Network formation models

Prof. Pascal Frossard

Signal Processing Laboratory (LTS4)

Some slides are taken from Prof. Barabási's class on Network Science (www.BarabasiLab.com)

Outline

- Growth and preferential attachment
- Barabási-Albert model
- Degree distribution
- Origins of preferential attachment

Network formation models

- Networks can be characterised by different parameters
  - Degree distribution, size, etc
  - Network models, like random or scale-free networks

- Many networks seem to have the same properties, but they capture very different data: how comes?
  - WWW and biological cellular networks are both scale-free

- It is important to understand how networks get formed
  - network synthesis models explaining the properties of networks
  - mechanisms explaining the emergence of scale-free properties

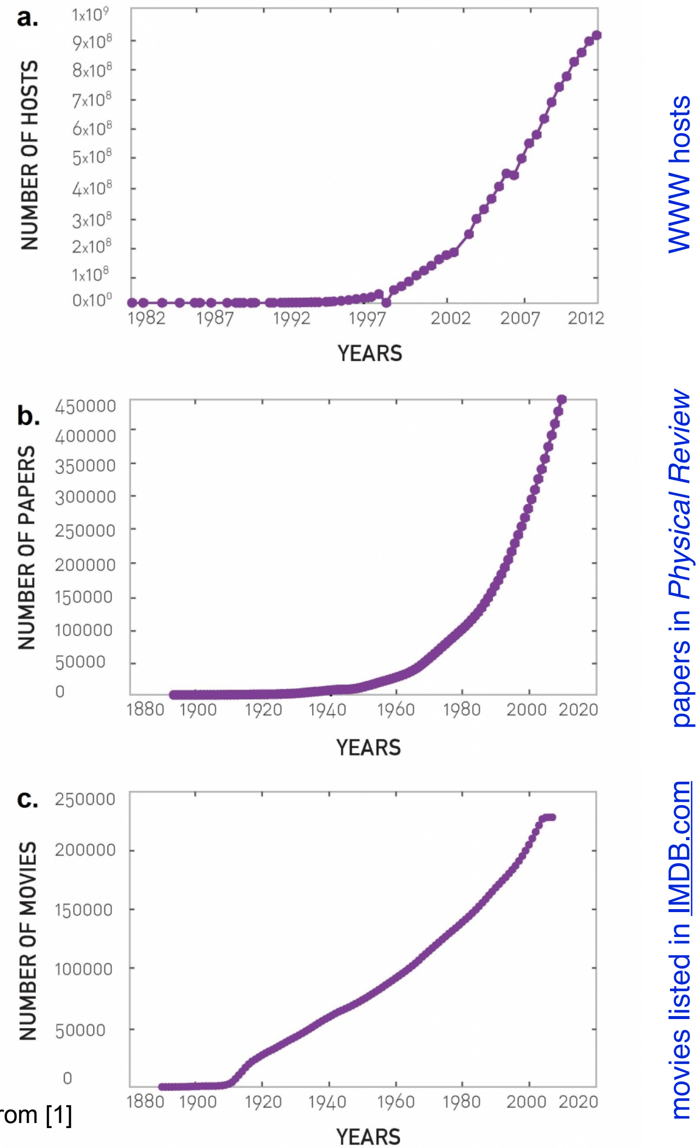
- We focus on the Barabási-Albert model
  - numerous other models exist in the literature

Growth and preferential attachment

- Real networks grow with the addition of new nodes
  - ✓ networks are the product of a steady growth process
  - x random networks assume a fixed number of nodes

- In most real networks, new nodes tend to link to the more connected nodes
  - ✓ networks follow a preferential attachment process
  - x random networks assume a random choice of connections

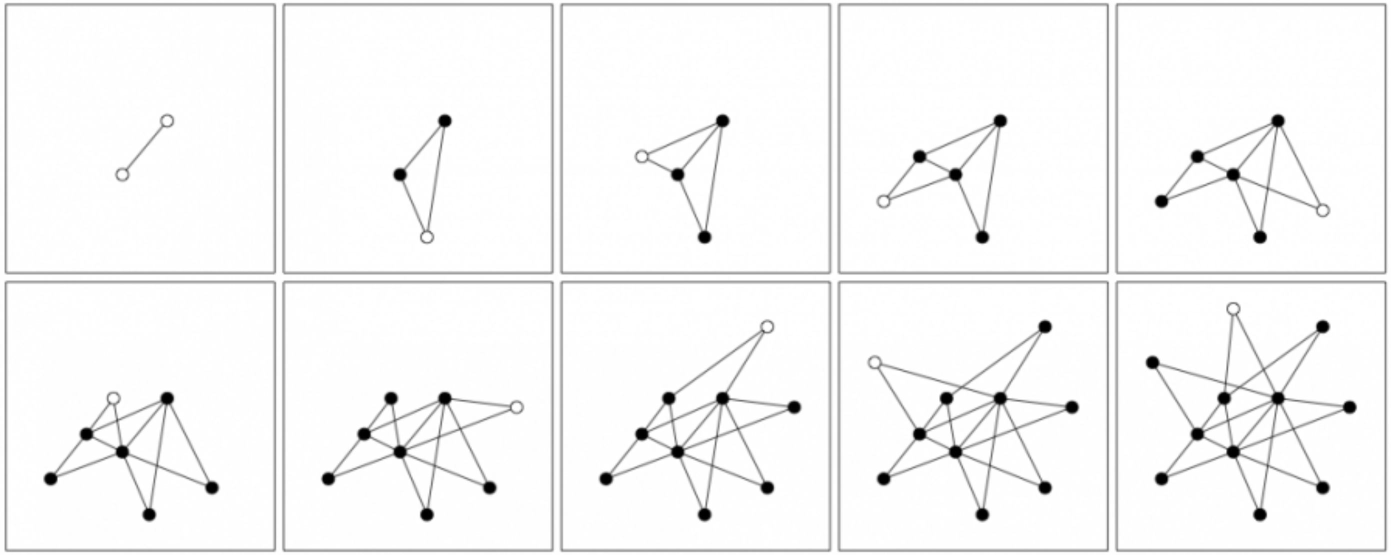
- *Growth* and *preferential attachment* cannot be captured by random networks
  - \- the degree distribution of real networks is different from the random network ones



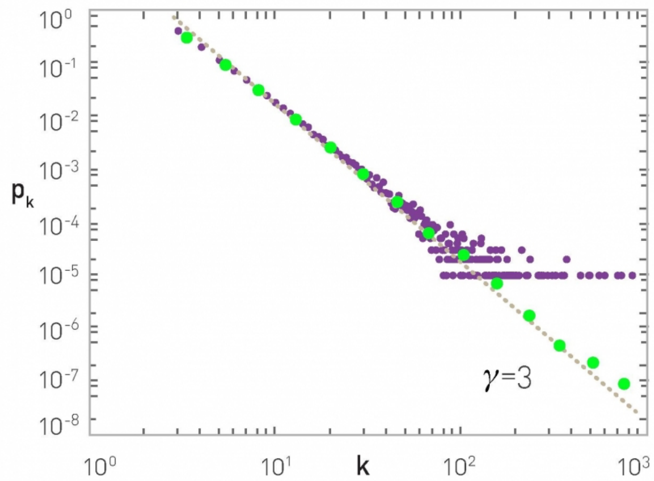
From [1]

The Barabási-Albert Model

- The BA model generates scale-free networks
  - start with m_0 nodes, and choose links arbitrarily (with at least one link per node)
  - then develop the networks with growth and preferential attachment
    - Growth: add a new node with $m \leq m_0$ links that connects to *m* nodes already in the network
    - Preferential attachment: probability that the new node connects to node *i* depends on $\Pi(k_i) = \frac{k_i}{\sum_j k_j}$
  - after *t* steps, the network has $N = t + m_0$ nodes and $mt + m_0$ links, and a power-law distribution

