Network Science

#4 Preferential attachment

Preferential attachment

Network expansion model

Q

- ☐ How to generate a network with arbitrary distribution p_k?
- ☐ How to provide a justification to the power law?

A

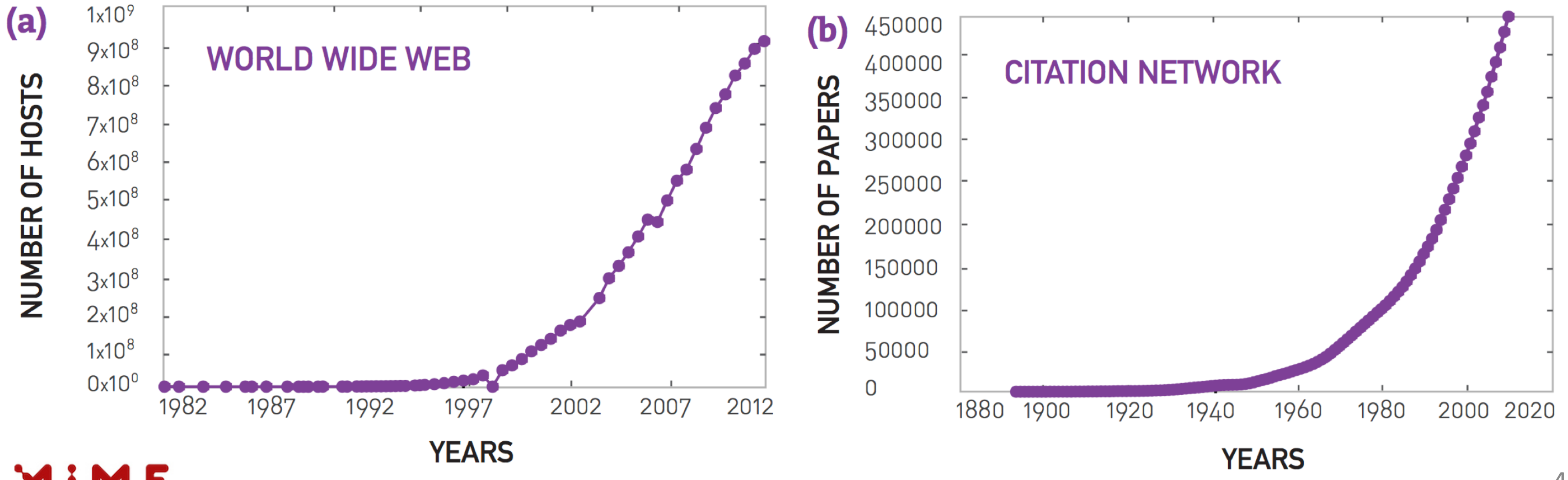
- ☐ Use the network expansion model

Network expansion model

#1. Growth

- Erdös-Rényi is a static network
- Real networks are dynamic: they expand through the addition of new nodes

e.g., www, citation network, actor network



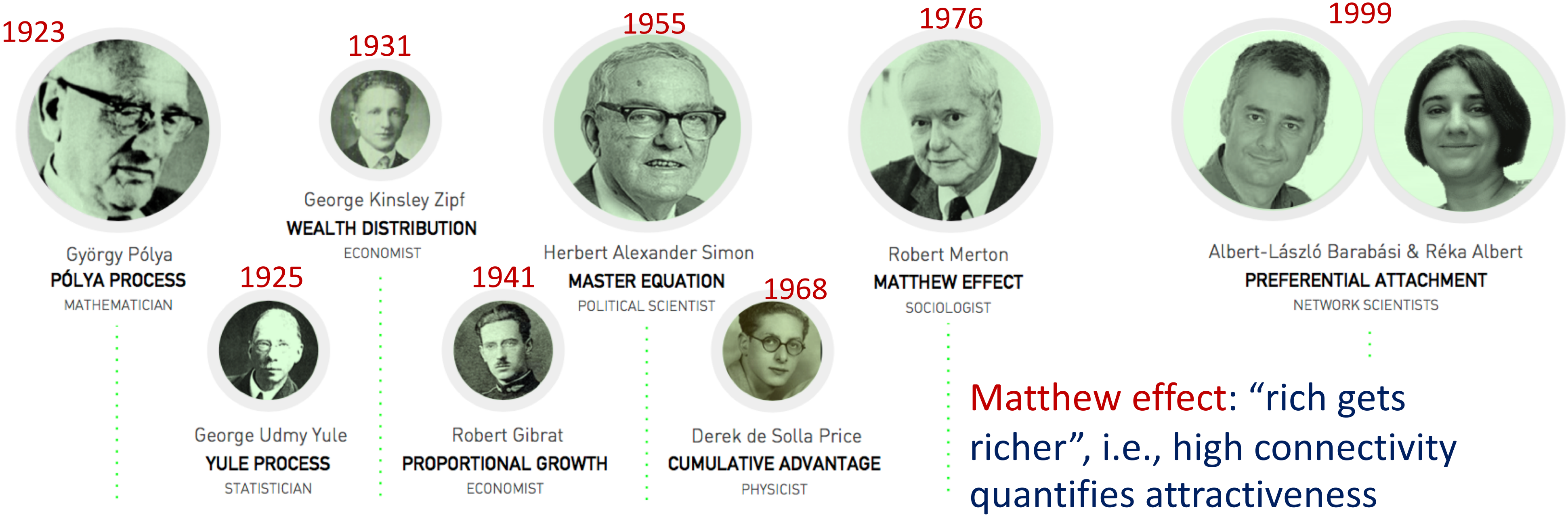
Network expansion model

#2. Preferential attachment

- ❑ Nodes link to the more connected nodes

  e.g., think of www

- ❑ This idea has a long history



Matthew effect: “rich gets richer”, i.e., high connectivity quantifies attractiveness

Explaining preferential attachment

- ☐ Citation network

  researchers decide what papers to read and cite by "copying" references from papers they have read → papers with more citations are more likely to be cited

- ☐ Social network

  the more acquaintances an individual has, the higher the chancer of getting new friends, i.e., we "copy" the friends of friends → difficult to get friends if you have none

This is called the copying model

Barabási-Albert model [1999]

