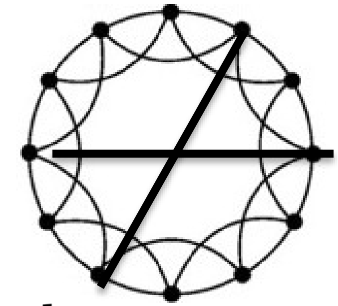
Circle models
are not “small-world”

Random models
have small clustering
coefficients

Watts-Strogatz Model (1998)

- Given circle models with *n* nodes
- Go through all edges, remove each with probability *p* and then add new edge uniformly at random

Small-World Model (without edge removal)



If c is a degree in circle model and p is a prob of an edge then

- $\frac{1}{2}ncp$ short-cuts in new graph
- cp ends of new short-cuts at each node on the average
- s number of short-cuts is Poisson distributed over all nodes

$$p_s = e^{-cp}\frac{(cp)^s}{s!}$$

- node degree $k = s + c$, then degree distribution of small-world models is

$$p_k = e^{-cp}\frac{(cp)^{k-c}}{(k-c)!},$$

where $p_k = 0$ if $k < c$.

Small-World Model (without edge removal)

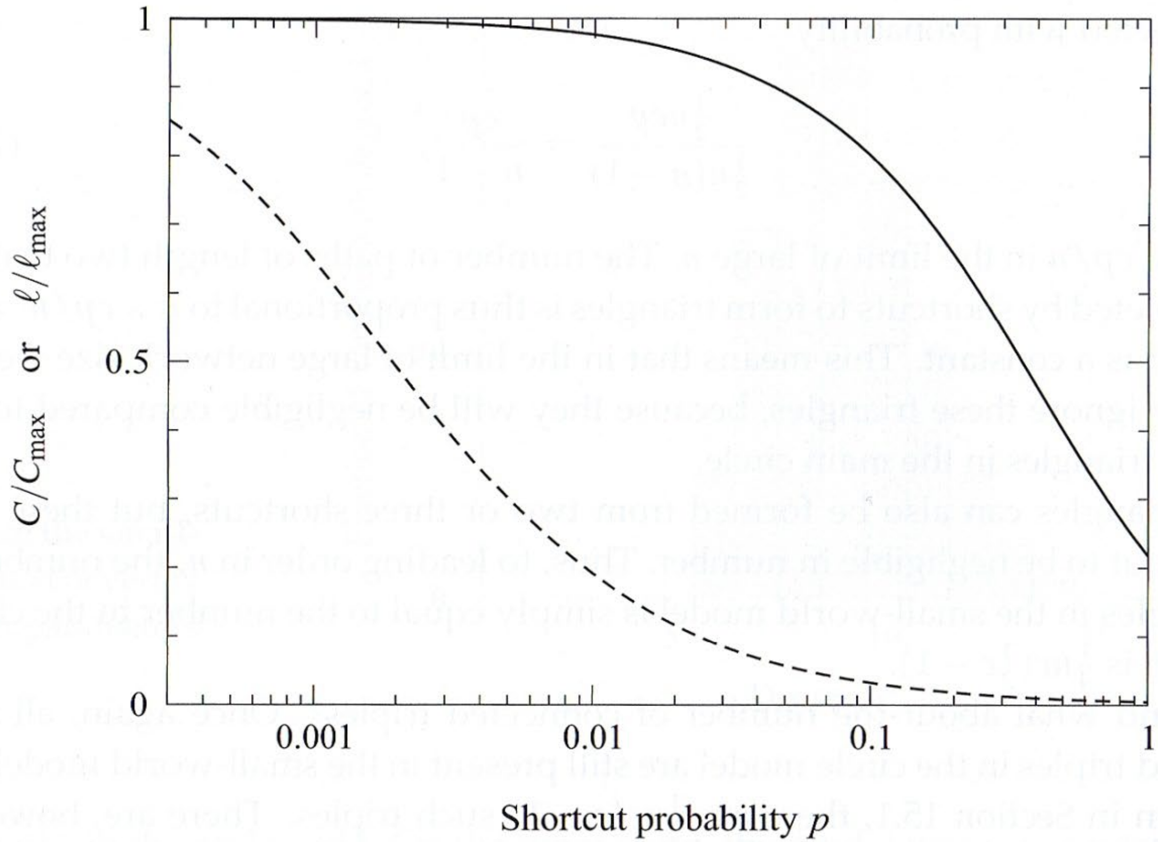


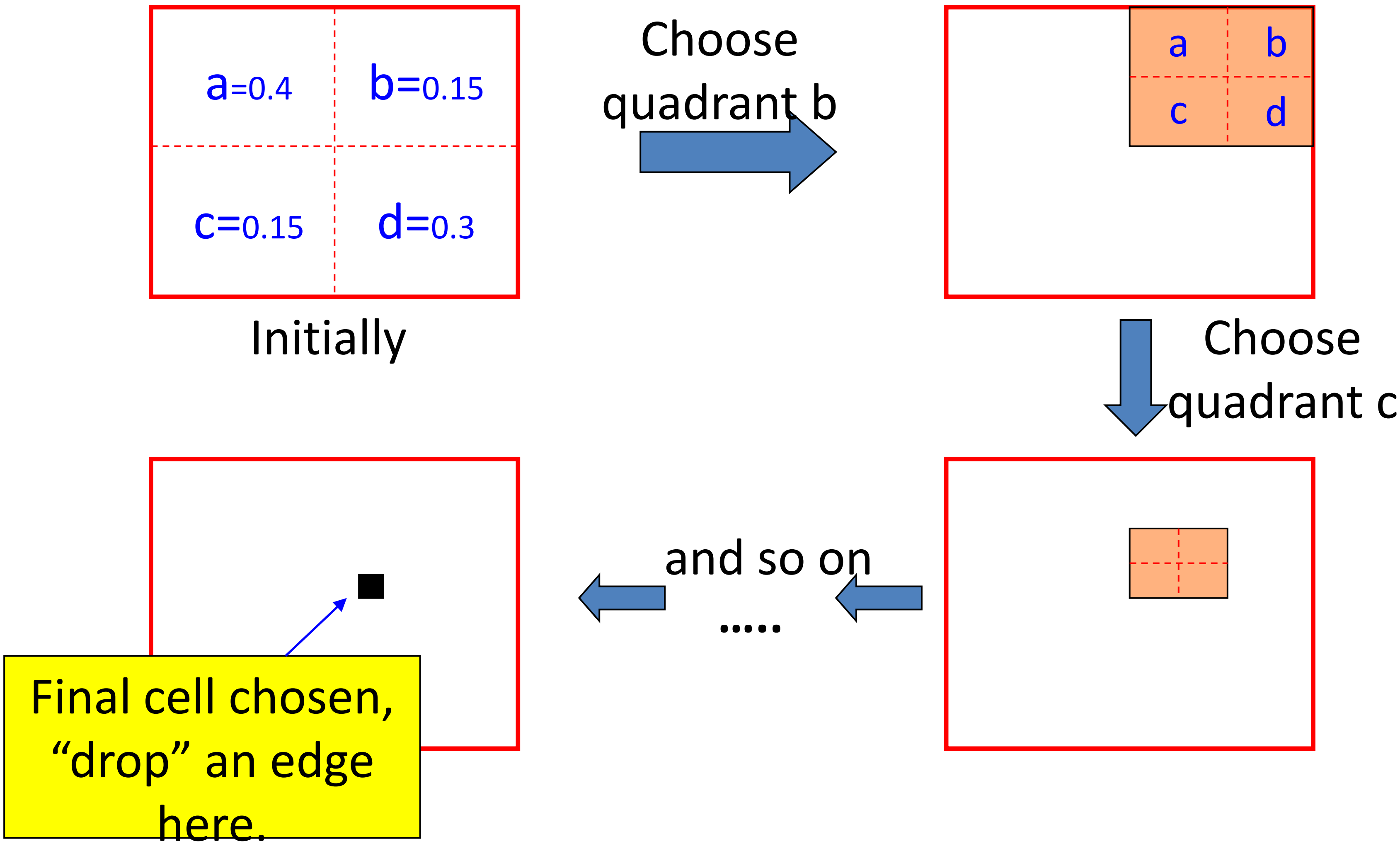
Figure 15.5: Clustering coefficient and average path length in the small-world model. The solid line shows the clustering coefficient, Eq. (15.7), for a small-world model with $c = 6$ and $n = 600$, as a fraction of its maximum value $C_{max} = \frac{3}{4}(c-2)/(c-1) = 0.6$, plotted as a function of the parameter p. The dashed line shows the average geodesic distance between vertices for the same model as a fraction of its maximum value $\ell_{max} = n/2c = 50$, calculated from the mean-field solution, Eq. (15.14). Note that the horizontal axis is logarithmic.