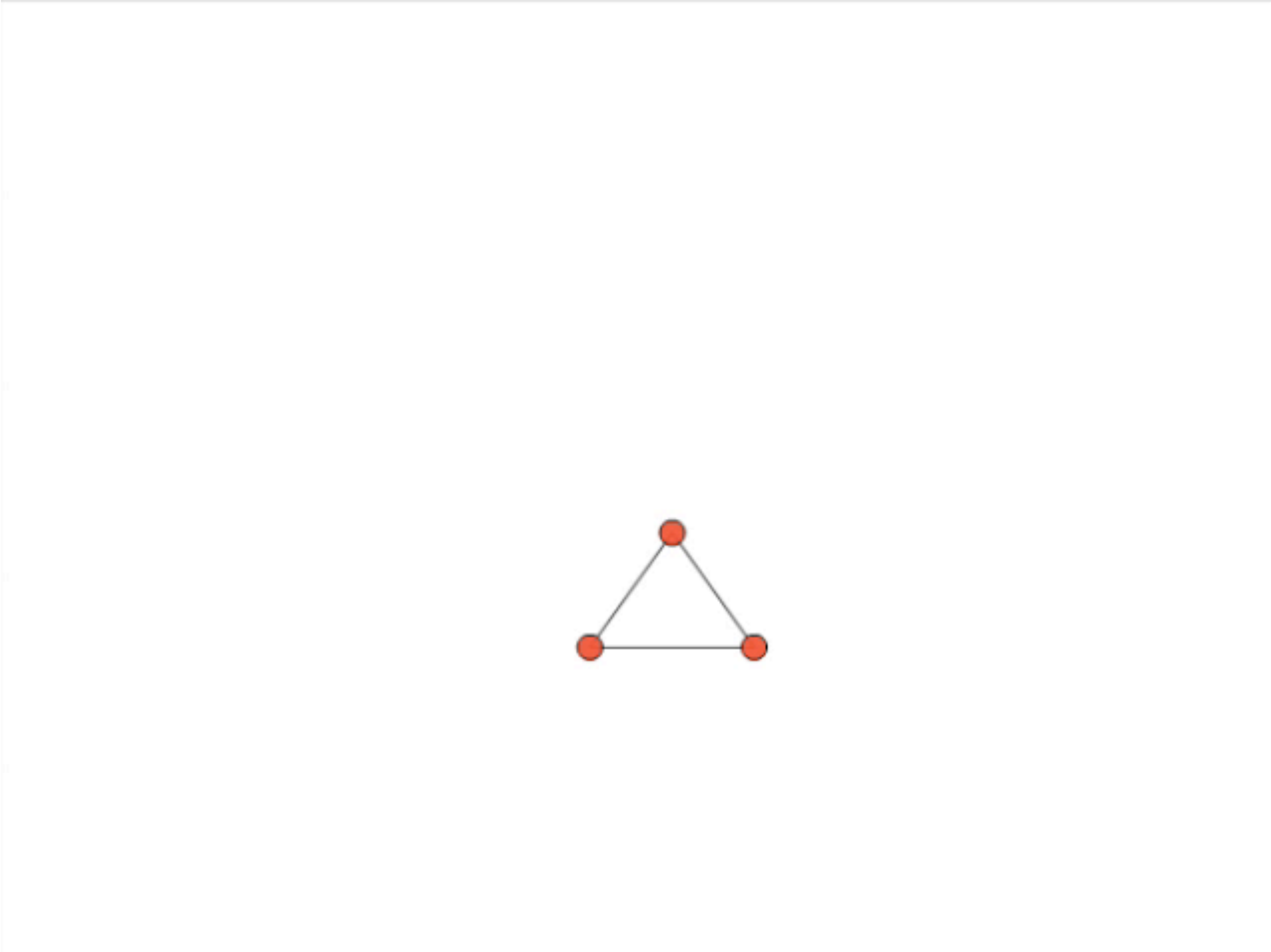


From [1]



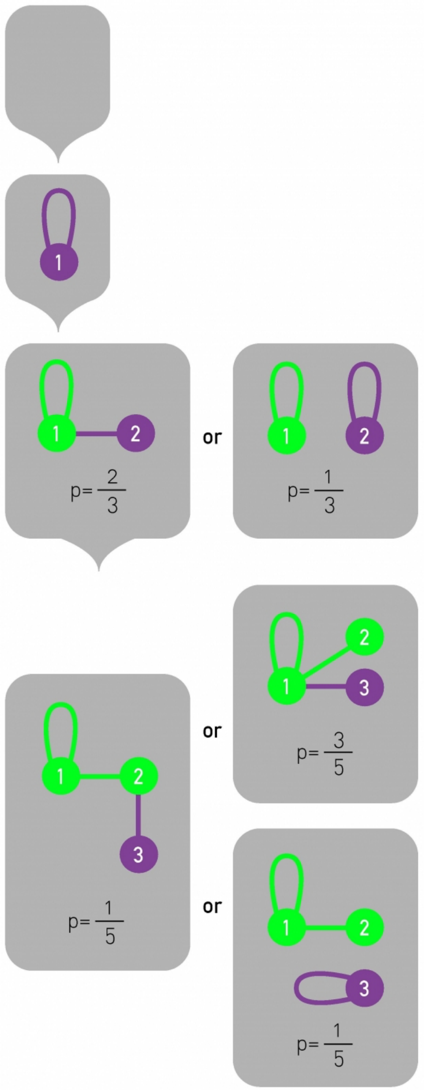
$N = 100000 \quad m = 3$

Emergence of a Scale-free Network



From [1]

Linearized Chord Diagram



From [1]

- LCD: alternative model to the BA model

more amenable to mathematical analysis

- For $m = 1$
  - start with $G_1(0)$, an empty graph
  - generate $G_1(t)$ by adding the node v_t with a single link to v_i chosen with probability

$$p = \begin{cases} \frac{k_i}{2t-1}, & \text{if } 1 \leq i \leq t-1 \\ \frac{1}{2t-1}, & \text{if } i = t \end{cases}$$

  - nodes can link to themselves, and permit multi-links and self-loops (negligible for large t)

- For $m > 1$
  - $G_m(t)$ is built by adding *m* links one by one
  - the outward half of each newly added link is counted in the degrees.

B. Bollobás, O. Riordan, J. Spencer, and G. Tusnády. The degree sequence of a scale-free random graph process. Random Structures and Algorithms, 18:279-290, 2001.

Degree evolution

- Consider the time evolution of a single node
  - Approximate k_i with a continuous variable (expectation of many realisations)
  - Rate of link aquisition: $$\frac{dk_i}{dt} = m\Pi(k_i) = mk_i / \left[\sum_{j=1}^{N-1} k_j\right]$$
  - with $\sum_{j=1}^{N-1} k_j = 2mt - m$, we have

$$\frac{dk_i}{dt} = \frac{k_i}{2t-1}$$ or, when $t \to \infty$, $$\frac{dk_i}{k_i} = \frac{1}{2}\frac{dt}{t}$$

  - after integration and with $k_i(t_i) = m$, we can write

$$k_i(t) = m\left(\frac{t}{t_i}\right)^{\beta}$$ with $\beta = \frac{1}{2}$ the dynamical exponent

All nodes follow the same dynamical law.

The growth is sublinear: each node has more nodes to link than earlier nodes.

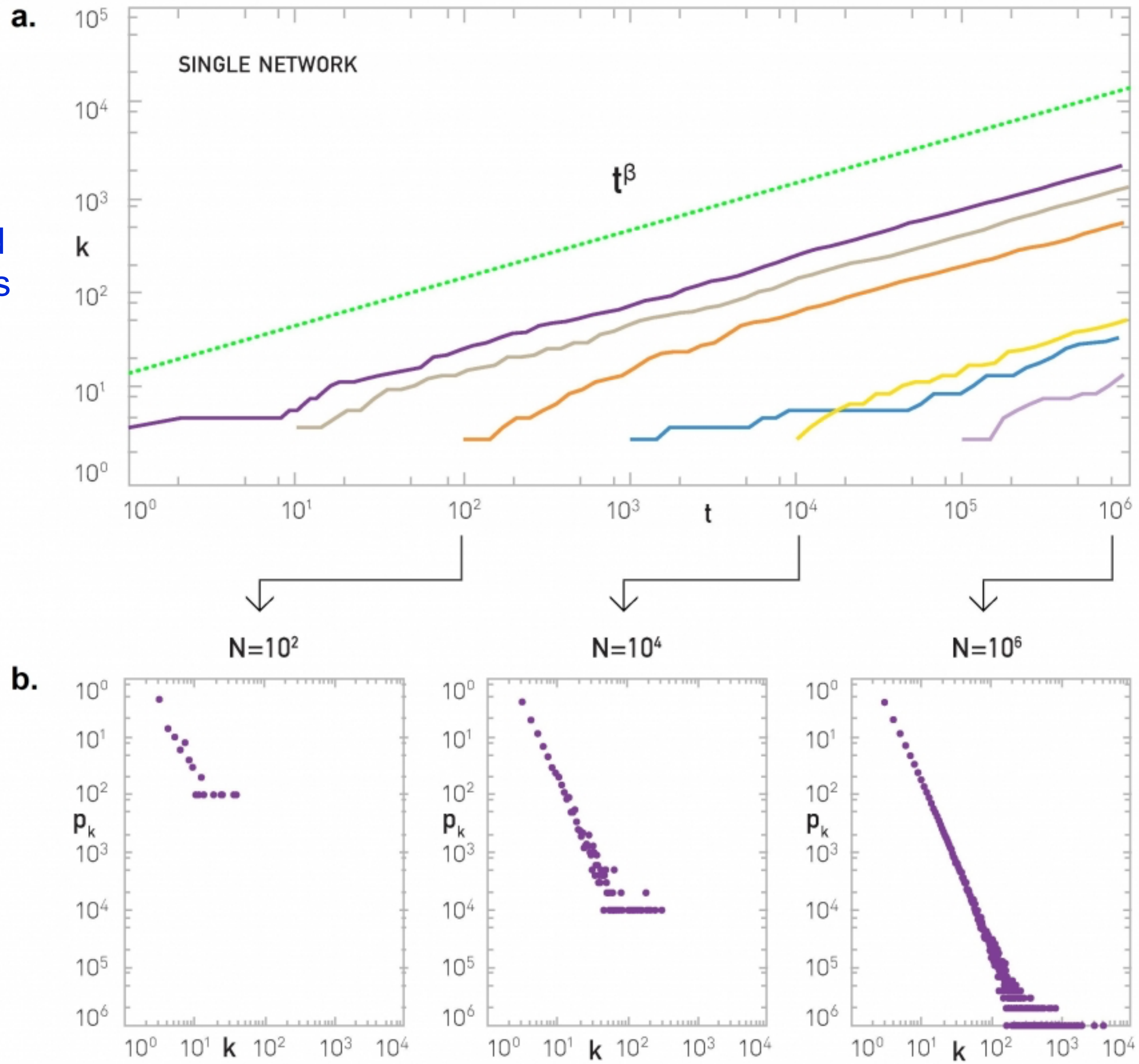
Hubs correspond to earlier nodes (*first-mover advantage*)

