Network Science

Class 5: BA model

Albert-László Barabási

With

Emma K. Towlson, Sebastian Ruf, Michael Danziger and Louis Shekhtman

www.BarabasiLab.com

Measuring preferential attachment

Measuring preferential attachment

$$\frac{\partial k_i}{\partial t} \propto \Pi(k_i) \sim \frac{\Delta k_i}{\Delta t}$$

Plot the change in the degree Δk during a fixed time Δt for nodes with degree k.

To reduce noise, plot the integral of $\Pi(k)$ over

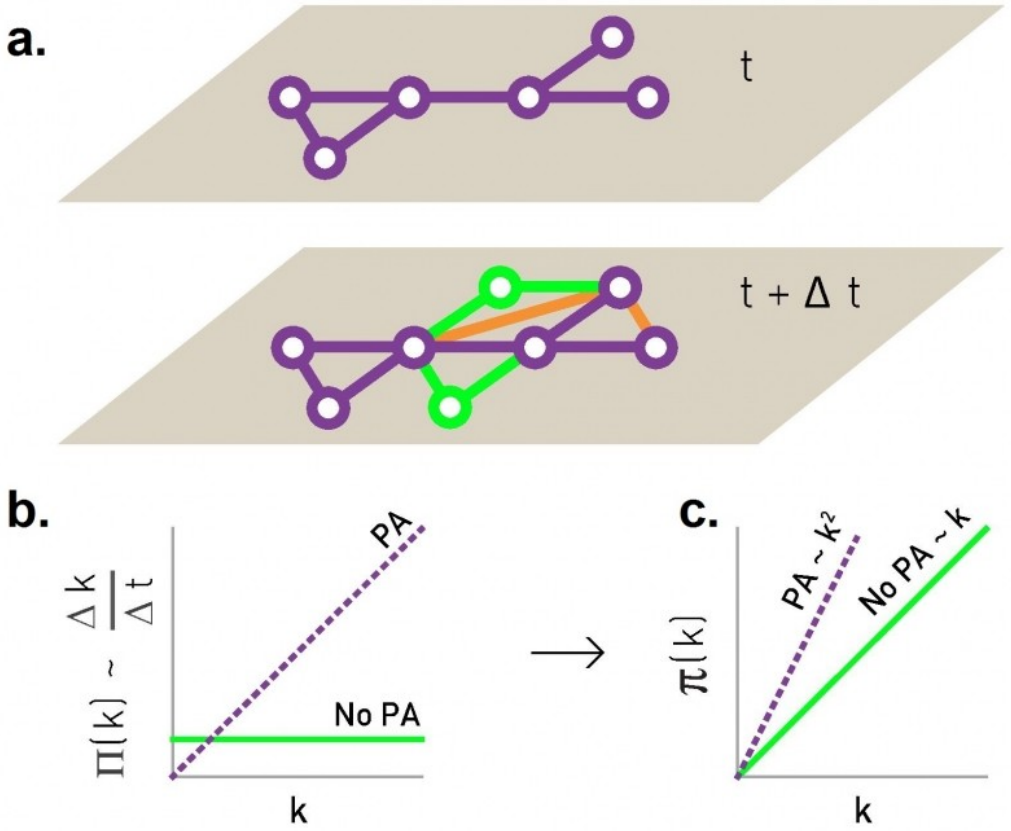
$$\kappa(k) = \sum_{K<k} \Pi(K)$$

No pref. attach:

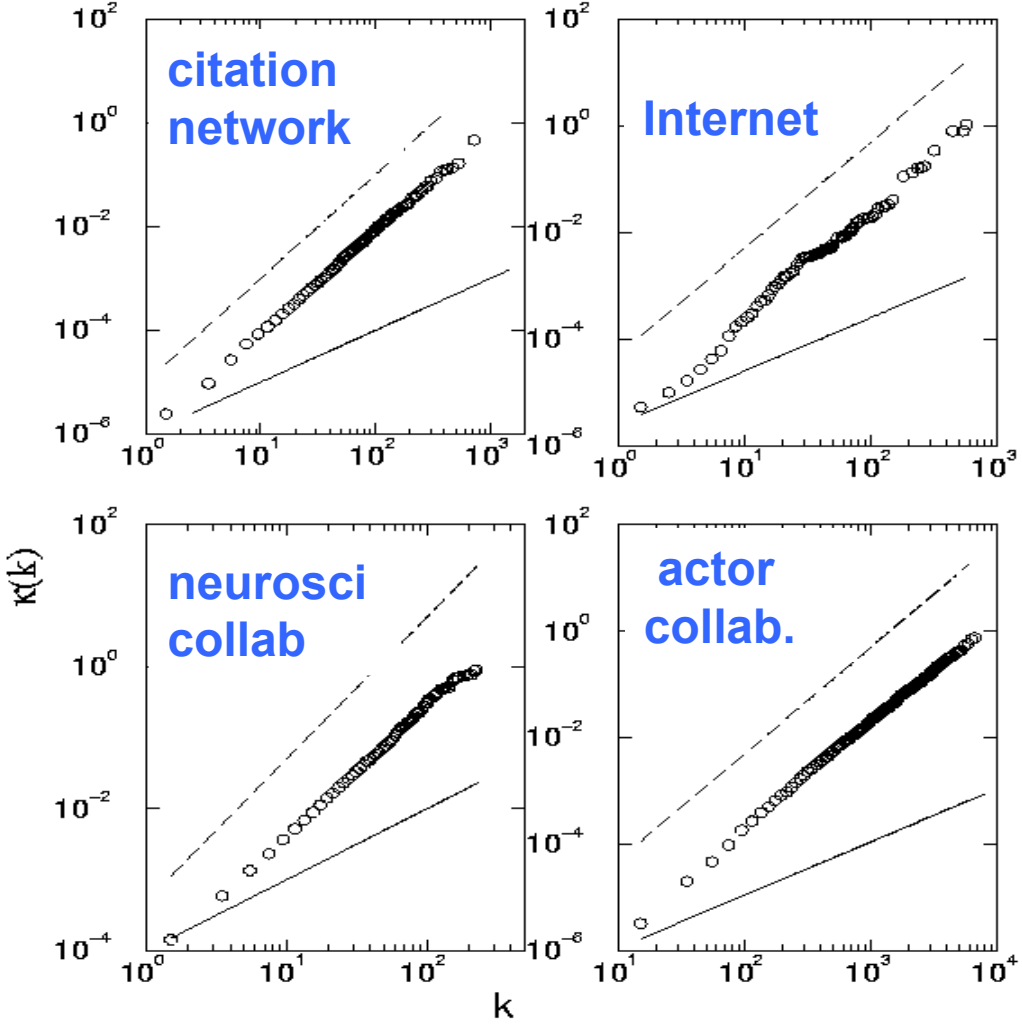
$\kappa \sim k$ ——

Linear pref. attach:

$\kappa \sim k^2$ – – – –



(Jeong, Neda, A.-L. B, Europhys Letter 2003; cond-mat/0104131)



Plots shows the integral of Π(k) over k:

$$\kappa(k) = \sum_{K<k} \Pi(K)$$

No pref. attach:
$\kappa \sim k$ ——

Linear pref. attach:
$\kappa \sim k^2$ - - - -

$$\Pi(k) \approx A + k^{\alpha}, \ \alpha \leq 1$$

Nonlinear preferential attachment

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

α=0: Reduces to Model A discussed in Section 5.4. The degree distribution follows the simple exponential function.

α=1: Barabási-Albert model, a scale-free network with degree exponent 3.

0<α<1: Sublinear preferential attachment. New nodes favor the more connected nodes over the less connected nodes. Yet, for the bias is not sufficient to generate a scale-free degree distribution. Instead, in this regime the degrees follow the stretched exponential distribution:

$$p_k \sim k^{-\alpha} \exp\left(\frac{2\mu(\alpha)}{\langle k \rangle (1-\alpha)} k^{1-\alpha} \right)$$

$$k_{\max} \sim (\ln t)^{1/(1-\alpha)}$$

Nonlinear preferential attachment

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

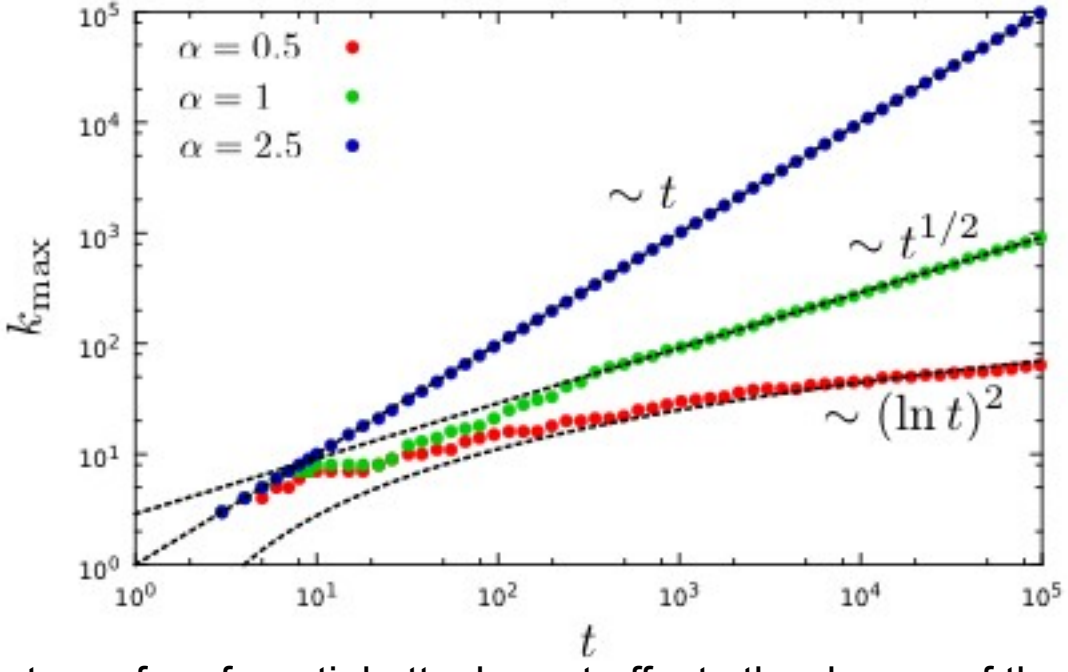
α=0: Reduces to Model A discussed in Section 5.4. The degree distribution follows the simple exponential function.

α=1: Barabási-Albert model, a scale-free network with degree exponent 3.