Homework (reviews by 5/6/2021):

Watts and Strogatz “Collective dynamics of ‘small-world’ networks”, 1998
Kleinberg “Small-world phenomenon: an algorithmic perspective”, 2000

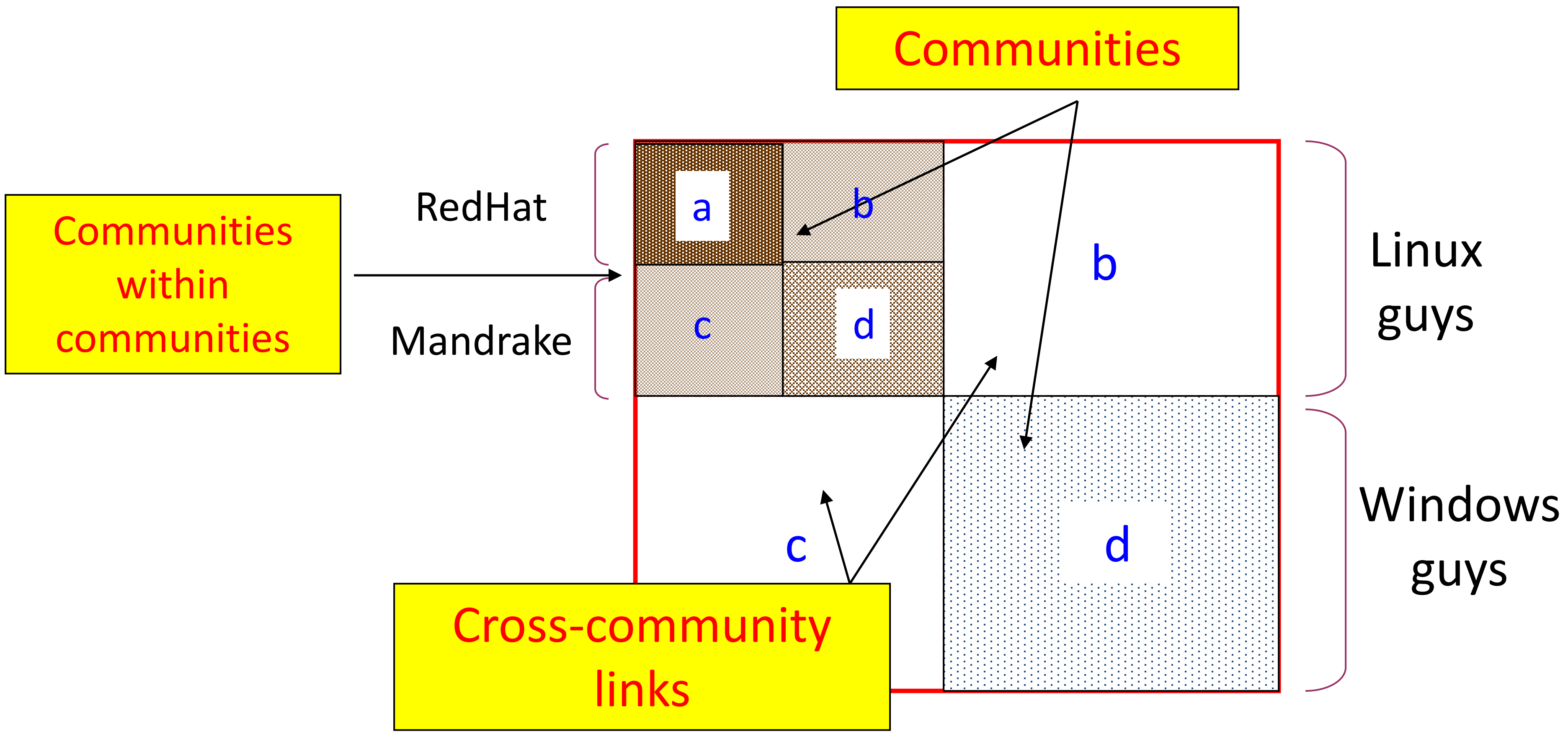
R-Mat Generator

by Chakrabarti, Zhang, Faloutsos



R-Mat Generator

by Chakrabarti, Zhang, Faloutsos



Presentation by C. Faloutsos at SIAM DM04

Kronecker Graphs

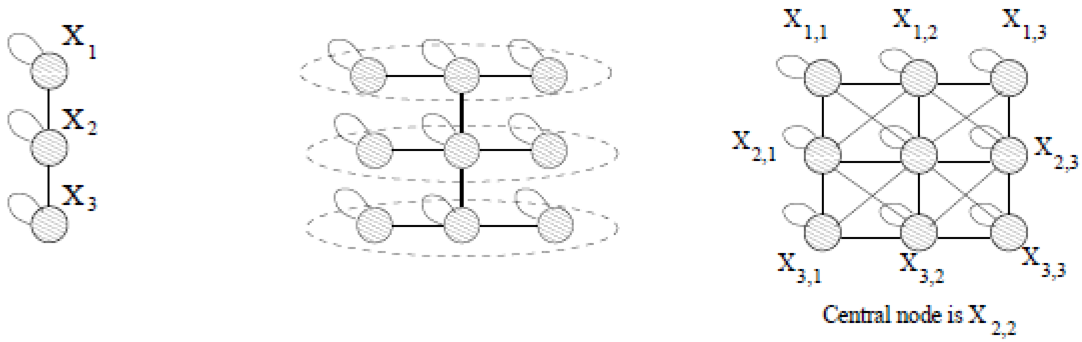
Definition 1 (Kronecker product of matrices) *Given two matrices* $\mathbf{A} = [a_{i,j}]$ *and* $\mathbf{B}$ *of sizes* $n \times m$ *and* $n' \times m'$ *respectively, the Kronecker product matrix* $\mathbf{C}$ *of dimensions* $(n \cdot n') \times (m \cdot m')$ *is given by*

$$\mathbf{C} = \mathbf{A} \otimes \mathbf{B} \doteq \begin{pmatrix} a_{1,1}\mathbf{B} & a_{1,2}\mathbf{B} & \dots & a_{1,m}\mathbf{B} \\ a_{2,1}\mathbf{B} & a_{2,2}\mathbf{B} & \dots & a_{2,m}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1}\mathbf{B} & a_{n,2}\mathbf{B} & \dots & a_{n,m}\mathbf{B} \end{pmatrix}$$

We then define the Kronecker product of two graphs simply as the Kronecker product of their corresponding adjacency matrices.

Definition 2 (Kronecker product of graphs (Weichsel, 1962)) *If* G *and* H *are graphs with adjacency matrices* $A(G)$ *and* $A(H)$ *respectively, then the Kronecker product* $G \otimes H$ *is defined as the graph with adjacency matrix* $A(G) \otimes A(H)$.

from paper Kronecker Graphs: An approach to modeling networks
by J. Leskovec, D. Chakrabarti, J. Kleinberg, C. Faloutsos, Z. Ghahramani



