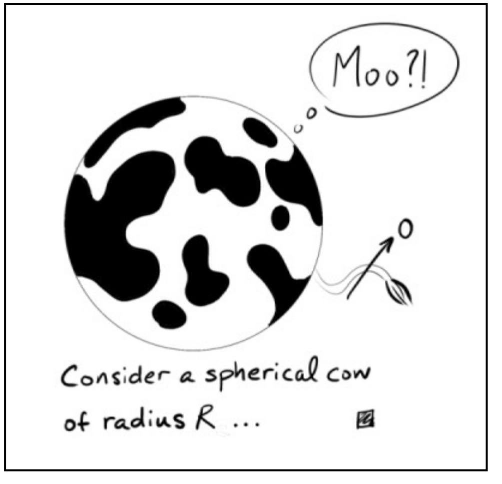
DEI
DEPARTAMENTO
DE ENGENHARIA INFORMÁTICA
TÉCNICO LISBOA

Let a million of variants bloom!

Models of evolving networks



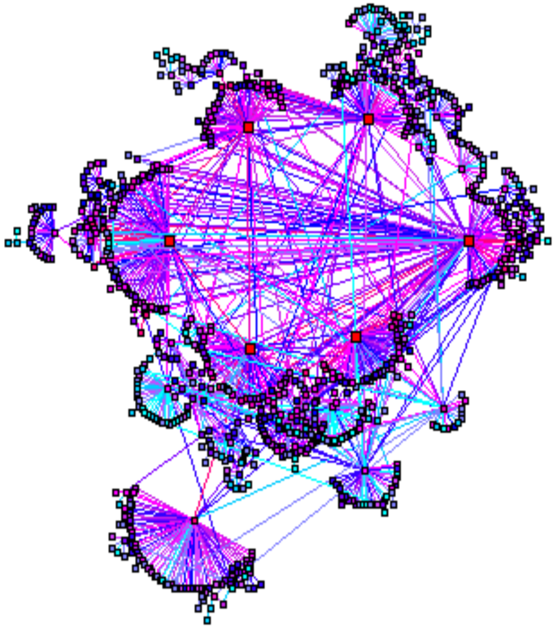
Network Science, 2023/2024

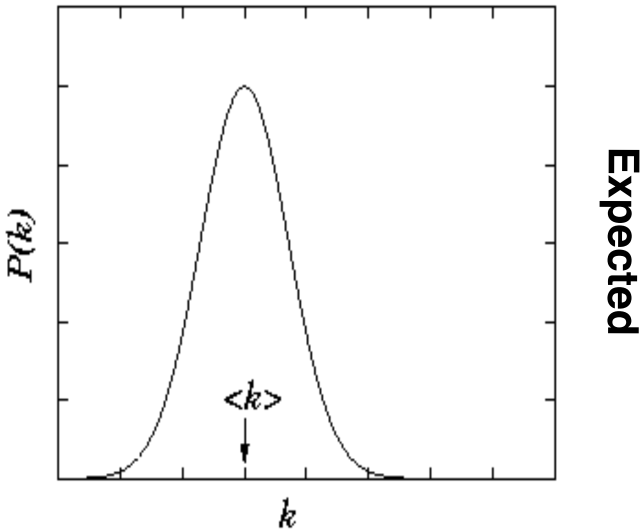
World-Wide-Web

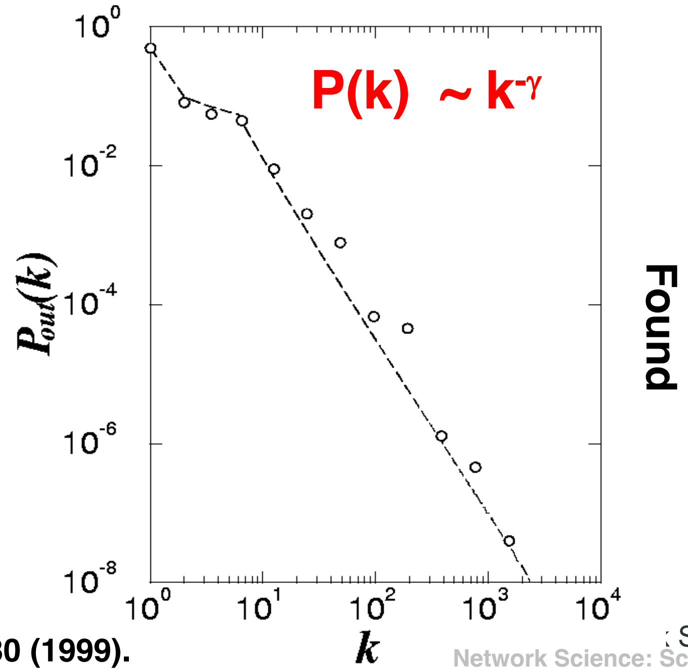
Nodes: **WWW documents**
Links: **URL links**

Over 3 billion documents

ROBOT: collects all URL's found in a document and follows them recursively



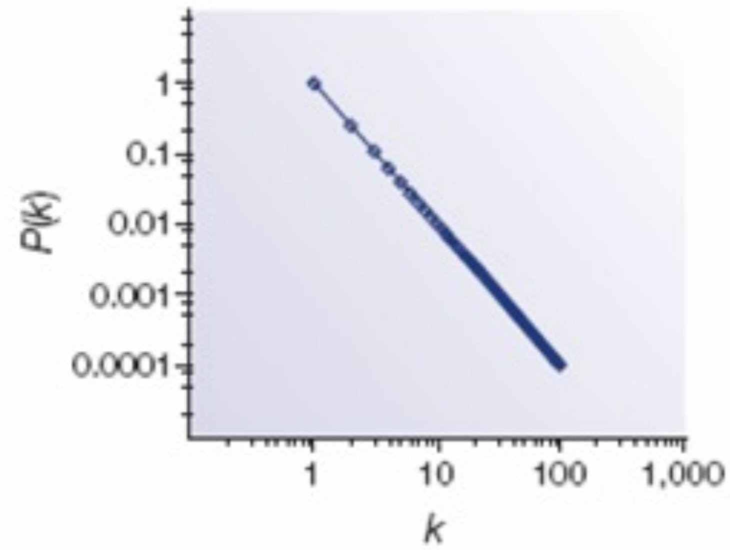




R. Albert, H. Jeong, A-L Barabasi, *Nature*, 401 130 (1999).

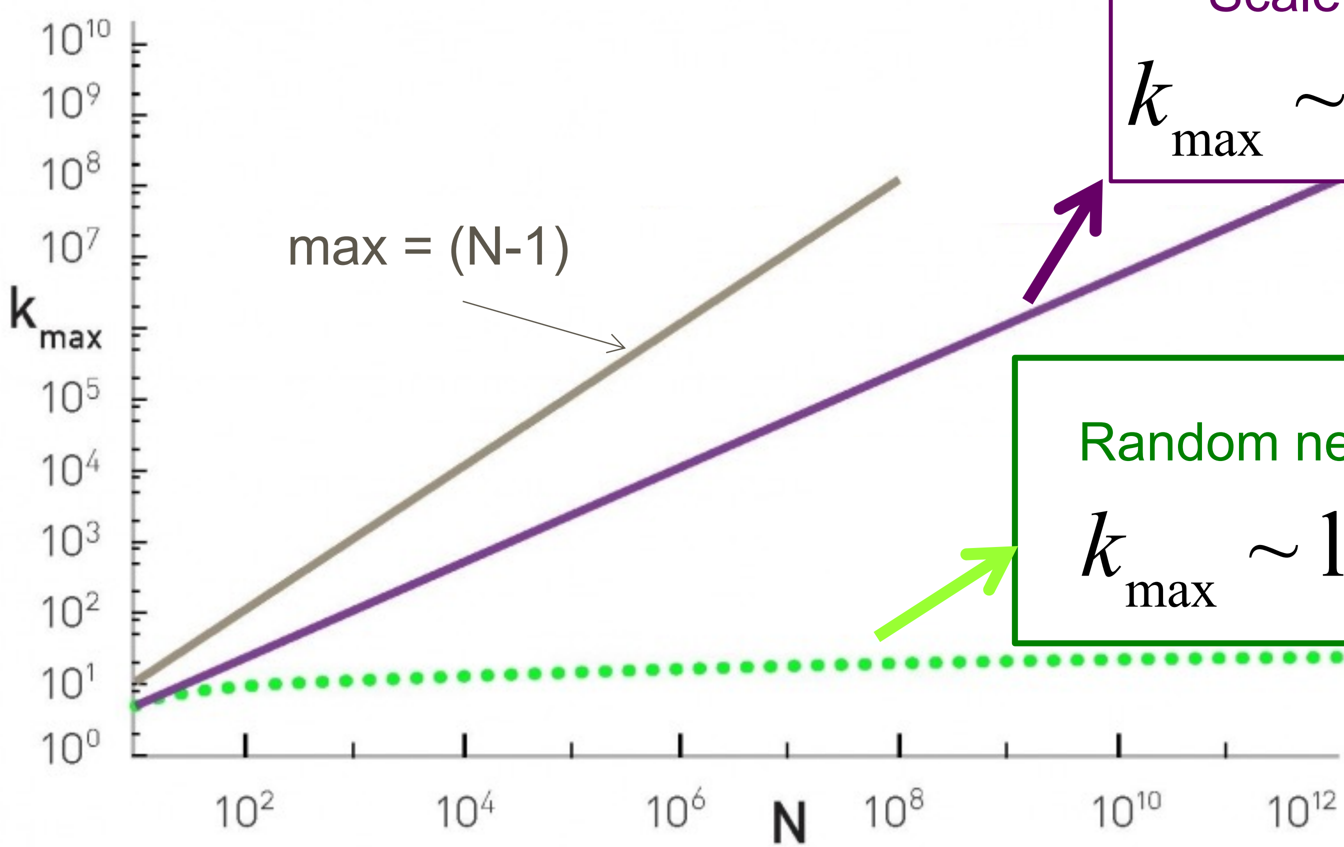
Power-law degree distributions

$$P_k \sim k^{-\gamma}$$



$$\ln P_k \sim \ln k^{-\gamma} \quad \Rightarrow \quad \ln P_k \sim -\gamma \ln k$$

Can we estimate the largest hub?



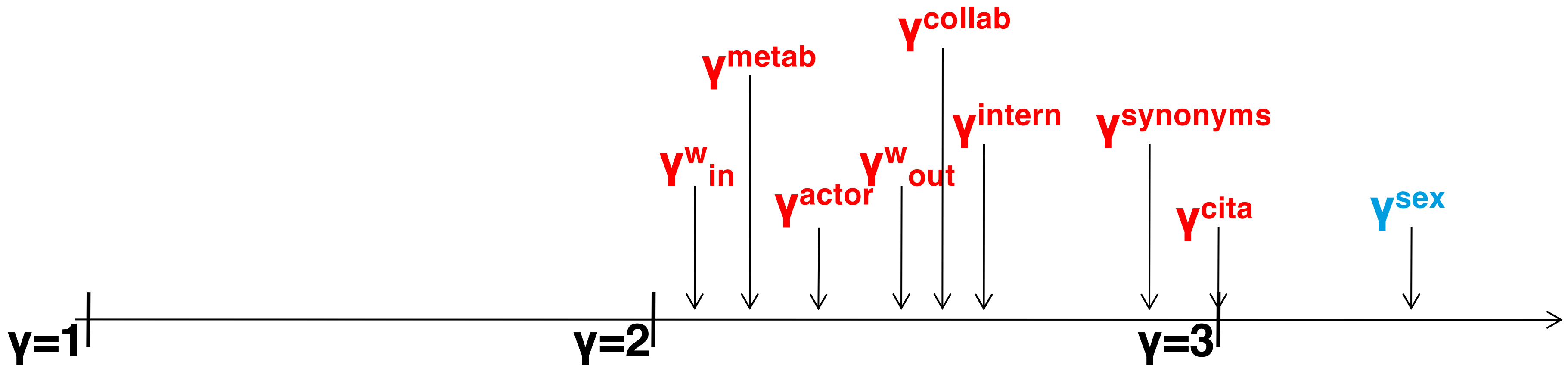
Scale-free

$$k_{max} \sim N^{\frac{1}{\gamma - 1}}$$

Random network

$$k_{max} \sim \ln N$$

The universe of scale-free networks



<k^2> diverges	<k^2> finite

<k> diverges	<k> finite

Ultra small world behavior	Small world

Regime full of anomalies…	The scale-free behavior is relevant	Behaves like a random network

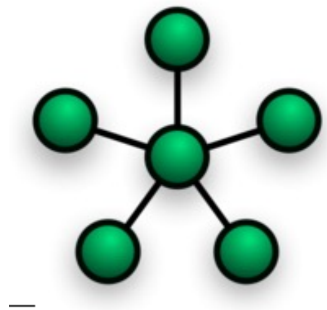
Why are most exponents in this regime?

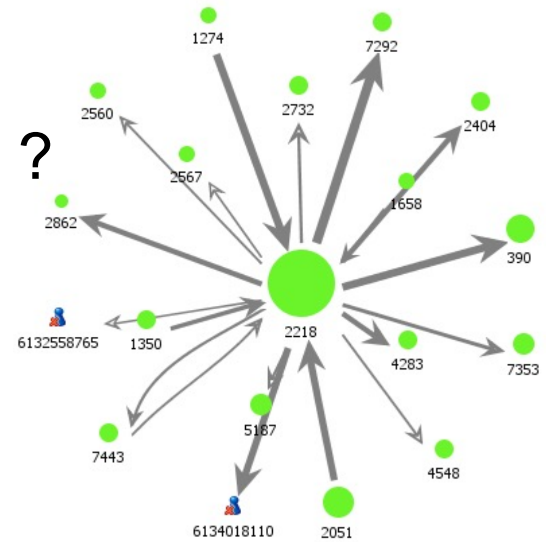
Magic exponents?

- Why is it hard to find networks with γ≤2 ?

$$k_{\max} = k_{\min} N^{\frac{1}{\gamma-1}}$$

K_{max} grows faster than *N*





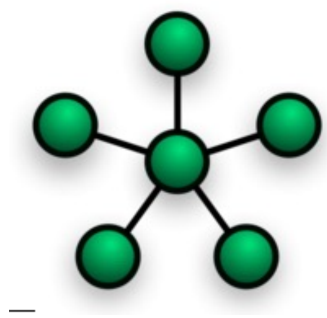
- Why is it hard to find networks with γ>>3 ?

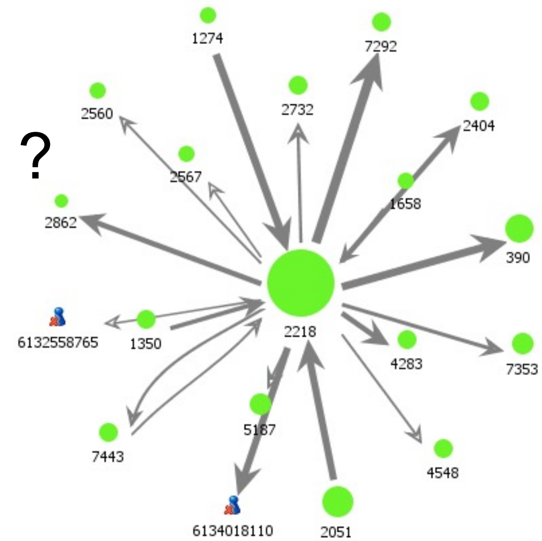
Magic exponents?

- Why is it hard to find networks with γ≤2 ?

$$k_{max} = k_{min} N^{\frac{1}{\gamma-1}}$$

K_{max} grows faster than *N*

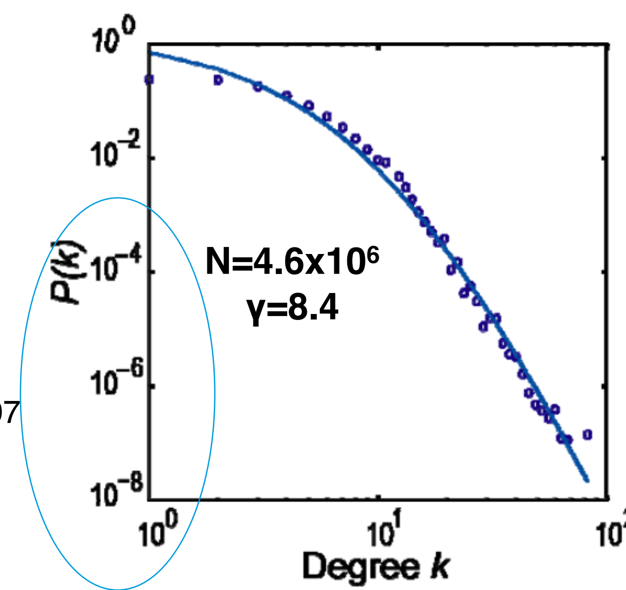




- **Why is it hard to find networks with γ>>3 ?**

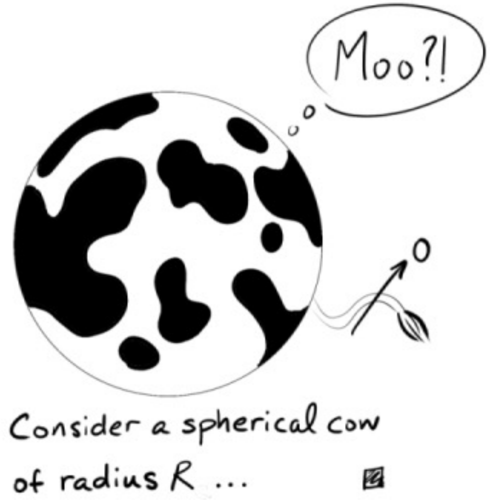
Mobile Call Network

Onella et al. PNAS 2007



Let’s try to be even more ambitious...

Can we identify the main principles leading to the emergence of scale-free networks?



Complex systems

Microscopic interactions

Emergent collective phenomena / properties

local

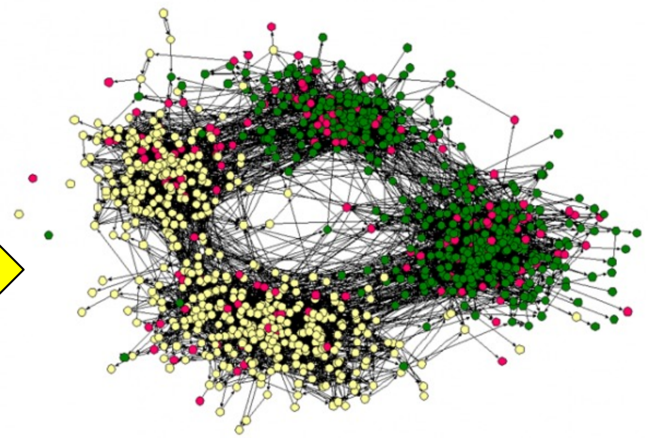
global

Complex systems

Ex: Social sciences

Microscopic interactions

local



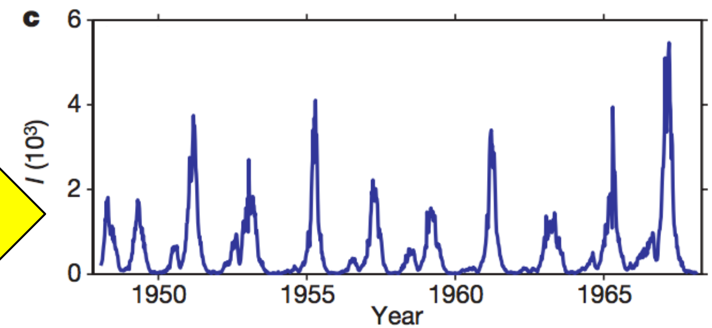
Collective/emergent patterns

global

Complex systems

Ex: Theoretical epidemiology

Microscopic interactions



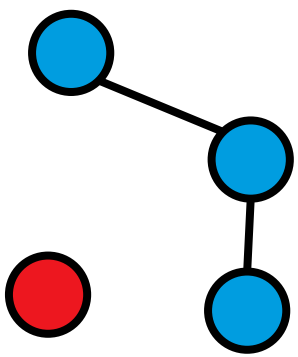
Collective/emergent patterns

local

global

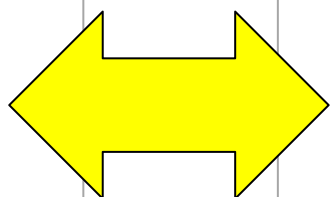
Complex systems

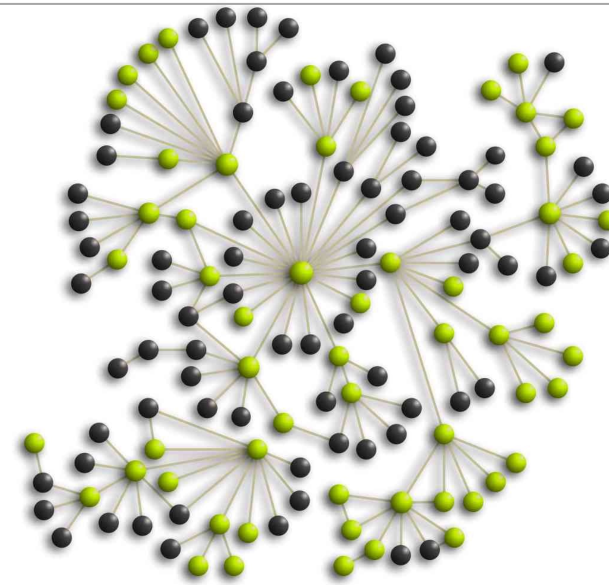
Ex: Scale-free networks



Microscopic interactions

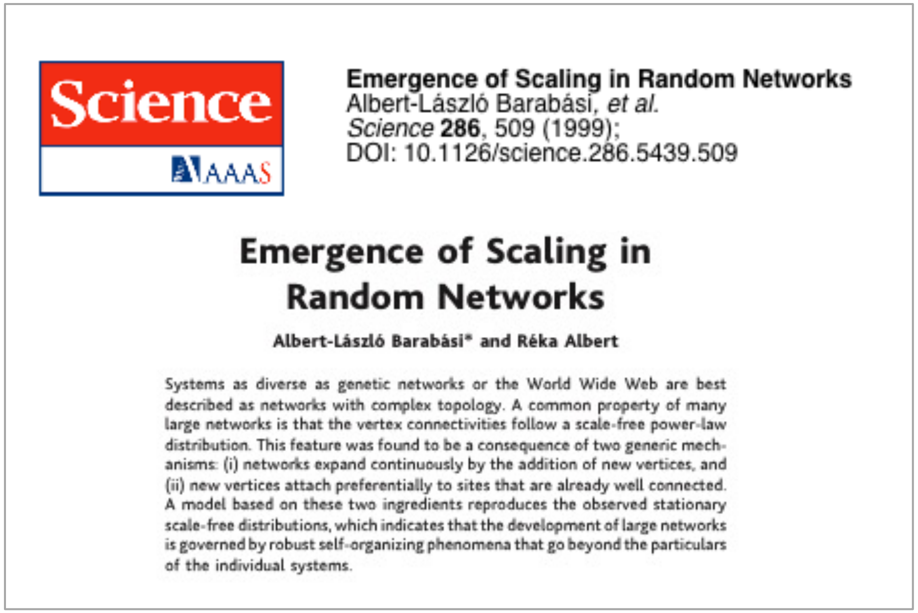
local





Collective/emergent patterns

global

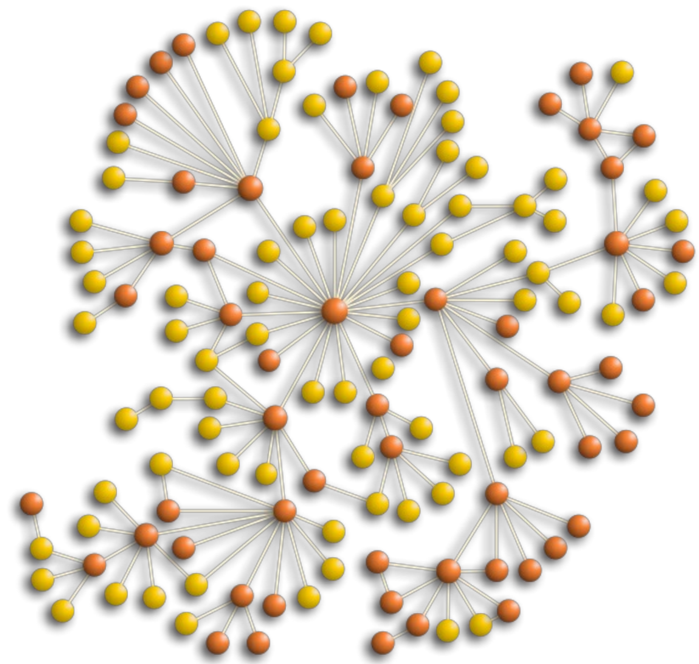


Emergence of Scaling in Random Networks
Albert-László Barabási, *et al.*
Science **286**, 509 (1999);
DOI: 10.1126/science.286.5439.509

Emergence of Scaling in Random Networks

Albert-László Barabási* and Réka Albert

Systems as diverse as genetic networks or the World Wide Web are best described as networks with complex topology. A common property of many large networks is that the vertex connectivities follow a scale-free power-law distribution. This feature was found to be a consequence of two generic mechanisms: (i) networks expand continuously by the addition of new vertices, and (ii) new vertices attach preferentially to sites that are already well connected. A model based on these two ingredients reproduces the observed stationary scale-free distributions, which indicates that the development of large networks is governed by robust self-organizing phenomena that go beyond the particulars of the individual systems.



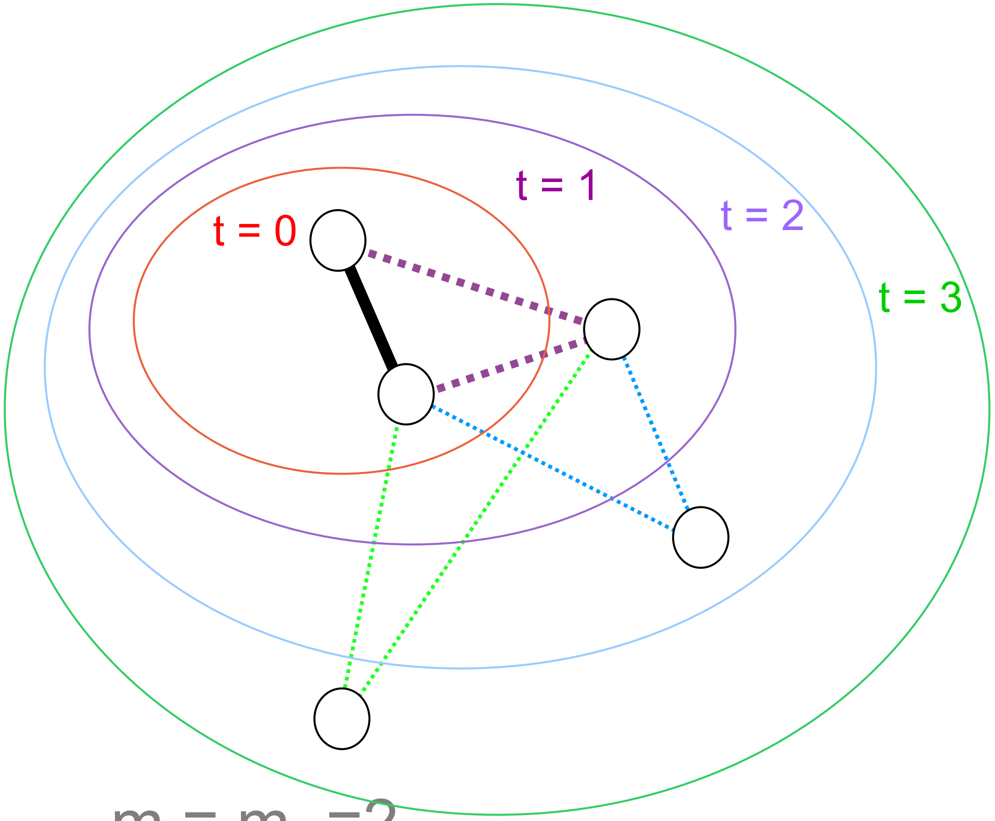
Reka Albert
Penn State University

Albert-László Barabási
Northeastern University

The Barabási-Albert model

- ***Growth*** & ***Preferential attachment***

n	k
1	2
2	4
3	4
4	2
5	2



$m = m_0 = 2$

The Barabási-Albert model

- ***Growth***: add nodes sequentially.