Link acquisition rate higher for older nodes, as $\frac{dk_i(t)}{dt} = \frac{m}{2}\frac{1}{\sqrt{t_i t}}$

Degree dynamics

Degree of nodes added at different time instants

Degree distribution of the network at different time instants



From [1]

Degree distribution

- The degree distribution in the BA model is a power-law distribution:

$$p(k) \approx 2m^{1/\beta} k^{-\gamma} \qquad \text{with} \qquad \gamma = \frac{1}{\beta} + 1 = 3$$

link between topology and dynamics

- Using continuum theory tools, we approximate
  - the number of nodes with degree $k_i(t) < k$ is $t\left(\frac{m}{k}\right)^{1/\beta}$, since BA adds one node at a time, and

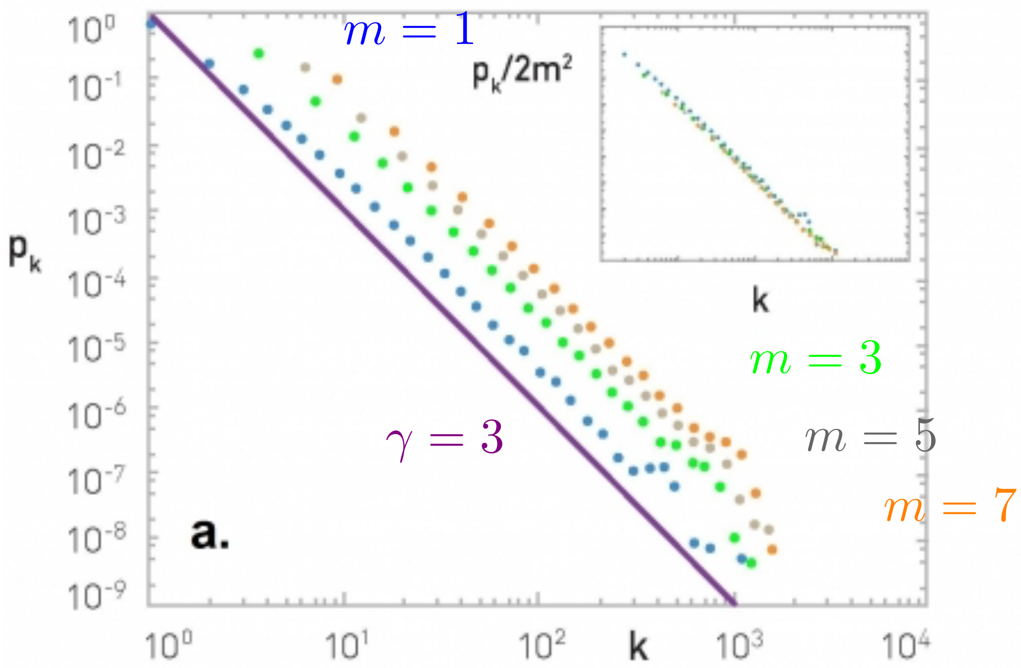
$$t_i < t\left(\frac{m}{k}\right)^{1/\beta} \quad \text{from} \quad k_i(t) = m\left(\frac{t}{t_i}\right)^{\beta}$$

  - the total number of nodes is $N = m_0 + t$ or $N \approx t$ when $t \to \infty$
  - the probability that a node has degree *k* or smaller is $P(k) = 1 - \left(\frac{m}{k}\right)^{1/\beta}$
  - the degree distribution is finally given by the derivative

$$p_k = \frac{\partial P(k)}{\partial k} = \frac{1}{\beta}\frac{m^{1/\beta}}{k^{1/\beta+1}} = 2m^2 k^{-3}$$

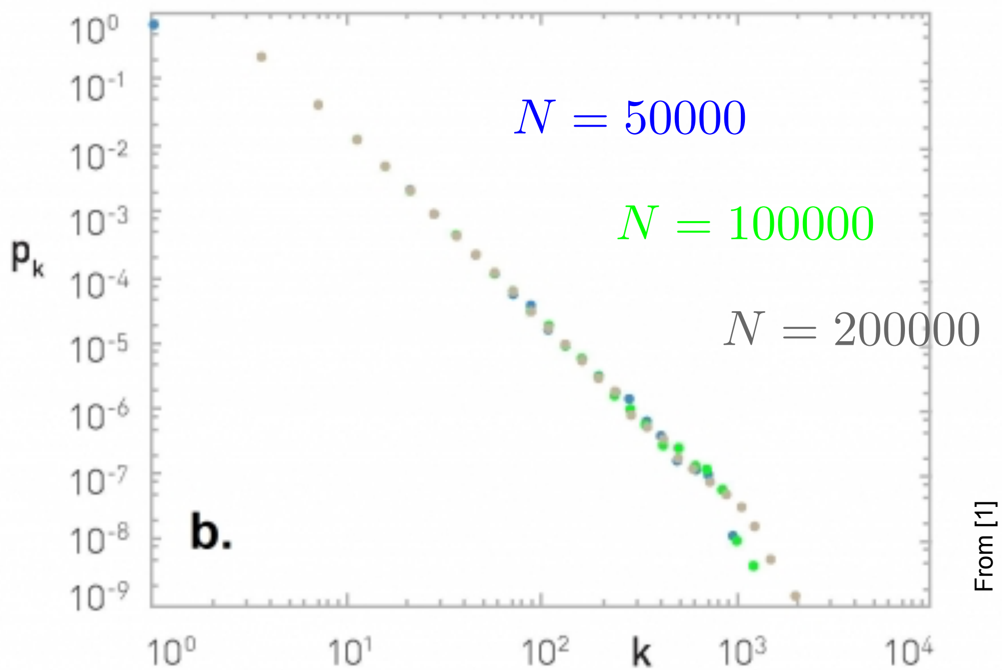
Kumar, Raghavan, Rajagopalan, Sivakumar, Tomkis and Upfal - Stochastic models for the Web graph, Proc. FOCS, 2000

Analytical predictions



$N = 100000$

$m_0 = m$



$m = m_0 = 3$

Slope independent of m and m_0
$p_k/(2m^2)$ independent of m

Degree distribution
independent of time and size

Exact degree distribution

- Continuum theory permits to compute only the degree exponent
- The exact degree distribution of the BA model is

$$p_k = \frac{2m(m+1)}{k(k+1)(k+2)}$$

  - for large *k* , $p_k \sim k^{-3}$ as approximated earlier
  - the degree exponent is independent of *m*
  - the degree distribution is independent of *t* and *N*: *stationary scale-free state.*
  - the degree distribution is proportional to m^2 for large m , as approximated earlier

B. Bollobás, O. Riordan, J. Spencer, and G. Tusnády. The degree sequence of a scale-free random graph process. Random Structures and Algorithms, 18:279-290, 2001.
S.N. Dorogovtsev, J.F.F. Mendes, and A.N. Samukhin. Structure of growing networks with preferential linking. Phys. Rev. Lett., 85:4633-4636, 2000.
P.L. Krapivsky, S. Redner, and F. Leyvraz. Connectivity of growing random networks. Phys. Rev. Lett., 85:4629-4632, 2000.

Growth-only model

- Let us consider a model A without preferential attachment
  - Growth: at each time step, add a new node with $m \leq m_0$ links
  - Attachment: probability of linking to previous nodes: $\Pi(k_i) = \dfrac{1}{(m_0 + t - 1)}$

independent of degree

- Using continuum theory, we approximate the degree distribution
  - the degree increases logarithmically with time,

$$k_i(t) = m \ln\left(e \frac{m_0 + t - 1}{m_0 + t_i - 1}\right)$$

much slower growth than the power law increase in the BA model

  - the degree distribution becomes

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

much faster decay than the power law in the BA model

The lack of preferential attachment eliminates the network scale-free character and the hubs!