(a) Graph K_1 (b) Intermediate stage (c) Graph $K_2 = K_1 \otimes K_1$

1	1	0
1	1	1
0	1	1

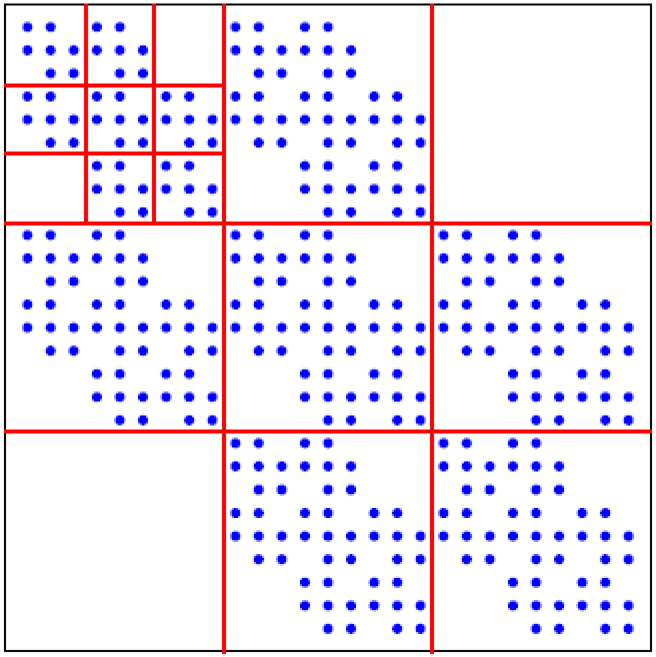
(d) Adjacency matrix of K_1

K_1	K_1	0
K_1	K_1	K_1
0	K_1	K_1

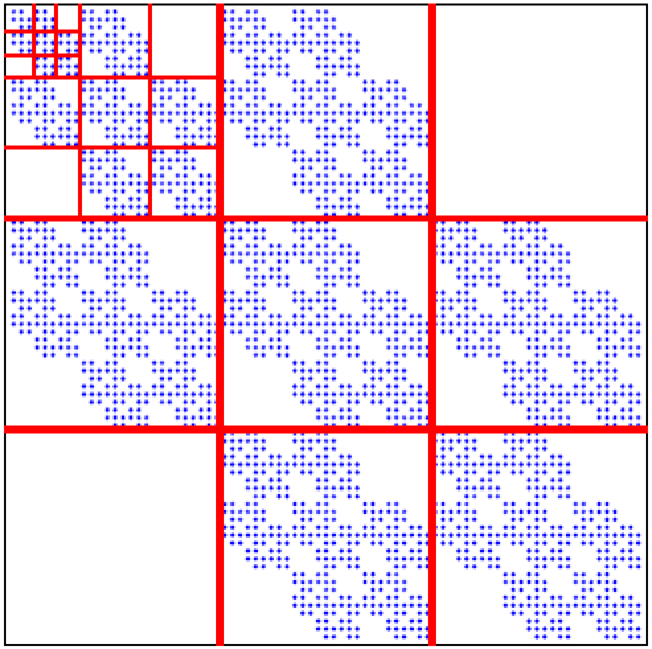
(e) Adjacency matrix of $K_2 = K_1 \otimes K_1$

Example of Kronecker multiplication: Top: a "3-chain" initiator graph and its Kronecker product with itself. Each of the X_i nodes gets expanded into 3 nodes, which are then linked using Observation 1. Bottom row: the corresponding adjacency matrices. See figure 2 for adjacency matrices of K_3 and K_4.

from paper [Kronecker Graphs: An approach to modeling networks]
by J. Leskovec, D. Chakrabarti, J. Kleinberg, C. Faloutsos, Z. Ghahramani



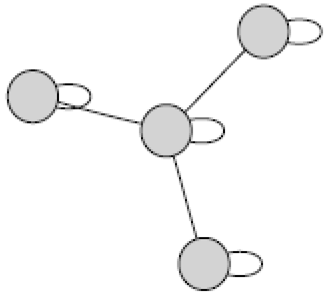
(a) K_3 adjacency matrix (27×27)



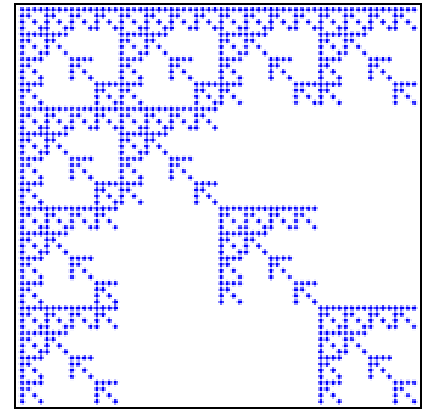
(b) K_4 adjacency matrix (81×81)

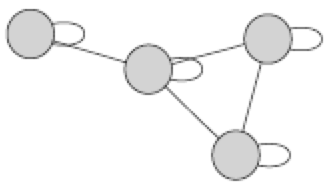
Figure 2: Adjacency matrices of K_3 and K_4, the 3^{rd} and 4^{th} Kronecker power of K_1 matrix as defined in Figure 1. Dots represent non-zero matrix entries, and white space represents zeros. Notice the recursive self-similar structure of the adjacency matrix.

from paper [Kronecker Graphs: An approach to modeling networks]
by J. Leskovec, D. Chakrabarti, J. Kleinberg, C. Faloutsos, Z. Ghahramani