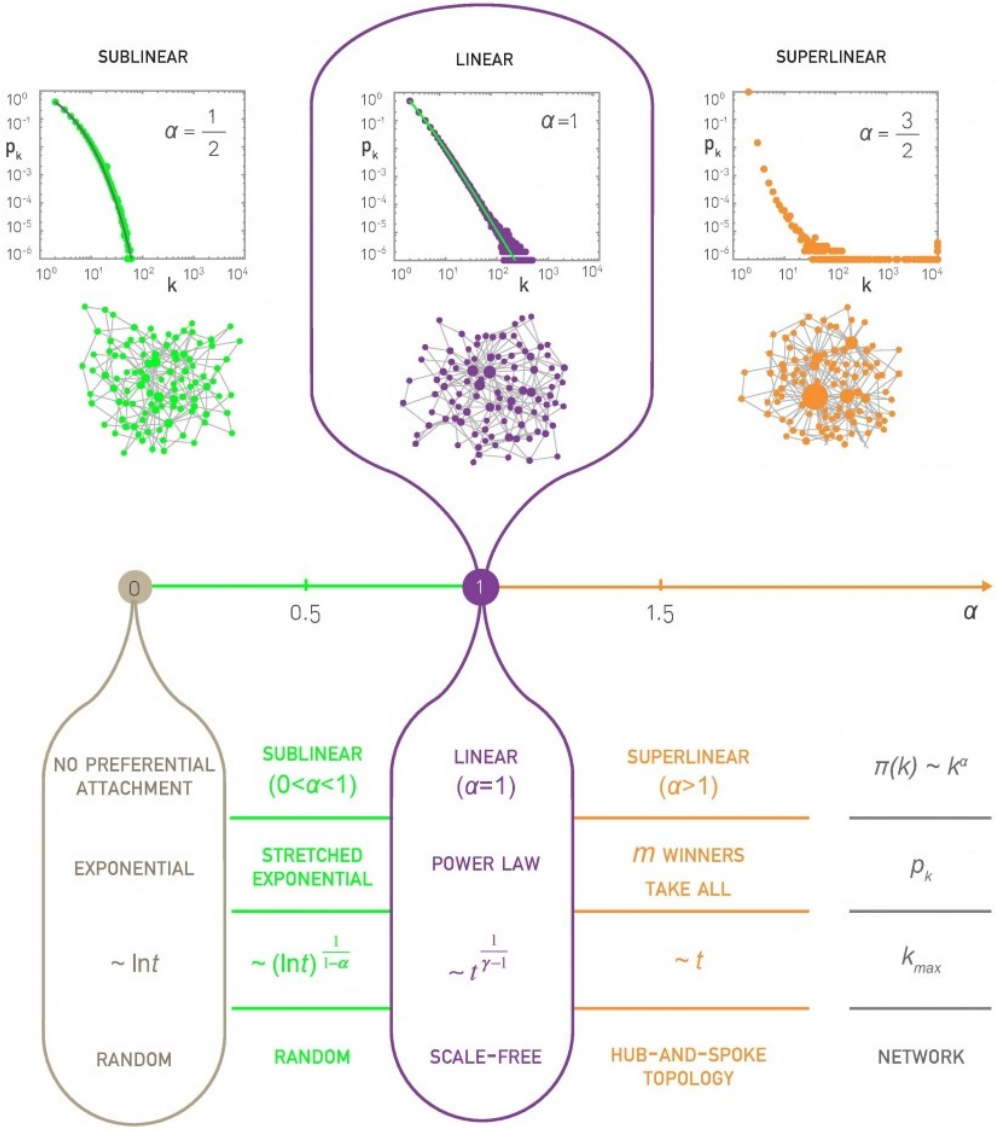
α>1: Superlinear preferential attachment. The tendency to link to highly connected nodes is enhanced, accelerating the "rich-gets-richer" process. The consequence of this is most obvious for **α>2**, when the model predicts a *winner-takes-all phenomenon: almost all nodes connect to a single or a few super-hubs.*

$$k_{\max} \sim t.$$

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



The growth of the hubs. The nature of preferential attachment affects the degree of the largest node. While in a scale-free network the biggest hub grows as (green curve), for sublinear preferential attachment this dependence becomes logarithmic (red curve). For superlinear preferential attachment the biggest hub grows linearly with time, always grabbing a finite fraction of all links (blue curve)). The symbols are provided by a numerical simulation; the dotted lines represent the analytical predictions.


SUBLINEAR
LINEAR
SUPERLINEAR
$\alpha = \frac{1}{2}$
$\alpha = 1$
$\alpha = \frac{3}{2}$
p_k
k
0
0.5
1
1.5
α
NO PREFERENTIAL ATTACHMENT
EXPONENTIAL
~ ln$t$
RANDOM
SUBLINEAR ($0<\alpha<1$)
STRETCHED EXPONENTIAL
$\sim (\ln t)^{\frac{1}{1-\alpha}}$
RANDOM
LINEAR ($\alpha=1$)
POWER LAW
$\sim t^{\frac{1}{\gamma-1}}$
SCALE-FREE
SUPERLINEAR ($\alpha>1$)
m WINNERS TAKE ALL
~ t
HUB-AND-SPOKE TOPOLOGY
$\pi(k) \sim k^{\alpha}$
p_k
k_{max}
NETWORK

The origins of preferential attachment

Link selection model

Link selection model -- perhaps the simplest example of a local or random mechanism capable of generating preferential attachment.

***Growth**: at each time step we add a new node to the network.*

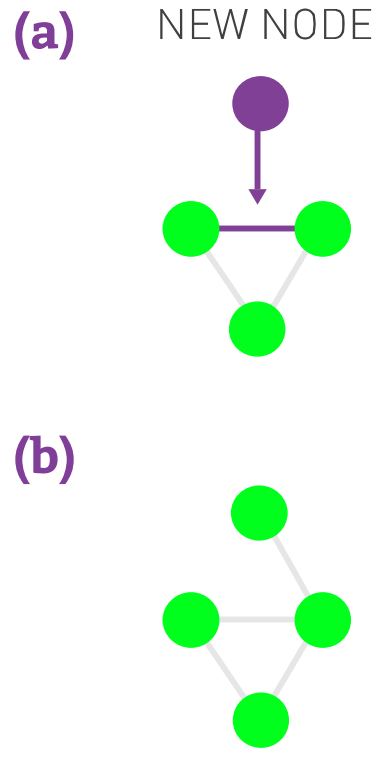
***Link selection**: we select a link at random and connect the new node to one of nodes at the two ends of the selected link.*

To show that this simple mechanism generates linear preferential attachment, we write the probability that the node at the end of a randomly chosen link has degree k as