At t=0 consider (e.g.) a ring with m_0 nodes; for each time-step, add 1 node and connect it via m new edges with existing nodes;

The Barabási-Albert model

- ***Growth***: add nodes sequentially.

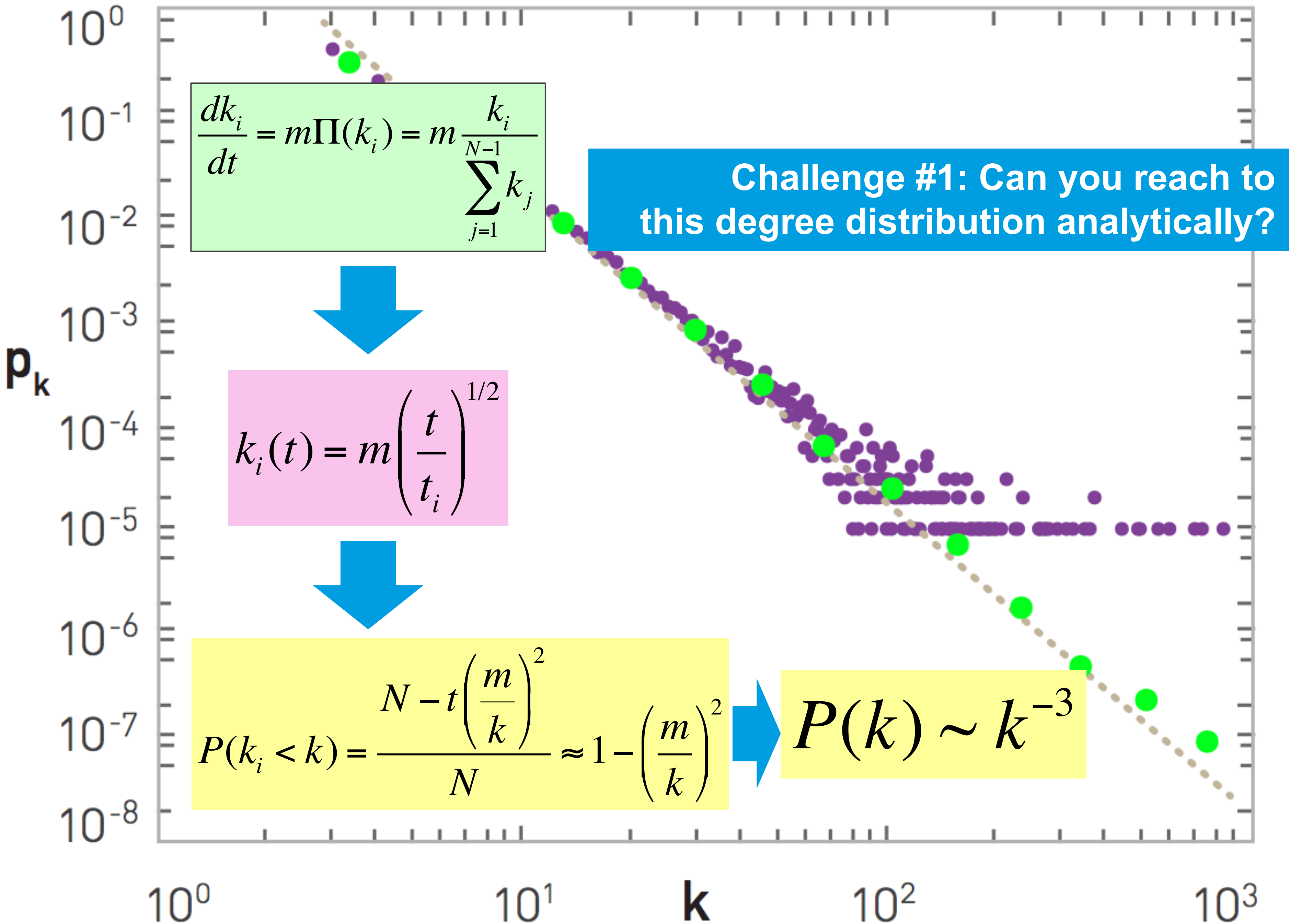
At t=0 consider (e.g.) a ring with m_0 nodes; for each time-step, add 1 node and connect it via m new edges with existing nodes;

- ***Preferential attachment***:

when establishing the connections, each new node is connected to older nodes with a probability proportional to the degree of the older nodes.

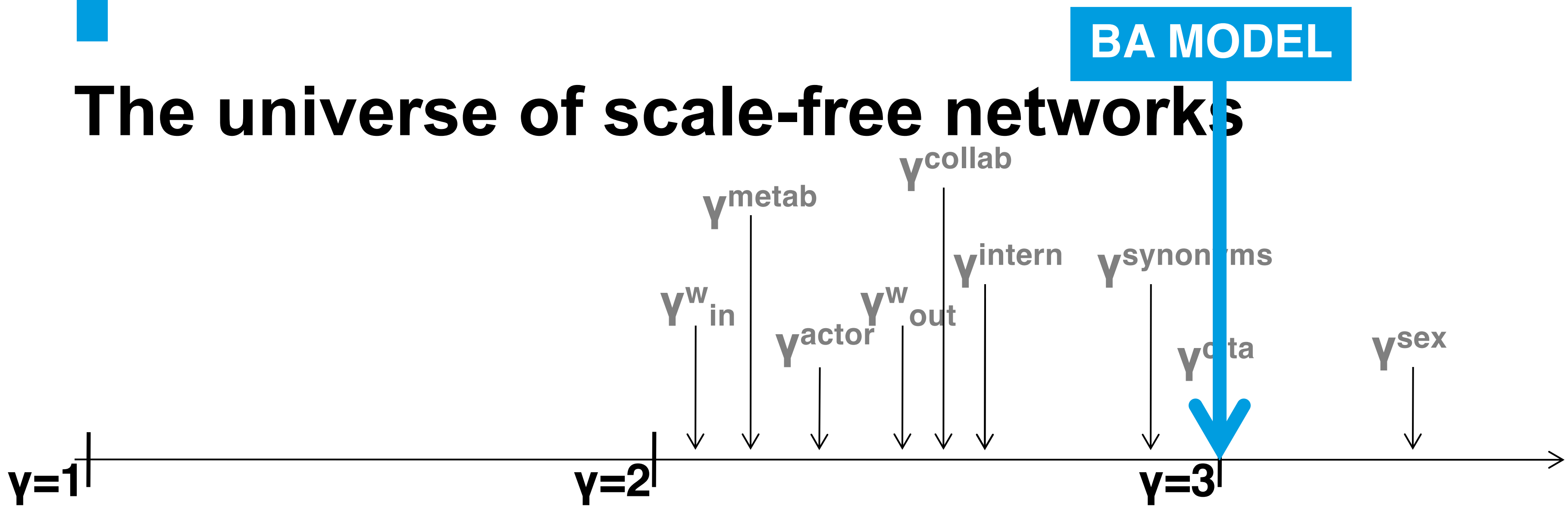
$$\Pi_i = \frac{k_i}{\sum_j k_j}$$

the more popular you are, the more popular you become



Challenge #1: Can you reach to this degree distribution analytically?
$\frac{dk_i}{dt} = m\Pi(k_i) = m\frac{k_i}{\sum_{j=1}^{N-1} k_j}$
$k_i(t) = m\left(\frac{t}{t_i}\right)^{1/2}$
$P(k_i < k) = \frac{N - t\left(\frac{m}{k}\right)^2}{N} \approx 1 - \left(\frac{m}{k}\right)^2$
$P(k) \sim k^{-3}$
p_k
k
10^0
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{-8}
10^0
10^1
10^2
10^3

The universe of scale-free networks

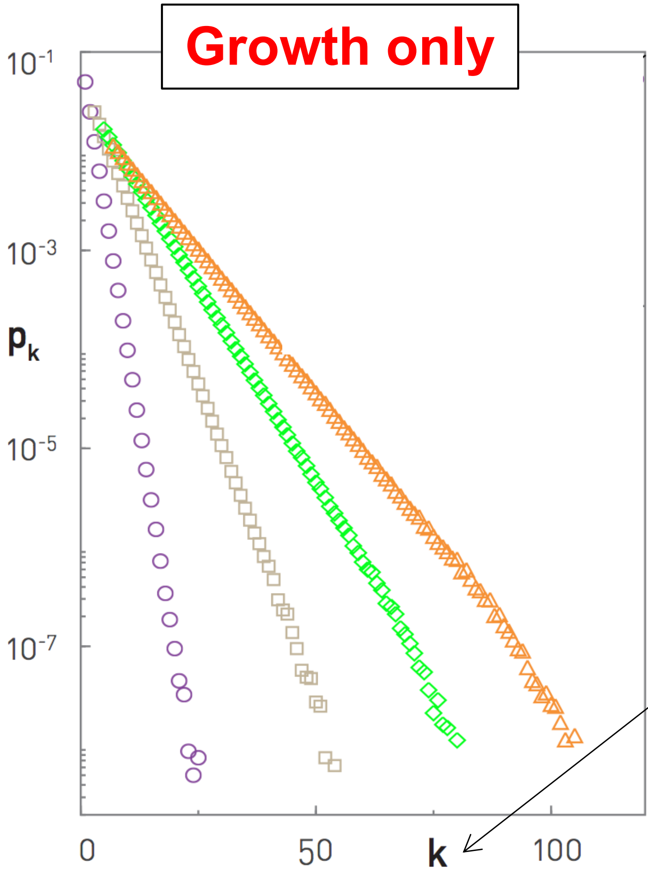


$\langle k^2 \rangle$ diverges	$\langle k^2 \rangle$ finite
$\langle k \rangle$ diverges	$\langle k \rangle$ finite
Ultra small world behavior	Small world

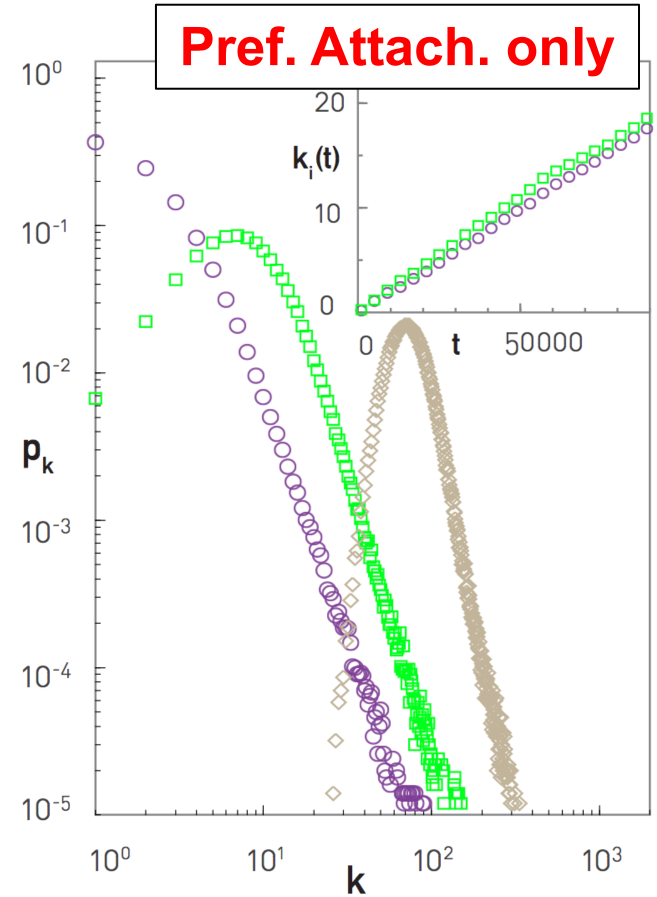
Regime full of anomalies…	The scale-free behavior is relevant	Behaves like a random network

Last class: Do we need both growth and preferential attachment? YES!!!

Barabási, Albert, & Jeong, (1999). Mean-field theory for scale-free random networks. *Physica A: Statistical Mechanics and its Applications*, *272*(1-2), 173-187.



linear scale



$$\frac{dk_i}{dt} = m\Pi(k_i) = m\frac{1}{N(t)-1} \sim m\frac{1}{t}$$

Can we be sure of having a preferential attachment mechanism?

- Changes in degree should follow

$$\Delta k_i = k_i(t+\Delta t) - k_i(t)$$

i.e.,

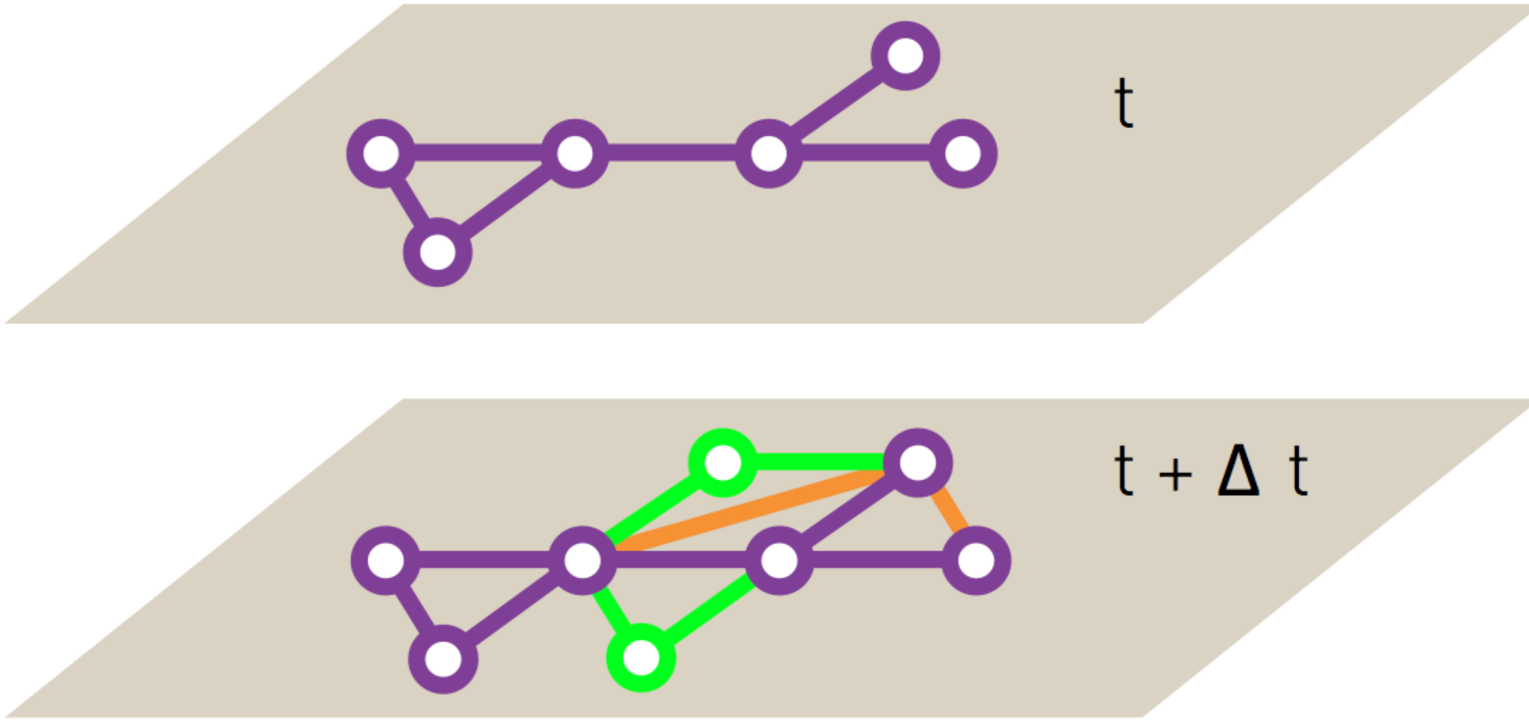
$$\frac{\Delta k_i}{\Delta t} \sim \frac{k_i}{\sum_j k_j}$$

The BA model assumes a linear form of preferential attachment

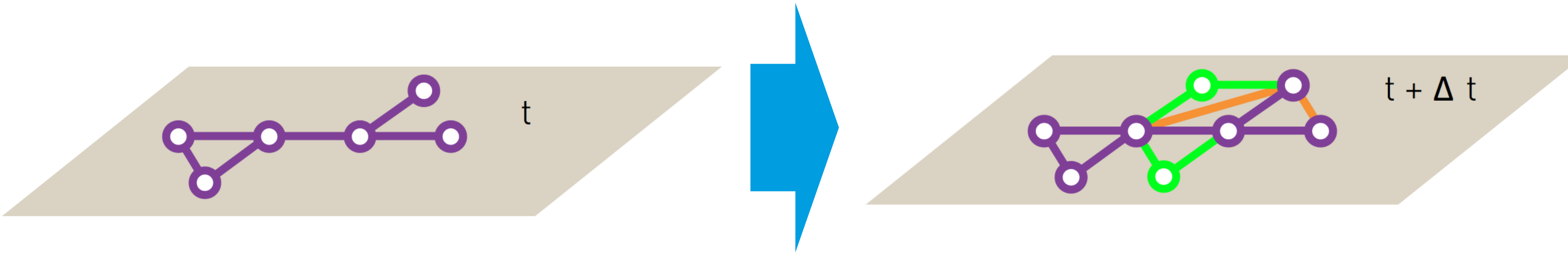
We can average this variation for empirical data and compare it with different scenarios

$$\frac{\Delta k_i}{\Delta t} \sim \frac{k_i^{\alpha}}{\sum_j k_j^{\alpha}}$$

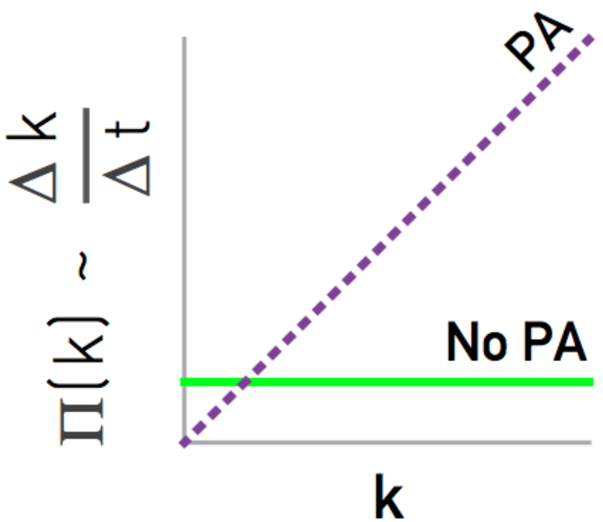
Can we be sure of having a preferential attachment mechanism?



Can we be sure of having a preferential attachment mechanism?

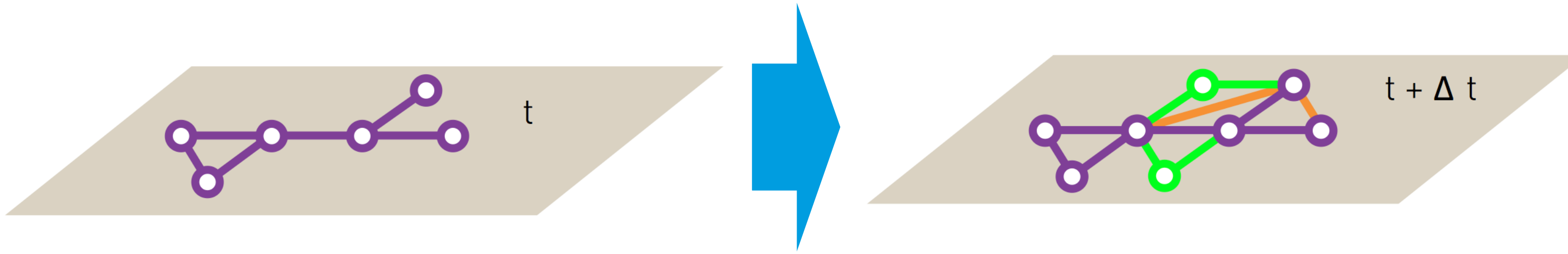


preferential attachment (PA)



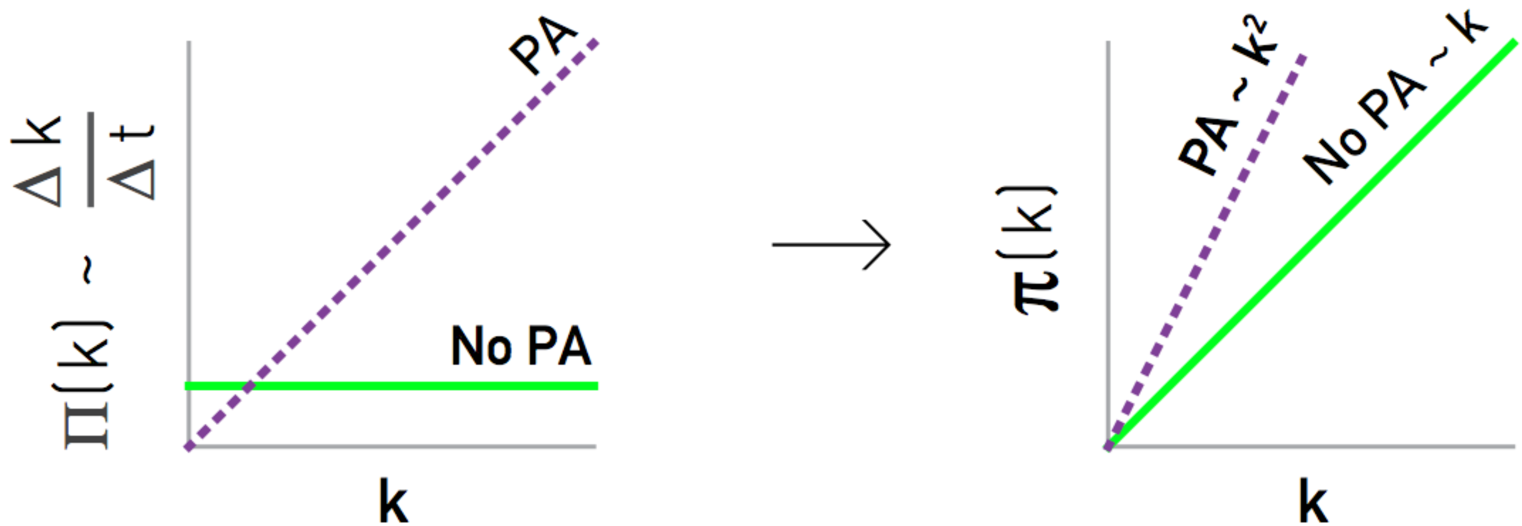
Log-log scales

Can we be sure of having a preferential attachment mechanism?



preferential attachment (PA)

cumulative preferential attachment function π(k)



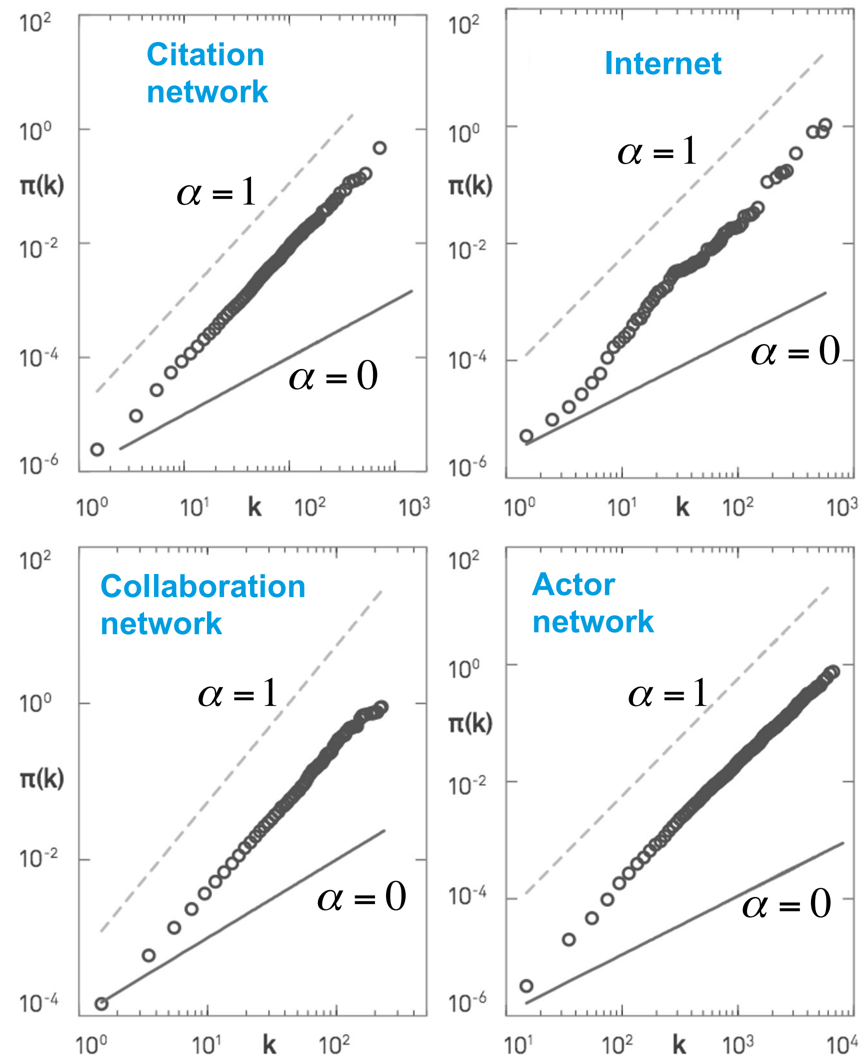
Log-log scales

Can we be sure of having a preferential attachment mechanism?

- Preferential attachment is present
- Yet, we may have a non-linear preferential attachment!!

$$\frac{\Delta k_i}{\Delta t} \sim \frac{k_i^{\alpha}}{\sum_j k_j^{\alpha}}$$

cumulative preferential attachment function π(k)



Non-linear preferential attachment:
Does it change anything?

- ***Growth***: add nodes sequentially.

At t=0 consider (e.g.) a ring with m_0 nodes; for each time-step, add 1 node and connect it via m new edges with existing nodes;

- ***Preferential attachment***:

when establishing the connections, each new node is connected to older nodes with a probability proportional to ***degree***$^{\alpha}$ of the older nodes.