A.-L. Barabási and R. Albert. Emergence of scaling in random networks. Science, 286:509-512, 1999
A.-L. Barabási, H. Jeong, R. Albert. Mean-field theory for scale free random networks. Physica A, 272:173-187, 1999

Preferential attachment-only model

- Let us consider a model B without growth
  - the network has *N* nodes and does not grow
  - preferential attachment: at each time step, a randomly selected node connects to node *i*, with probability

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}$$

nodes with no links gets $\Pi(1)$

- The number of links increases linearly with time
  - For large *t*, the degree also increases with time, $k_i(t) \approx \frac{2}{N} t$

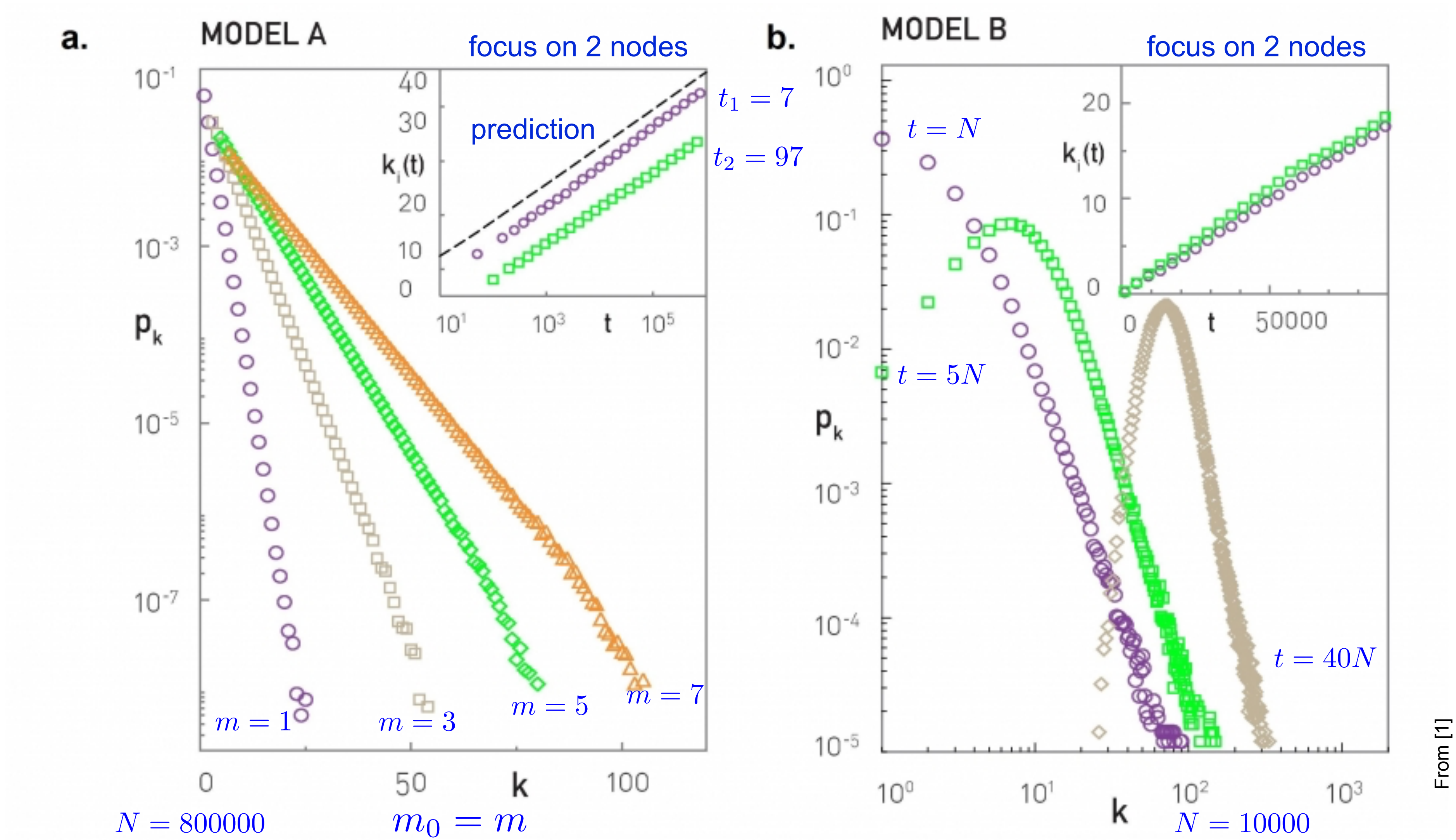
    the model adds links without changing the number of nodes
  - At early times, there are few links, and the model is close to BA with $m = 1$

    model with power-law tail
  - Yet, later the degrees converge to an average one and develops a peak. It becomes a complete graph for $t \to N(N-1)/2$ with degree $k_{\max} = N - 1$ and $p_k = \delta(N-1)$

The lack of growth eliminates the network scale-free character!

A.-L. Barabási and R. Albert. Emergence of scaling in random networks. Science, 286:509-512, 1999
A.-L. Barabási, H. Jeong, R. Albert. Mean-field theory for scale free random networks. Physica A, 272:173-187, 1999

Role of growth and preferential attachment



Measuring preferential attachment

- Test both hypotheses behind preferential attachment
  - A. connection probability depends on node's degree
  - B. the connection probability $\Pi(k)$ is linear in *k*

- Look at maps of a network at different time instants
  - for nodes that have changed degrees, measure: $\Delta k_i = k_i(t + \Delta t) - k_i(t)$
  - for sufficiently small Δt we should have an approximation of the functional form

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

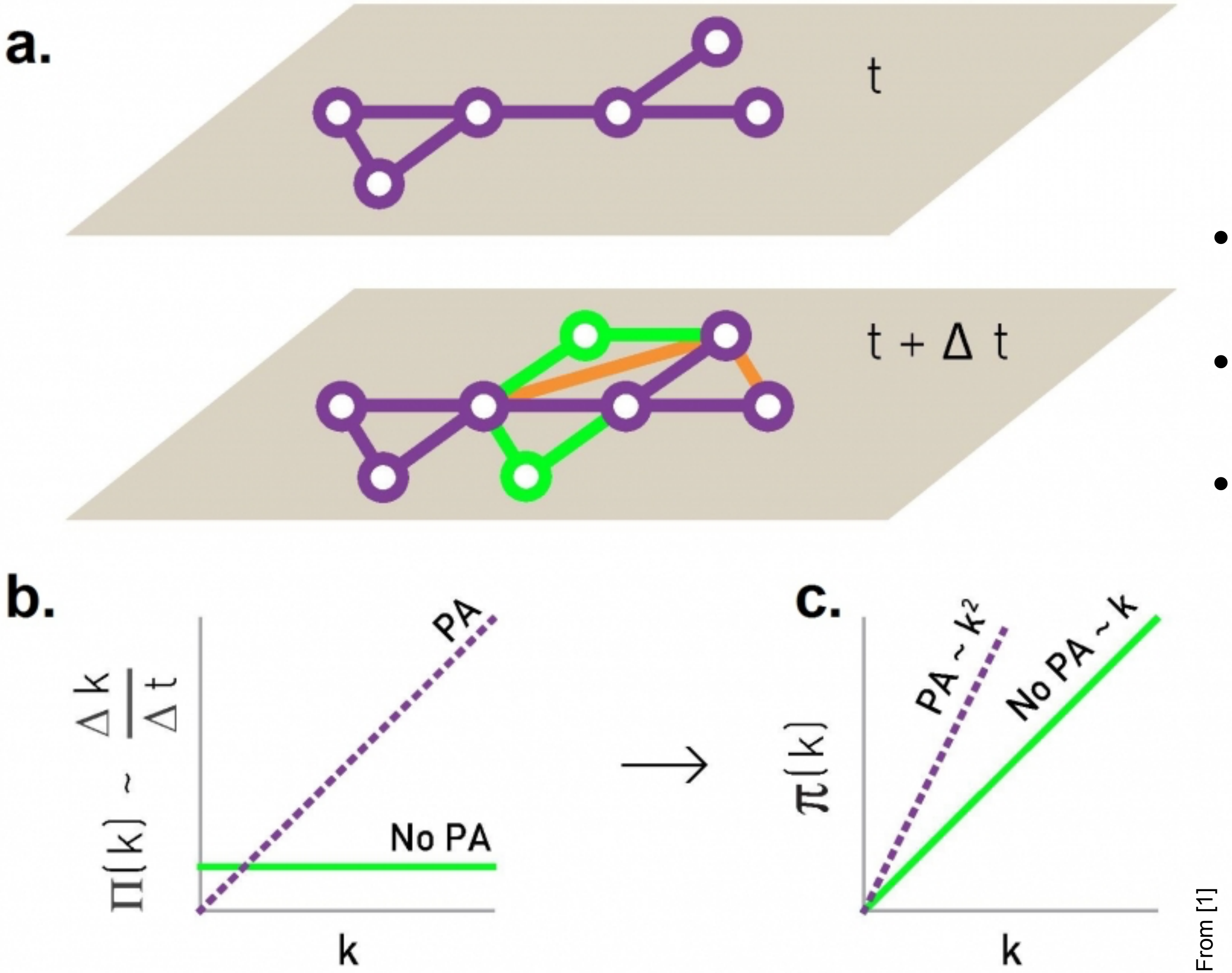
often noisy

  - in practice, it is better to measure the cumulative preferential attachment function