Start with m_0 nodes arbitrarily connected, with $\langle k \rangle$=m

- ☐ Growth:

  add a node (the Nth) with m links that connect the node to nodes in the network

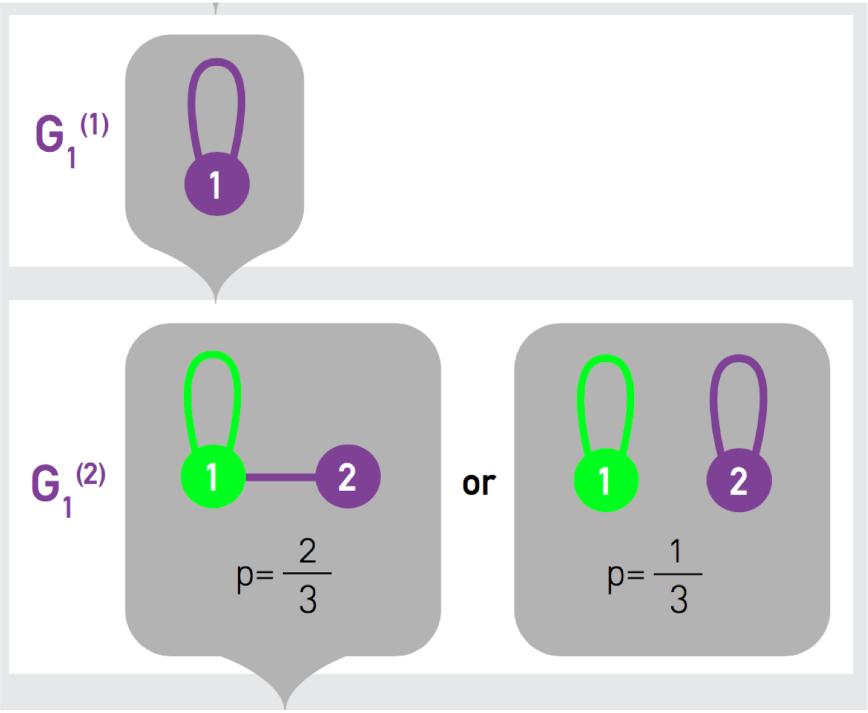
- ☐ Preferential attachment:

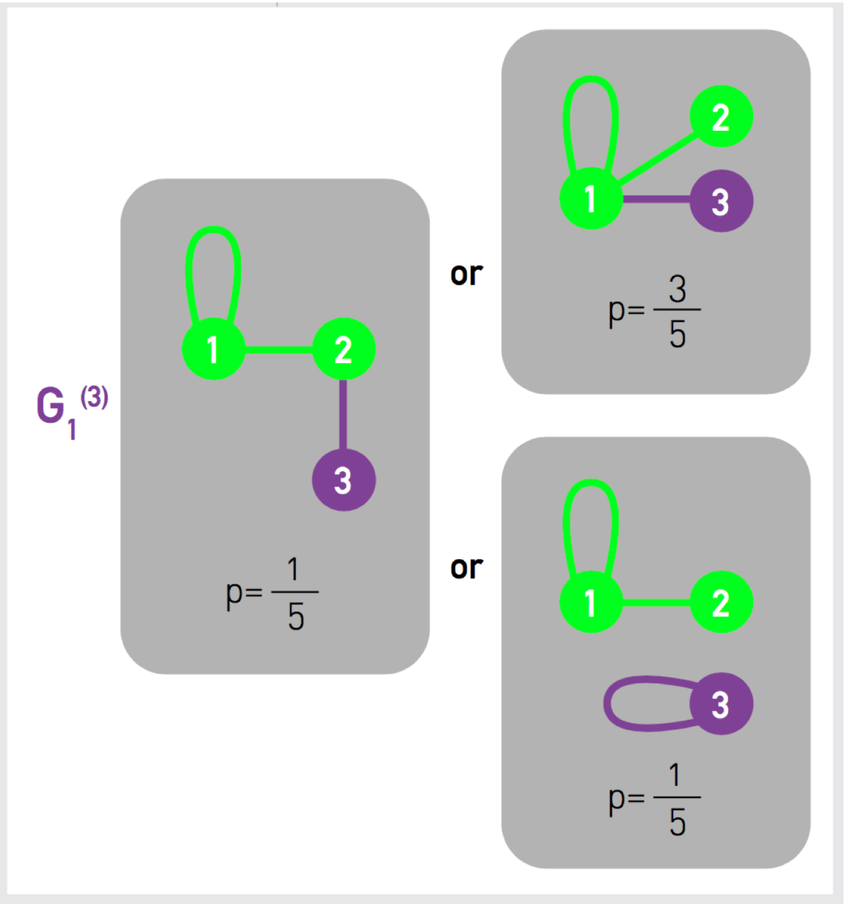
  $p_i = k_i / C$ probability of connecting to node i

  $p_i = 1/C$ for self-loops

$$C = 1 + \sum_i k_i = 1 + 2(N\text{-}1)m$$

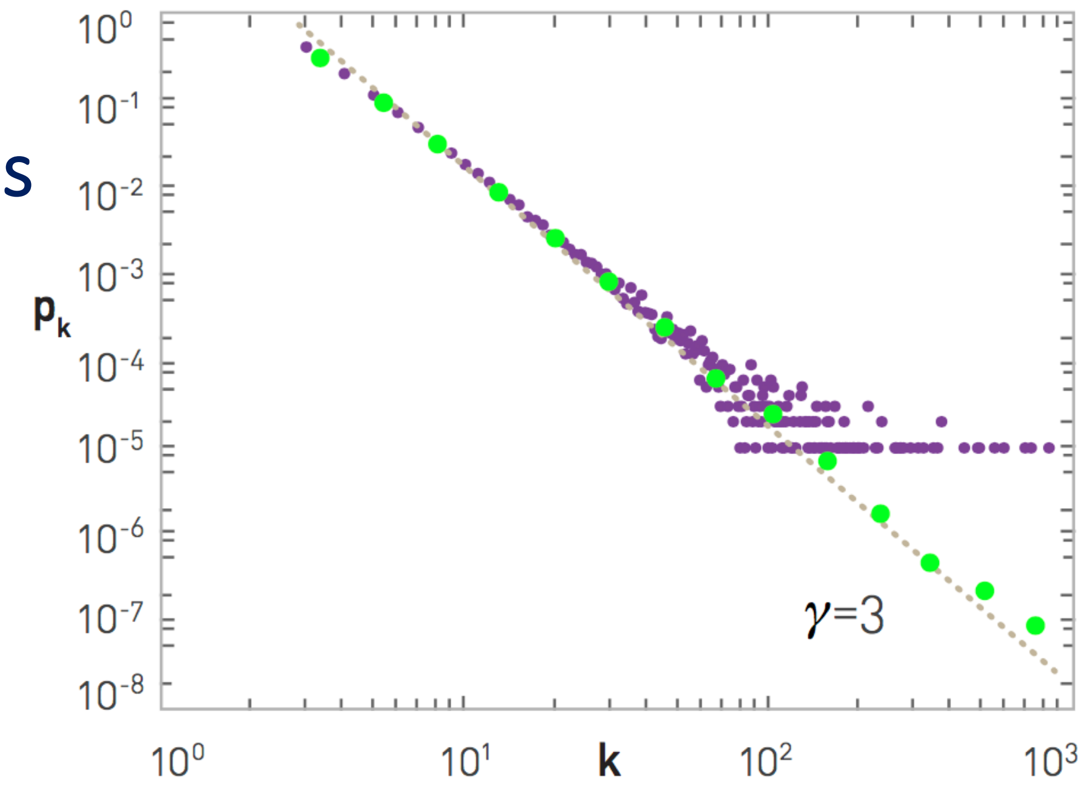
Example with *m*=1





Barabási-Albert model

- Depending on the implementation there might be self/multiple links
- Most nodes have a small degree (exactly *m* for the youngest ones)
- Hubs appear
- The average degree is $\langle k \rangle = 2m$, and in fact $L = Nm = \frac{1}{2}\langle k \rangle N$
- The resulting degree distribution is always a power-law with exponent $\gamma = 3$