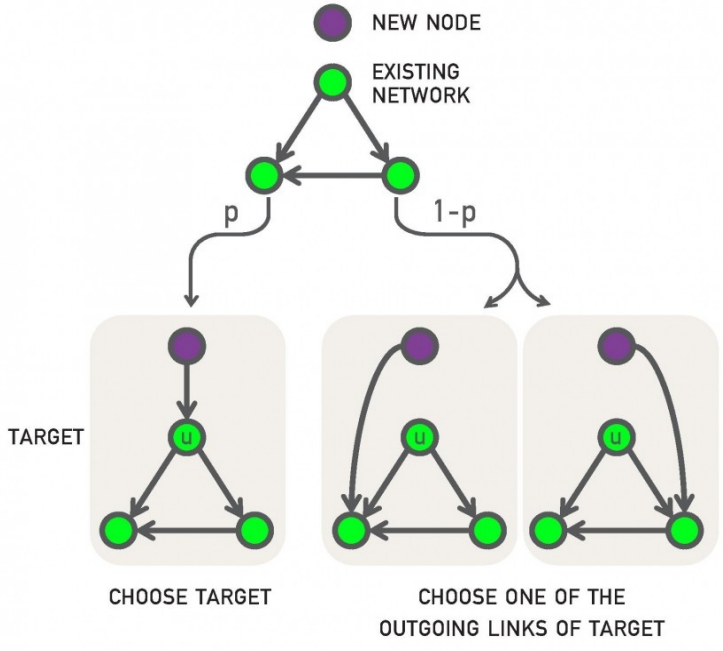
$$q_k = Ckp_k \qquad (5.26)$$

In (5.26) C can be calculated using the normalization condition $\Sigma q_k = 1$, obtaining $C = 1/\langle k \rangle$. Hence the probability to find a degree-k node at the end of a randomly chosen link is

$$q_k = \frac{kp_k}{\langle k \rangle}, \qquad (5.27)$$



Section 9 Copying model



(a) Random Connection: *with probability p the new node links to u.*

(b) Copying*: with probability 1-p we randomly choose an outgoing link of node u and connect the new node to the selected link's target. Hence the new node "copies" one of the links of an earlier node*

(a) the probability of selecting a node is 1/N.

(b) is equivalent with selecting a node linked to a randomly selected link. The probability of selecting a degree-k node through the copying process of step (b) is k/2L for undirected networks.

The likelihood that the new node will connect to a degree-k node follows preferential attachment

$$\Pi(k) = p/N + (1-p)k/(2L)$$

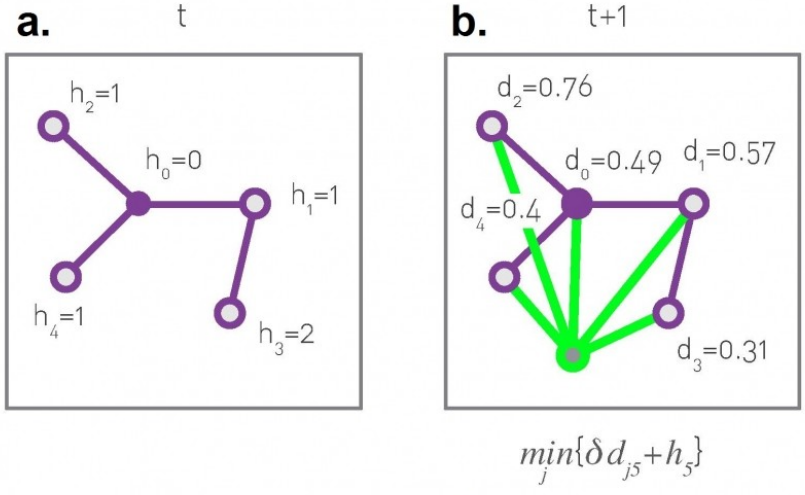
Social networks: *Copy your friend's friends.*

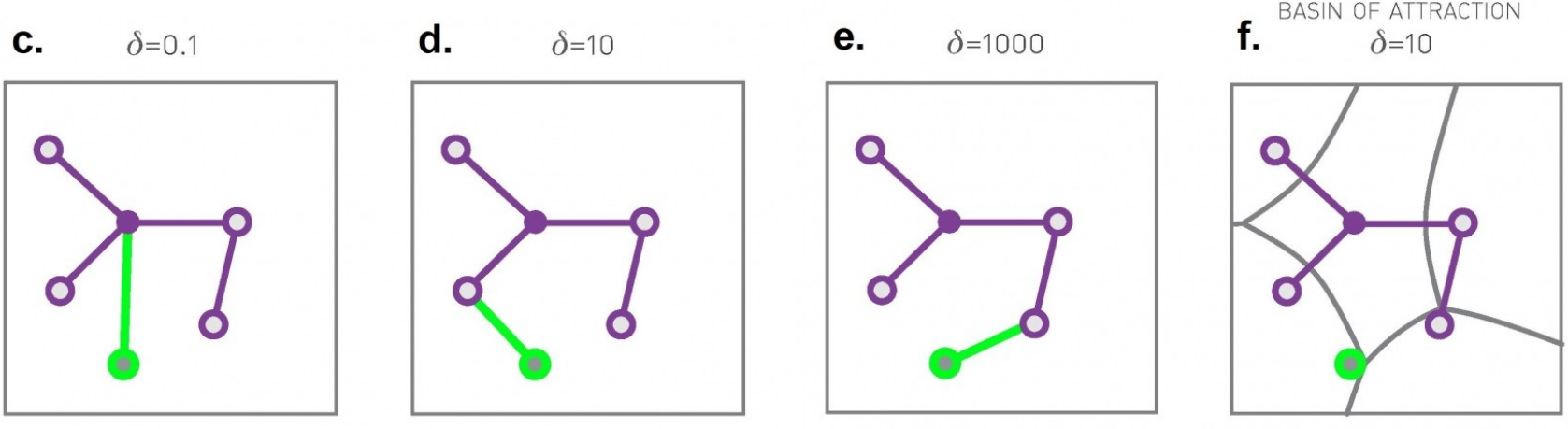
Citation Networks*: Copy references from papers we read.*