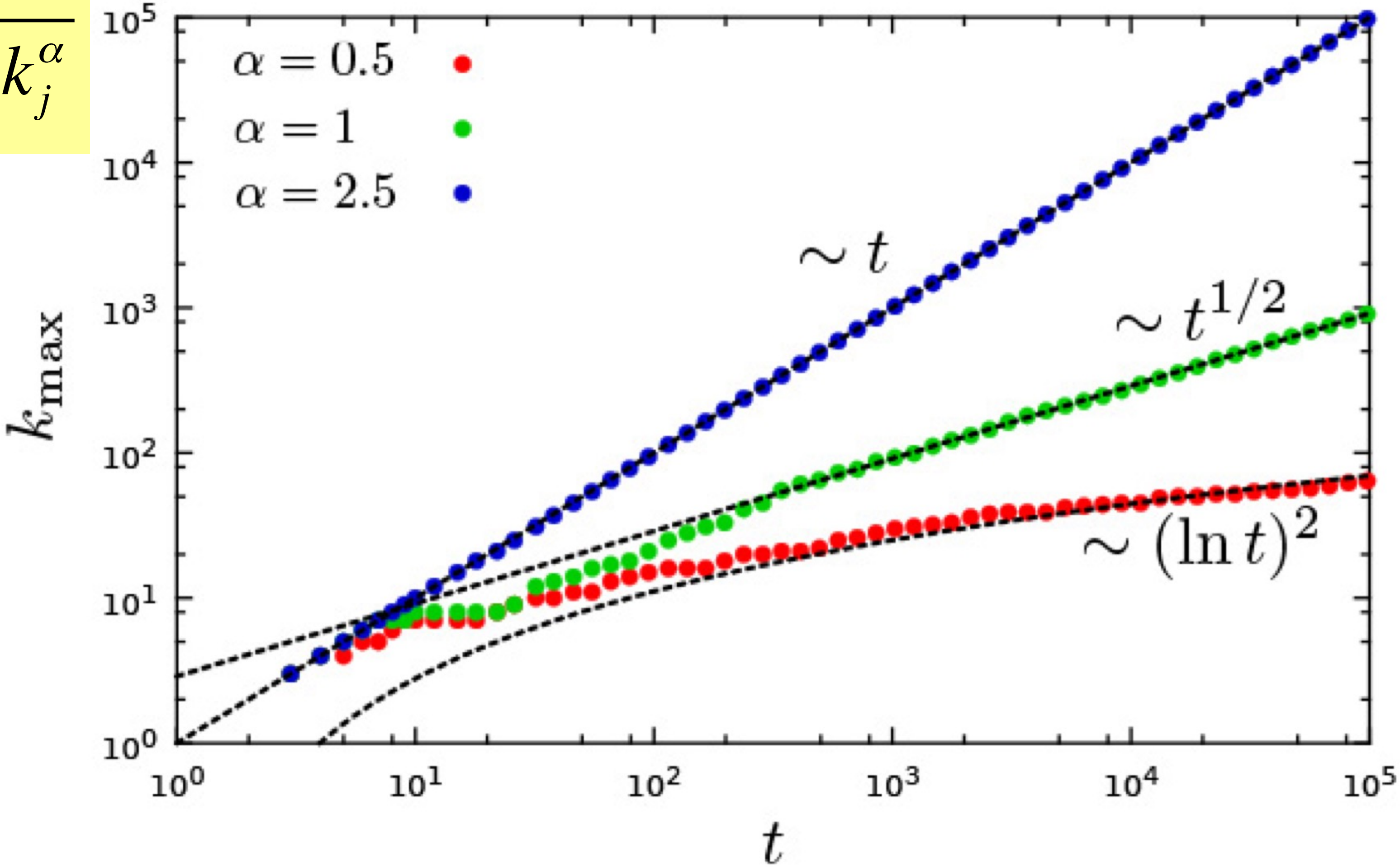
$$\Pi_i = \frac{k_i^{\alpha}}{\sum_j k_j^{\alpha}}$$

the more popular you are, the more popular you become, yet with different strengths

Hub's degree dynamics

$$\Pi_i = \frac{k_i^\alpha}{\sum_j k_j^\alpha}$$

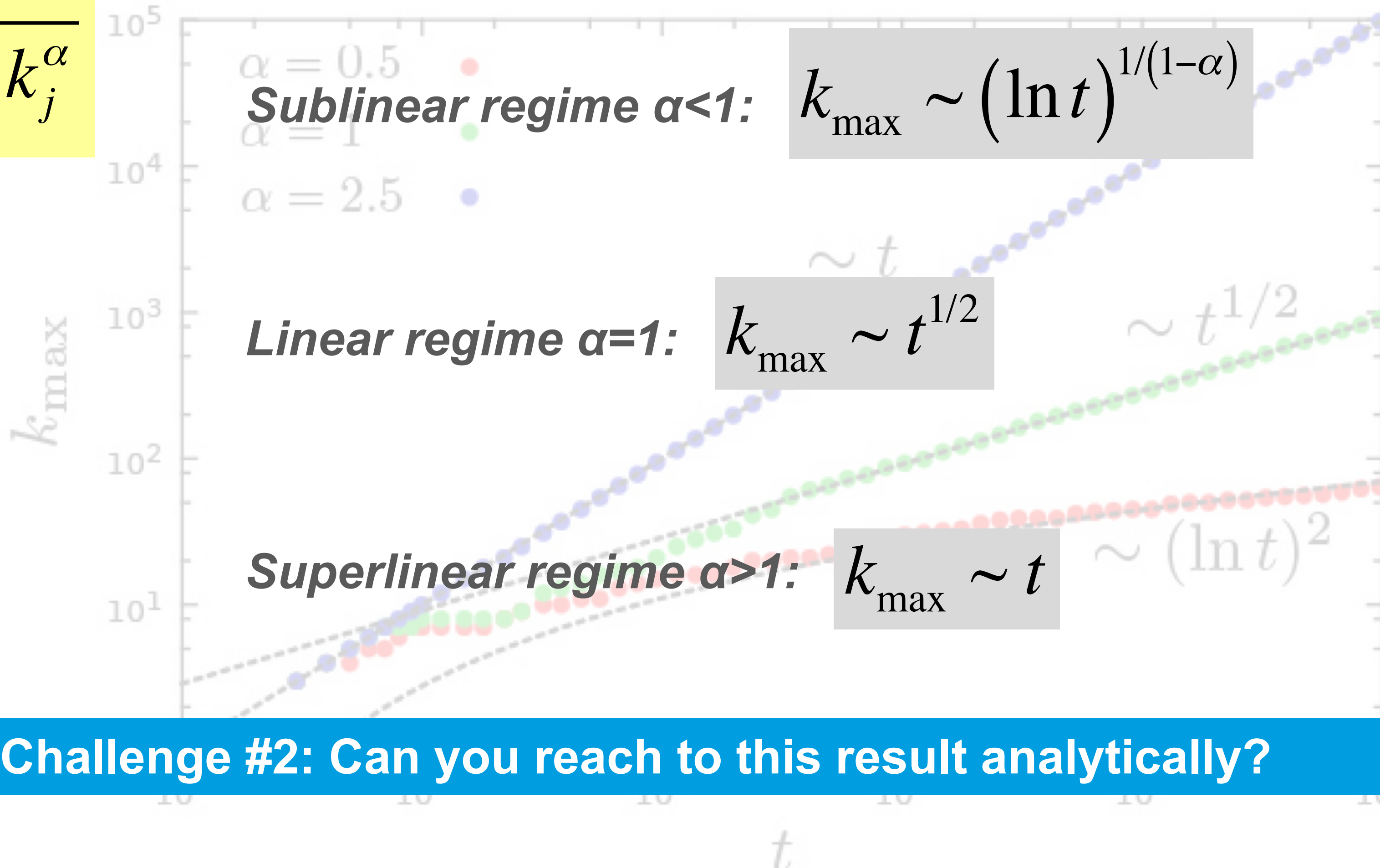
At *each time-step we add a node. What's the degree of the largest hub after t steps?*



Hub’s degree dynamics

$$\Pi_i = \frac{k_i^\alpha}{\sum_j k_j^\alpha}$$

At each time-step we add a node. What’s the degree of the largest hub after t steps?



Sublinear regime α<1: $k_{max} \sim (\ln t)^{1/(1-\alpha)}$

Linear regime α=1: $k_{max} \sim t^{1/2}$

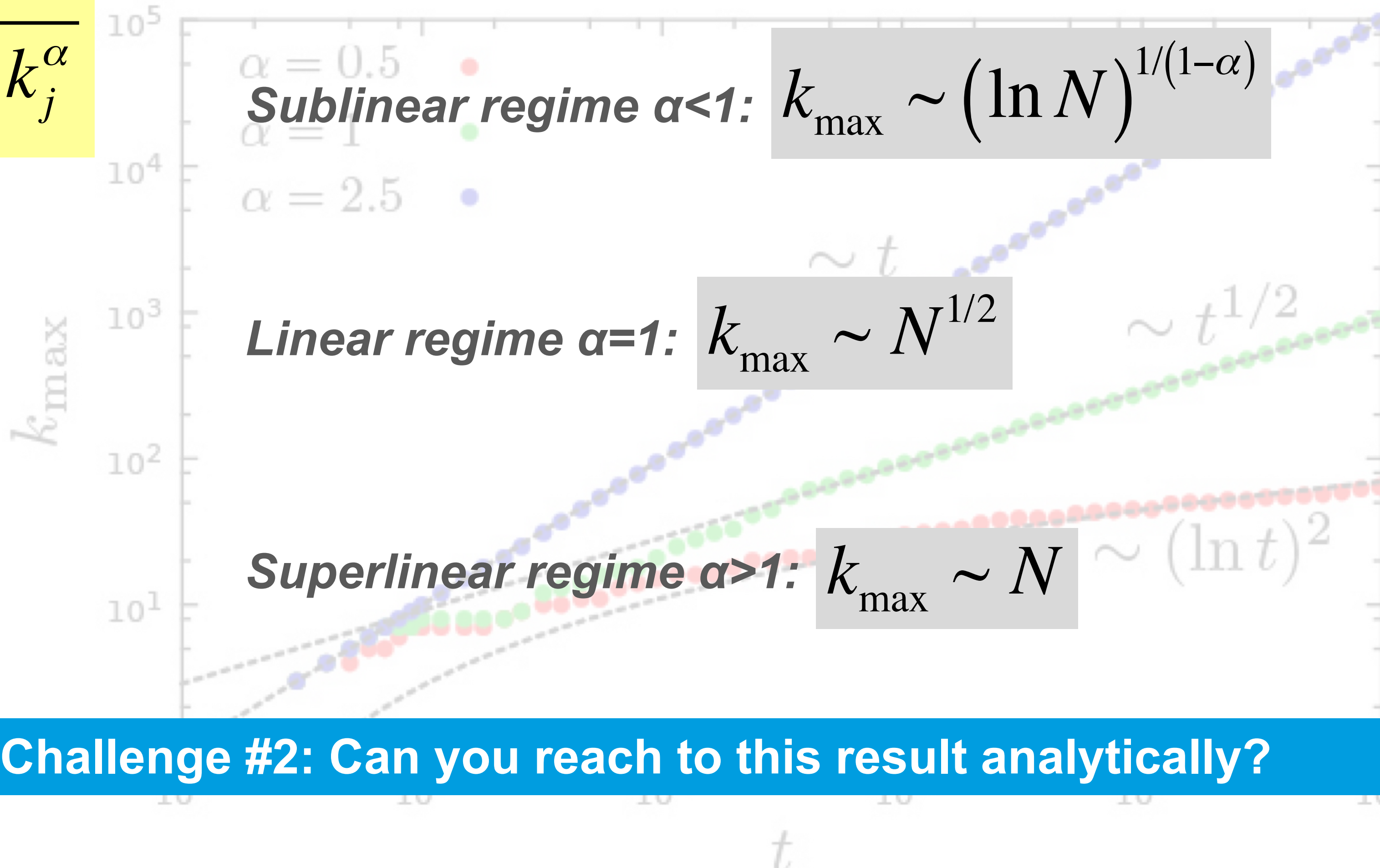
Superlinear regime α>1: $k_{max} \sim t$

Challenge #2: Can you reach to this result analytically?

Hub’s degree dynamics

$$\Pi_i = \frac{k_i^{\alpha}}{\sum_j k_j^{\alpha}}$$

***At** each time-step we add a node. What’s the degree of the largest hub after t steps?*



Sublinear regime α<1: $k_{\max} \sim (\ln N)^{1/(1-\alpha)}$

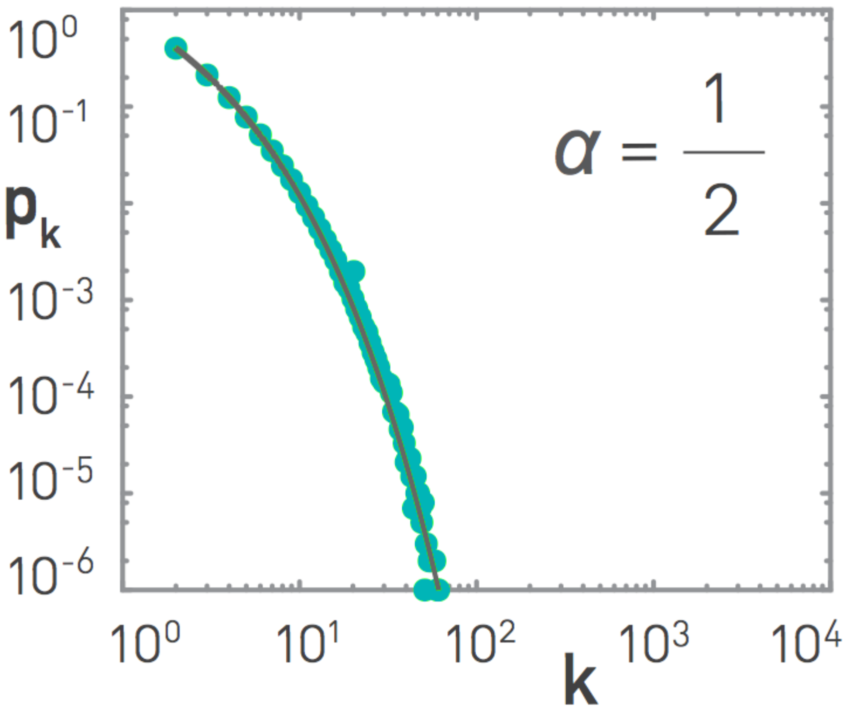
Linear regime α=1: $k_{\max} \sim N^{1/2}$

Superlinear regime α>1: $k_{\max} \sim N$

Challenge #2: Can you reach to this result analytically?

Degree distributions (sublinear regime)

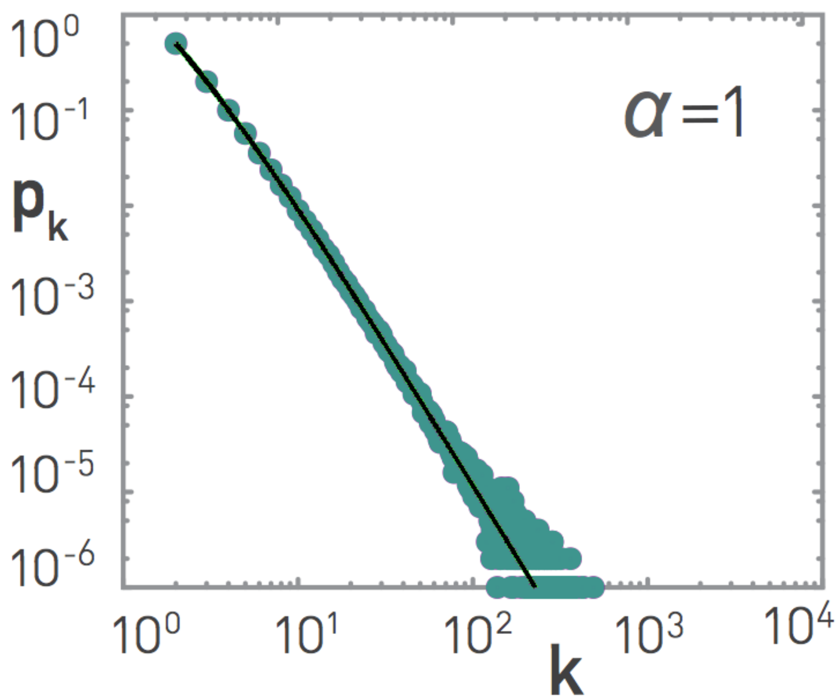
Sublinear



Stretched Exponential

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

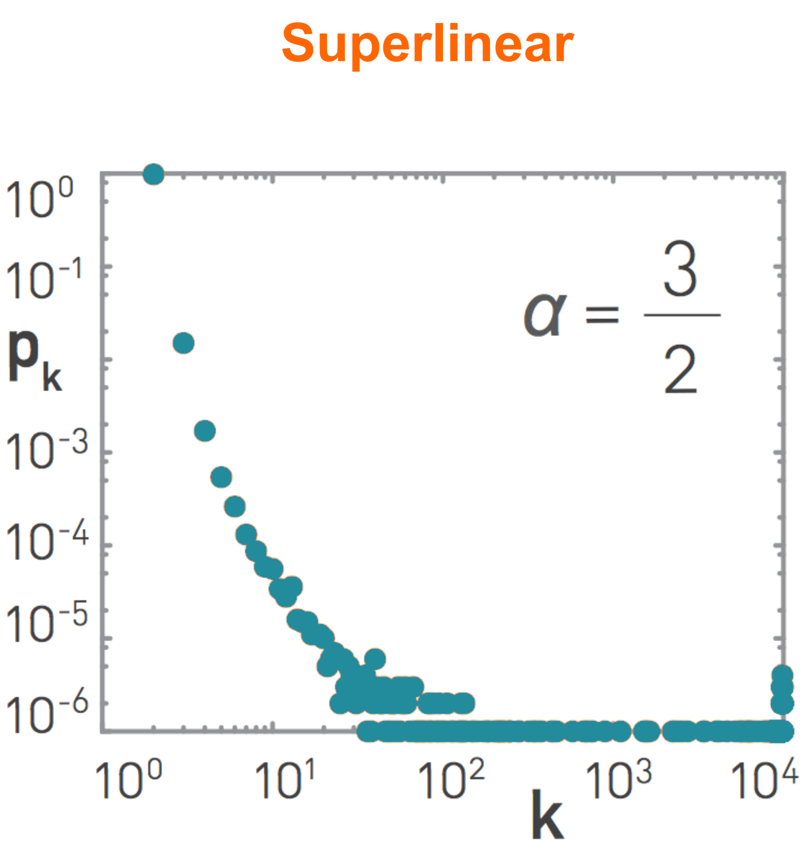
Linear



Power law

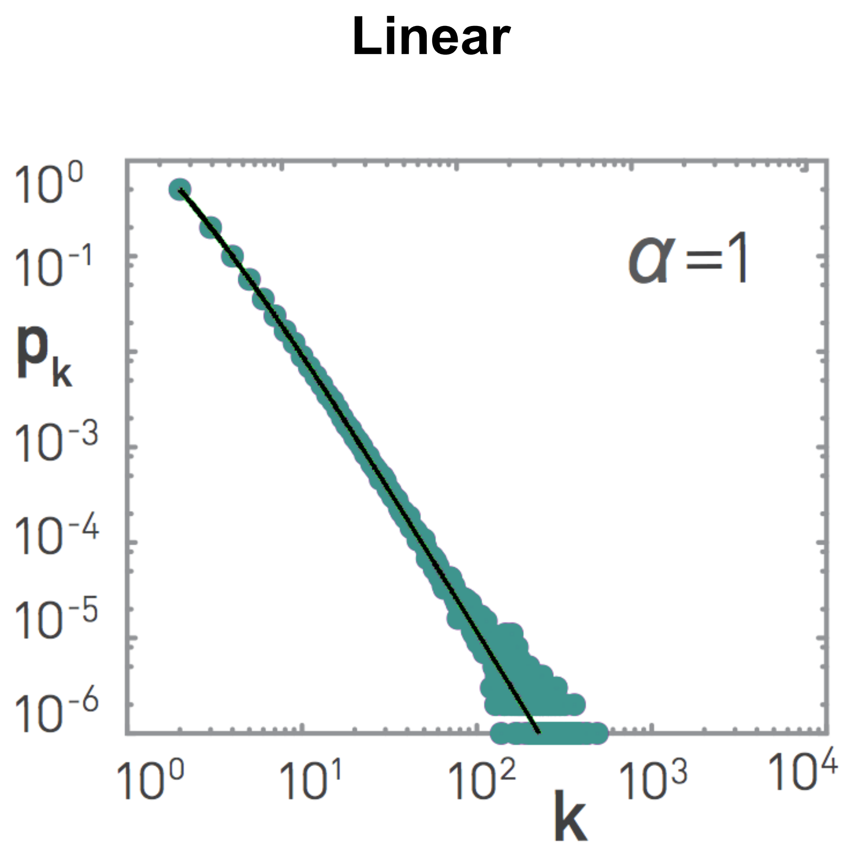
$$k_{\max} \sim t^{1/2}$$

Degree distributions (superlinear regime)



Winners take it all!

$k_{max} \sim t$



Power law

$k_{max} \sim t^{1/2}$

Non-linear preferential attachment

Conclusion:

Only linear preferential attachment produces scale-free growing networks. On the other hand, non-linear preferential attachment is more than sufficient to provide a wide spectrum of fat-tailed degree distributions.

The idea of "Scale-free networks" is a subtle and fragile concept

Classes of small-world networks

We can state (based on existing empirical analysis of real nets) that there are 3 main classes of graphs :

single-scale : individual degrees do not deviate appreciably from the average degree of the graph (type most compatible with WS-model);

broad-scale : degrees span a wider interval, with degree distribution falling off exponentially for large k ;

scale-free : those graphs in which the degree distributions decays with a power law, exhibiting the same behavior at all scales.

Luís Nunes Amaral
Northwestern University

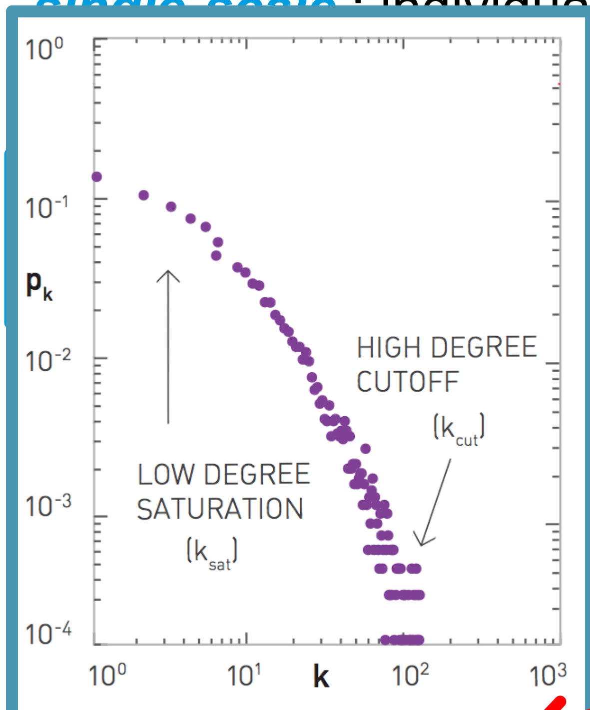
Amaral, Scala, Barthélémy, and Stanley,
Classes of small-world networks,
PNAS 97 (21) 11149 (2000).

Classes of small-world networks

We can state (based on existing empirical analysis of real nets) that there are 3 main classes of graphs :

single-scale : individual degrees do not deviate appreciably from the

raph (type most compatible with WS-model);



span a wider interval, with degree distribution

for large k ;

ns in

he s

Exponential cutoffs & finite size effects:

- A net is SF only in the limit of very large N (try it!).
- There's a maximum degree a node can have. This introduces natural (exponential) cutoffs in the degree dist.

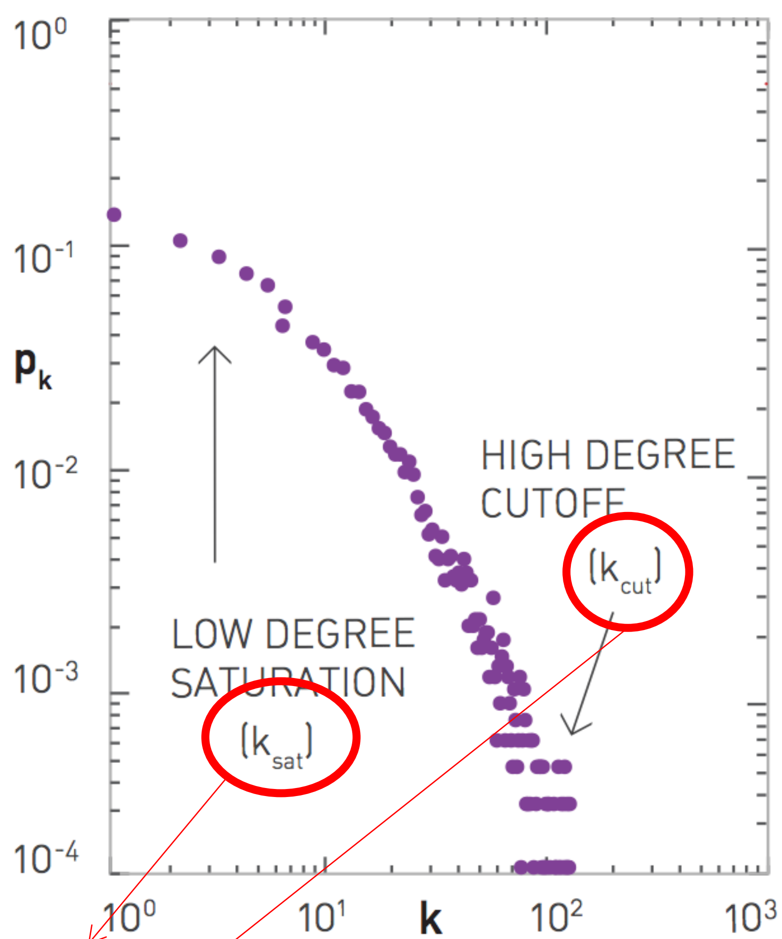
Dorogovtsev, Mendes, & Samukhin (2001). Size-dependent degree distribution of a scale-free growing network. Phys Rev E, 63(6), 062101.

A general picture…

In real systems we rarely find perfect power-laws

Low-degree saturation: is a common deviation from the power-law behavior. Its signature is a flattened P(k) for $k < k_{sat}$. This indicates that we have fewer small degree nodes than expected for a pure power law.

High-degree saturation: High-degree cutoff appears as a rapid drop in P(k) for $k > k_{cut}$. These cutoffs can emerge from different sources, some of which will be analyzed today!



Note however that these cutoffs are hard to determine…

Clauset et al. method: General idea

For details (and code) on this procedure, see: http://tuvalu.santafe.edu/~aaronc/powerlaws/

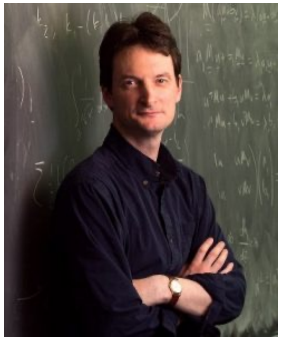
Goal: estimate γ from a discrete set of data points

1. Choose a k_{sat} and a k_{cut} between an interval of possible k_{min} and k_{max}. Optimize the best value of the degree exponent corresponding to this pair, using a statistical test to evaluate the quality of the fit.

1. Repeat 1, scanning the entire interval of possible values of k_{sat}, k_{cut} , keeping the best combination {γ, $k_{sat,}$ k_{cut}} in what concerns the “goodness” of the fit.

Aaron Clauset
Santa Fe Inst. & Univ. of Colorado

Mark E. J. Newman
Michigan Univ.

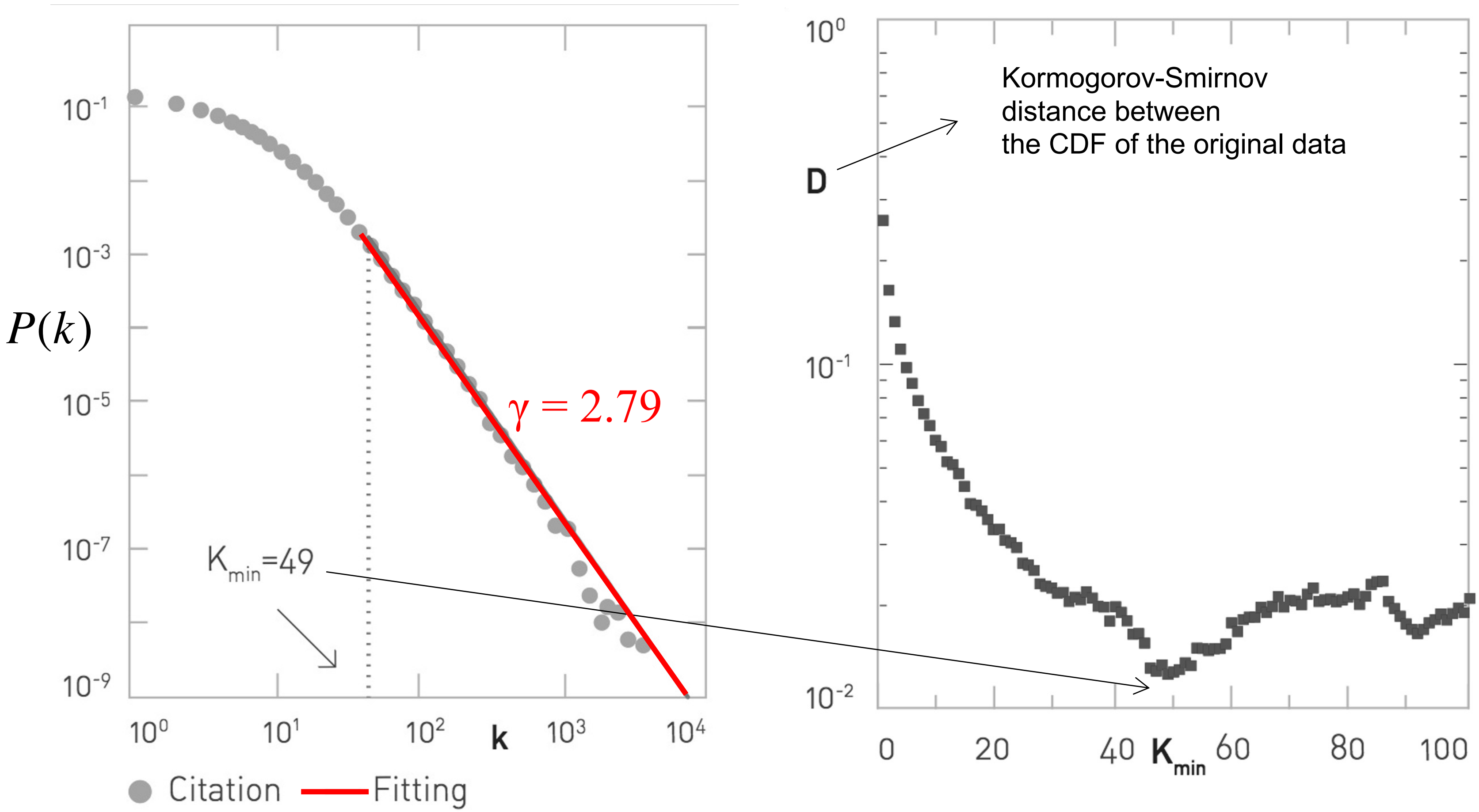


A. Clauset, et al. *Power-law distributions in empirical data*. SIAM Review S1: 661-703, 2009.

Clauset et al. method: General idea

For details (and code) on this procedure, see: http://tuvalu.santafe.edu/~aaronc/powerlaws/

Example: citation networks



Clauset et al. method: General idea

For details (and code) on this procedure, see: http://tuvalu.santafe.edu/~aaronc/powerlaws/

Example: email networks

Let's resort to the input file enron.outdegrees
(available here: https://dl.dropboxusercontent.com/s/e1kr9ri2btfwg9v/enron.outdegrees?dl=0) and
enron.outdegree
(available here: https://dl.dropboxusercontent.com/s/79n6w3p3cyat5lx/enron.outdegree?dl=0)
created with webgraph in exercise 6 of our problem set 1.