$$\pi(k) = \sum_{k_i=0}^{k} \Pi(k_i)$$

should grow quadratically with linear preferential attachment

Detecting Preferential Attachment



- Compare maps at different time instants
- Look at nodes that have new links due to new nodes
- With preferential attachment, there is a linear dependency

$$\Pi(k) \sim \frac{\Delta k}{\Delta t}$$

or a quadratic relationship in the cumulative PA function

$$\pi(k) \sim k^2$$

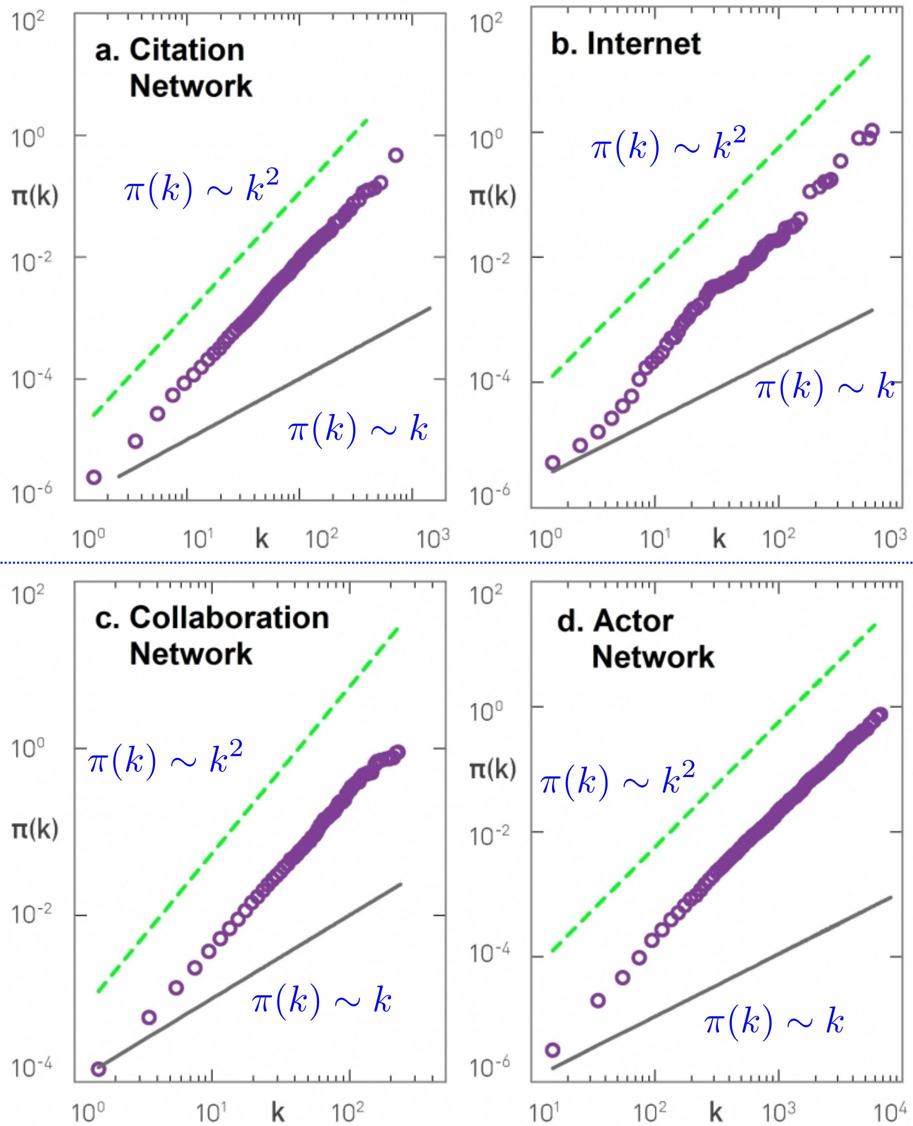
Evidence of Preferential Attachment

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

best fit with $\alpha \approx 1$

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

best fit with $\alpha \approx 0.9 \pm 0.1$



From [1]

Sublinear Preferential Attachment

- This mode corresponds to having $0 < \alpha < 1$ in $\Pi(k) \sim k^\alpha$
  - for $\alpha > 0$ new nodes favour the more connected nodes
  - yet, for $\alpha < 1$, the bias is not large, not sufficient to have a scale-free network

- The degree distribution follows a stretched exponential distribution

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

$\mu(\alpha)$ only depends weakly on α

  - the exponential cutoff limits the size and number of hubs

- Also, the size of the largest degree is limited

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

  - slower growth than the polynomial one in a scale-free network - hubs are therefore smaller

Examples: scientific collaboration network, or actor network

P.L. Krapivsky, S. Redner, and F. Leyvraz. Connectivity of growing random networks. Phys. Rev. Lett., 85:4629-4632, 2000

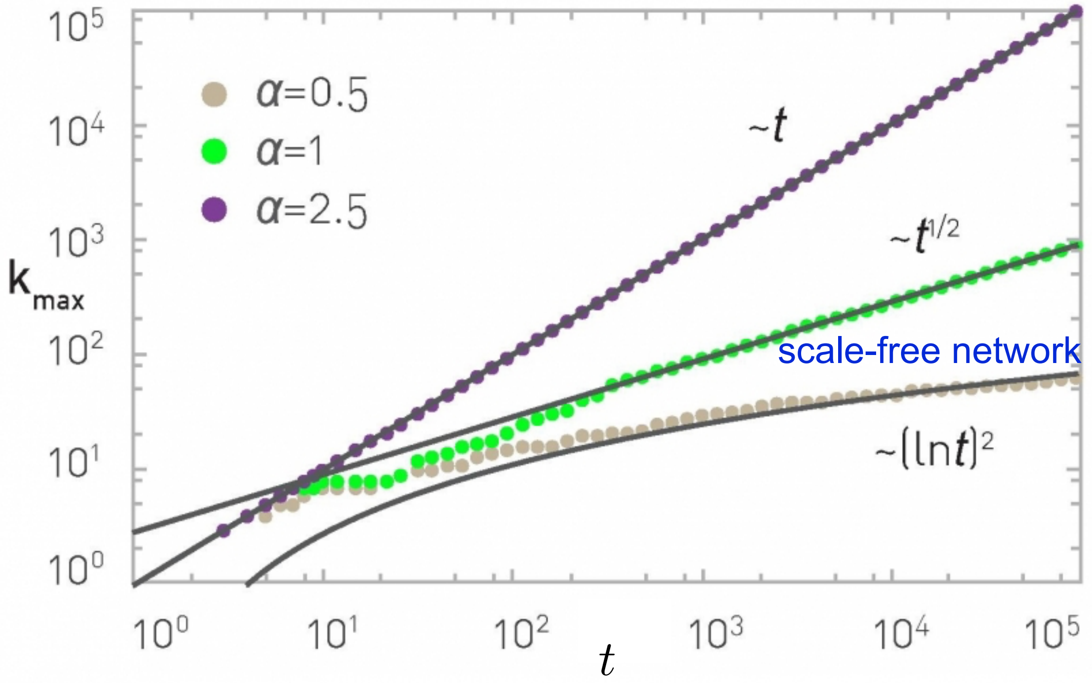
Superlinear Preferential Attachment

- This mode corresponds to having $\alpha > 1$ in $\Pi(k) \sim k^{\alpha}$
- High tendency to link to highly connected nodes
  - *rich-gets-richer* process
  - almost all nodes connect to a few super-hubs: obvious *winner-takes-all* for $\alpha > 2$ with the emergence of a *hub-and-spoke* network
  - the size of the largest hub gets $k_{\max} \sim t$

- Non-linear preferential attachment changes the degree distribution
  - To have a pure power-law, the preferential attachment has to be linear!