Protein interaction networks*: gene duplication,*

Optimization model

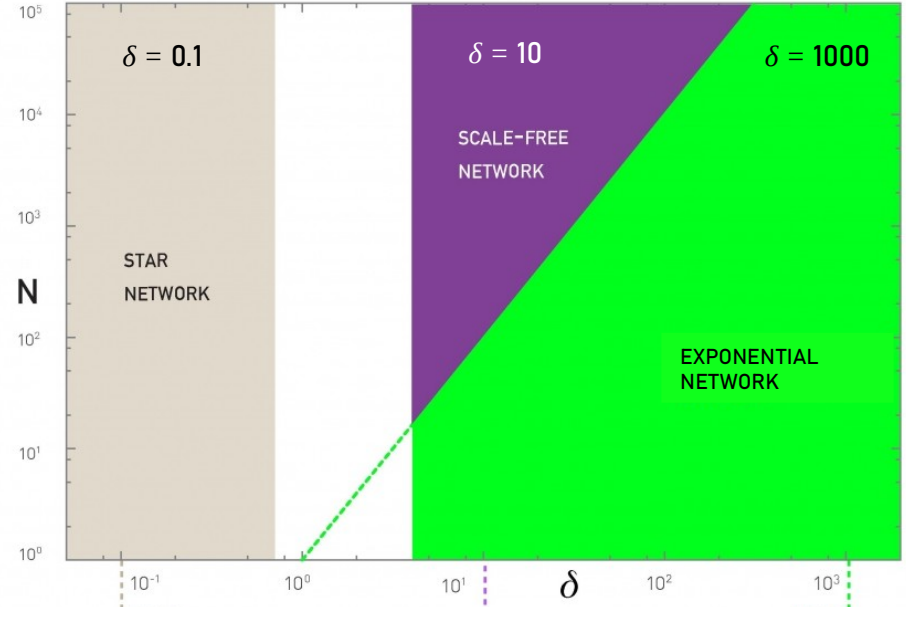
$$C_i = \min_j[\delta d_{ij} + h_j]$$

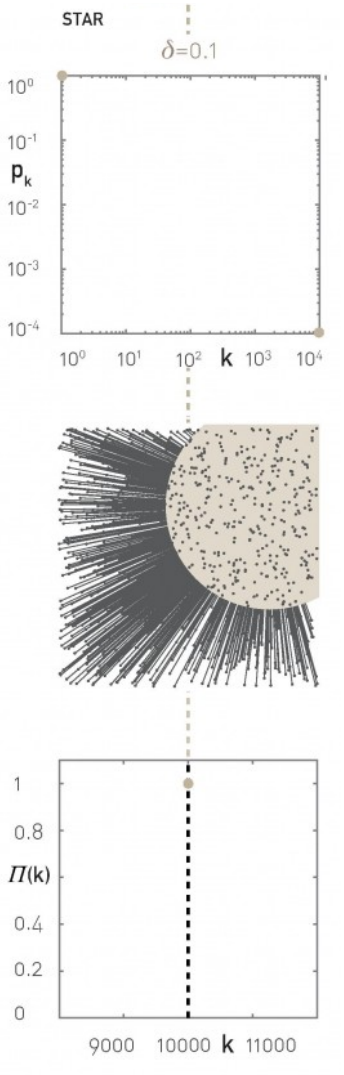




Section 9 Optimization model

$$C_i = \min_j[\delta d_{ij} + h_j]$$

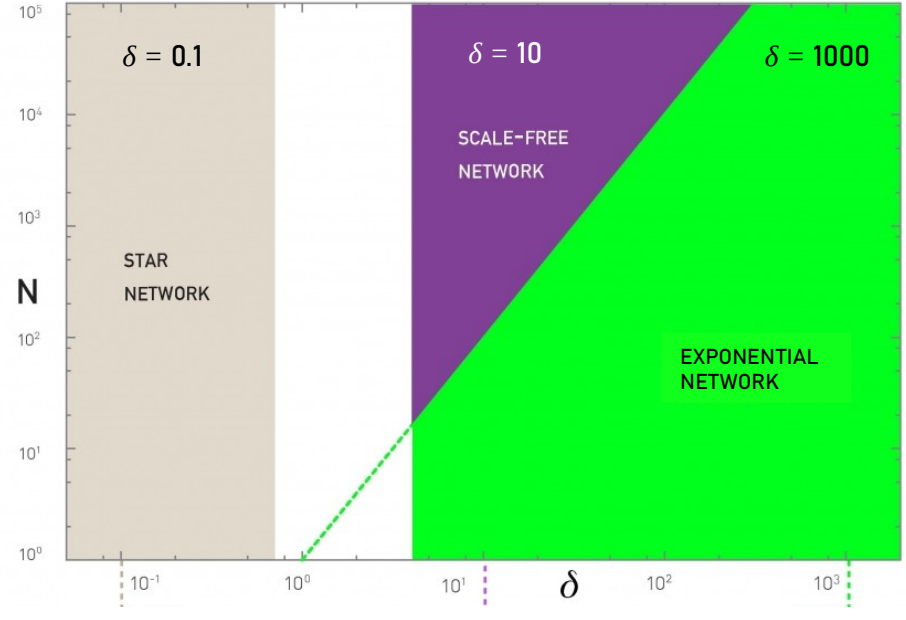


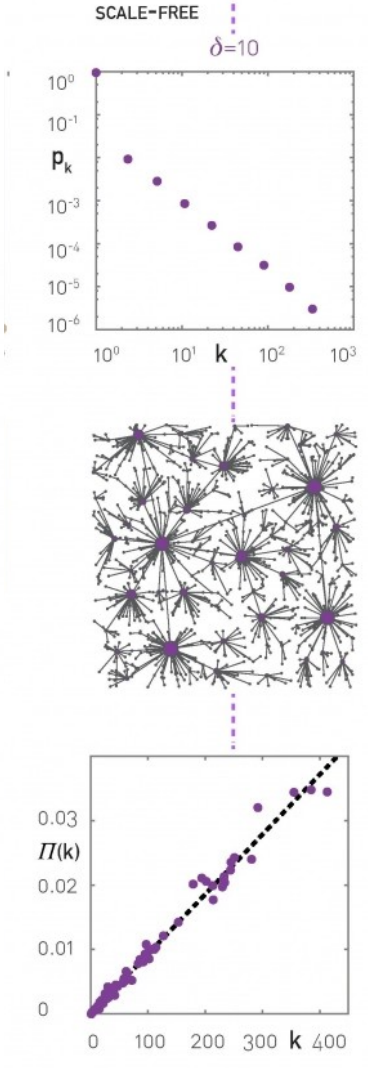


The vertical boundary of the star configuration is at $\delta=(1/2)^{1/2}$. This is the inverse of the maximum distance between two nodes on a square lattice with unit length, over which the model is defined. Therefore if $\delta < (1/2)^{1/2}$, for any new node $\delta d_{ij} < 1$ and the cost (5.28) of connecting to the central node is $C_i = \delta d_{ij}+0$, always lower than connecting to any other node at the cost of $f(i,j) = \delta d_{ij}+1$. Therefore for $\delta < (1/2)^{1/2}$ all nodes connect to node 0, resulting in a network dominated by a single hub (star-and-spoke network (c)).