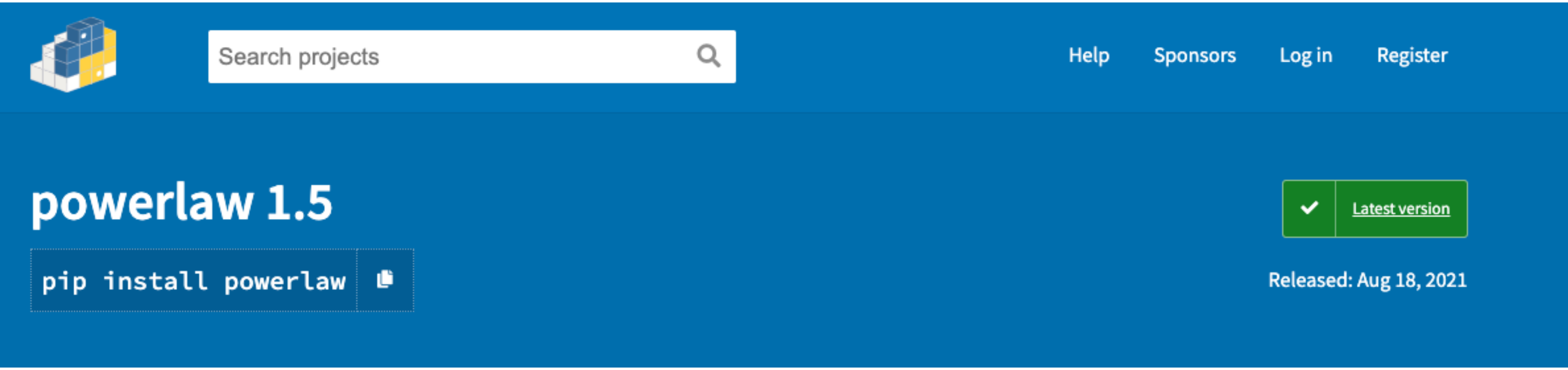
You may try the implementation in R of Clauset's algorithm (last problem in our 2nd set of exercises

```
[madonna@tecnico lab02]$ R
> install.packages("igraph")
> library('igraph')
> degs <- read.table("enron.outdegrees")
> deg_pmf <- read.table("enron.outdegree")
> degs_pl_fit <- power.law.fit(degs$V1)
> degs_pl_fit > plot(deg_pmf$V1, log="xy", xlab="degree", ylab="#vertices")
> plot(rev(cumsum(rev(deg_pmf$V1))), log="xy", xlab="degree", ylab="#vertices")
> plot(sort(degs$V1, decreasing = TRUE), 1:length(degs$V1), log="xy", xlab="degree",
ylab="rank")
```

Clauset et al. method: General idea

For details (and code) on this procedure, see: http://tuvalu.santafe.edu/~aaronc/powerlaws/

Same idea, but in Python:
https://pypi.org/project/powerlaw/



Conclusion:
Be careful with bold statements...

- The scale-free property is a rather fragile concept for small N.

- Moreover, it is unlikely to have a perfect ***linear*** preferential attachment, we may have constraints on the number of links and vertices, etc.

- **Take-home message:**

Perfect power-laws are useful concepts to have in mind as a reference point. Reality is often more complex, portraying different classes of complex networks.

Amaral, Scala, Barthélémy, and Stanley, PNAS 97 (21) 11149 (2000).

Broido, A. D., & Clauset, A.. Nature communications, 10(1), 1-10. (2019)

What are the mechanisms behind these variations?

Understanding topological variety

Power-laws: Growth & preferential attachment

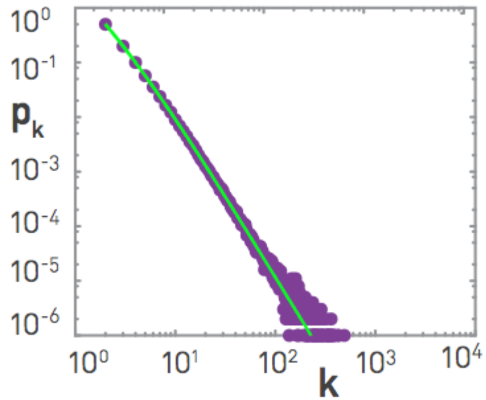
Stretched exponentials: If preferential attachment is sublinear you get a stretched exponential.

Fitness-induced corrections: ??

Small-degree saturations: ??

High degree cutoffs: ??

Hierarchical structure & power-law dep. in clustering: ??



Understanding topological variety

Power-laws: Growth & preferential attachment

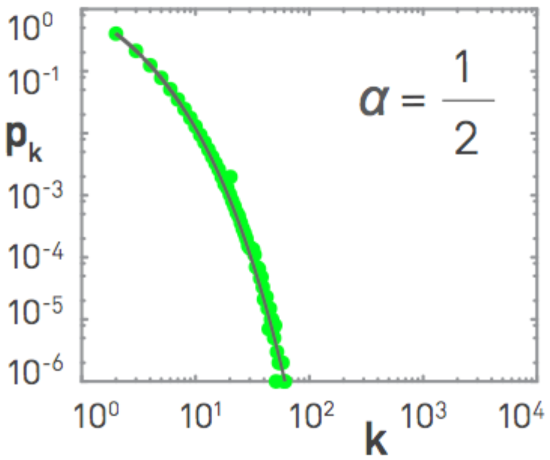
Stretched exponentials: If preferential attachment is sublinear you get a stretched exponential.

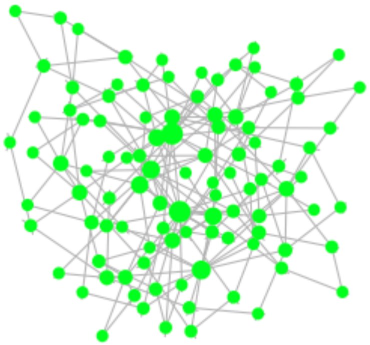
Fitness-induced corrections: ??

Small-degree saturations: ??

High degree cutoffs: ??

Hierarchical structure & power-law dep. in clustering: ??





Understanding topological variety

Power-laws: Growth & preferential attachment

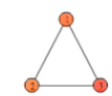
Stretched exponentials: If preferential attachment is sublinear you get a stretched exponential.

Fitness-induced corrections: ??

Small-degree saturations: ??

High degree cutoffs: ??

Hierarchical structure & power-law dep. in clustering: ??



Understanding topological variety

Power-laws: Growth & preferential attachment

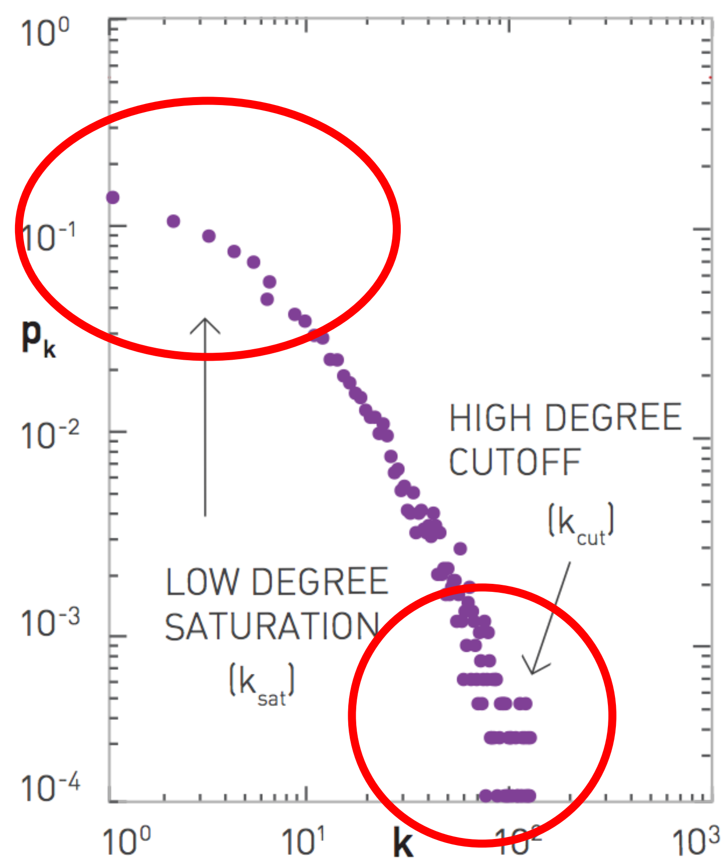
Stretched exponentials: If preferential attachment is sublinear you get a stretched exponential.

Fitness-induced corrections: ??

Small-degree saturations: ??

High degree cutoffs: ??

Hierarchical structure & power-law dep. in clustering: ??



Understanding the origins of topological variety

Power-laws: Growth & preferential attachment

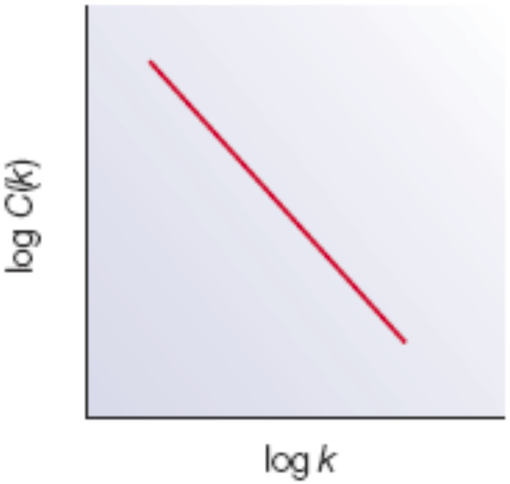
Stretched exponentials: If preferential attachment is sublinear you get a stretched exponential.

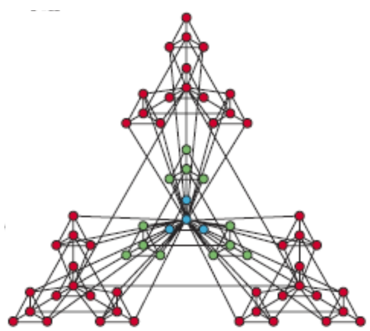
Fitness-induced corrections: ??

Small-degree saturations: ??

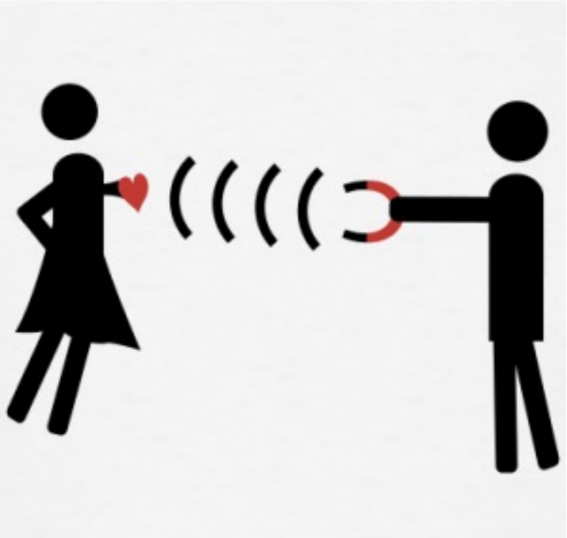
High degree cutoffs: ??

Hierarchical structure & power-law dep. in clustering: ??





Is the degree the only thing that matters?



Other variant: BA model with initial attractiveness

- ***Growth***: add nodes sequentially.

At t=0 consider (e.g.) a ring with m_0 nodes; for each time-step, add 1 node and connect it via m new edges with existing nodes;

- ***Preferential attachment***:

when establishing the connections, each new node is connected to m older nodes with a probability proportional to a constant plus the degree of the older nodes.

$$\Pi_i = \frac{A_i + k_i}{\sum_j \left(A_j + k_j\right)}$$

Irrespectively of your degree, you have always some attractiveness!

Dorogovtsev, et al. (2000), PRL 85: 4633.

Other variant: BA model with initial attractiveness

- ***Growth***: add nodes sequentially.

At t=0 consider (e.g.) a ring with m_0 nodes; for each time-step, add 1 node and connect it via m new edges with existing nodes;

- ***Preferential attachment***:

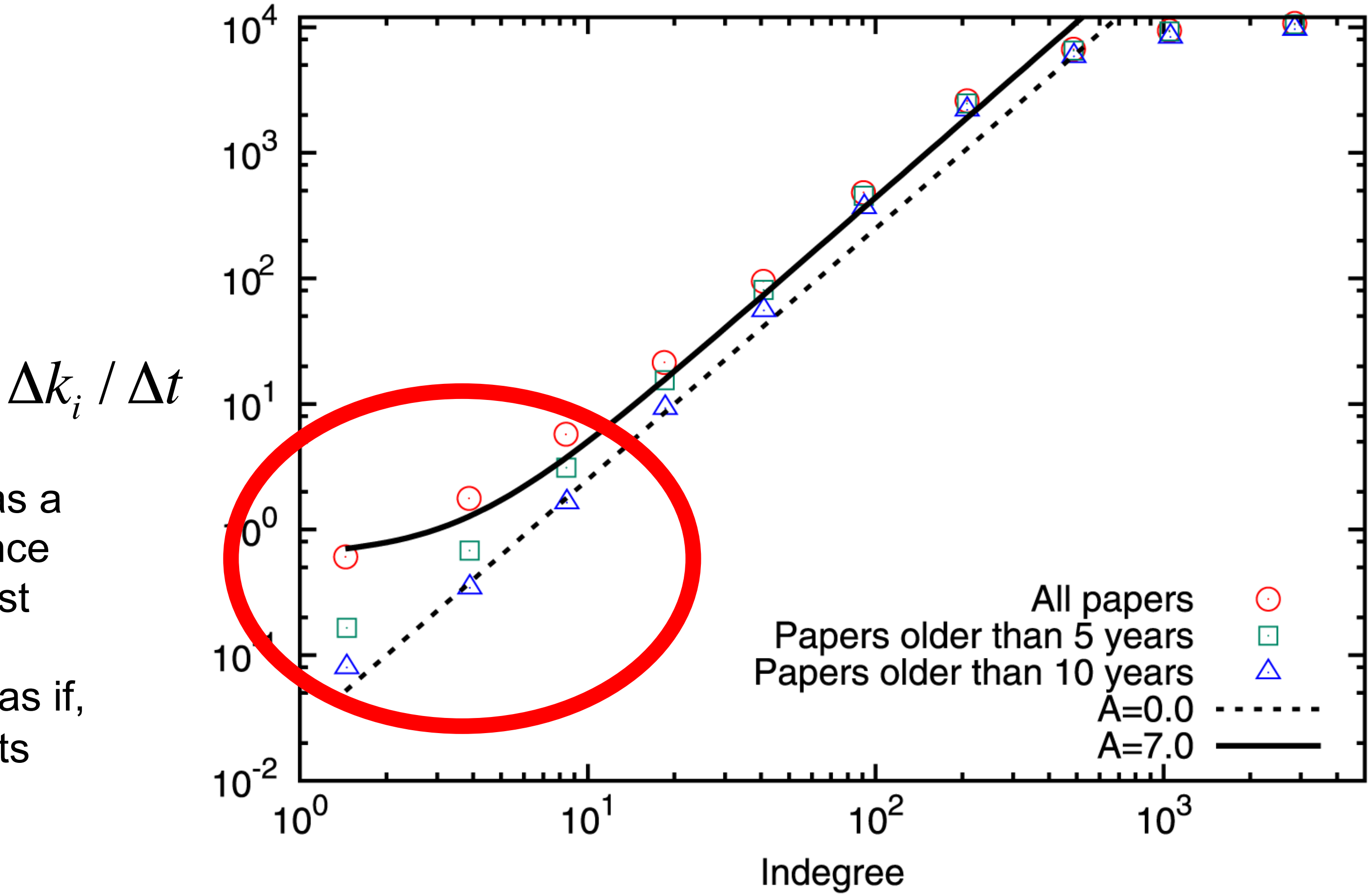
when establishing the connections, each new node is connected to m older nodes with a probability proportional to a constant plus the degree of the older nodes.

$$\Pi_i = \frac{A_i + k_i}{\sum_j \left(A_j + k_j\right)} \Rightarrow \gamma = 3 + \frac{A}{m} > 3$$

Dorogovtsev, et al. (2000), PRL 85: 4633.

Initial attractiveness & citation networks

Attractiveness can also be measured



Attractiveness has a significant influence only within the first few years after publication. It is, as if, a new paper starts with 7 citations…

Y.-H. Eom and S. Fortunato. *Characterizing and Modeling Citation Dynamics*. PLoS ONE, 6: e24926, 2011.

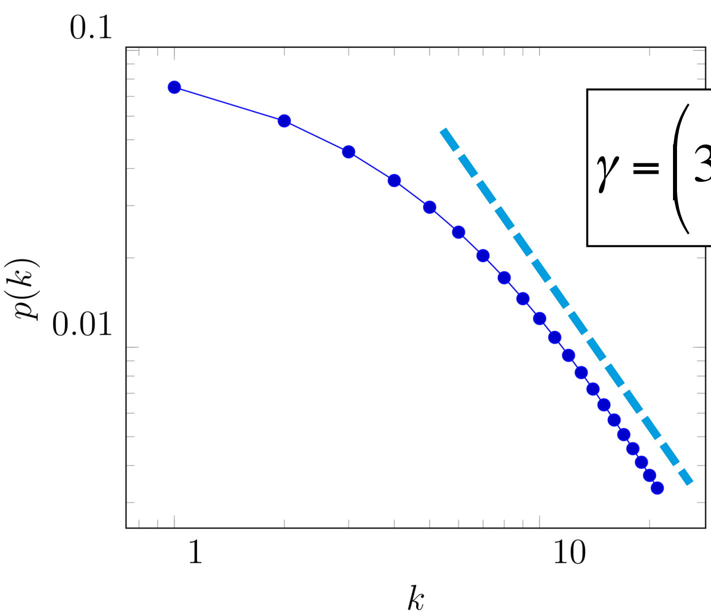
BA model with initial attractiveness

Dorogovtsev, et al. (2000), Phys Rev Lett 85: 4633.

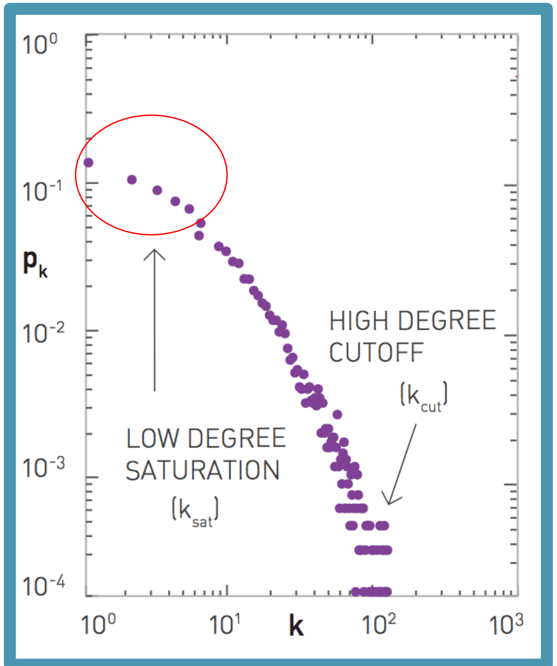
- It also generates a small degree saturation and increases the value of γ:

$$\Pi_i = \frac{A + k_i}{\sum_j \left(A + k_j\right)}$$

$$P(k) \sim \left(A + k\right)^{-\gamma}$$



$$\gamma = \left(3 + \frac{A}{m}\right)$$



Other variant: BA model with initial attractiveness

Challenge #3: Can you analyze either through computer simulations or analytically the impact of initial attractiveness?

What would be the result if attractiveness depends on the age of the node?

Dorogovtsev, et al. (2000), PRL 85: 4633.

Fitness models

Ginestra Bianconi et al. Bose-Einstein Condensation in Complex Networks, Phys. Rev. Lett., 86: 5632–5635, 2001.

- ***Growth***: add nodes sequentially.

At t=0 consider (e.g.) a ring with m_0 nodes; for each time-step, add 1 node and connect it via m new edges with existing nodes;

Fitness models

Ginestra Bianconi et al. Bose-Einstein Condensation in Complex Networks, Phys. Rev. Lett., 86: 5632–5635, 2001.

- ***Growth***: add nodes sequentially.

At t=0 consider (e.g.) a ring with m_0 nodes; for each time-step, add 1 node and connect it via m new edges with existing nodes;

- ***Preferential attachment***:

when establishing the connections, each new node is connected to older nodes with a probability proportional to the degree of the older nodes and to nodes intrinsic fitness η.

$$\Pi_i = \frac{\eta_i k_i}{\sum_j \eta_j k_j}$$

"fittest-get-richer" process

Fitness models

Ginestra Bianconi et al. Bose-Einstein Condensation in Complex Networks, Phys. Rev. Lett., 86: 5632–5635, 2001.

- ***Growth***: add nodes sequentially.

At t=0 consider (e.g.) a ring with m_0 nodes; for each time-step, add 1 node and connect it via m new edges with existing nodes;

- ***Preferential attachment***:

when establishing the connections, each new node is connected to older nodes with a probability proportional to the degree of the older nodes and to nodes intrinsic fitness η.

$$\Pi_i = \frac{\eta_i k_i}{\sum_j \eta_j k_j}$$

"fittest-get-richer" process

The intrinsic fitness is taken from a given distribution $\rho(\eta)$

New nodes may now become central, even if added at a later stage!

Result: The shape of the degree distribution depends on $\rho(\eta)$

Fitness models

"fittest-get-richer" process

Example: Let's consider a uniform distribution $\rho(\eta)$ taken from the interval [0,1].

$$\Pi_i = \frac{\eta_i k_i}{\sum_j \eta_j k_j}$$

$$P(k) \sim \frac{k^{-\gamma}}{\ln k}$$

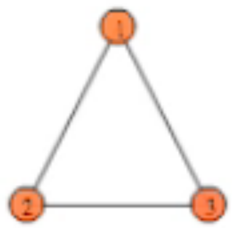
$$\gamma = 2.255 < 3 \text{ (BA-model)}$$

Fitness models

“fittest-get-richer” process

Challenge: Confirm this result simulating the growth of a network biased by a uniform fitness dist.

What if we do not have a uniform distribution of fitness values?



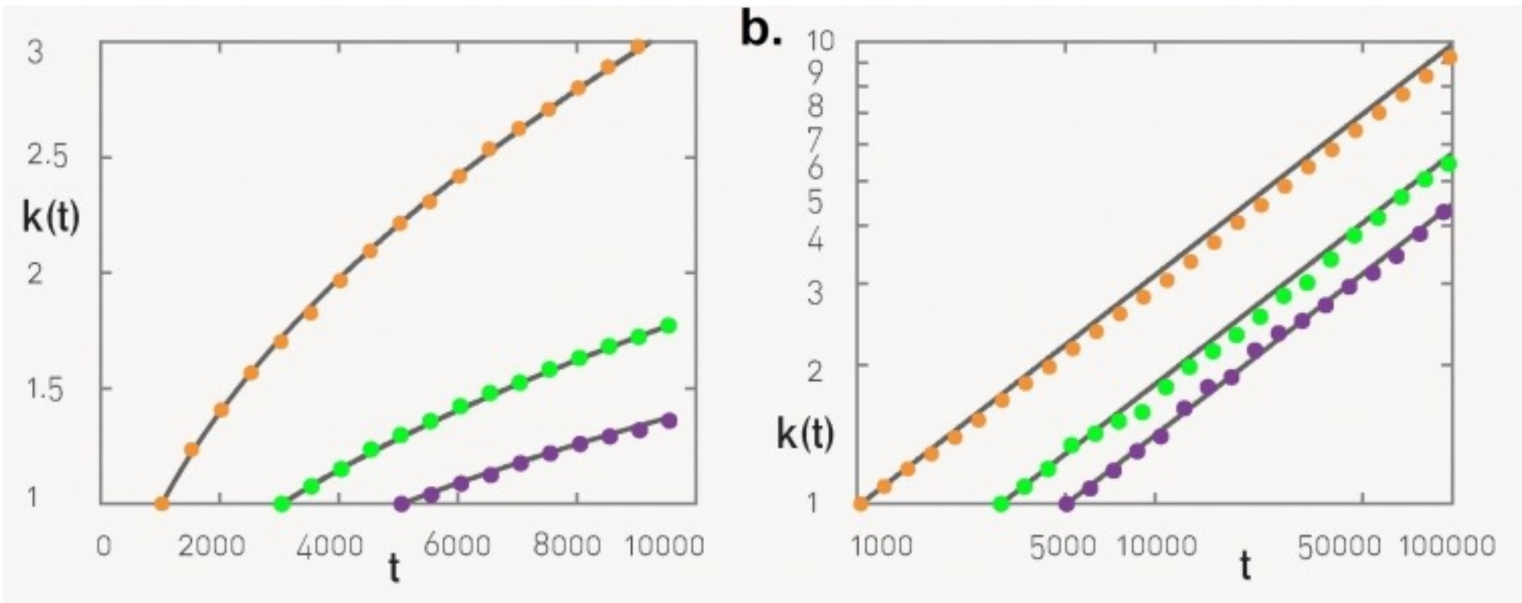
Growing network in which each new node acquires a randomly chosen fitness parameter at birth