Approximate analysis

- Increase in the degree (at each step)

$$\Delta k_i \simeq \underset{\substack{\uparrow\\ \text{trials}}}{m} \cdot \underset{\substack{\uparrow\\ \text{probability per trial}}}{k_i} / (1+2m(N-1)) \simeq k_i / 2N$$

- Approximation in the continuous domain

$$\Delta k_i \simeq \mathrm{d}k_i/\mathrm{d}N \rightarrow \boxed{\mathrm{d}k_i/k_i \simeq \tfrac{1}{2}\,\mathrm{d}N/N}$$

- Integration

$$\ln(k_i) = \tfrac{1}{2} \ln(N) + \text{cost.} \rightarrow k_i = c\, N^{\frac{1}{2}}$$

- Recalling that node *i* joins the network at time $N = i$

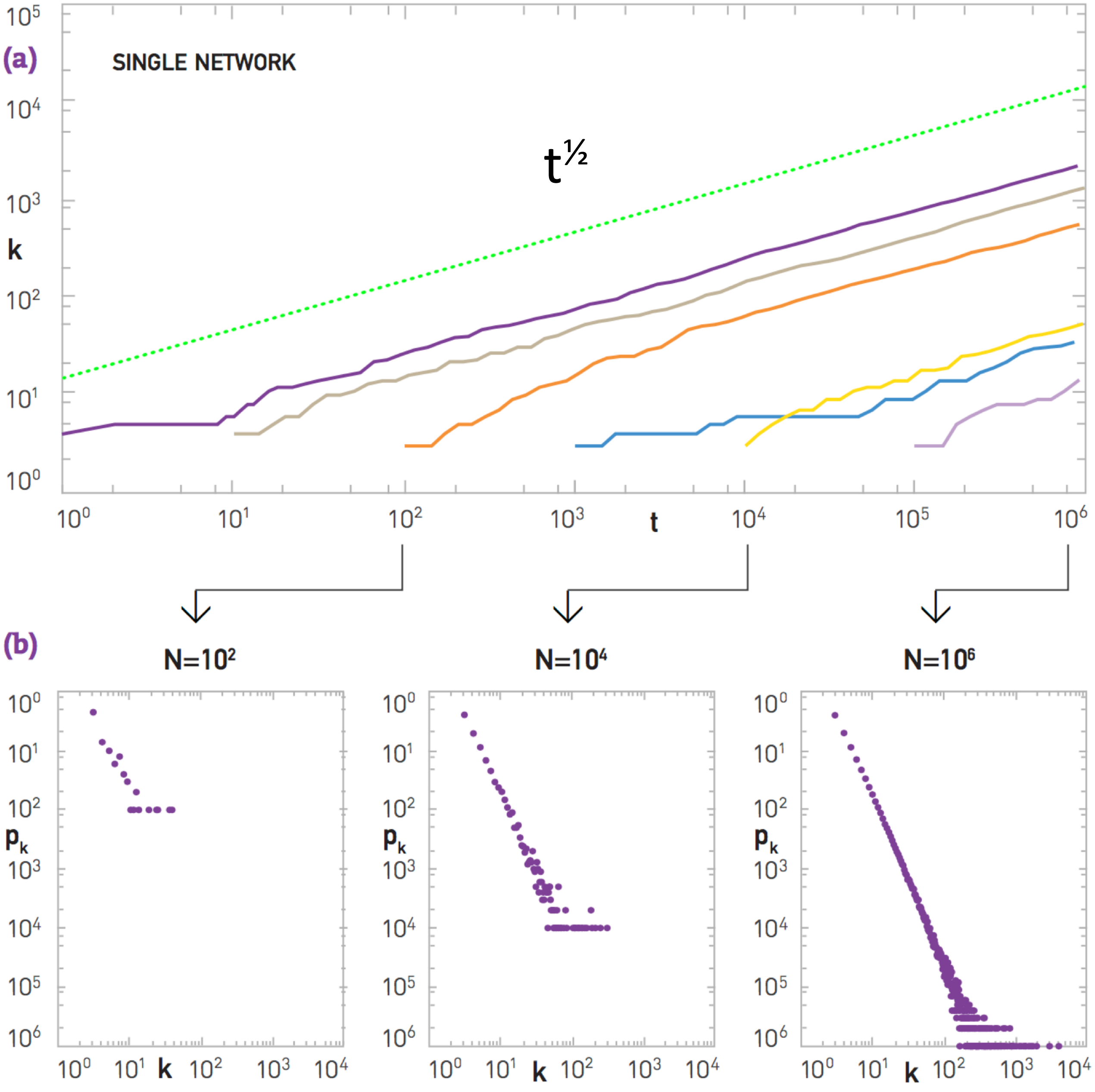
$$k_i\,(N=i) = m \rightarrow \boxed{k_i\,(N) = m\,(\mathrm{N/i})^{\frac{1}{2}}}$$

½ is the dynamic exponent

Implications of $k_i = m\,(N/i)^{½}$

- ☐ k_i sub-linearly increases as a power law with exponent ½ – all nodes follow the same dynamics
- ☐ The growth is sub-linear, due to the fact that nodes are competing with the others
- ☐ The earlier the node is added, the higher the degree – "first-mover advantage" in marketing and business
- ☐ The rate of acquiring new links $dk_i/dN = ½m/(N\,i)^{½}$ indicates that older nodes acquire more links
- ☐ This explains the hub formation in the Barabási-Albert model

Example


(a)
SINGLE NETWORK
t½
k
t
10⁰
10¹
10²
10³
10⁴
10⁵
10⁶
(b)
N=10²
N=10⁴
N=10⁶
pₖ

Approximate analysis (cont’d)