P.L. Krapivsky, S. Redner, and F. Leyvraz. Connectivity of growing random networks. Phys. Rev. Lett., 85:4629-4632, 2000

Growth of the hubs



From [1]

- Let us replace $\Pi(k) = \dfrac{k}{\sum_j k_j}$ by $\Pi(k) \sim k^{\alpha}$

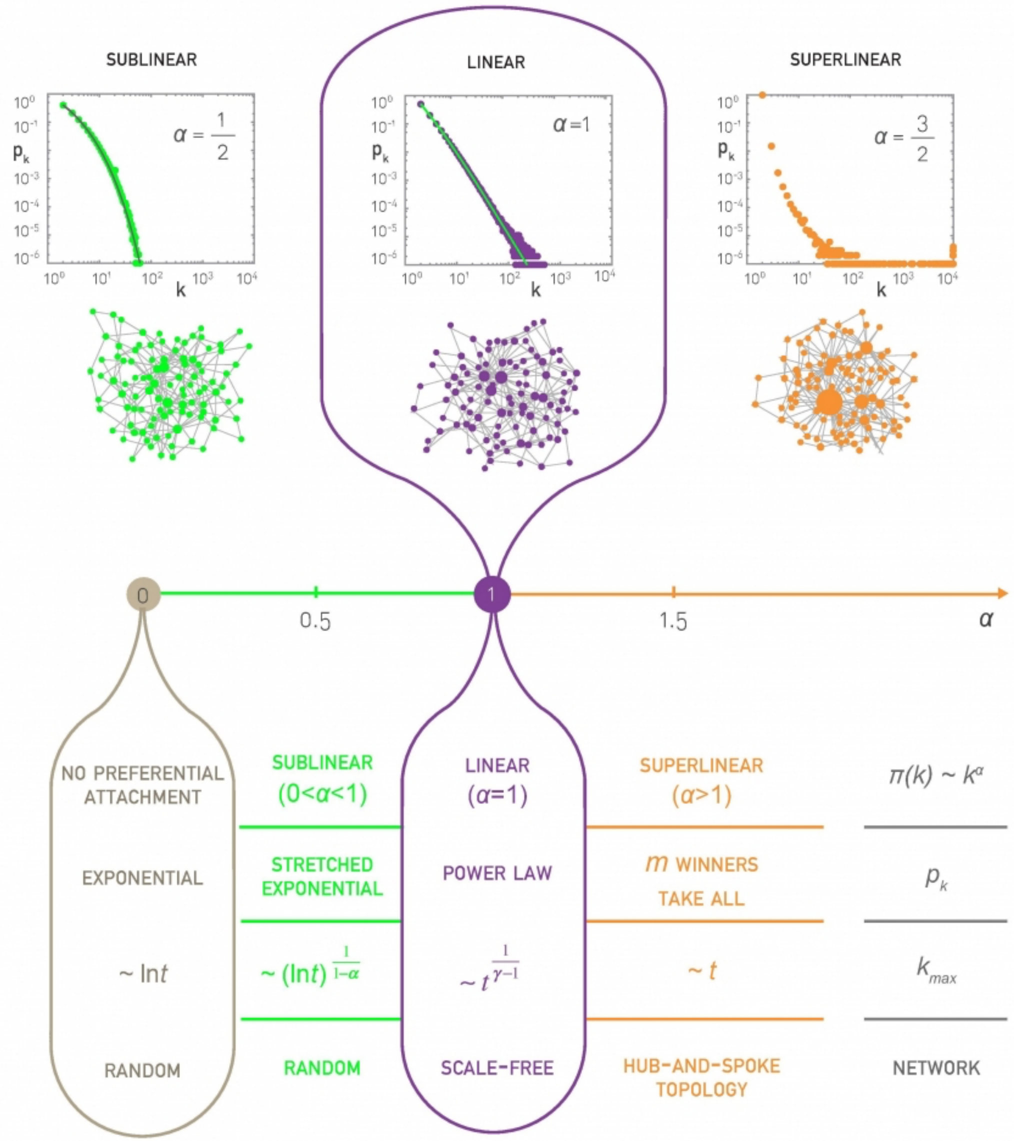
- For
  - $\alpha = 0$: model A (growth only)
  - $\alpha = 1$: BA model, scale-free network with

    $$P(k) = 1 - \left(\frac{m}{k}\right)^{1/\beta}$$

  - $0 < \alpha < 1$: sublinear regime with fewer and smaller hubs
  - $\alpha > 1$: super-linear regime, with hub-and-spoke topology

More details in [1], Chapter 5.

Pref. attachment regimes: summary



$N = 10^4$

$N = 100$

From [1]

The origins of preferential attachment

- Philosophy A: preferential attachment results of the interplay between random events and some structural properties
  - local or random mechanisms
  - e.g., link selection model, or copying model
- Philosophy B: preferential attachment results from nodes balancing conflicting needs following a cost-benefit analysis
  - global or optimised mechanisms, that assume the knowledge of the whole network

- The tension between randomness and optimisation philosophies is not new
  - sort of agreement that randomness drives power-laws in complex networks, while optimisation is important in explaining the origins of preferential attachment…
- Most complex systems are driven by a bit of reason and a bit of luck
  - preferential attachment wins either way: it is present in many different systems!

Link selection model

- At each time step, a new node is added, it randomly selects a link, and then randomly picks one of its end nodes for creating a new connection
  - probability that connected node has degree *k*

$$q_k = Ckp_k$$

increases with node degree
increases with p_k

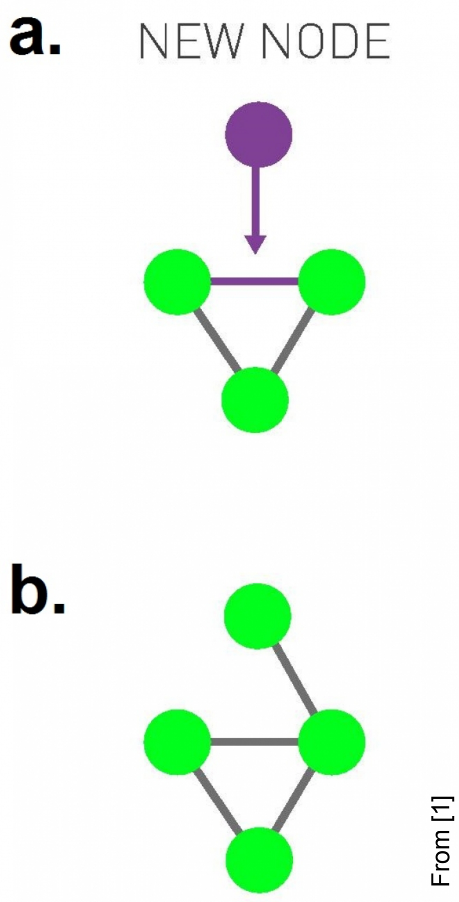
  - by normalization

$$\sum q_k = 1 \quad \text{and} \quad C = \frac{1}{\langle k \rangle}$$

so that $$q_k = \frac{kp_k}{\langle k \rangle}$$

linear preferential attachment

- Simplest local mechanism that generates a scale-free network without explicit preferential attachment



S.N. Dorogovtsev and J.F.F. Mendes. Evolution of networks. Oxford Clarendon Press, 2002

Copying model

- A new node connects with probability *p* to a randomly chosen target node *u*, or with probability *1-p* to one of the nodes the target *u* points to (= *copy*).

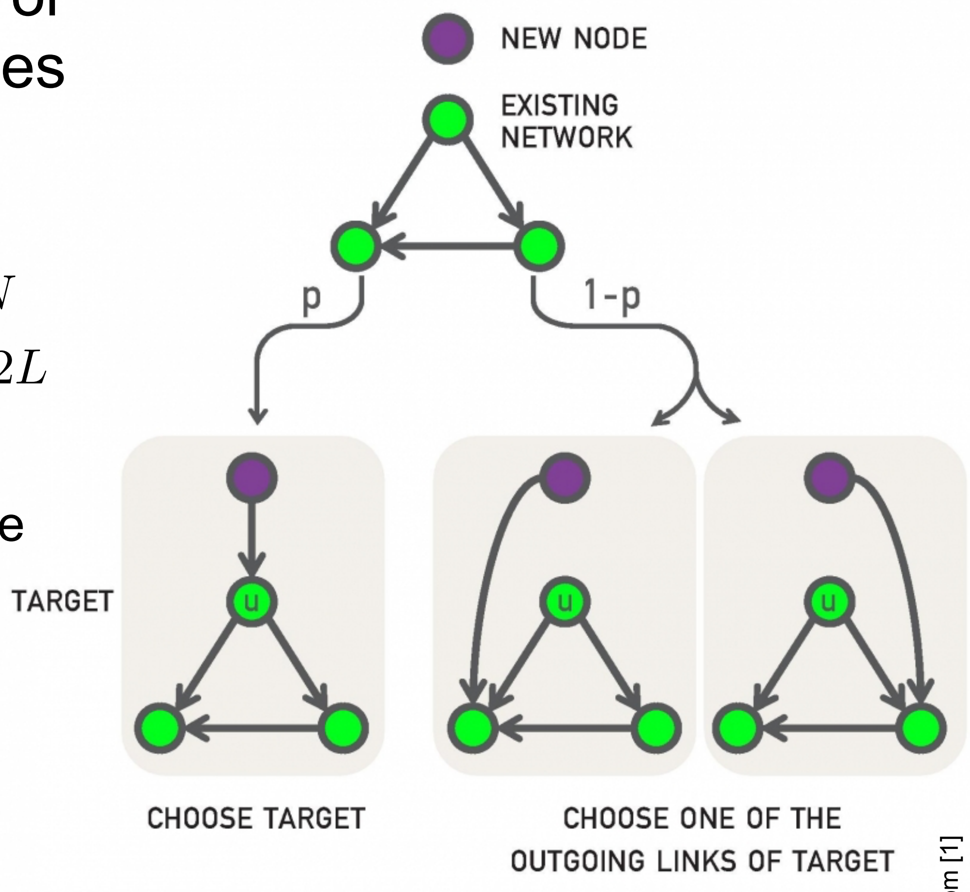
  - probability of selecting a random node: $1/N$
  - probability of copying a degree-k node: $k/2L$

equivalent to selecting a node linked to a randomly selected link

  - probability of connecting to a degree-k node

$$\Pi(k) = \frac{p}{N} + \frac{1-p}{2L}k$$

linear preferential attachment



From [1]