Fitness models

"fittest-get-richer" process

Linear plot

Log-log plot



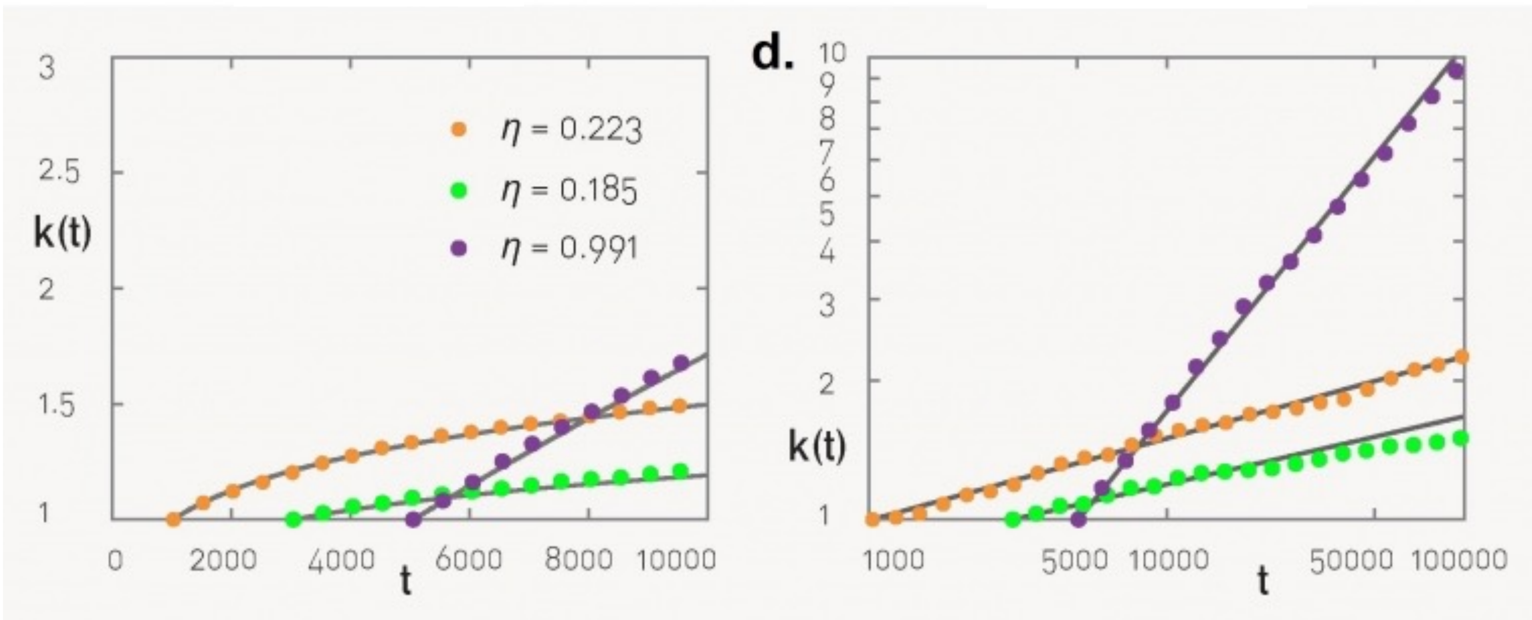
In the Barabasi-Albert model, nodes that get into the network later are unable to pass the earlier nodes

Fitness models

“fittest-get-richer” process

Linear plot

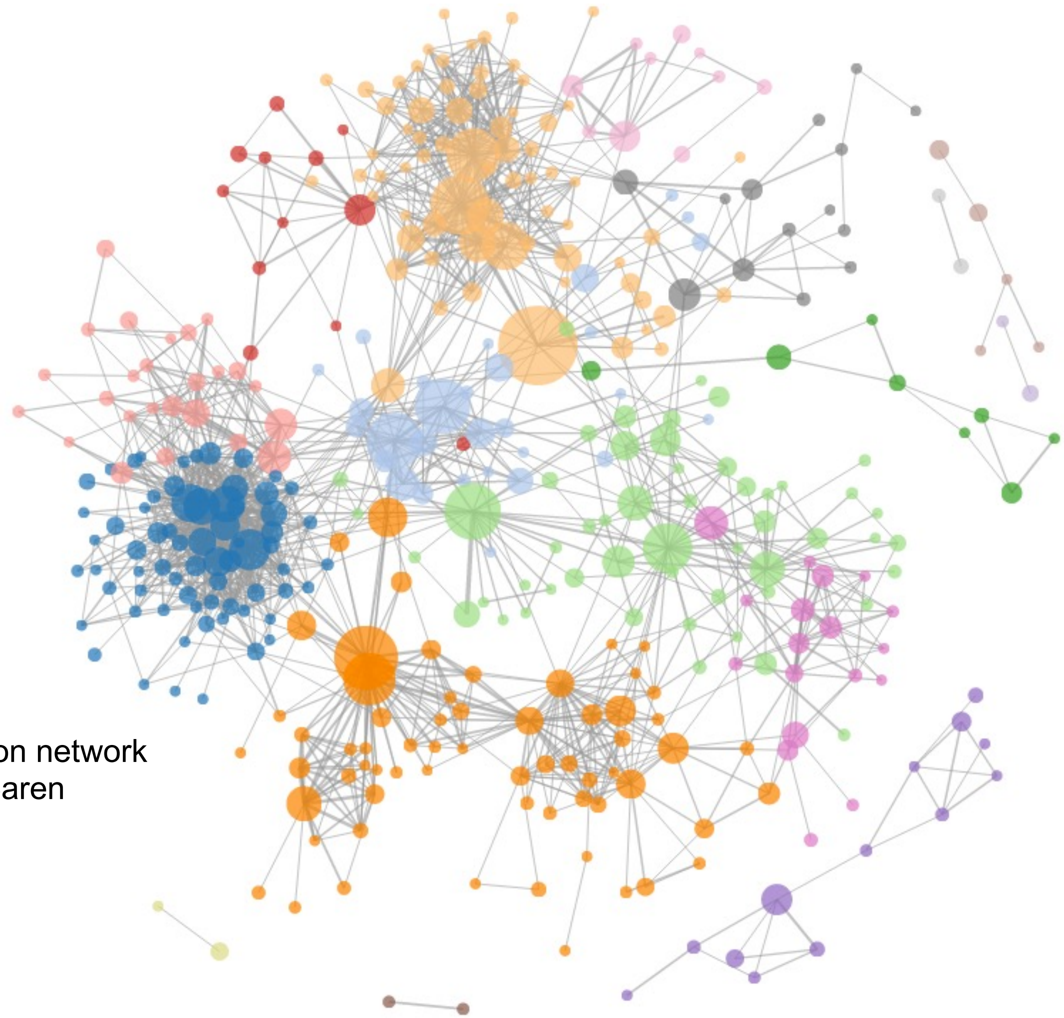
Log-log plot



If we add fitness, a latecomer node with a higher fitness (purple symbols) can overcome the earlier nodes.

Fitness models are problem dependent

- Example: Modeling citation networks



Sociology citation network
by Neal Caren

Fitness models are problem dependent

- Example: Modeling citation networks

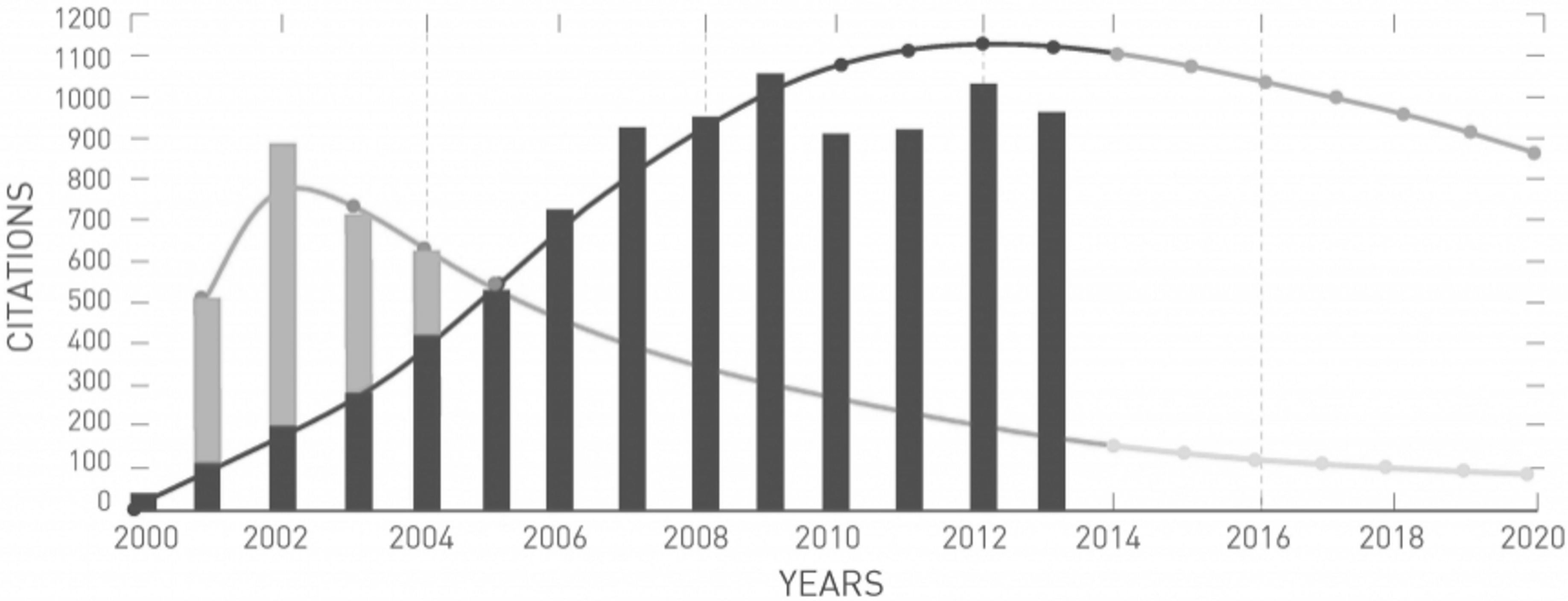
$$\Pi_i \sim \eta_i k_i P_i(t)$$

Fitness of a paper (novelty, importance, etc)

Number of citations at time t

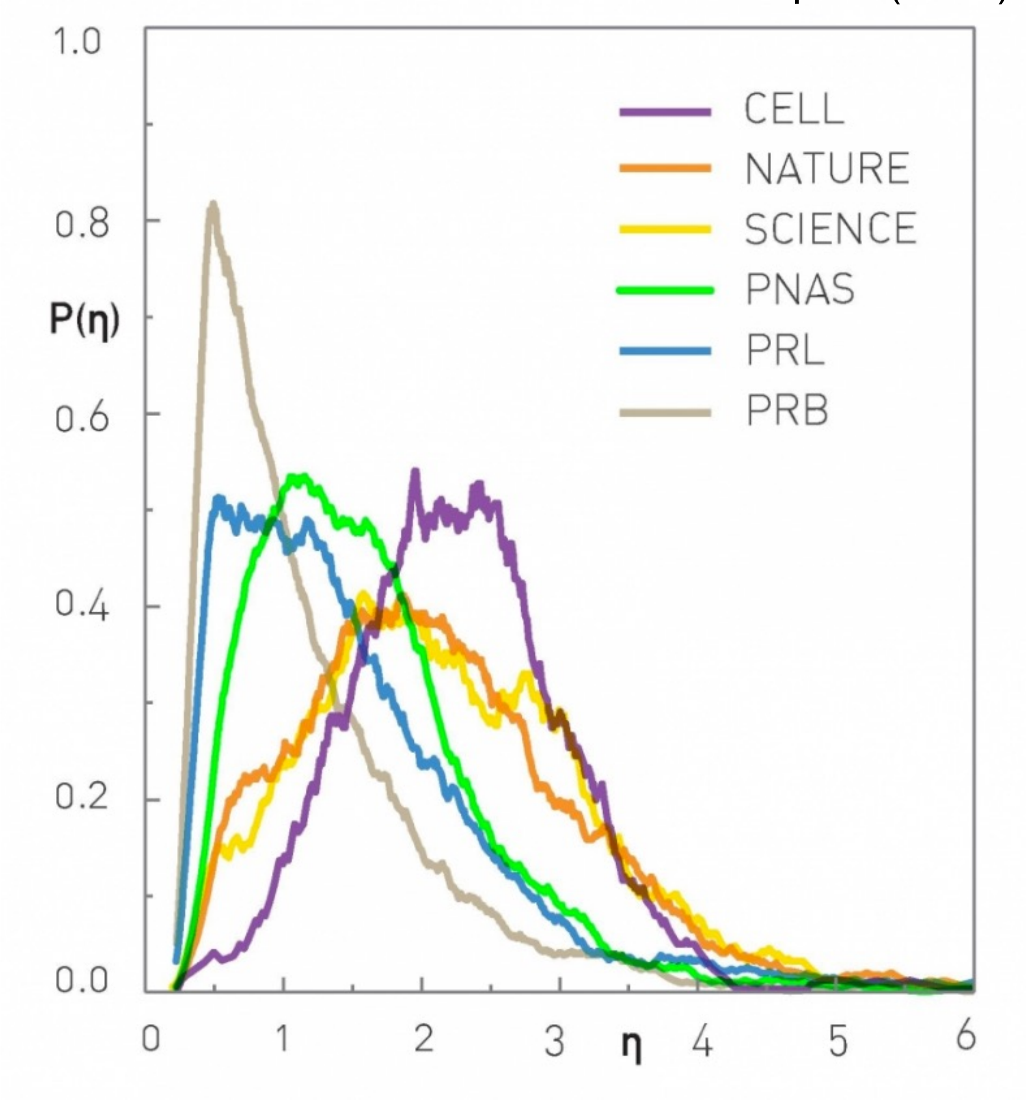
new ideas are integrated into subsequent work, hence the novelty fades with time

Fitness models are problem dependent

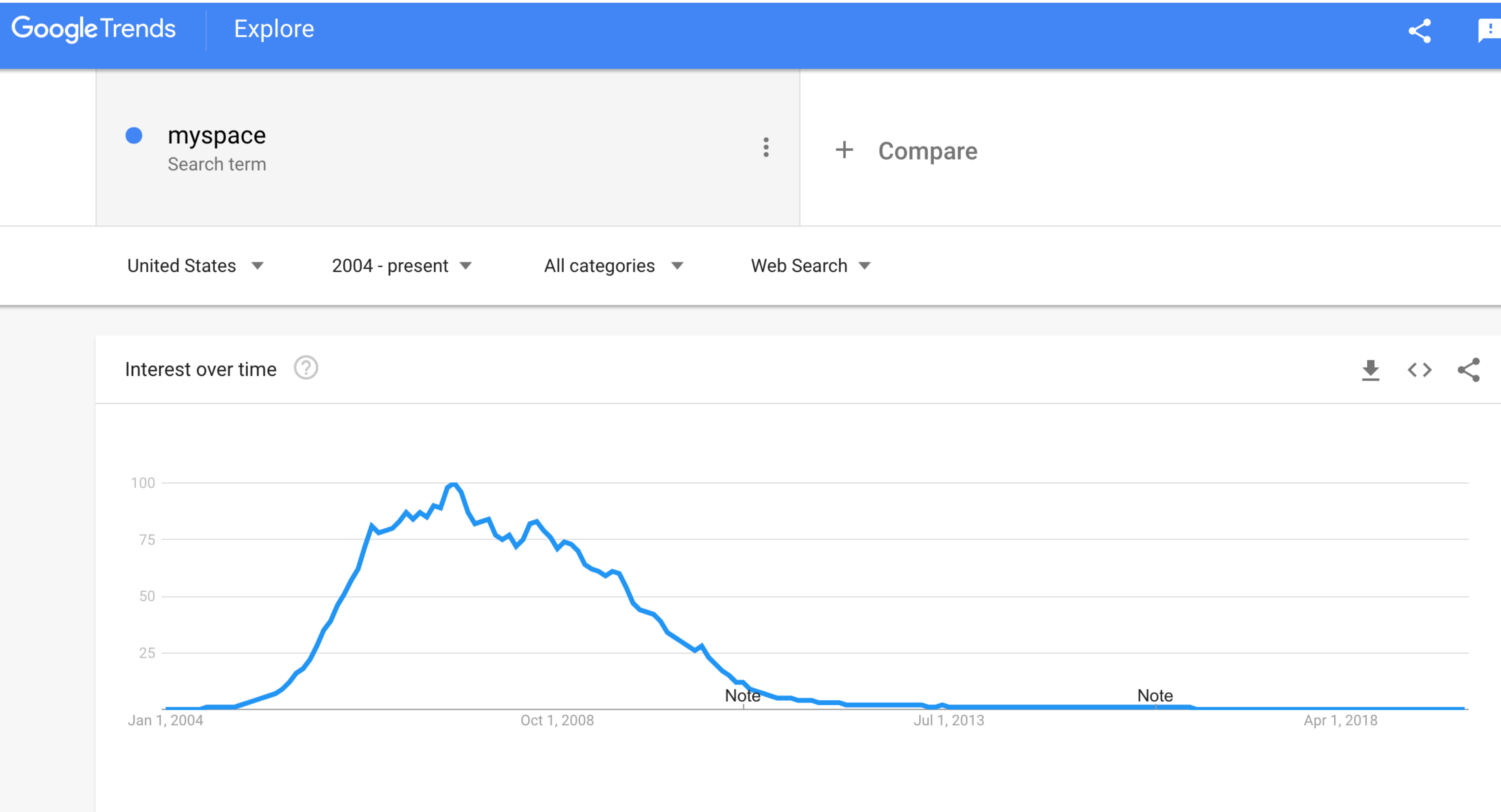


Fitness models are problem dependent

Fitness Distribution of Research Papers (1990)



Another example: WWW and MySpace



Fitness of the WWW

- Example: Modeling the WWW

$$\Pi_i \sim \eta_i k_i$$

Fitness of a document
(which can be measured)

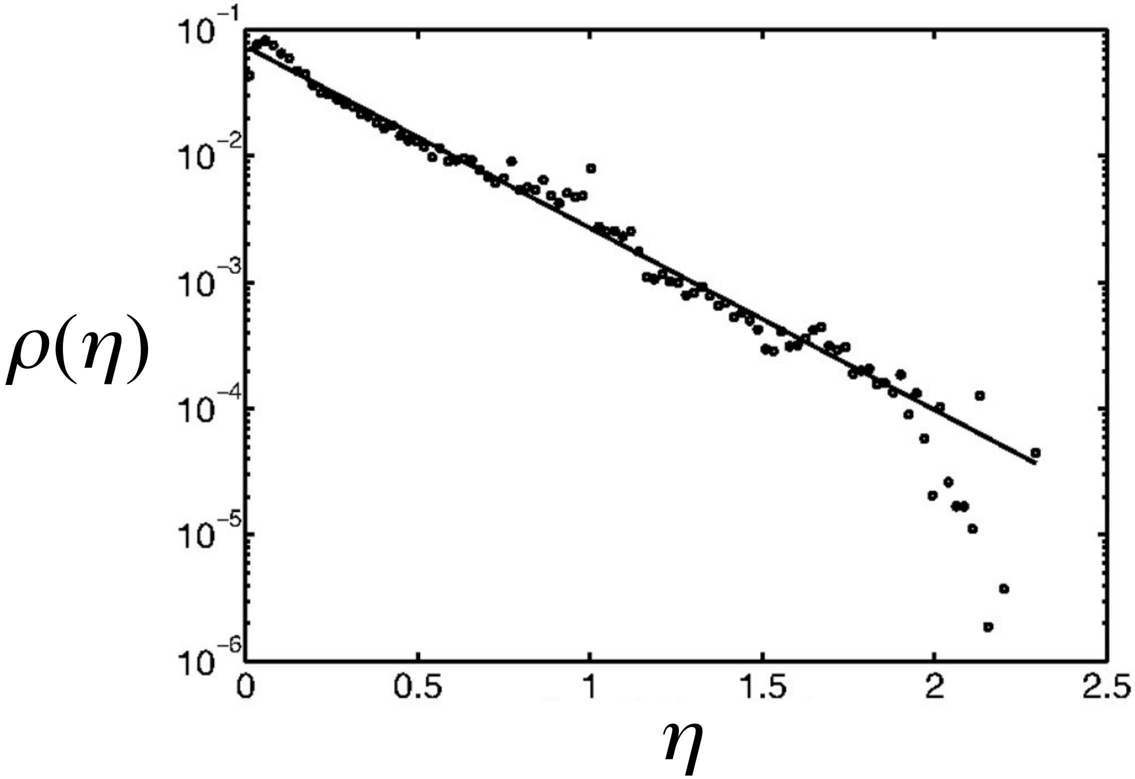
Number of in-links

It's also neat to show that, small differences in fitness lead, far into the future, to large differences in degrees.

J. S. Kong, N. Sarshar, and V. P. Roychowdhury.
Experience versus talent shapes the structure of the Web.
PNAS, 105:13724-9, 2008.

Fitness of the WWW

- Distribution of fitness values:



- Fitness distribution is exponential
- Fitness of Web documents varies in a relatively narrow range.
- Growth+preferential attachment amplifies the small fitness differences, turning nodes with slightly higher fitness into much bigger nodes.

J. S. Kong, N. Sarshar, and V. P. Roychowdhury. Experience versus talent shapes the structure of the Web. PNAS, 105:13724-9, 2008.

Do only new nodes create links?

Other variant: BA model with internal links

- new links do not only arrive with new nodes but are added between pre-existing nodes (e.g., WWW).
- Consider an extension of the BA-model, where in each time step we ***add a new node with m links***, followed by ***n internal links***, each selected with probability

$$\Pi(k_i, k_j) \sim k_i \times k_j$$

(double preferential attachment)

$$\gamma = 2 + \frac{m}{m+2n} < 3$$

lowers the degree exponent from 3 to 2, hence increasing the network's heterogeneity

Other variant: BA model with internal links

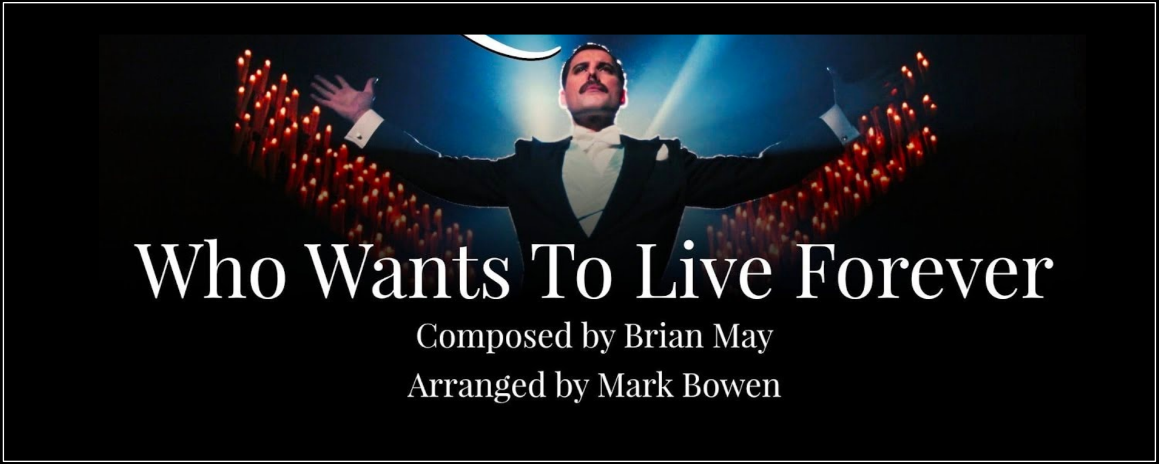
- new links do not only arrive with new nodes but are added between pre-existing nodes (e.g., WWW).
- Consider an extension of the BA-model, where in each time step we add a new node with *m* links, followed by ***n* internal links**, each selected with ***random probability***

$$\gamma = 3 + \frac{2n}{m} > 3$$

the resulting network will be more homogenous than the network without internal links

What about age?

Can you propose a model where age matters?



Other variant: BA model *with ageing* (model 1)

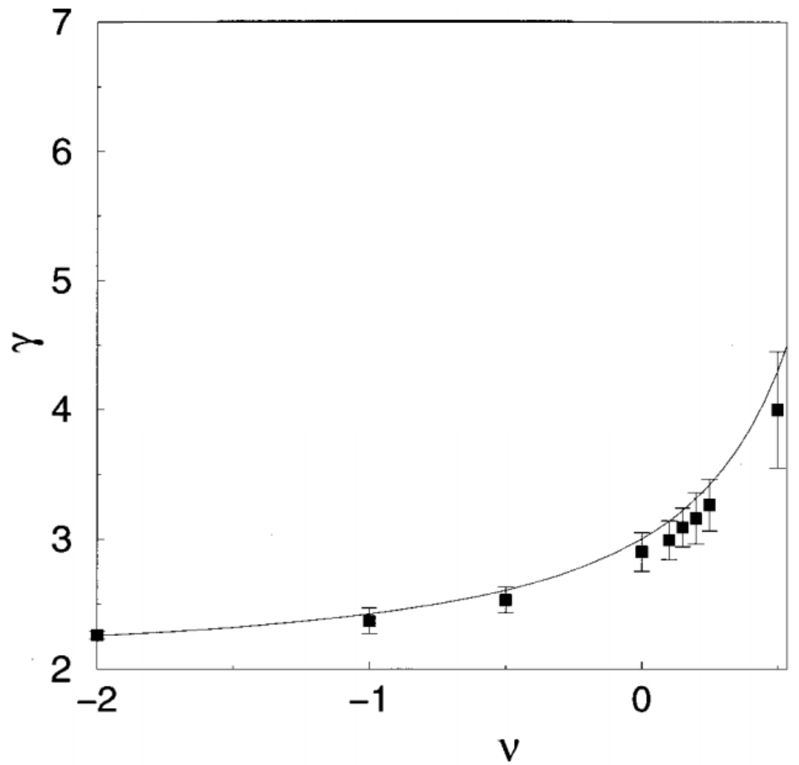
Amaral et al. (2000)
Dorogovtsev, et al. (2000)

Preferential attachment with preference for old/ young nodes

$$\Pi(k,t,t_i) \sim k(t-t_i)^{-\nu}$$

v>0 : young nodes get more attractive

v<0 : old nodes get (even) more attractive



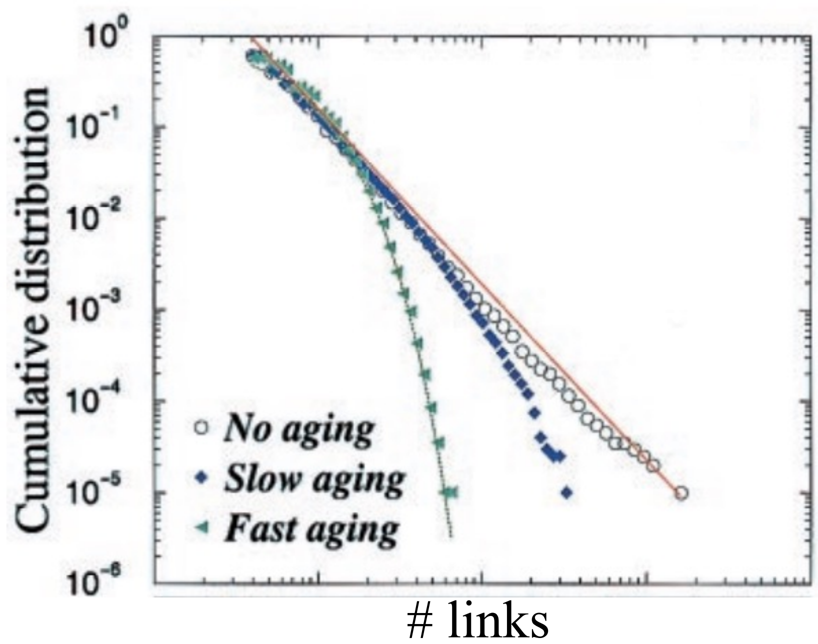
Other variant: BA model *with ageing* (model 2)

Amaral, Barthélémy, Stanley,
Classes of small-world networks, PNAS 2000

Same as the linear preferential attachment. Yet, with a probability that scales with

$$(t - t_i)^{-\nu}$$

an old node becomes inactive and cannot receive links from new nodes.



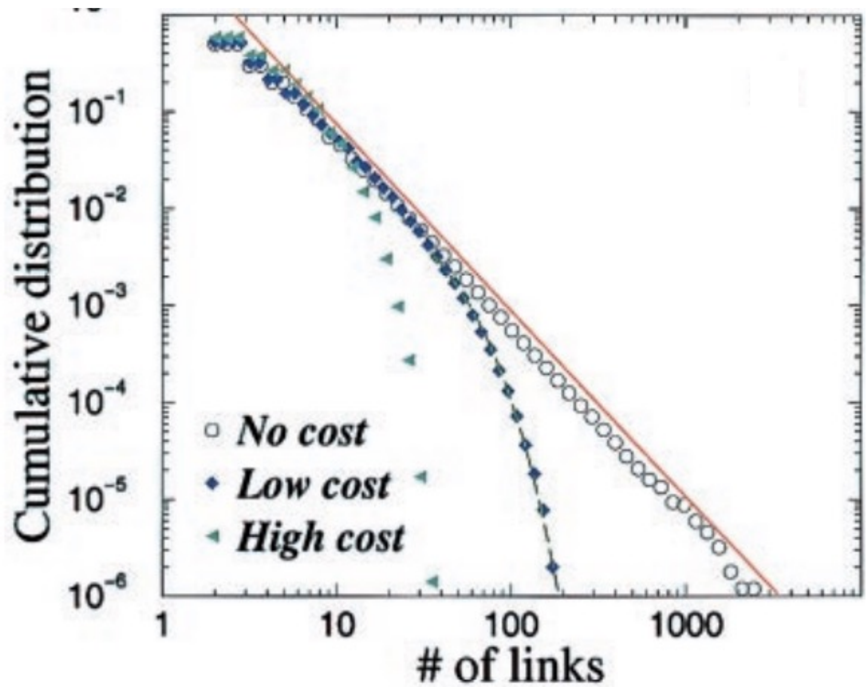
Other variant: BA model *with costs*

Amaral, Barthélémy, Stanley,
Classes of small-world networks, PNAS 2000

physical costs of adding links limits the number of possible links attaching to a given node.

a vertex becomes inactive when it reaches a maximum number of links k_{max}.