- ☐ Recall $k_i = m\ (N/i)^{½}$
- ☐ The number of nodes with degree smaller than k is

$$k_i < k \rightarrow m\ (N/i)^{½} < k$$

$$\rightarrow i > N\ (m/k)^2 \rightarrow \boxed{N - N\ (m/k)^2}$$

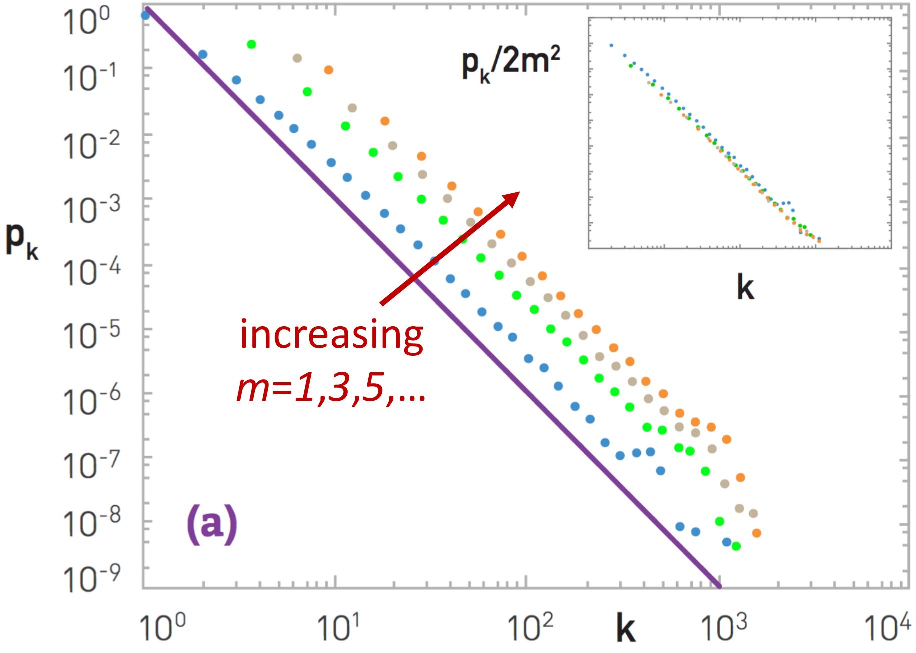
- ☐ CDF is $P_k = \mathrm{P}[k_i \leq k] = 1 - (m/k)^2$
- ☐ The degree distribution is

$$\mathrm{d}P_k\ /\mathrm{d}k = \boxed{p_k = 2\ m^2\ /\ k^3}$$

but the correct expression is
$2m(m+1)/k(k+1)(k+2)$
(tedious)

Implications of $p_k = 2m^2/k^3$

- ❑ The exponent $\gamma = 3$ is correctly guessed
- ❑ The degree exponent is independent of *m* ✓
- ❑ The scaling coefficient is proportional to m^2 ✓
- ❑ The degree distribution is independent of *N* ✓



Measuring preferential attachment

We measure $\Pi(k_i) = \Delta k_i / \Delta N$

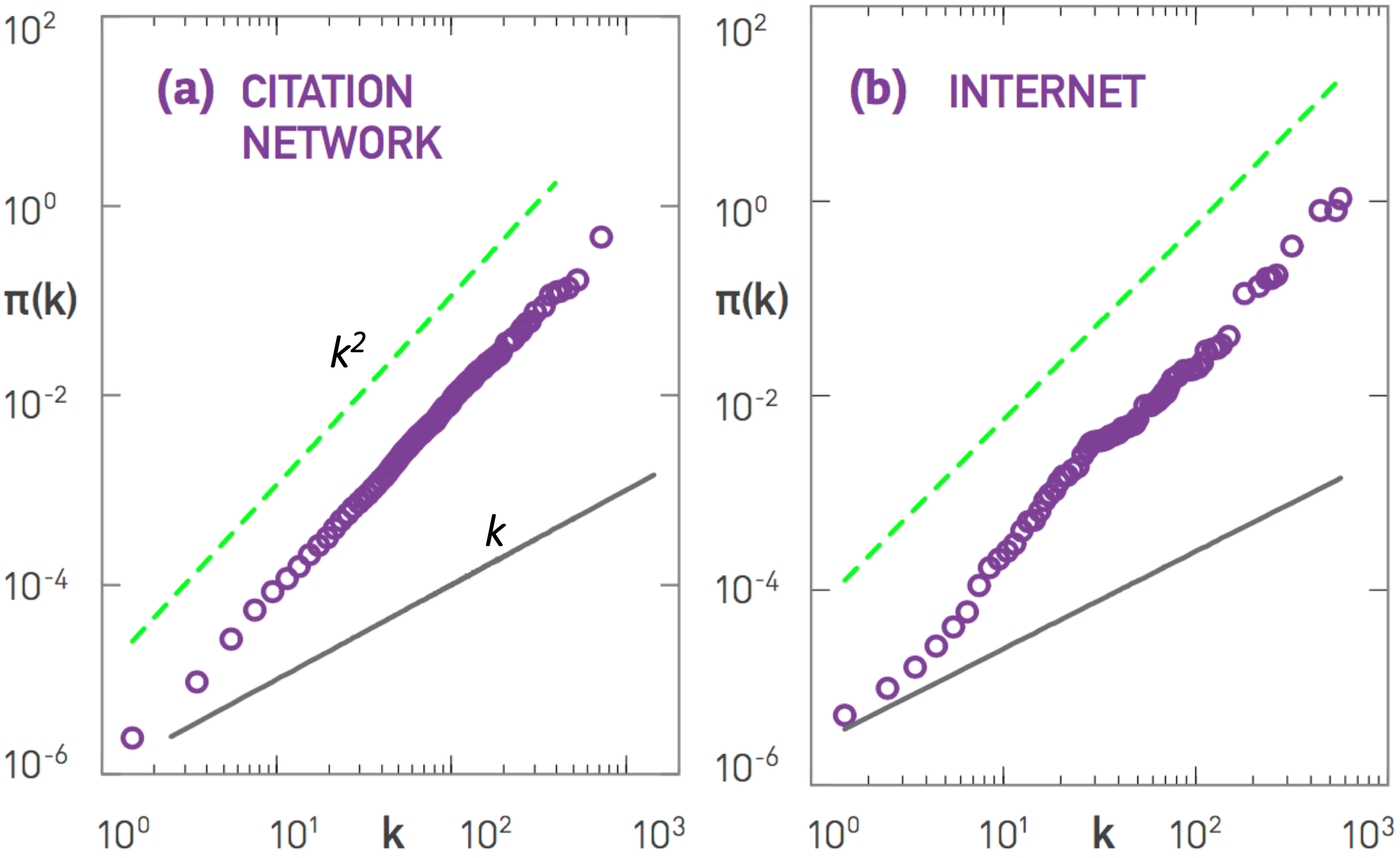
- It is expected to be equal to $k_i / \sum_j k_j$

To minimize the noise effect we plot $\pi(k) = \sum_{k_i=1}^{k} \Pi(k_i)$

We expect that:

- Under preferential attachment $\pi(k) \sim k^2$
- In the absence of p.a $\pi(k) \sim k$

Measuring preferential attachment



Pros and cons of the Barabási-Albert model

- ☐ Better explanations than Erdös-Renyi
- ☐ Generates a scale-free network
- ☐ Predicts $\gamma = 3$ but real networks have $2 < \gamma < 5$
- ☐ Generates an undirected network (many real networks are directed)
- ☐ It is a minimal proof-of-principle model

Attractiveness

Attractiveness

- A modelling flaw of Barabási-Albert: oldest nodes have an inherent advantage and cannot be defeated (*first mover's advantage*)
- They become hubs, $k_i \simeq m\,(N/i)^{½}$
- This is in contrast with intuition and evidence

e.g., Altavista [90's] → Google [2000] → Facebook [2011]

add the idea of "attractivenes"

Bianconi-Barabási model

- Basic idea: let us try to model the innate ability of a node to attract links
  - it does not depend on age
  - just a quality assessment of the individual that we assume to be determined at birth
- Call it the fitness η of a node
- Motivation
  - some people have an innate charisma
  - some websites attract immediate interest
  - some companies are good at alliances

Bianconi-Barabási model

The model:

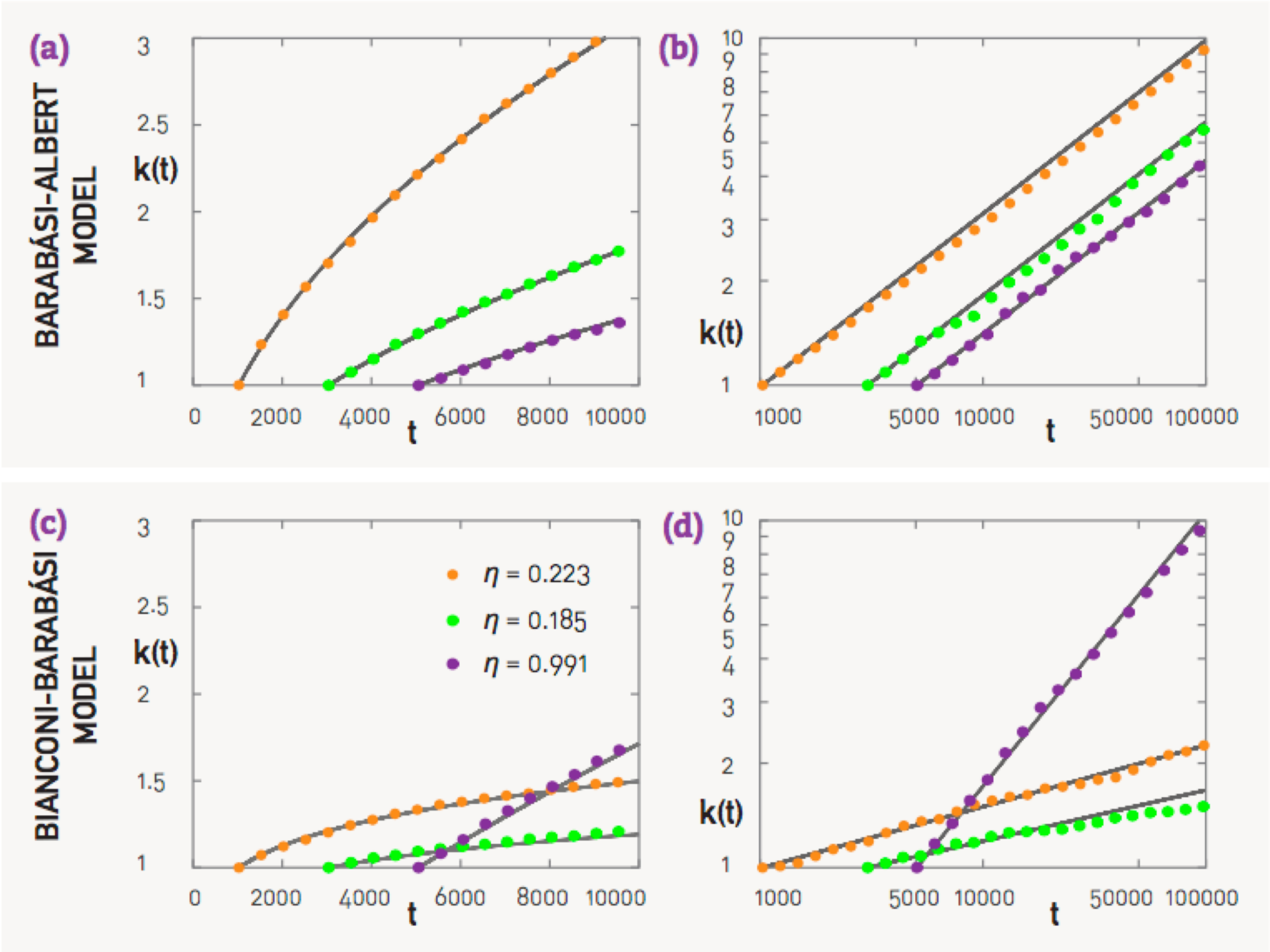
- ☐ Growth – at time step N a new node $i=N$ is added with m links and fitness η_i
- ☐ Fitness is a random number drawn from a given fitness distribution $\rho(\eta)$
- ☐ Preferential attachment - probability of linking to node i is proportional to both the degree and the fitness, i.e., $p_i = k_i\eta_\mathrm{i} / \sum_j k_j\eta_\mathrm{j}$

Bianconi-Barabási model

The Bianconi-Barabási model (a.k.a. the fitness model) tries to capture that:

- ☐ Nodes with higher degree have higher visibility and can better exploit fitness
- ☐ Among nodes with the same degree, the one with highest fitness is preferred
- ☐ A newcomer can conquer more connections than older nodes if it has a better fitness

Bianconi-Barabási model



We guess $k_i \simeq m\,(N/i)^{\beta(\eta_i)}$ for some $\beta(\eta)$

Approximate analysis

- We guess $\boxed{k_i \simeq m\,(N/i)^{\beta(\eta_i)}}$

- Increase in the degree $\Delta k_i \simeq \overset{\text{trials}}{\overset{\downarrow}{m}} \cdot k_i \overset{\text{probability per trial}}{\overset{\downarrow}{\eta_i}} / \sum k_j \eta_j$
- It is $\sum k_j \eta_j \simeq m\, N \cdot C$ (see proof)

Hence:

1. By inspection of the above

$$\Delta k_i \simeq m\,(N/i)^{\beta(\eta_i)}\, \eta_i / N\, C$$

2. By continuum theory

$$\Delta k_i \simeq \mathrm{d}k_i/\mathrm{d}N \simeq m\, \beta(\eta_i)\, N^{\beta(\eta_i)-1}\, i^{-\beta(\eta_i)}$$