Section 9 Optimization model

$$C_i = \min_j[\delta d_{ij} + h_j]$$

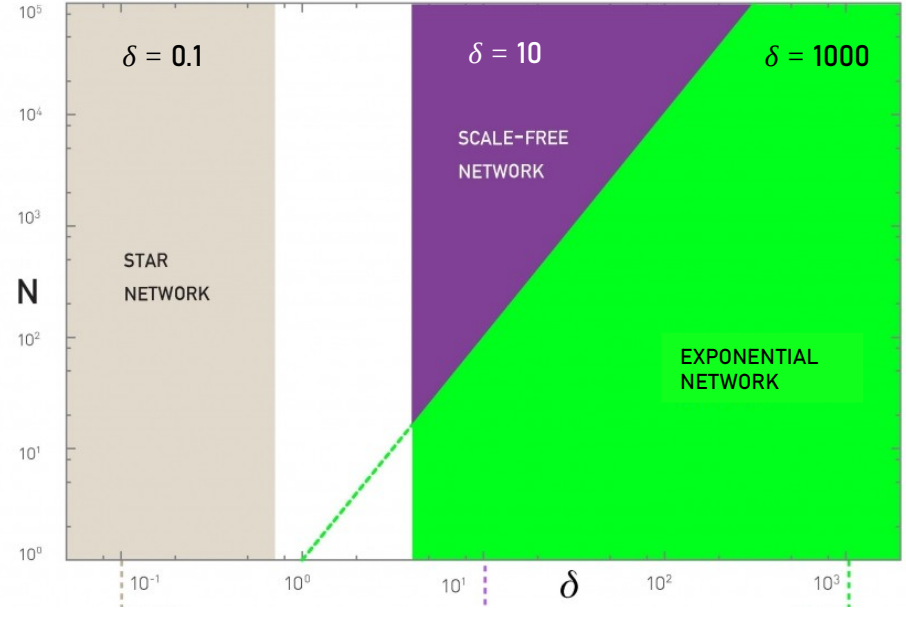


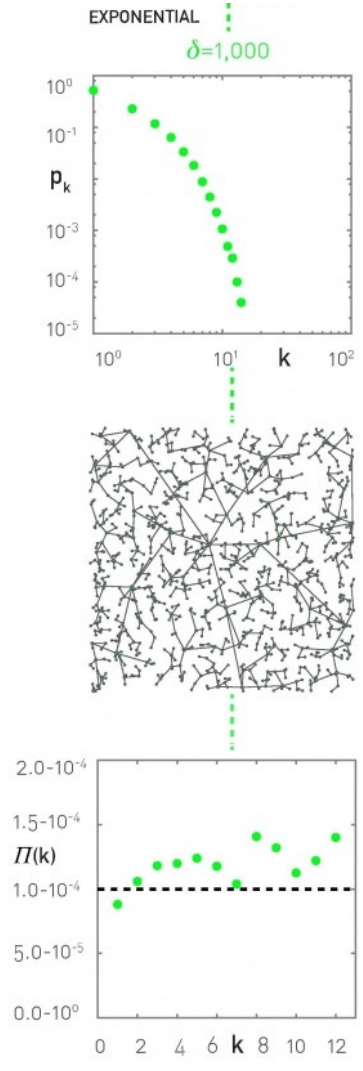


The oblique boundary of the scale-free regime is $\delta = N^{1/2}$. Indeed, if nodes are placed randomly on the unit square, then the typical distance between neighbors decreases as $N^{-1/2}$. Hence, if $d_{ij} \sim N^{-1/2}$ then $\delta d_{ij} \geq h_{ij}$ for most node pairs. Typically the path length to the central node h_j grows slower than N (in small-world networks $h_j \sim \log N$, in scale-free networks $h_j \sim \ln\ln N$). Therefore C_i is dominated by the δd_{ij} term and the smallest C_i is achieved by minimizing the distance-dependent term. Note that strictly speaking the transition only occurs in the $N \to \infty$ limit.