- Particularly relevant for real systems
  - social networks, citation networks, protein interaction network

J.M. Kleinberg, R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins. The Web as a graph: measurements, models and methods. Intern.Conf. on Combinatorics and Computing, 1999.
R. Kumar, P. Raghavan, S. Rajalopagan, D. Divakumar, A.S. Tomkins, and E. Upfal. The Web as a graph. Proceedings of the 19th Symposium on principles of database systems, 2000.

Optimisation model

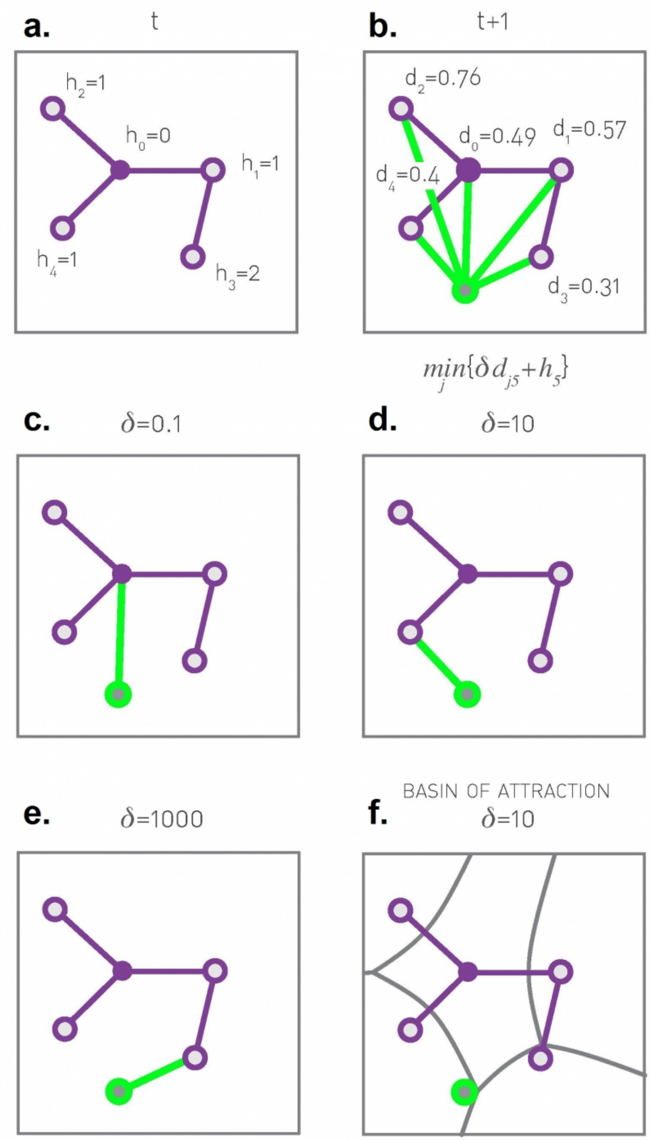
- Assumption: connection by balancing costs and benefits
- Example cost function (unit square)

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

Euclidian distance distance for *first* node

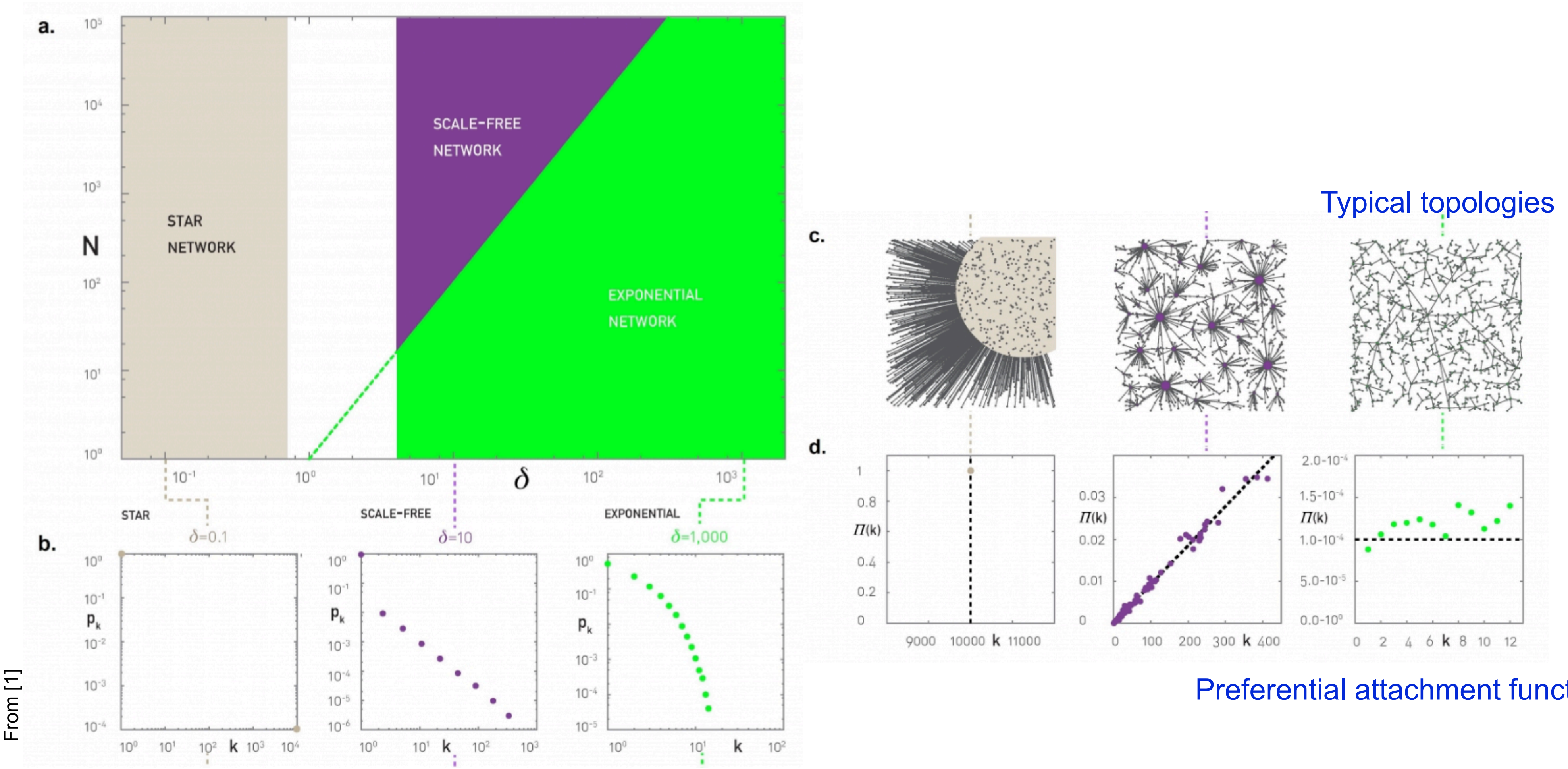
- $\delta < (1/2)^{1/2}$ - *star network:* the Euclidian distance is essentially irrelevant
- $\delta \geq N^{1/2}$ - *random network:* each node connects to the closest node
- $4 \leq \delta \leq N^{1/2}$ - *scale-free network*: power law distribution development by
  - optimisation: basin of attraction for each node (size correlated with h_j and k_j)
  - random location selection: larger basin leads to largest probability

linear preferential attachment



From [1]