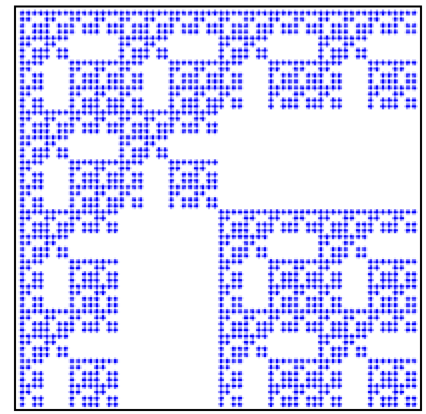


1	1	1	1
1	1	0	0
1	0	1	0
1	0	0	1





1	1	1	1
1	1	0	0
1	0	1	1
1	0	1	1



Initiator K_1 | K_1 adjacency matrix | K_3 adjacency matrix

Figure 3: Two examples of Kronecker initiators on 4 nodes and the self-similar adjacency matrices they produce.

from paper [Kronecker Graphs: An approach to modeling networks]
by J. Leskovec, D. Chakrabarti, J. Kleinberg, C. Faloutsos, Z. Ghahramani

Exponential Random Graphs Model

Instead of analyzing one network with fixed parameters, it is useful to consider ensembles of networks that are similar to the original.

Let us fix *average* values of some network properties (such as clustering and modularity). Possible property of an ensemble: values closer to the averages have higher probability. Define

$$\sum_{G \in \mathcal{G}} \Pr(G) = 1$$

graphs with *n* nodes

For network measure x_i, $1 \leq i \leq M(\ll 2^{n(n-1)/2})$

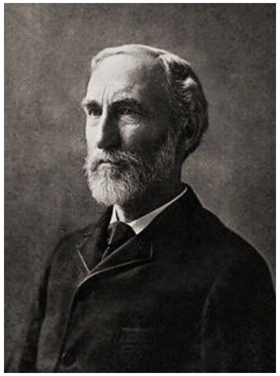
$$\langle x_i \rangle = \sum_{G \in \mathcal{G}} \Pr(G) x_i(G)$$

i.e., if $\Pr(G)$ are variables then such systems do not describe the system completely.

How to choose $\Pr(G)$?

Best choice of probability distribution given a small number of constraints maximizes Gibbs entropy

$$S = -\sum_{G \in \mathcal{G}} \Pr(G) \ln \Pr(G)$$



J. Willard Gibbs
1839-1903

Maximization of entropy with Lagrange multipliers

$$\max \quad -\sum_{G \in \mathcal{G}} \Pr(G) \ln \Pr(G) - \alpha\Big(1 - \sum_{G \in \mathcal{G}} \Pr(G)\Big) - \sum_i \beta_i\Big(\langle x_i \rangle - \sum_{G \in \mathcal{G}} \Pr(G) x_i(G)\Big)$$

Differentiate wrt $P(G)$ of a particular G

$$-\ln \Pr(G) - 1 + \alpha + \sum_i \beta_i x_i(G) = 0$$

or

$$\Pr(G) = \exp\Big(\alpha - 1 + \sum_i \beta_i x_i(G)\Big) \;\Rightarrow\; \Pr(G) = \frac{e^{H(G)}}{Z},$$

where $Z = e^{1-\alpha}$ and $H(G) = \sum_i \beta_i x_i(G)$ is the graph Hamiltonian.

Z is solved by normalization

$$\sum_{G\in\mathcal{G}} \Pr(G) = \frac{1}{Z}\sum_{G\in\mathcal{G}} e^{H(G)} = 1$$

β_i are solved by substituting $\Pr(G) = \frac{e^{H(G)}}{Z}$ into $\sum_{G\in\mathcal{G}} \Pr(G)x_i(G) = \langle x_i\rangle$

In general β_i can play a role of importance coefficients.

Practice

If we have $\Pr(G)$ over graphs let us estimate useful quantities. For property y

$$\langle y\rangle = \sum_{G\in\mathcal{G}} \Pr(G)y(G) = \frac{1}{Z}\sum_{G\in\mathcal{G}} e^{H(G)}y(G)$$

Example: Fix the expected number of edges only. Then $H = \beta m$ and individual graphs appear with prob

$$\Pr(G) = \frac{e^{\beta m}}{Z}, \text{ where } Z = \sum_{G} e^{\beta m} \;\Rightarrow\; \text{higher } \beta \text{ correspond to denser networks}$$