Section 9 Optimization mode

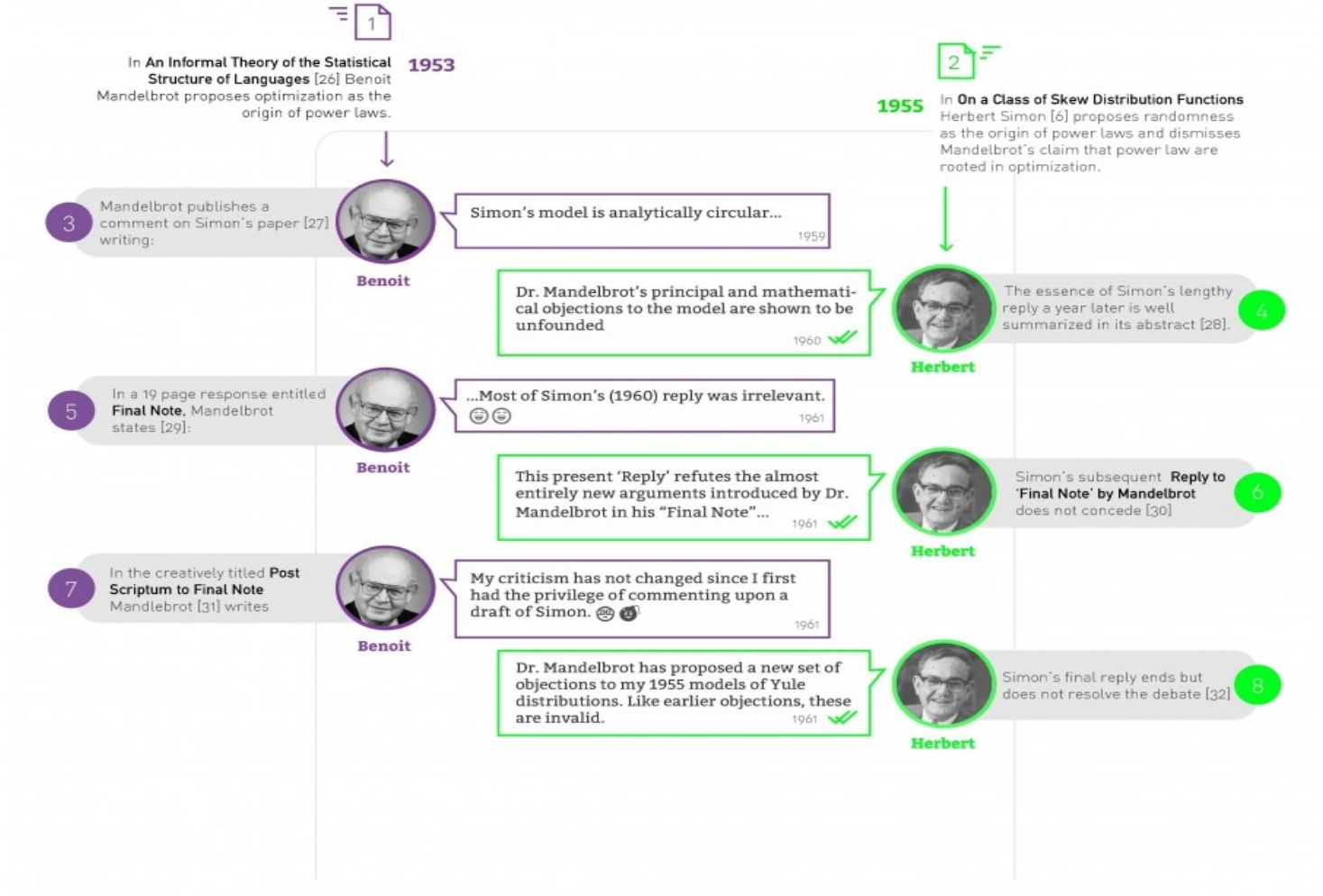
$$C_i = \min_j[\delta d_{ij} + h_j]$$





For very large δ the contribution provided by the distance term δd_{ij} overwhelms h_j in (5.28). In this case each new node connects to the node closest to it.
The resulting network will have a exponential-like, bounded degree distribution, resembling a random network.
In the white regime we lack an analytical form for the degree distribution.
We used the method described in SECTION 5.6. Starting from a network with N=10,000 nodes we added a new node and measured the degree of the node that it connected to. We repeated this procedure 10,000 times, obtaining Π(k).
The plots document the presence of linear preferential attachment in the scale-free phase, but its absence in the star and the exponential phases.

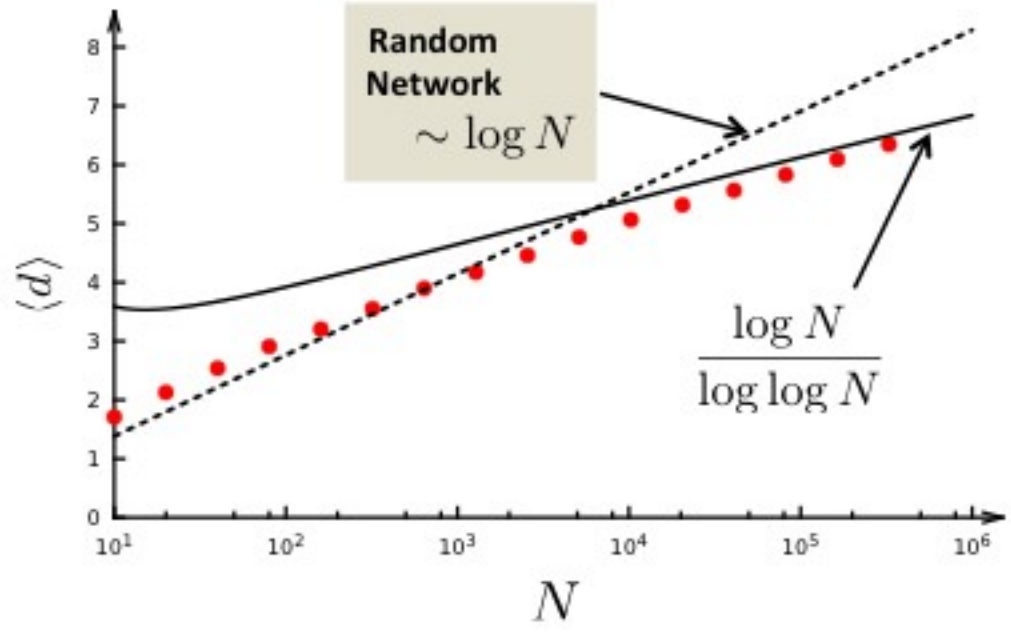
Section 9


1
In An Informal Theory of the Statistical Structure of Languages [26] Benoit Mandelbrot proposes optimization as the origin of power laws.
1953
2
1955
In On a Class of Skew Distribution Functions Herbert Simon [6] proposes randomness as the origin of power laws and dismisses Mandelbrot's claim that power law are rooted in optimization.
3
Mandelbrot publishes a comment on Simon's paper [27] writing:
Simon's model is analytically circular...
1959
Benoit
Dr. Mandelbrot's principal and mathematical objections to the model are shown to be unfounded
1960
Herbert
The essence of Simon's lengthy reply a year later is well summarized in its abstract [28].
4
5
In a 19 page response entitled Final Note. Mandelbrot states [29]:
...Most of Simon's (1960) reply was irrelevant.
1961
Benoit
This present 'Reply' refutes the almost entirely new arguments introduced by Dr. Mandelbrot in his "Final Note"...
1961
Herbert
Simon's subsequent Reply to 'Final Note' by Mandelbrot does not concede [30]
6
7
In the creatively titled Post Scriptum to Final Note Mandlebrot [31] writes
My criticism has not changed since I first had the privilege of commenting upon a draft of Simon.
1961
Benoit
Dr. Mandelbrot has proposed a new set of objections to my 1955 models of Yule distributions. Like earlier objections, these are invalid.
1961
Herbert
Simon's final reply ends but does not resolve the debate [32]
8