3. By combining the results $\boxed{\beta(\eta_i) \simeq \eta_i / C}$

Proof

- Analysis of denominator $\sum k_i \eta_i$

  → average value wrt η

  → hypothesis $k_i \simeq m\,(N/i)^{\beta(\eta_i)}$

- $A = E[\sum_i k_i \eta_i] = \sum E[k_i \eta_i] \simeq \int_1^N E[k_i \eta_i]\, di$
- $E[k_i \eta_i] = \int m (N/i)^{\beta(\eta)}\, \eta \cdot \rho(\eta)\, d\eta$
- Swap integrals

$$A \simeq \int m\, N^{\beta(\eta)} \left[\int_1^N i^{-\beta(\eta)}\, di\right]\ \eta \cdot \rho(\eta)\, d\eta$$

- Integrate

constant C

$$A \simeq m\, N \cdot \boxed{\int \left(\frac{1 - \cancel{N^{\beta(\eta)-1}}}{1-\beta(\eta)}\right) \eta\, \rho(\eta)\, d\eta}$$

negligible for large N if $0 < \beta < 1$

On the constant *C*

Check consistency of

growth with exponent <1

$$0 < \beta(\eta) = \eta \,/\, C \; < 1$$

$$C = \int \frac{\eta\, \rho(\eta)\, d\eta}{1-\beta(\eta)} \quad \rightarrow \quad \boxed{1 = \int_0^{\eta_{\max}} (C/\eta - 1)^{-1}\, \rho(\eta)\, d\eta}$$

this identifies C for a given $\rho(\eta)$

Since we assumed $\beta<1$, it is $C > \eta_{\max}$ → the integral makes sense

Approximate analysis (cont'd)

Want to identify $P_k = \mathrm{P}[k_i \leq k] = 1 - \mathrm{P}[k_i > k]$

- $k_i > k$ and $k_i = m\,(N/i)^{\eta_i/C} \rightarrow i < N\,(m/k)^{C/\eta_i}$
- Hence $\mathrm{P}[k_i > k \,|\, \eta_\mathrm{i}] = (m/k)^{C/\eta_i}$
- We have $P_k = 1 - \int (m/k)^{C/\eta}\,\rho(\eta)\,\mathrm{d}\eta$

The degree distribution is

$$p_k = P_k' = C \int_0^{\eta_{\max}} \eta^{-1}\, m^{C/\eta}\, k^{-(C/\eta+1)}\, \rho(\eta)\,\mathrm{d}\eta$$

weighted combination of power laws with exponent in $[2,\infty)$ since $\eta_{\max} < C$

Bianconi Barabasi wrap-up

- Preferential attachment proportional to both the degree and the fitness, i.e., $p_i = k_i\eta_i / \sum k_j\eta_j$
- Node growth $k_i \simeq m\,(N/i)^{\eta_i/C}$
- Constant C defined by $\int_0^{\eta_{max}} (C/\eta - 1)^{-1}\, \rho(\eta)\, d\eta = 1$

fitness distribution

The degree distribution is

$$p_k = P_k' = \boxed{C \int_0^{\eta_{max}} \eta^{-1}\, m^{C/\eta}\, k^{-(C/\eta+1)}\, \rho(\eta)\, d\eta}$$

weighted combination of power laws with exponent in [2,∞) since $\eta_{max} < C$

Equal fitness

What if $\rho(\eta) = \delta(\eta\text{-}1)$?

- ☐ Coefficient $C = 2$ since

$$\int_0^{\eta_{\max}} (C/\eta - 1)^{-1}\, \delta(\eta\text{-}1)\, d\eta = (C - 1)^{-1} = 1$$

- ☐ Exponential degree $k_i \simeq m\,(N/i)^{½}$

Degree distribution

$$p_k = C \int_0^{\eta_{\max}} \eta^{-1}\, m^{C/\eta}\, k^{-(C/\eta+1)}\, \delta(\eta\text{-}1)\, d\eta = 2\, m^2\, k^{-3}$$

Back to Barabási-Albert model !!!

Uniform fitness

What if $\rho(\eta) = 1$ and $\eta_{\max} = 1$?

- Coefficient $C = 1.255$ since

$$\int_0^1 (C/\eta - 1)^{-1}\, d\eta = 1 \rightarrow e^{-2/C} = 1-1/C$$

- Exponential degree $k_i \simeq m\,(N/i)^{\eta_i/C}$
- Each node has its own dynamic exponent !!!