This creates natural exponential cut-offs in degree distributions.

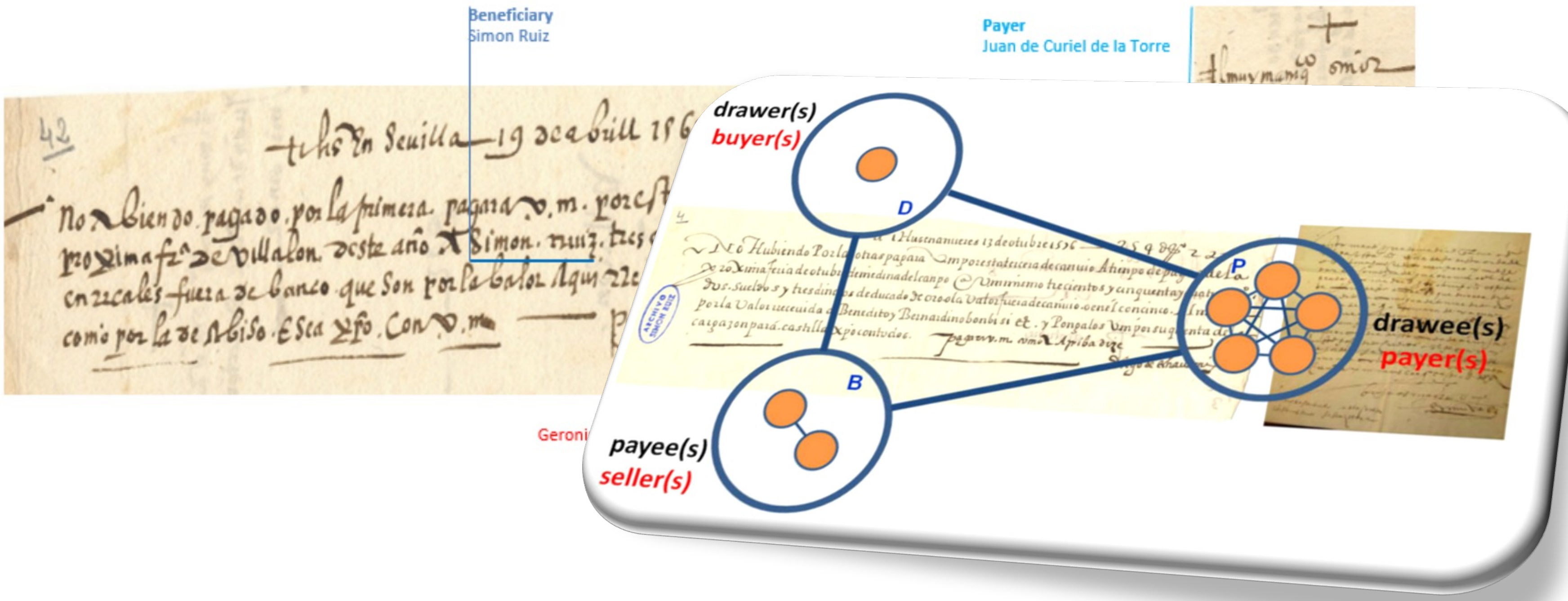


One more challenge

- **Networks grow, but also shrink…**
- Suggest a model in which the impact of this effect is tested?

Example #1

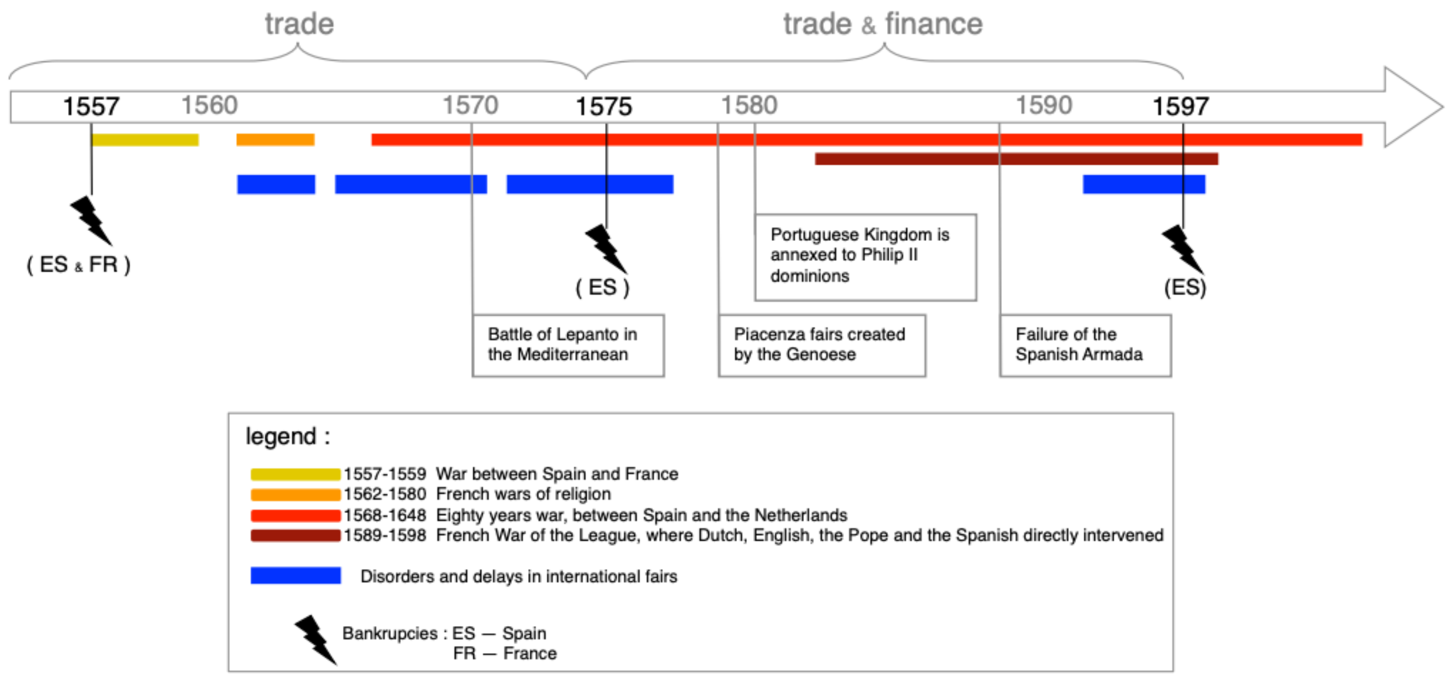
First global trading market using information contained in ~9000 Bills of Exchange



Ribeiro et al. 2018 Structural and temporal patterns of the first global trading market. R. Soc. open sci. 5: 180577.

Example #1

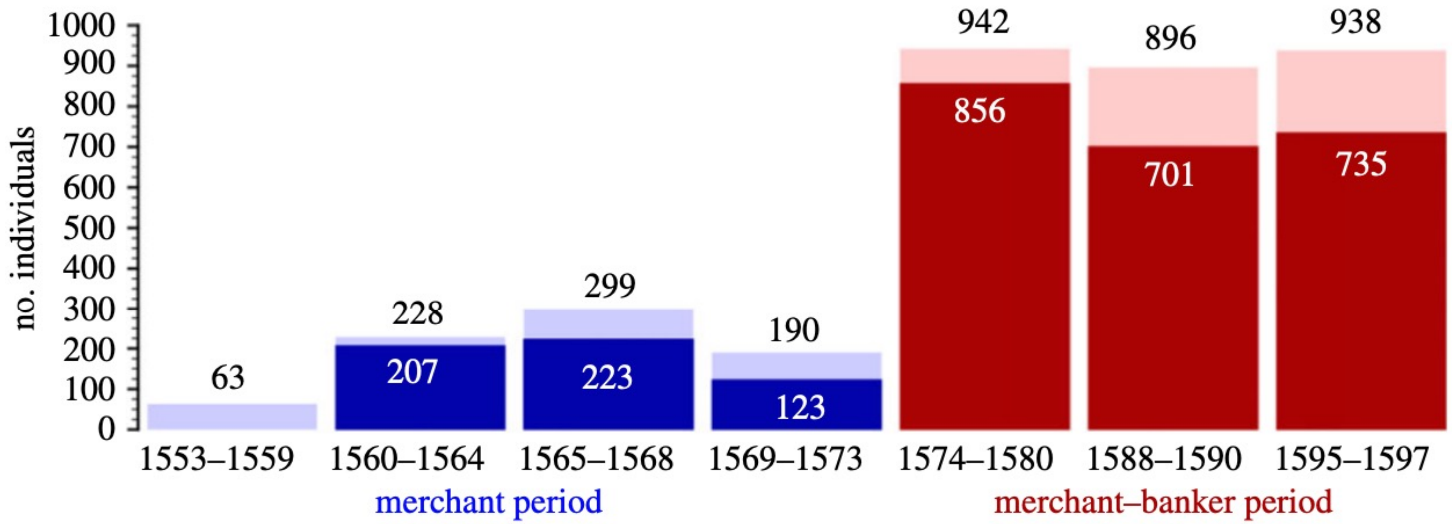
First global trading market using information contained in ~9000 Bills of Exchange



Ribeiro et al. 2018 Structural and temporal patterns of the first global trading market. R. Soc. open sci. 5: 180577.

Example #1

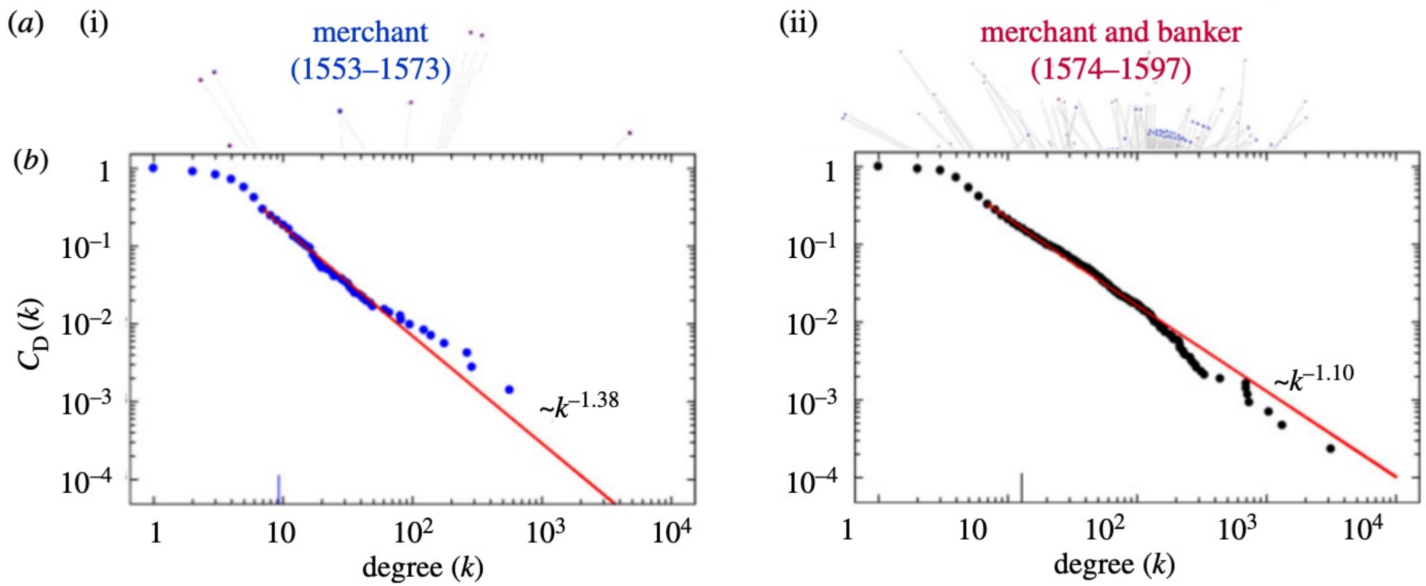
First global trading market using information contained in ~9000 Bills of Exchange



Ribeiro et al. 2018 Structural and temporal patterns of the first global trading market. R. Soc. open sci. 5: 180577.

Example #1

First global trading market using information contained in ~9000 Bills of Exchange



Ribeiro et al. 2018 Structural and temporal patterns of the first global trading market. R. Soc. open sci. 5: 180577.

Example #2

- Networks grow, but also shrink… Nodes and links can disappear.

Another example: NY fashion industry.

Nodes = designers and contractors

Links = annual co-production of lines of clothing

The industry has decayed persistently during the 90s:

$$N_{1985} = 3249 \text{ nodes}$$

$$N_{2003} = 190 \text{ nodes}$$

PNAS

Asymmetric disassembly and robustness in declining networks

Serguei Saavedra*, Felix Reed-Tsochas†‡, and Brian Uzzi§

*Department of Engineering Science and CABDyN Complexity Centre, Oxford University, Oxford OX1 3PJ, United Kingdom; †James M
CABDyN Complexity Centre, Saïd Business School, Oxford University, Oxford OX1 1HP, United Kingdom; and §Kellogg School of Man
Northwestern Institute on Complex Systems, Northwestern University, Evanston, IL 60208

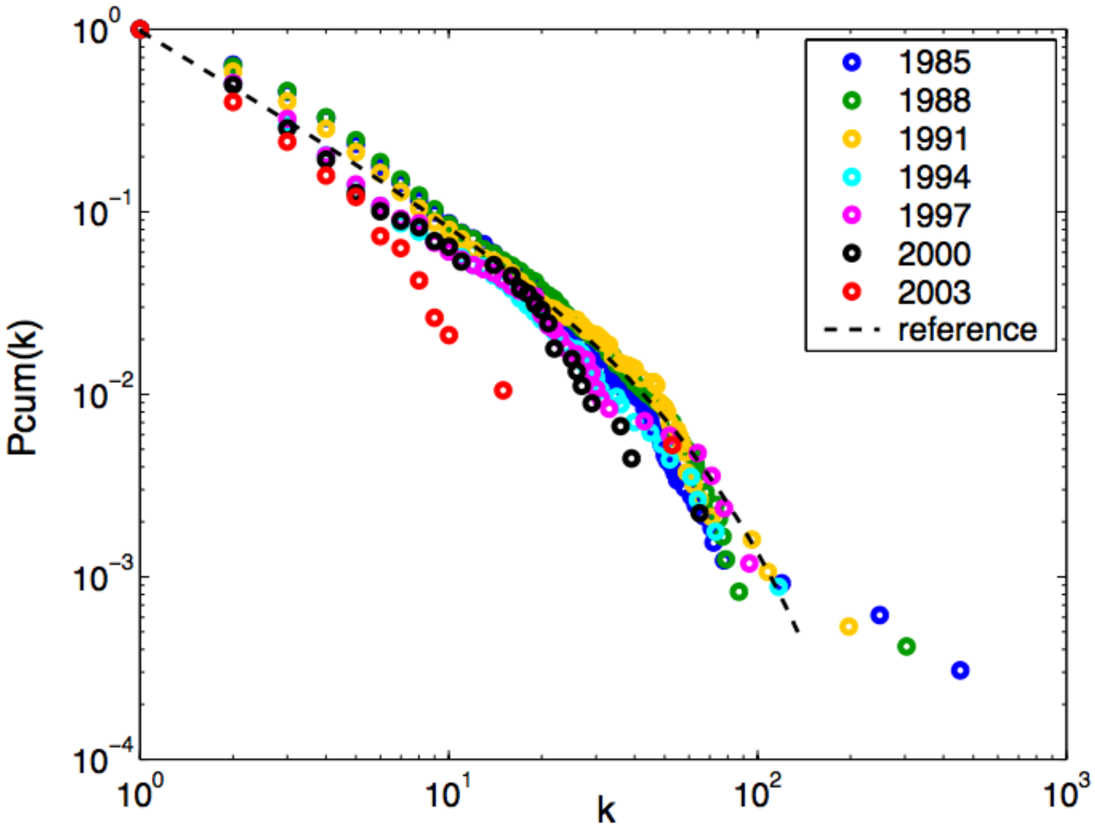
Edited by Simon A. Levin, Princeton University, Princeton, NJ, and approved September 3, 2008 (received for review May 17, 2008)

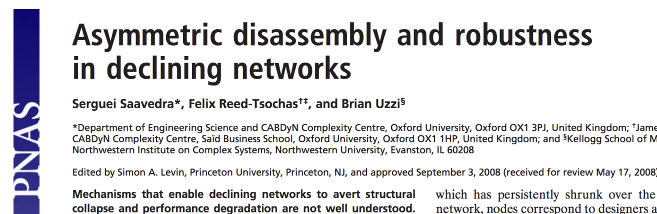
Mechanisms that enable declining networks to avert structural collapse and performance degradation are not well understood. This knowledge gap reflects a shortage of data on declining

which has persistently shrunk over the la network, nodes correspond to designers and linked through coproductions of annual run

Declining networks

Interestingly the network's degree dist. remained unchanged





PNAS

Asymmetric disassembly and robustness in declining networks

Serguei Saavedra*, Felix Reed-Tsochas†‡, and Brian Uzzi§

*Department of Engineering Science and CABDyN Complexity Centre, Oxford University, Oxford OX1 3PJ, United Kingdom; †James M
CABDyN Complexity Centre, Saïd Business School, Oxford University, Oxford OX1 1HP, United Kingdom; and §Kellogg School of Man
Northwestern Institute on Complex Systems, Northwestern University, Evanston, IL 60208

Edited by Simon A. Levin, Princeton University, Princeton, NJ, and approved September 3, 2008 (received for review May 17, 2008)

Mechanisms that enable declining networks to avert structural collapse and performance degradation are not well understood. This knowledge gap reflects a shortage of data on declining

which has persistently shrunk over the la
network, nodes correspond to designers and
linked through coproductions of annual run

Let's create a model with death of nodes…

Example:

- With rate r_{death} we remove a node.
- With rate r_{birth} we add a node as in the BA model

For simplicity let's reduce the number of parameters such that

$$r = \frac{r_{death}}{r_{birth}} = 1$$

What would be the outcome of this model?

Let's create a model…

Results are not surprising:

$r<1$

the number of removed nodes is smaller than the number of new nodes, hence the network continues to grow. In this case the network is scale-free with degree exponent

$$\gamma = 3 + \frac{2r}{r-1}$$

$r=1$

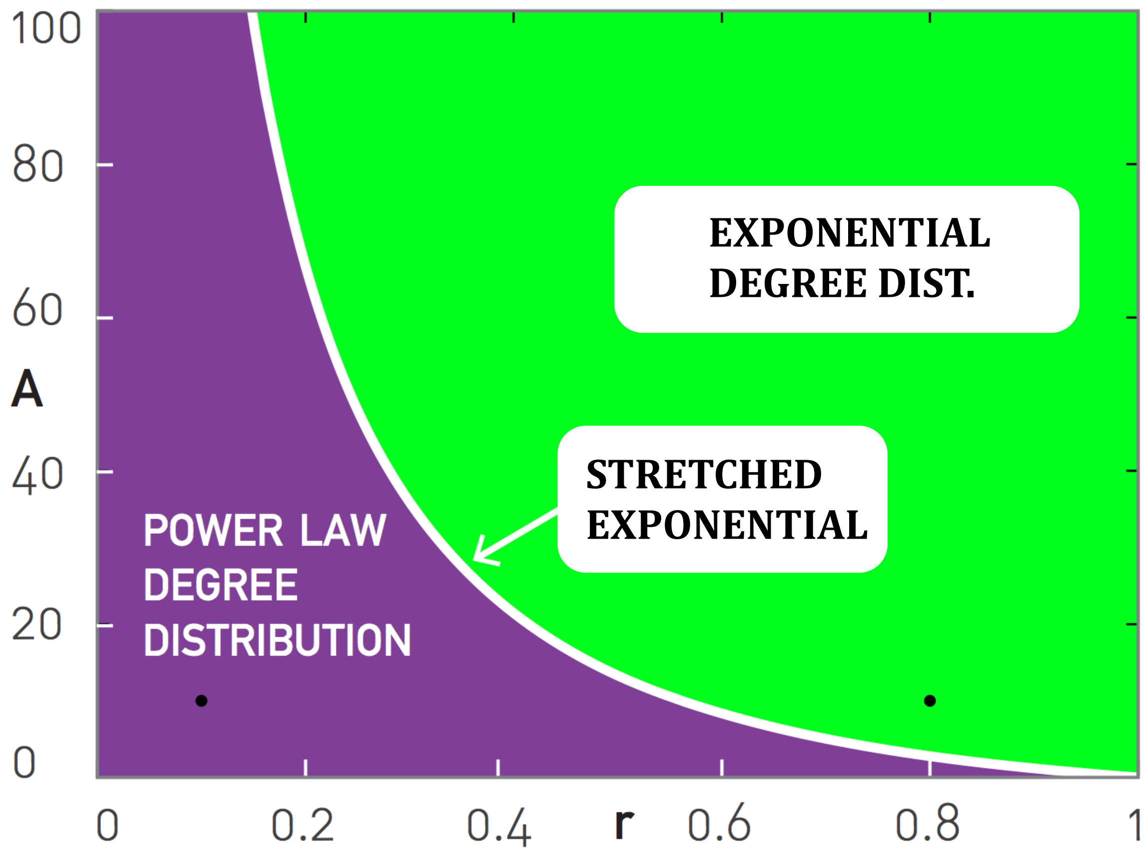
nodes arrive and are removed at the same rate, hence the network has a fixed size and would lose its scale-free nature.

$r>1$

number of removed nodes exceeds the number of new nodes, hence the network declines.

Let’s create a model…

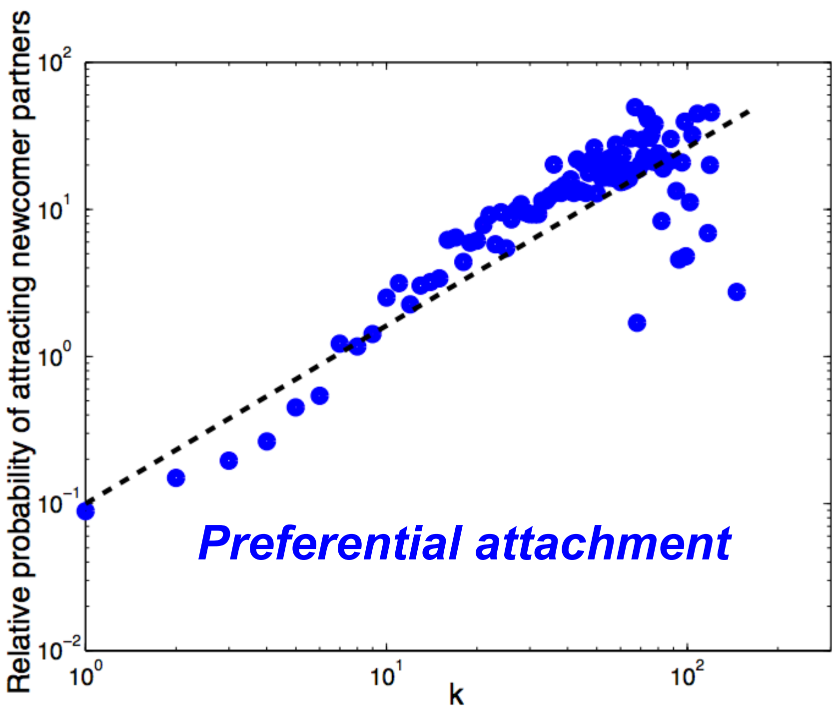
All gets counter-intuitive when combined with other real-world constraints (example: initial attractiveness, A):

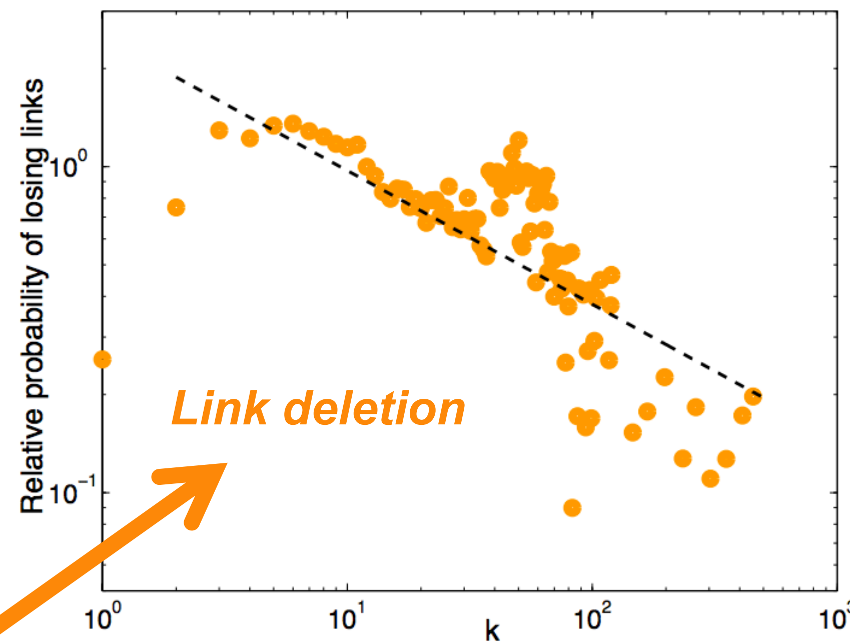


H. Bauke, C. Moore, J. Rouquier, and D. Sherrington. Topological phase transition in a network model with preferential attachment and node removal. The European Physical Journal B, 83: 519-524, 2011.

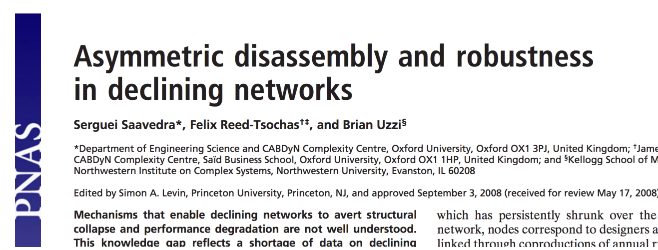
Open challenge

Now, let's return back to the fashion industry… Empirical results show





What would be the result of removing links (instead of nodes), taking into account the degrees of the nodes involved? Would you get the observed degree dists?


PNAS

Asymmetric disassembly and robustness in declining networks

Serguei Saavedra*, Felix Reed-Tsochas†‡, and Brian Uzzi§

*Department of Engineering Science and CABDyN Complexity Centre, Oxford University, Oxford OX1 3PJ, United Kingdom; †James M… CABDyN Complexity Centre, Saïd Business School, Oxford University, Oxford OX1 1HP, United Kingdom; and §Kellogg School of Man… Northwestern Institute on Complex Systems, Northwestern University, Evanston, IL 60208

Edited by Simon A. Levin, Princeton University, Princeton, NJ, and approved September 3, 2008 (received for review May 17, 2008)

Mechanisms that enable declining networks to avert structural collapse and performance degradation are not well understood. This knowledge gap reflects a shortage of data on declining …

which has persistently shrunk over the la… network, nodes correspond to designers and… linked through coproductions of annual run…

Conclusions so far…

Models with Pref. Attachment

Barabási-Albert Model
$\Pi \sim k \rightarrow \gamma = 3$

Non-linear Pref. Attachment
$$\Pi_i = \frac{k_i^\alpha}{\sum_j k_j^\alpha}$$

- $\alpha < 1 \rightarrow$ exponential dist
- $\alpha = 1 \rightarrow$ power-law
- $\alpha > 1 \rightarrow$ winners-take all

Growth with removal of nodes and/or links

Initial attractiveness
$\Pi \sim A + k \rightarrow \gamma = 3 + A/m$

n internal links
i) double pref. attachment

$$\gamma = 2 + \frac{m}{m+2n} < 3$$

ii) random attachment

$$\gamma = 3 + \frac{2n}{2m} > 3$$

Growth w/costs
$$k > k_{\max}$$

Fitness models
$$\Pi_i = \frac{\eta_i k_i}{\sum_j \eta_j k_j}$$

Ageing models
$$(t_i - t)^{-\nu}$$

Do we need degree-based preferential attachment to get to scale-free (or strongly heterogeneous) networks?