Scaling in the Optimization Model



Preferential attachment function

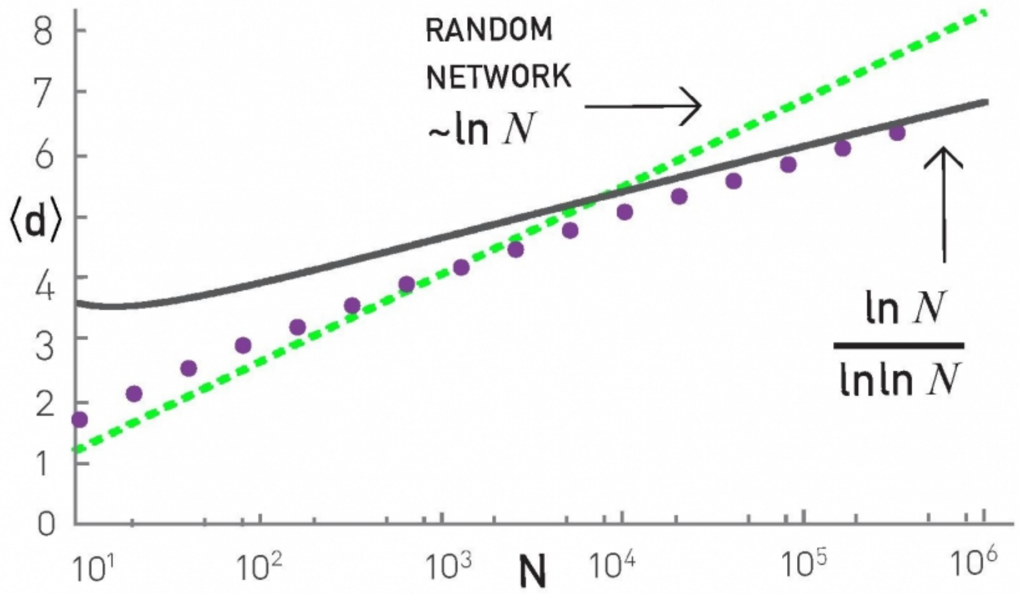
From [1]

Degree distribution

$N = 10^4$

A. Fabrikant, E. Koutsoupias, and C. Papadimitriou. Heuristically optimized trade-offs: a new paradigm for power laws in the internet. In Proceedings of the 29th International Colloquium on Automata, Languages, and Programming (ICALP), pages 110-122, Malaga, Spain, July 2002

Diameter



10 runs and $m = 2$

- The network diameter in the BA model follows

$$d_{\max} \sim \frac{\ln N}{\ln \ln N}$$

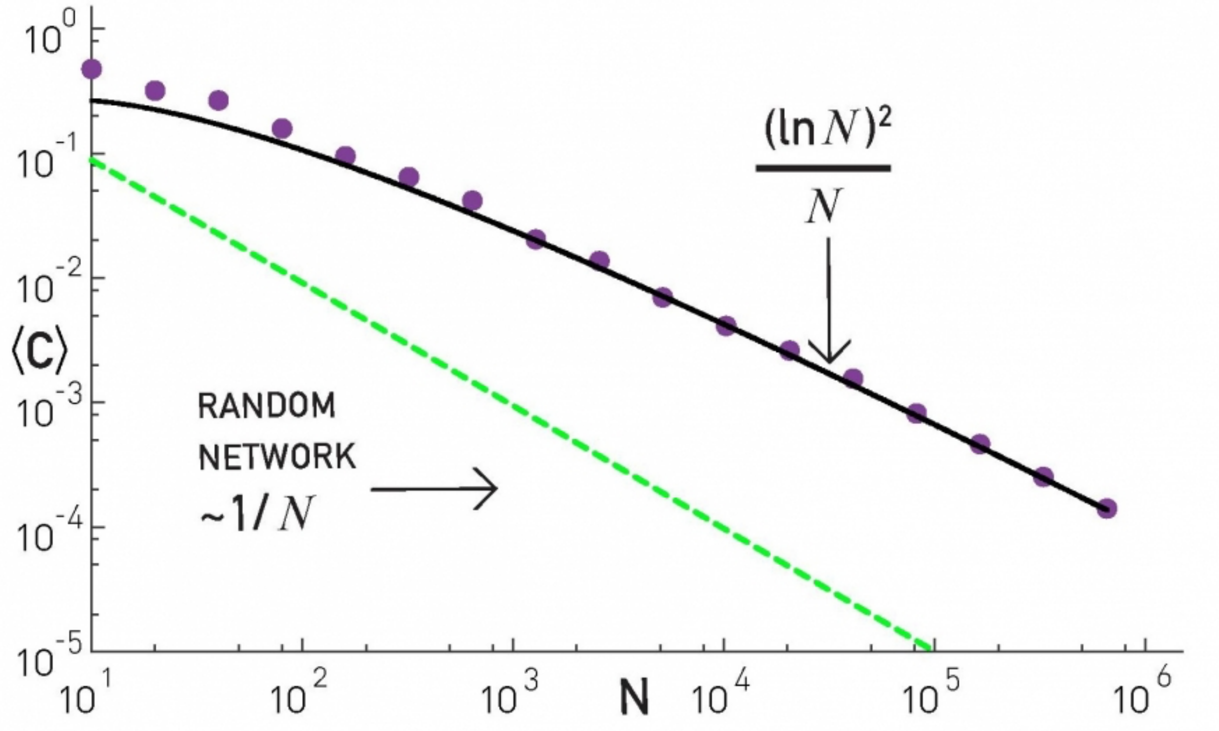
$m > 1$ and large N

- The network grows slower than *ln(N)*
  - distances are smaller than in random networks
- The average distance $\langle d \rangle$ scales in a similar way
  - *ln(N)* for small *N*
  - similar to diameter for large *N*

R. Cohen and S. Havlin. Scale-free networks are ultra small. Phys. Rev. Lett., 90:058701, 2003.
B. Bollobás and O.M. Riordan. The diameter of a scale-free random graph. Combinatorica, 24:5-34, 2004.

Clustering coefficient

- The clustering coefficient in the BA model follows $\langle C \rangle \sim \dfrac{(\ln N)^2}{N}$
  - different from *1/N* dependence in random networks - grows larger for large *N*



10 runs and $m = 2$

K. Klemm and V.M. Eguluz. Growing scale-free networks with small-world behavior. Phys. Rev. E, 65:057102, 2002.
B. Bollobás and O.M. Riordan. Mathematical results on scale-free random graphs. In Handbook of Graphs and Networks, edited by S. Bormholdt and A. G. Schuster, Wiley, 2003.

Clustering coefficient computation

- Number of triangles in the model with node *l* of degree k_l

continuous degree approximation

$$Nr_l(\triangleleft) = \int_{i=1}^{N} di \int_{j=1}^{N} dj P(i,j)P(i,l)P(j,l)$$

probability for 3 nodes to form a triangle

- Probability of a link between *i* and *j* (assuming node *j* arrived with $t_j = j$)

$$P(i,j) = m\Pi(k_i(j)) = m\frac{k_i(j)}{\sum_{l=1}^{j} k_l(j)} = m\frac{k_i(j)}{2mj}$$

- Using $k_i(t) = m\left(\frac{t}{t_i}\right)^{\beta}$ we have

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

$$P(i,j) = \frac{m}{2}(ij)^{-\frac{1}{2}}$$

$$Nr_i(\triangleleft) = \int_{i=1}^{N} di \int_{j=1}^{N} dj P(i,j)P(i,l)P(j,l)$$
$$= \frac{m^3}{8}\int_{i=1}^{N} di \int_{j=1}^{N} dj (ij)^{-\frac{1}{2}}(il)^{-\frac{1}{2}}(jl)^{-\frac{1}{2}}$$
$$= \frac{m^3}{8l}\int_{i=1}^{N} \frac{di}{i} \int_{j=1}^{N} \frac{dj}{j} = \frac{m^3}{8l}(\ln N)^2$$

- The clustering coefficient is

$$C_l = \frac{2Nr_l(\triangleleft)}{k_l(k_l-1)} \quad \text{or} \quad C_l = \frac{\frac{m^3}{4l}(\ln N)^2}{k_l(N)(k_l(N)-1)}$$

with $k_l(N) = m\left(\frac{N}{l}\right)^{\frac{1}{2}}$

$$k_l(N)(k_l(N)-1) \approx k_l^2(N) = m^2\frac{N}{l}$$

for large k_l

Finally: $C_l = \frac{m}{4}\frac{(\ln N)^2}{N}$

Summary - BA model

- Network structure and evolution are inseparable
  - to understand the topology of a complex system, we need to describe how it came into being.
- Growth and preferential attachment are necessary to create scale-free networks
  - some network formation processes do not have explicit preferential attachment
  - the BA model appear to capture the origin of the scale-free topology of many real systems
- The model still has limitations
  - undirected links, disappearance of nodes, intrinsic characteristics of nodes, etc, are missing

At a glance: Barabási-Albert Model

- 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 $\langle d \rangle \sim \dfrac{\ln N}{\ln \ln N}$
- Clustering coefficient $\langle C \rangle \sim (\ln N)^2/N$

References

[1] Network Science, by Albert-László Barabási, 2016 - Chapter 5

[2] Networks: An Introduction, by M. Newman, 2010