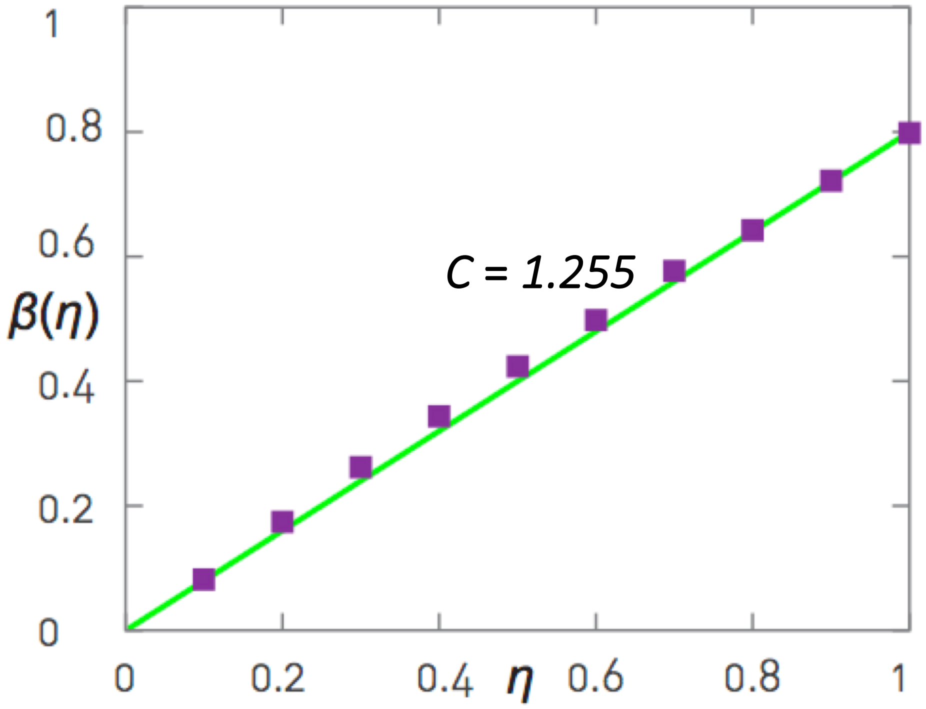
Degree distribution

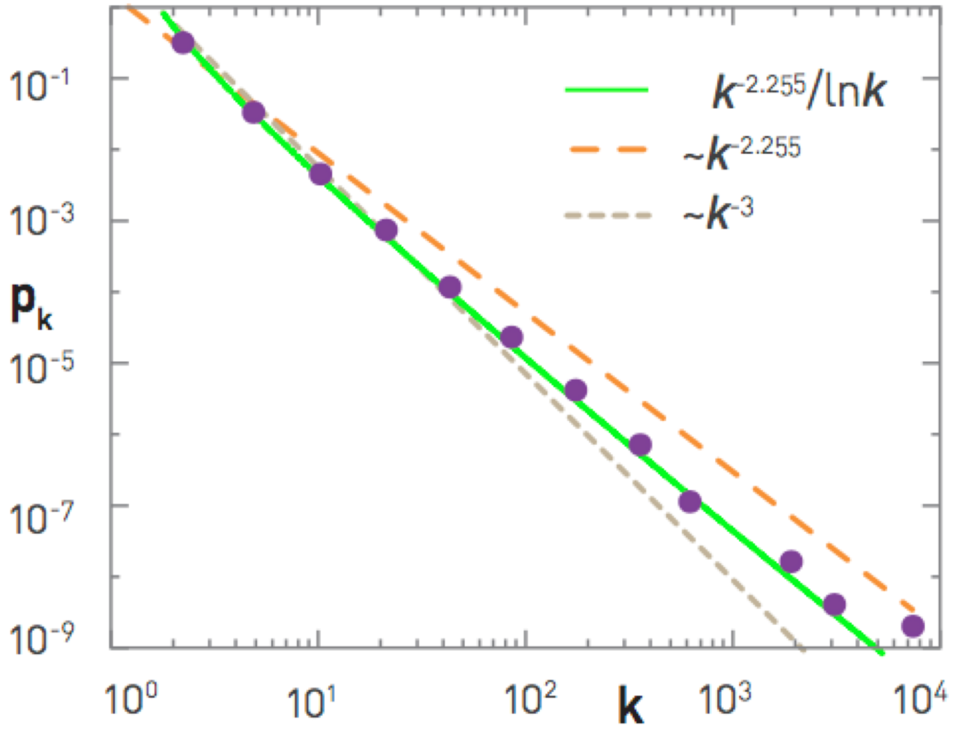
$$p_k = C/\eta k \int_0^1 e^{-C\ln(k/m)/\eta}\, d\eta \sim \boxed{k^{-(1+C)} / \ln(k)}$$