Can we get to power-laws following different principles?

Scale-free growth by ranking

Fortunato et al. PRL (2006)

- Growth + preferential attachment model in which attachment probability of a new node to an old vertex *s* given by the form

$$\Pi_s = \frac{R_s^{-\alpha}}{\sum_j R_j^{-\alpha}}$$

where R_s denotes the rank of the node *s* for some specific attribute and where α is a positive parameter.

Scale-free growth by ranking

Fortunato et al. PRL (2006)

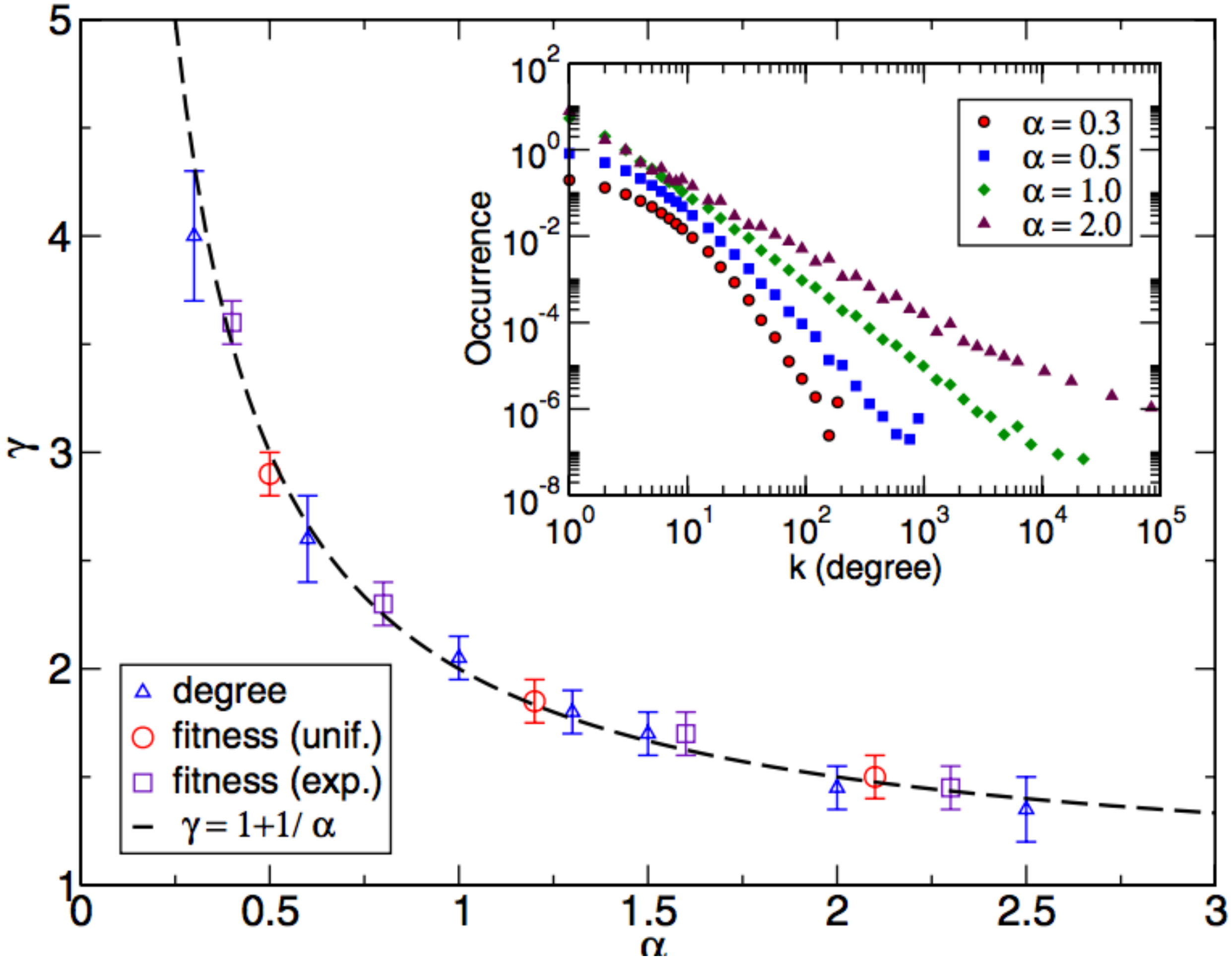
- *Example: R_s* denotes denotes the age-ranking of the nodes of a growing network.

$$P(k) \sim k^{-\left(1+\frac{1}{\alpha}\right)}$$

- Similar behavior will occur if, for instance, we consider the ranking in terms of the in-degree of a node (think for instance on the WWW).
- We get the same type of scaling if we assign a fitness value taken to each node from a given distribution. If we rank the nodes based on this, we get the same behavior.

Scale-free growth by ranking (simulations)

Fortunato et al. PRL (2006)



$$P(k) \sim k^{-\left(1+\frac{1}{\alpha}\right)}$$

Scale-free growth by ranking

Fortunato et al. PRL (2006)

- *Example: R_s* denotes denotes the age-ranking of the nodes of a growing network.

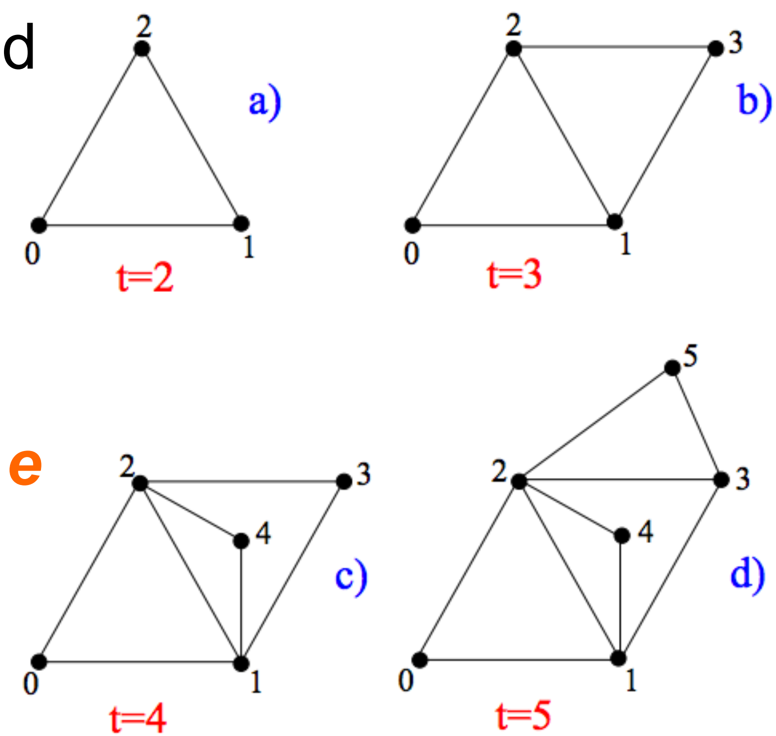
Challenge : Imagine a growth model Page Rank. Would you get to a power-law degree distribution?

- Similar behavior will occur if, for instance, we consider the ranking in terms of the in-degree of a node (think for instance on the WWW).
- A similar pattern will occur if we assign a fitness value taken to each node from a given distribution. If we rank the nodes based on this, we get the same behavior.

DMS minimal model or link-selection model

Dorogovtsev, ***M***endes, ***S***amukhin Phys. Rev. E 63, 062101 (2001)

- ***Growth***: At each time step we add a new node to the network.
- ***Link selection***: each new node selects a link ***e*** at random, and connects itself to the two ends of ***e***



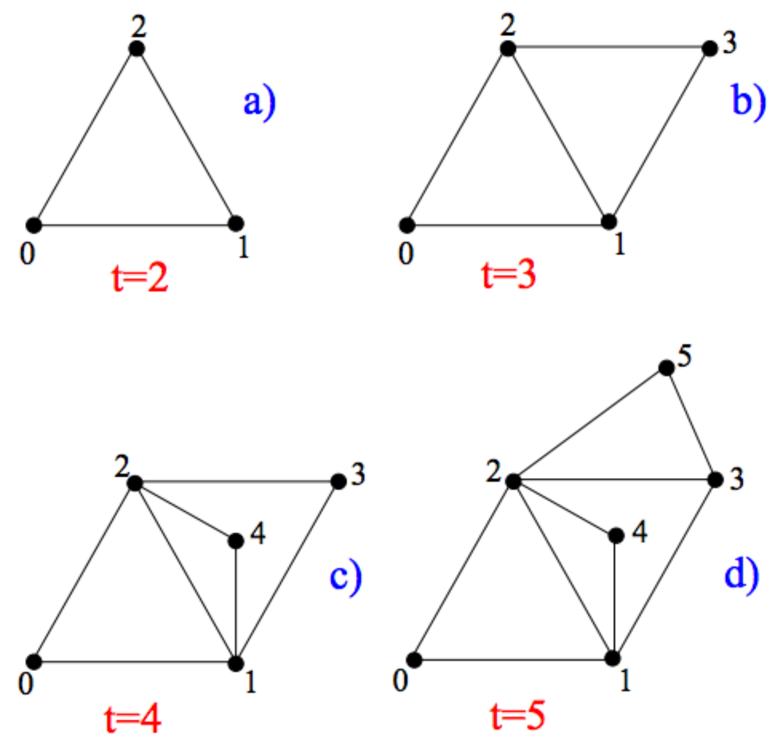
Note: you may also assume that it connects to only one of the two ends of ***e***.

DMS minimal model: degree distribution

Dorogovtsev, **M**endes, **S**amukhin Phys. Rev. E 63, 062101 (2001)

What's the probability that a node i, of degree k_i, gets a link from the new node?

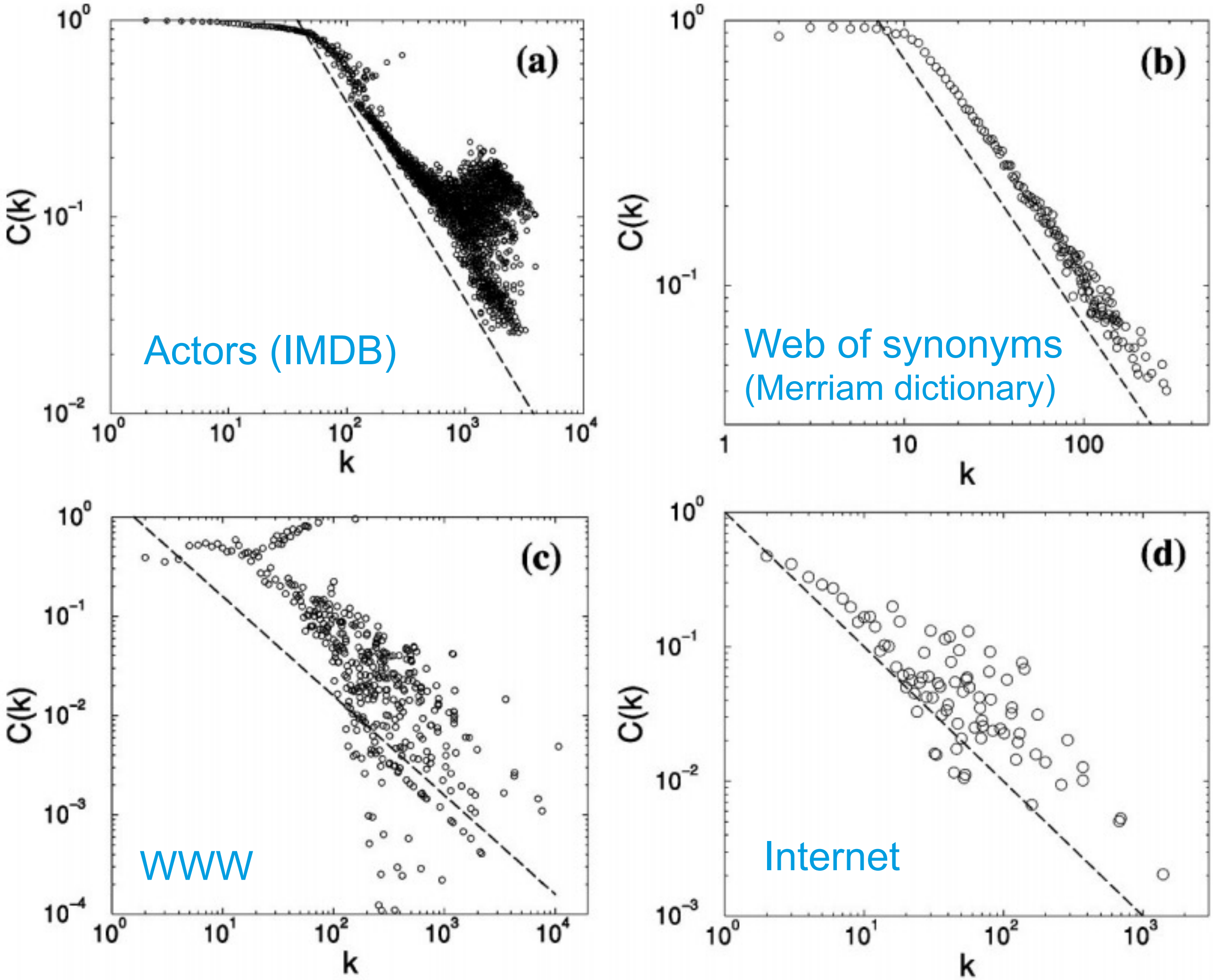
$$\Pi_i = \frac{k_i}{E} = \frac{k_i}{2t-1} \sim \frac{k_i}{\sum_j k_j}$$



i.e., linear preferential attachment, as the Barabási-Albert model ☺

$$P(k) \sim k^{-3}$$

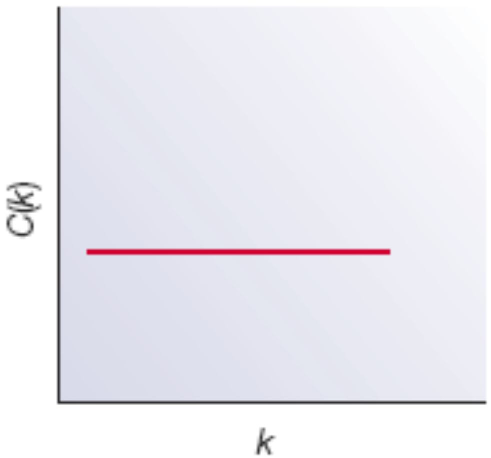
Scaling of clustering coefficient



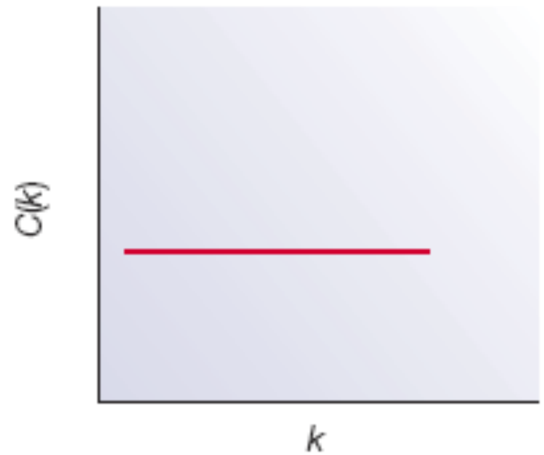
$$C(k) \sim k^{-\beta}$$

DMS Minimal model: Clustering coeff. distribution

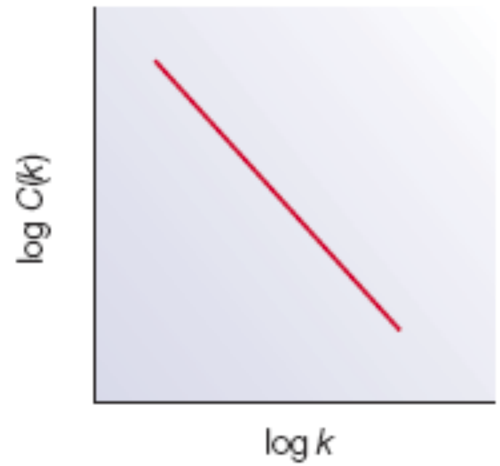
Random Networks

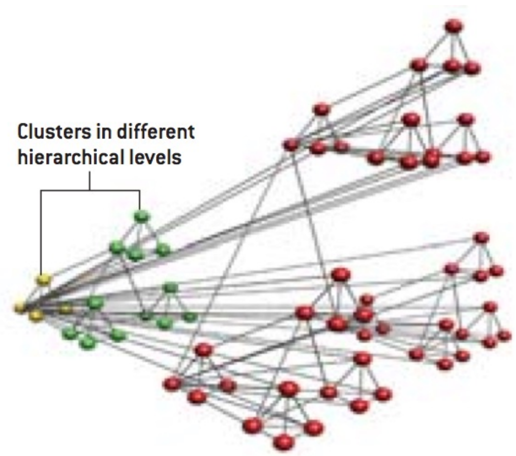


Barabási-Albert model

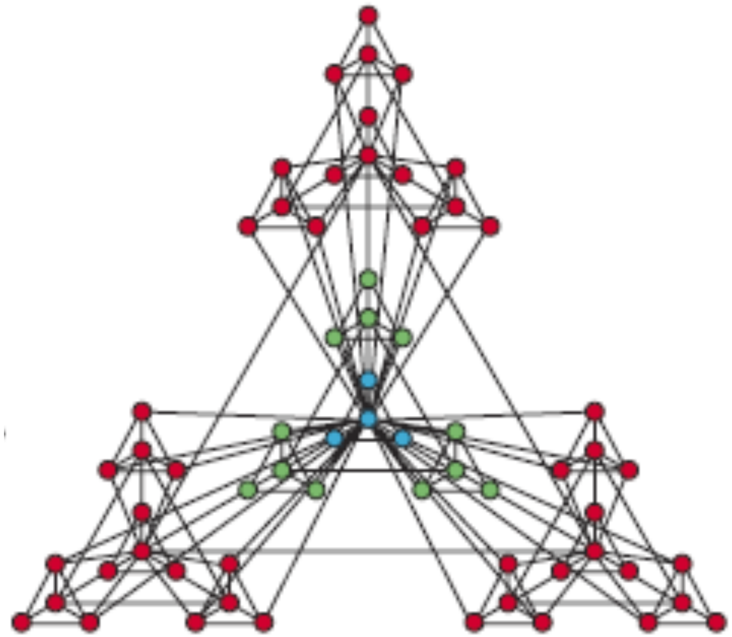


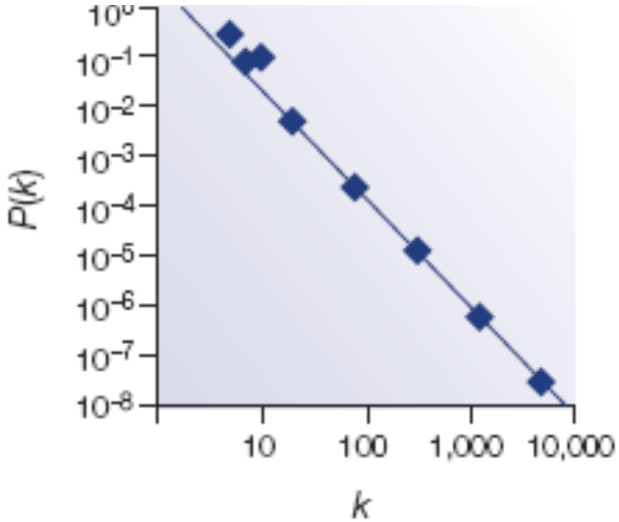
DMS Minimal model

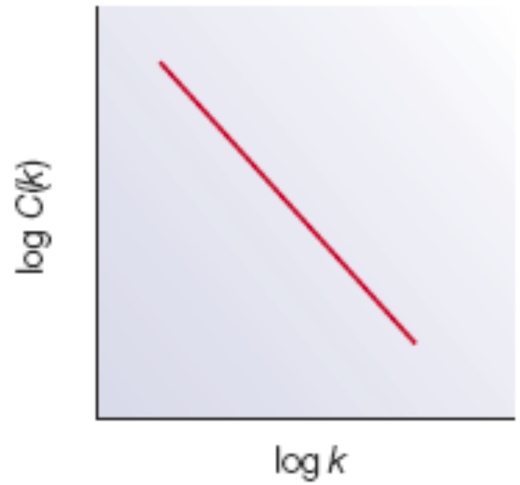




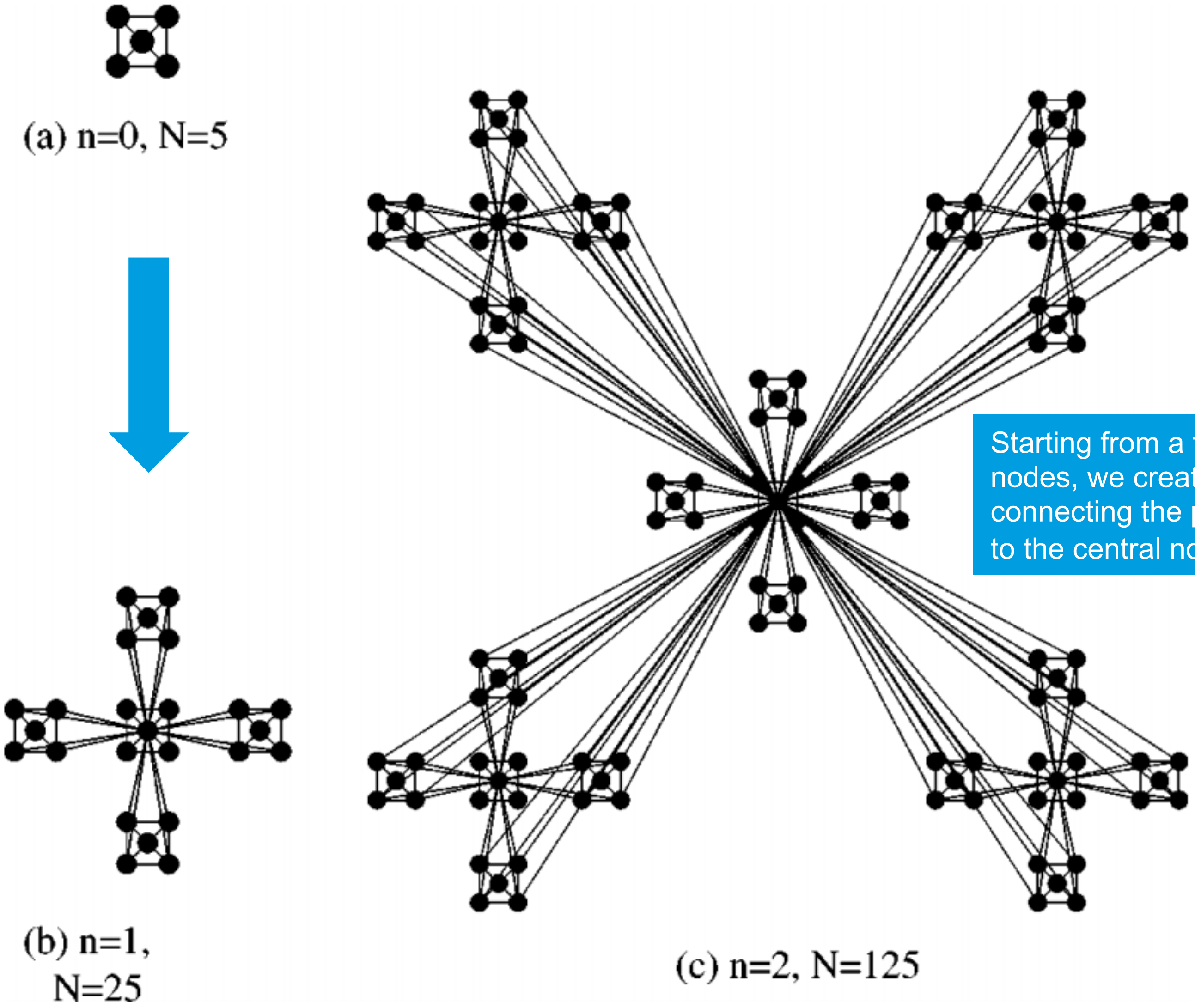
Hierarchical networks





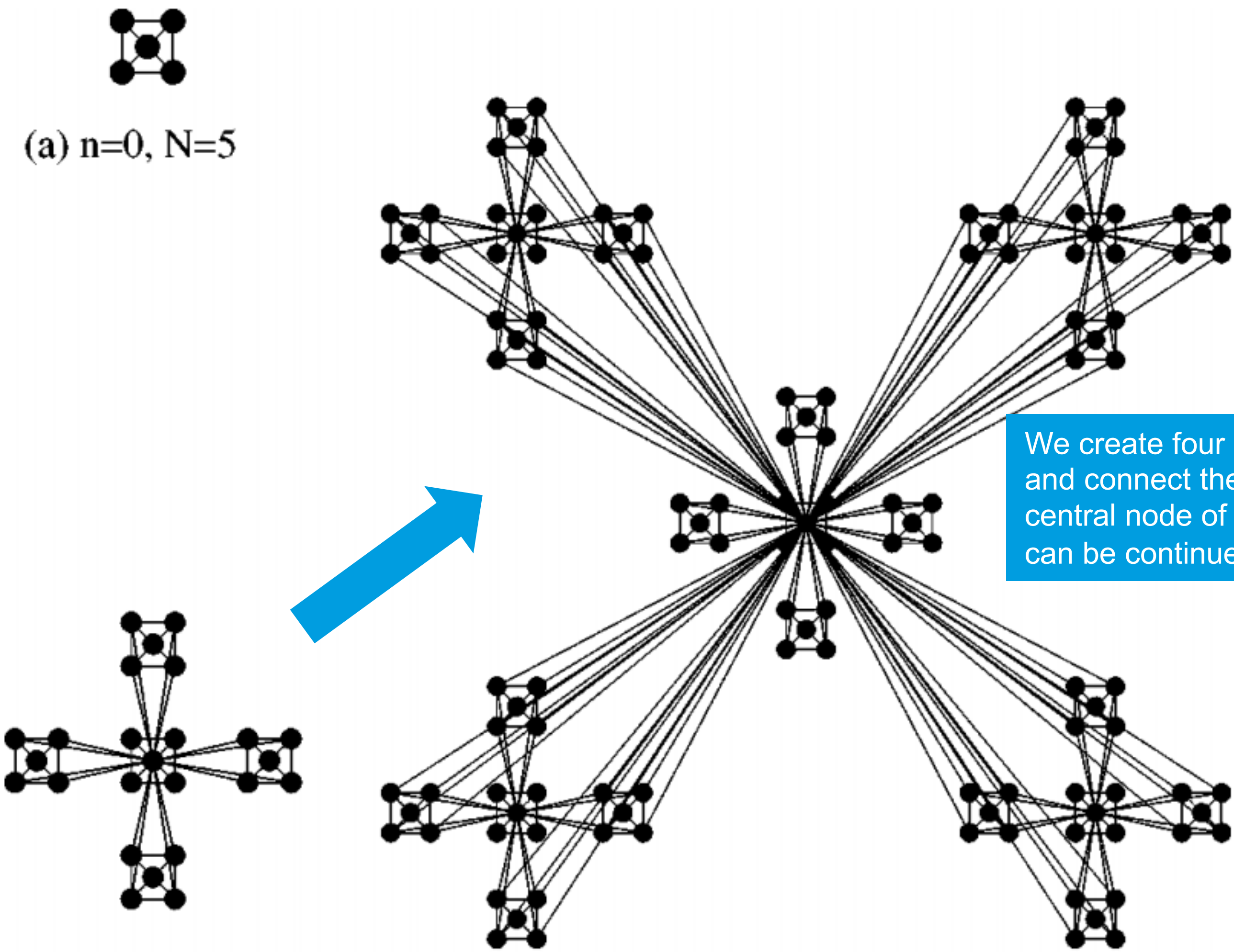


Deterministic Models of Hierarchical Networks



Starting from a fully connected cluster of five nodes, we create four identical replicas, connecting the peripheral nodes of each cluster to the central node of the original cluster.