Diameter and clustering coefficient

Diameter

$$D \sim \frac{\log N}{\log\log N}$$



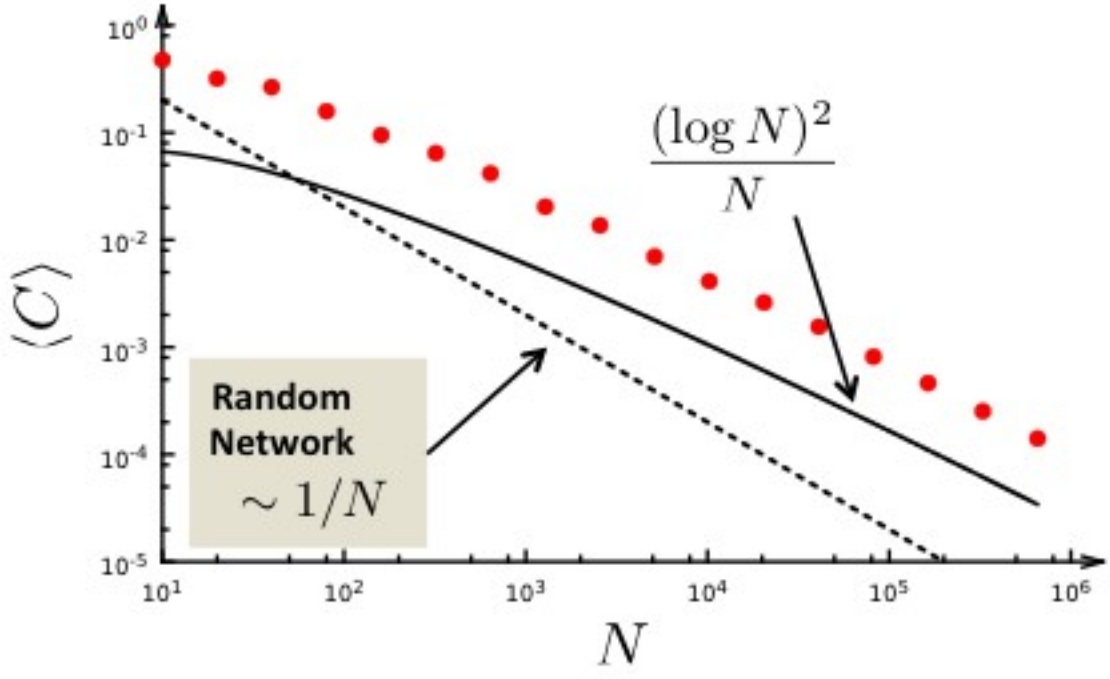
Bollobas, Riordan, 2002

Reminder: for a random graph we have:

$$C_{rand} = \frac{< k >}{N} \sim N^{-1}$$

What is the functional form of C(N)?

$$C = \frac{m}{8}\frac{(\ln N)^2}{N}$$



Konstantin Klemm, Victor M. Eguiluz,
Growing scale-free networks with small-world behavior,
Phys. Rev. E 65, 057102 (2002), cond-mat/0107607

Section 11: Summary

Number of Nodes

$N = t$

Number of Links

$N = mt$

Average Degree

$\langle k \rangle = 2m$

Degree Dynamics

$k_i(t) = m\,(t/t_i)^{\beta}$

Dynamical Exponent

$\beta = 1/2$

Degree Distribution

$p_k \sim k^{\gamma}$

Degree Exponent

$\gamma = 3$

Average Distance

$\langle d \rangle \sim \log N / \log\log N$

Clustering Coefficient

$\langle C \rangle \sim (\ln N)^2 / N$

Consequently, the modeling philosophy behind the model is simple: *to understand the topology of a complex system, we need to describe how it came into being.*

The network grows, but the degree distribution is stationary.

Section 11: Summary

Number of Nodes
$N = t$

Number of Links
$N = mt$

Average Degree
$\langle k \rangle = 2m$

Degree Dynamics
$k_i(t) = m\,(t/t_i)^{\beta}$

Dynamical Exponent
$\beta = 1/2$

Degree Distribution
$p_k \sim k^{\gamma}$

Degree Exponent
$\gamma = 3$

Average Distance
$\langle d \rangle \sim \log N / \log\log N$

Clustering Coefficient
$\langle C \rangle \sim (\ln N)^2 / N$

- The model predicts γ=3 while the degree exponent of real networks varies between 2 and 5 (Table 4.2).

- Many networks, like the WWW or citation networks, are directed, while the model generates undirected networks.

- Many processes observed in networks, from linking to already existing nodes to the disappearance of links and nodes, are absent from the model.

- The model does not allow us to distinguish between nodes based on some intrinsic characteristics, like the novelty of a research paper or the utility of a webpage.

- While the Barabási-Albert model is occasionally used as a model of the Internet or the cell, in reality it is not designed to capture the details of any particular real network. It is a minimal, proof of principle model whose main purpose is to capture the basic mechanisms responsible for the emergence of the scale-free property. Therefore, if we want to understand the evolution of systems like the Internet, the cell or the WWW, we need to incorporate the important details that contribute to the time evolution of these systems, like the directed nature of the WWW, the possibility of internal links and node and link removal.

Preliminary project presentation (Apr. 28th)

5 slides

Discuss:

What are your nodes and links

How will you collect the data, or which dataset you will study

Expected size of the network (# nodes, # links)

What questions you plan to ask (they may change as we move along with the class).

Why do we care about the network you plan to study.