$e^{-b} - b\,E_1(b)$, $b = C\ln(k/m)$

exponential integral E_1

Uniform fitness

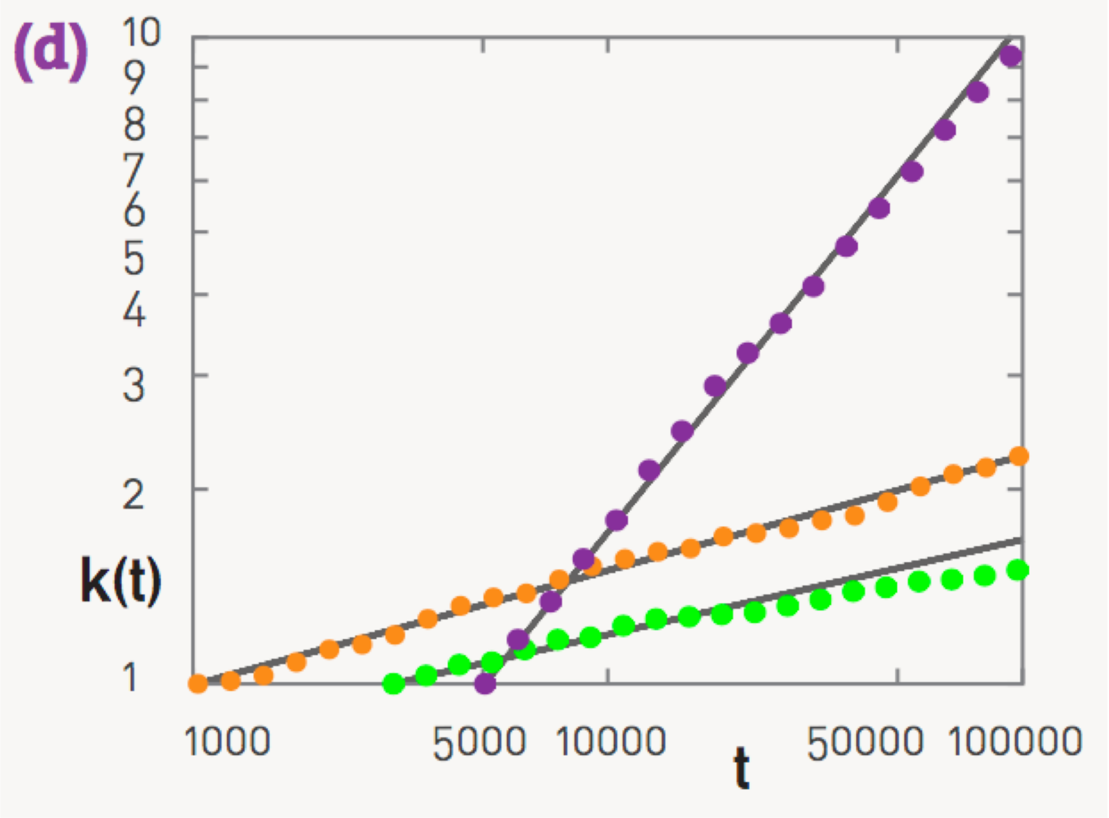




Degree distribution $p_k \sim k^{-(1+C)} / \ln(k)$

corrective term

Measuring fitness



Idea: compare the node's degree dynamics to that of other nodes

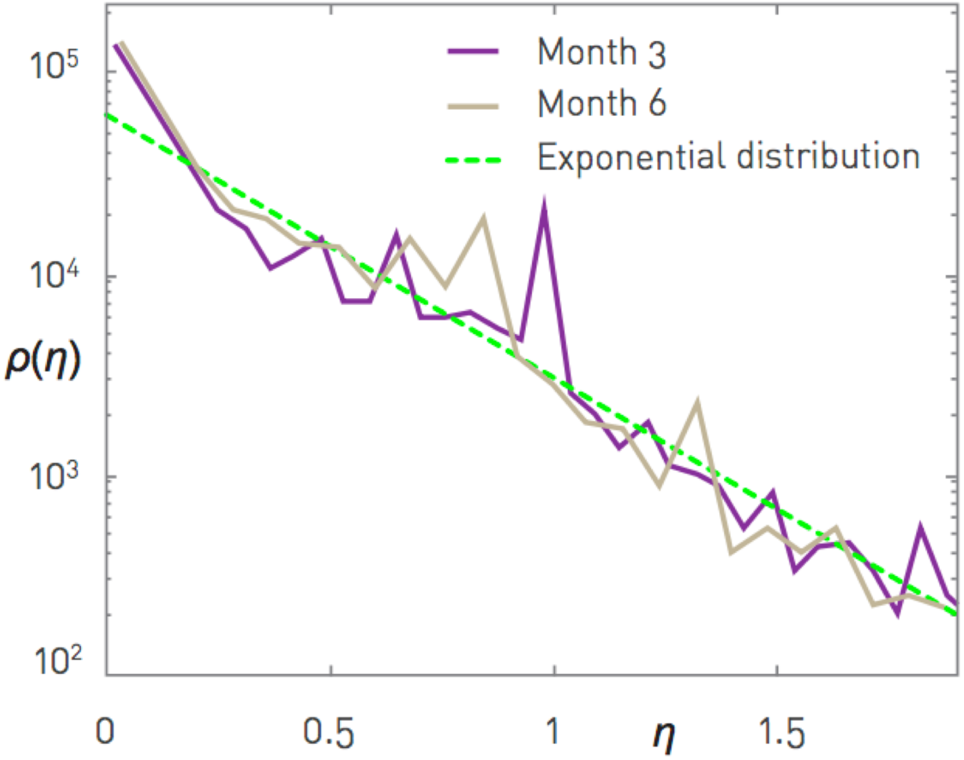
Recall: $\ln(k_i) = \eta_i \ln(N)/C + \ln(m\, i^{-\eta_i/C})$

$\uparrow$ linear in $\ln(N)$

$\uparrow$ constant in N

MIME.

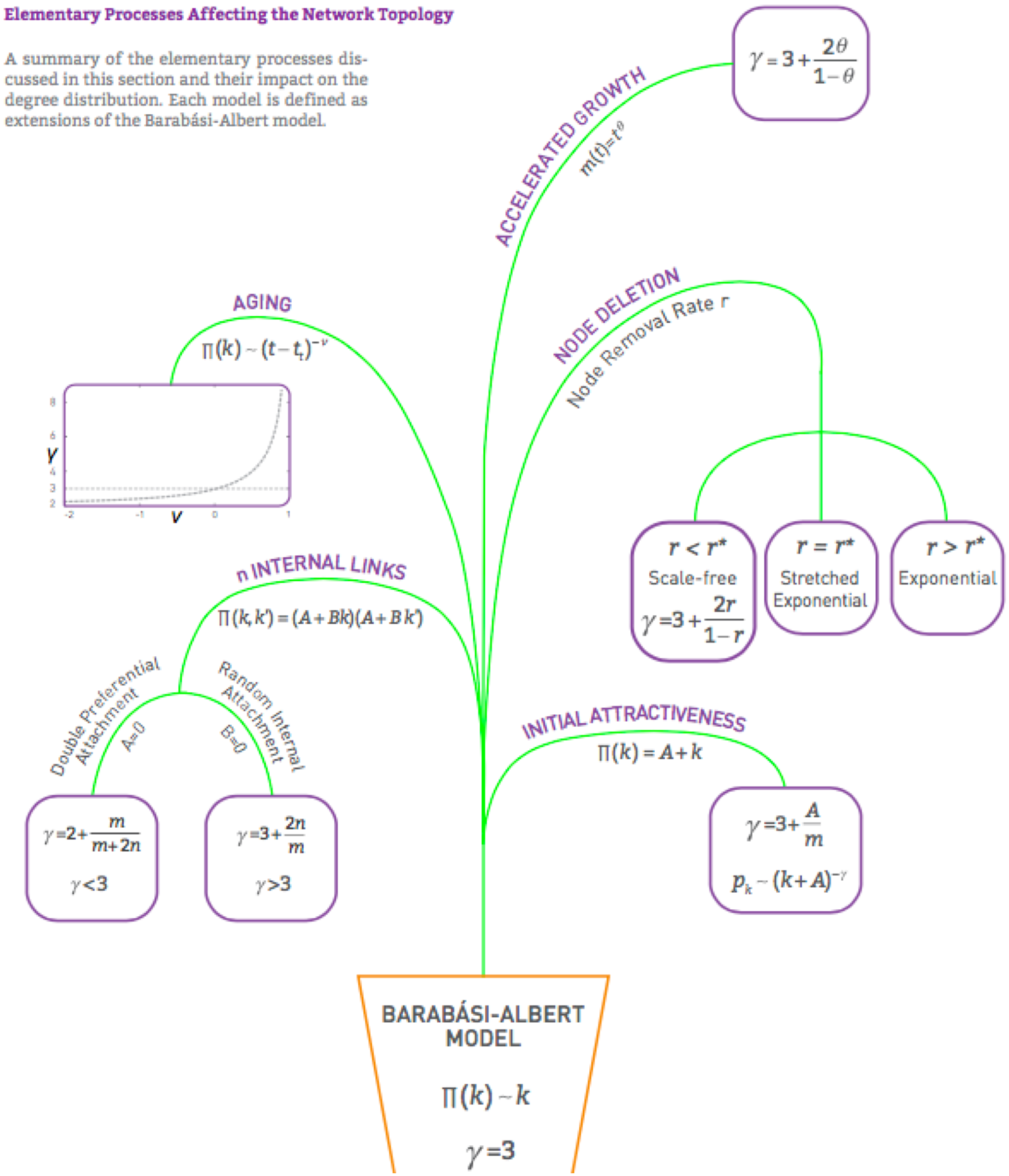
Fitness of the www



Fitness well approximated with an exponential: most nodes have small attractiveness, only a few large hubs

- ☐ $\rho(\eta) = a\, e^{-a\eta} / (1\text{-}e^{-a}), \eta_{max}=1$

Many other ideas for extension



Elementary Processes Affecting the Network Topology

A summary of the elementary processes discussed in this section and their impact on the degree distribution. Each model is defined as extensions of the Barabási-Albert model.

Lessons learned

MODEL CLASS	EXAMPLES	CHARACTERISTICS
Static Models	Erdős–Rényi Watts-Strogatz	• N fixed • p_k exponentially bounded • Static, time independent topologies
Generative Models	Configuration Model Hidden Parameter Model	• Arbitrary pre-defined p_k • Static, time independent topologies
Evolving Network Models	Barabási–Albert Model Bianconi-Barabási Model Initial Attractiveness Model Internal Links Model Node Deletion Model Accelerated Growth Model Aging Model	• p_k is determined by the processes that contribute to the network's evolution. •Time-varying network topologies

Readings

- ☐ A.L. Barabási, Network science

  http://barabasi.com/networksciencebook

  Ch.5 “The Barabási-Albert model”

  Ch.6 “Evolving networks”