Deterministic Models of Hierarchical Networks



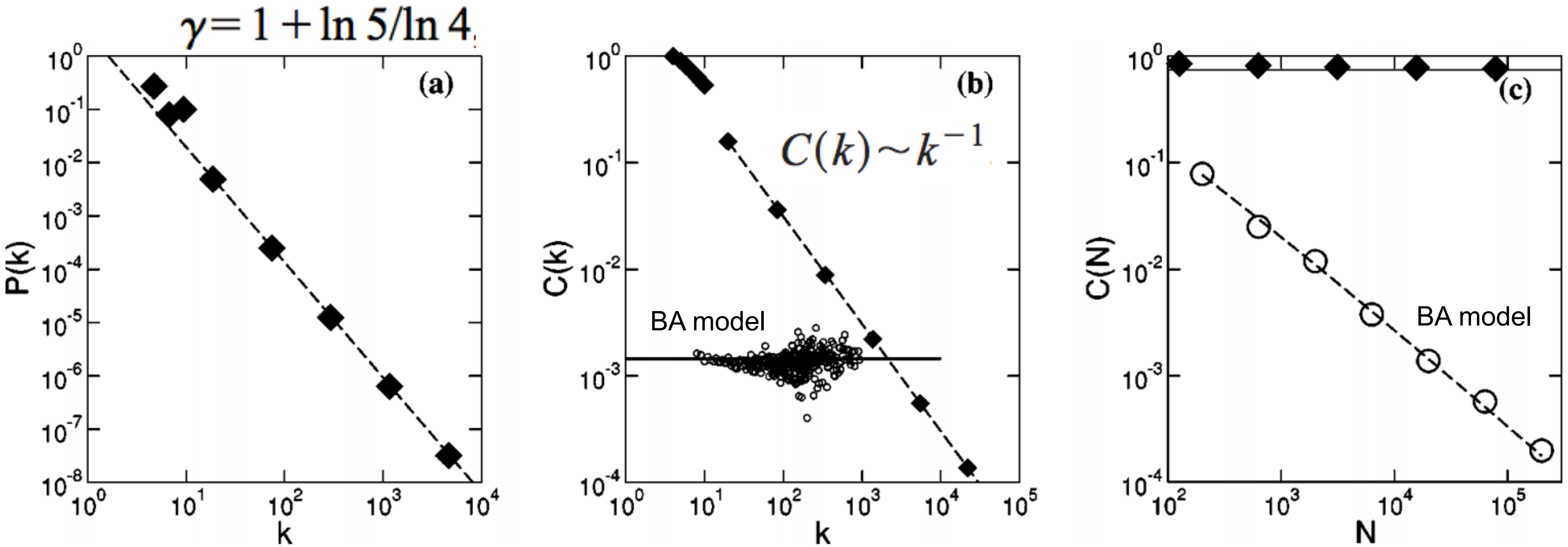
(a) n=0, N=5

(b) n=1, N=25

(c) n=2, N=125

We create four replicas of the obtained cluster, and connect the peripheral nodes again, to the central node of the original module. This process can be continued indefinitely

Hierarchical networks



Ravasz et al., Phys Rev E 67, 026112 (2003)

Duplication/copying models

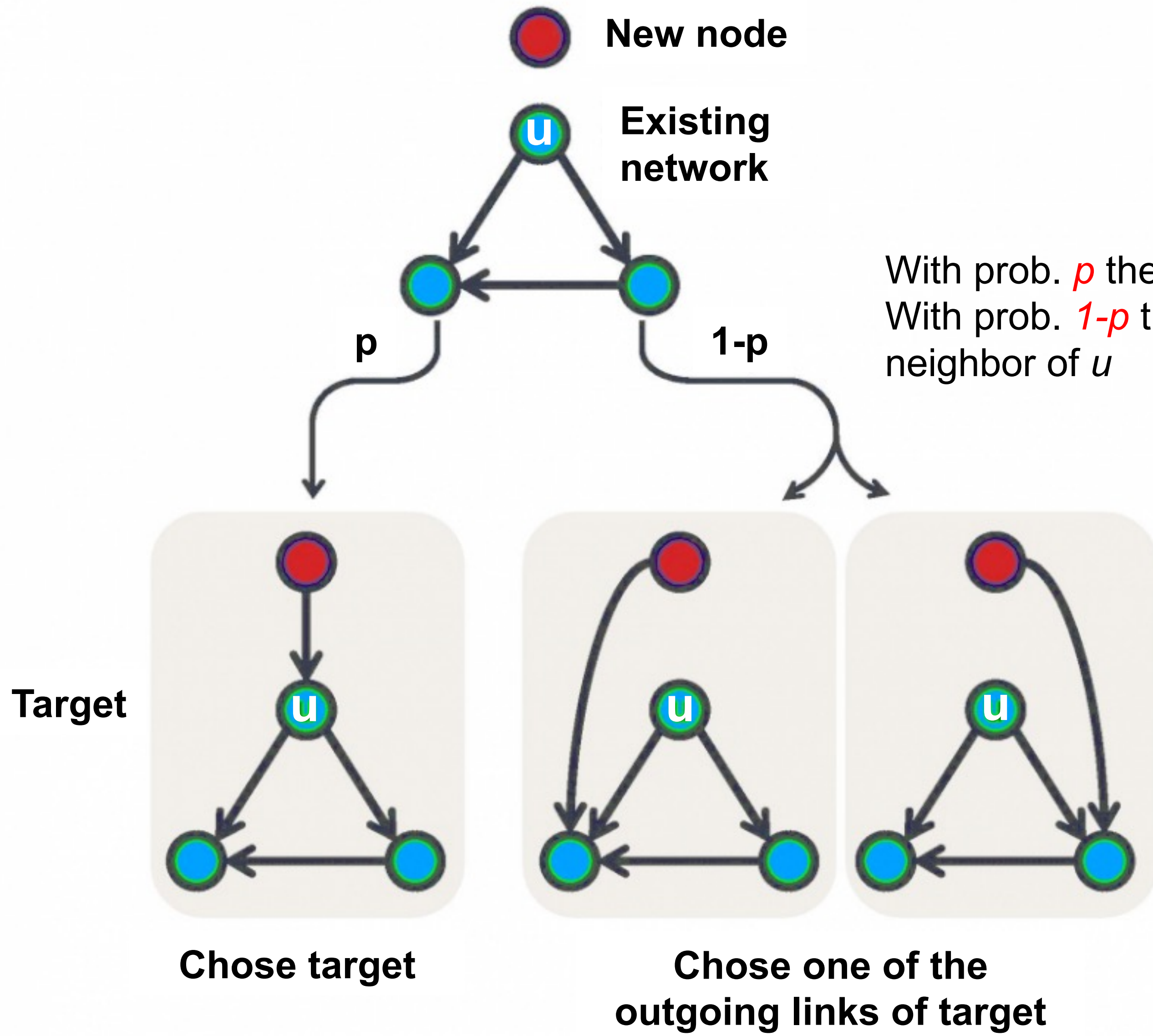
Wagner, Kleinberg, & others

Analytic results for copying models have been given by Chung et al.

Examples

- The authors of a new webpage tend to borrow links from other webpages on related topics.
- Similar arguments may be used for social networks.
- Genes that code for proteins duplicate. Since the proteins coded for by each copy are the same, their interactions are also the same, i.e., the new gene copies its edges in the interaction network from the old.
- Similar arguments have been used for Metabolic networks and other.
- Etc.

Copying model (simplest version)

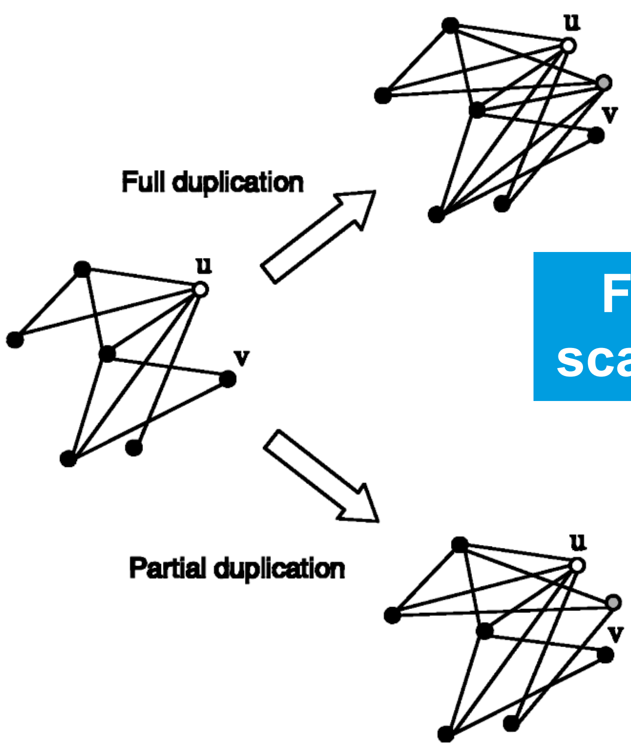


With prob. *p* the new node links to *u*
With prob. *1-p* the new node links to a neighbor of *u*

Other version: Partial duplication model

(Vazquez, Flammini, Maritan and Vespignani, 2003)

- At every time-step a randomly chosen vertex is duplicated at random creating a new node *s*.
- Each of *s* links is either kept with probability ***1−α*** or it is rewired (or removed) with probability ***α*** (equivalent to a mutation).



For low "mutation" probabilities (α), scale-free behavior ($\langle k^2\rangle \rightarrow \infty$) emerges.

Conclusion: Understanding topological variety

Models with Pref. Attachment

Non-linear Pref. Attachment

$$\Pi_i = \frac{k_i^\alpha}{\sum_j k_j^\alpha}$$

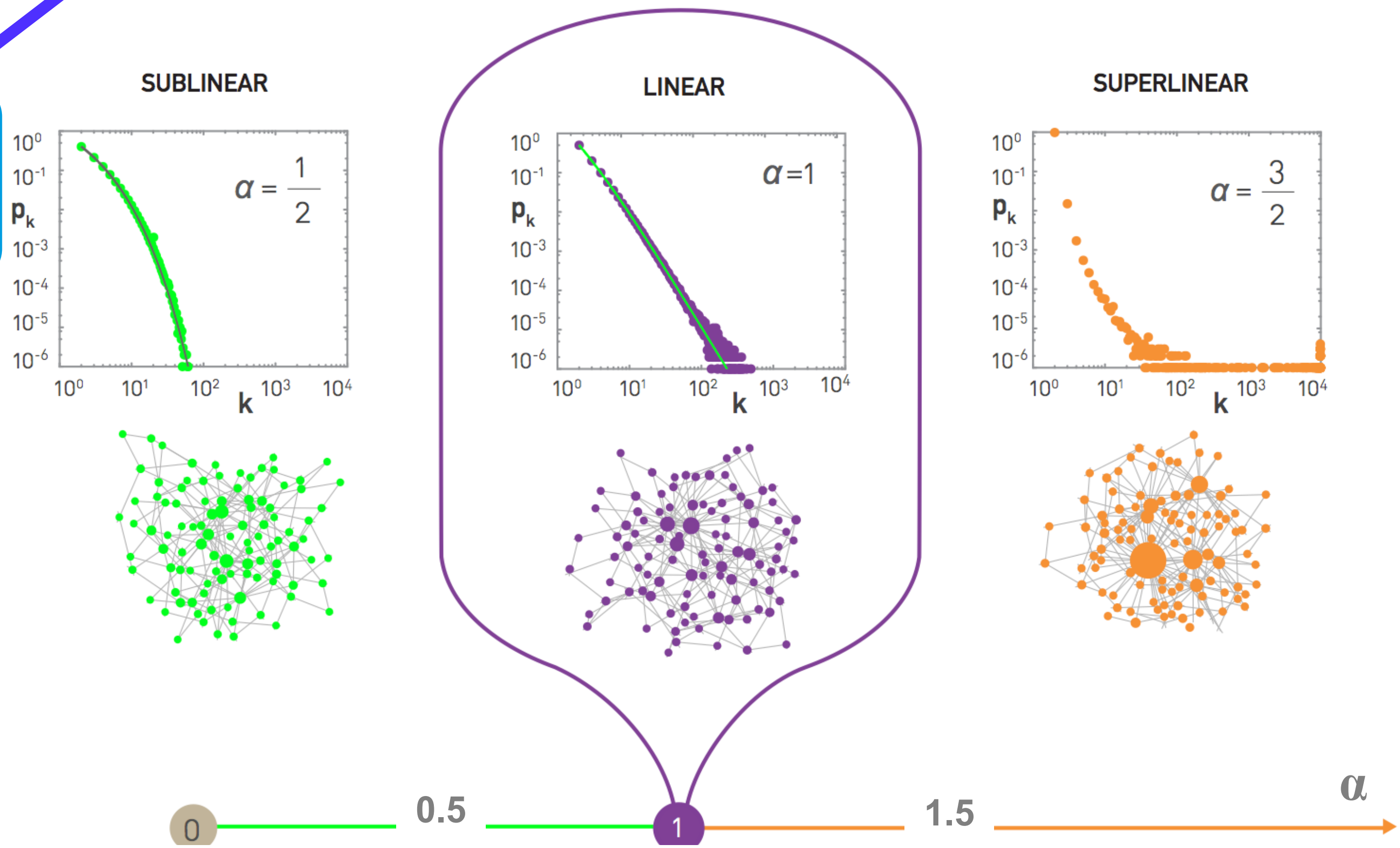
α<1 ➔ exponential dist

α=1 ➔ power-law

α>1 ➔ winners-take all

Barabási-Albert Model

Π~k ➔ γ=3



Conclusion: Understanding topological variety

Models with Pref. Attachment

Barabási-Albert Model
$\Pi \sim k \rightarrow \gamma=3$

Non-linear Pref. Attachment

$$\Pi_i = \frac{k_i^{\alpha}}{\sum_j k_j^{\alpha}}$$

- $\alpha<1 \rightarrow$ exponential dist
- $\alpha=1 \rightarrow$ power-law
- $\alpha>1 \rightarrow$ winners-take all

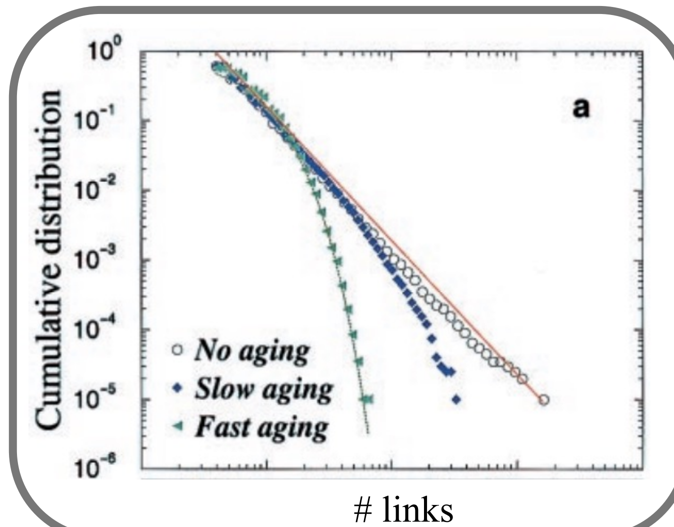
Initial attractiveness
$\Pi \sim A+k \rightarrow \gamma=3+A/m$

Fitness models

$$\Pi_i = \frac{\eta_i k_i}{\sum_j \eta_j k_j}$$

Ageing models

$$(t_i - t)^{-\nu}$$



Conclusion: Understanding topological variety

Models with Pref. Attachment

Barabási-Albert Model
$\Pi \sim k \rightarrow \gamma=3$

Non-linear Pref. Attachment

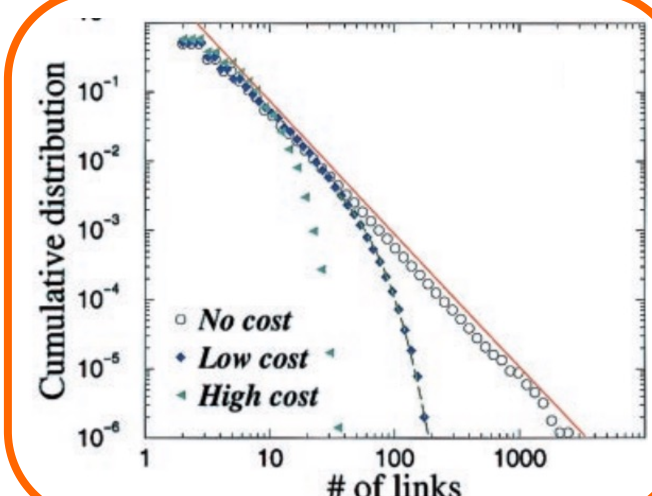
$$\Pi_i = \frac{k_i^\alpha}{\sum_j k_j^\alpha}$$

$\alpha<1 \rightarrow$ exponential dist
$\alpha=1 \rightarrow$ power-law
$\alpha>1 \rightarrow$ winners-take all

Initial attractiveness
$\Pi \sim A+k \rightarrow \gamma=3+A/m$

Growth w/costs

$$k > k_{\max}$$



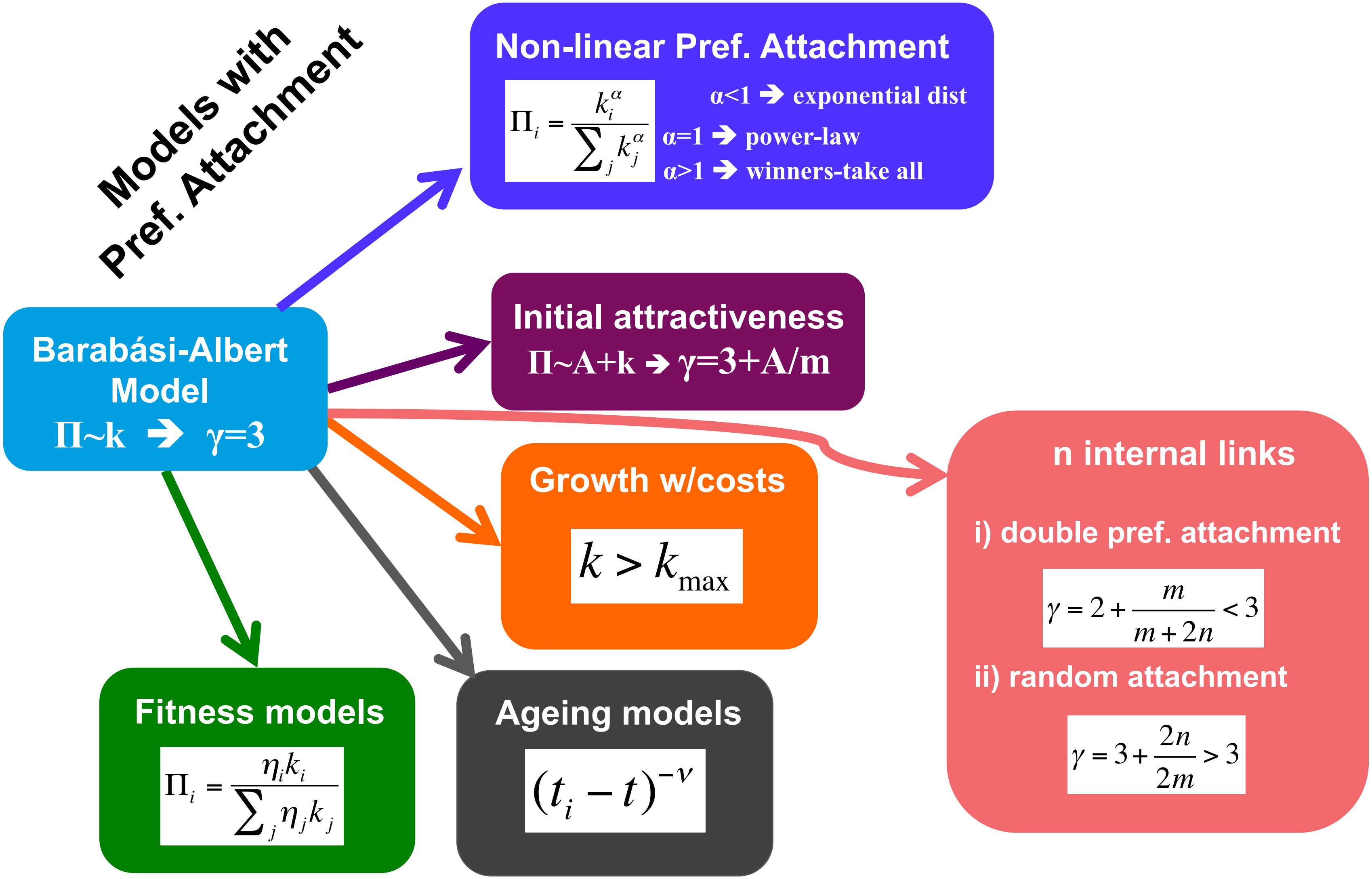
Fitness models

$$\Pi_i = \frac{\eta_i k_i}{\sum_j \eta_j k_j}$$

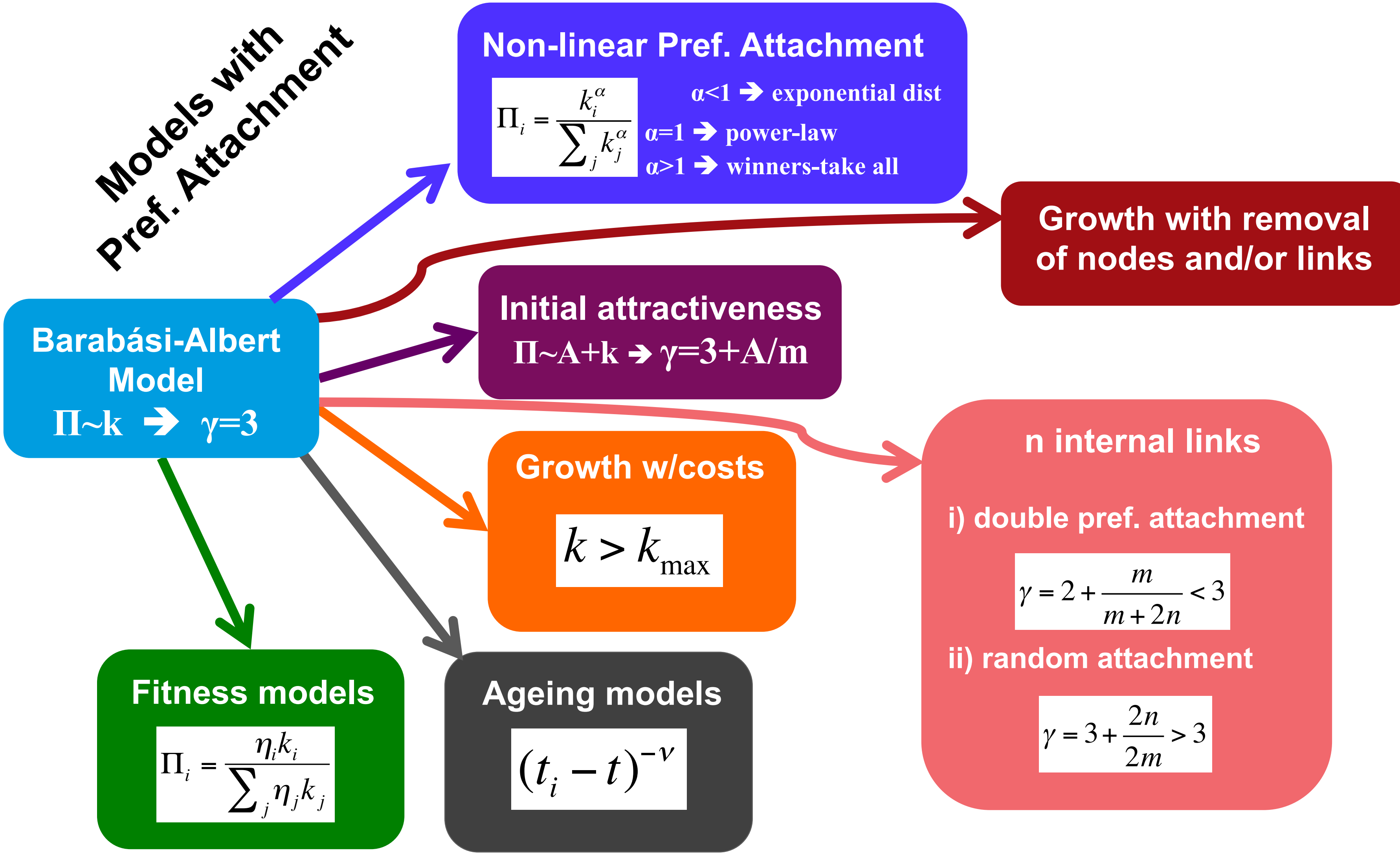
Ageing models

$$(t_i - t)^{-\nu}$$

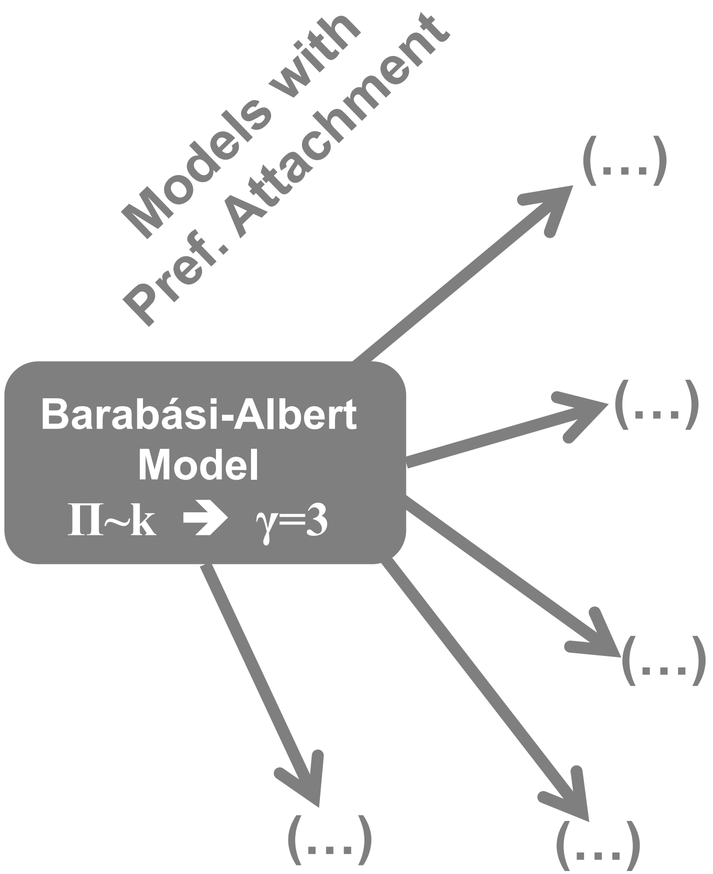
Conclusion: Understanding topological variety



Conclusion: Understanding topological variety



Conclusion: Understanding topological variety



Models without Pref. Attachment?

Conclusion: Understanding topological variety

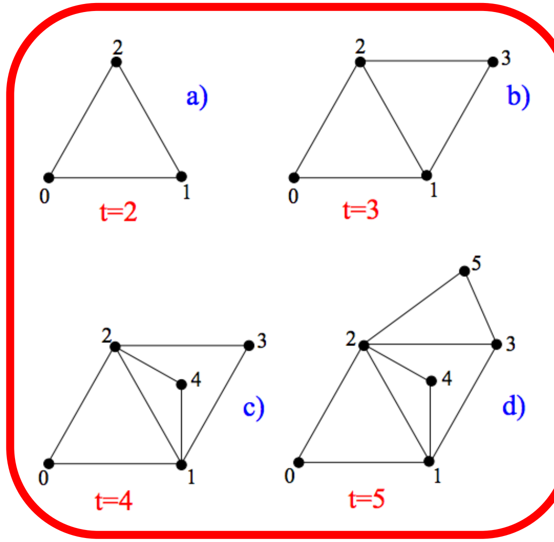
Models with Pref. Attachment

Barabási-Albert Model
$\Pi \sim k \rightarrow \gamma=3$

(...)
(...)
(...)
(...)
(...)

Models without Pref. Attachment

Minimal or link-selection model
$\Pi \sim k \rightarrow \gamma=3$



Conclusion: Understanding topological variety

Models with Pref. Attachment

Barabási-Albert Model

$\Pi \sim k \rightarrow \gamma = 3$

(…)

(…)

(…)

Models without Pref. Attachment

Ranking models

$$\Pi_s = \frac{R_s^{-\alpha}}{\sum_j R_j^{-\alpha}}$$

$$P(k) \sim k^{-\left(1+\frac{1}{\alpha}\right)}$$

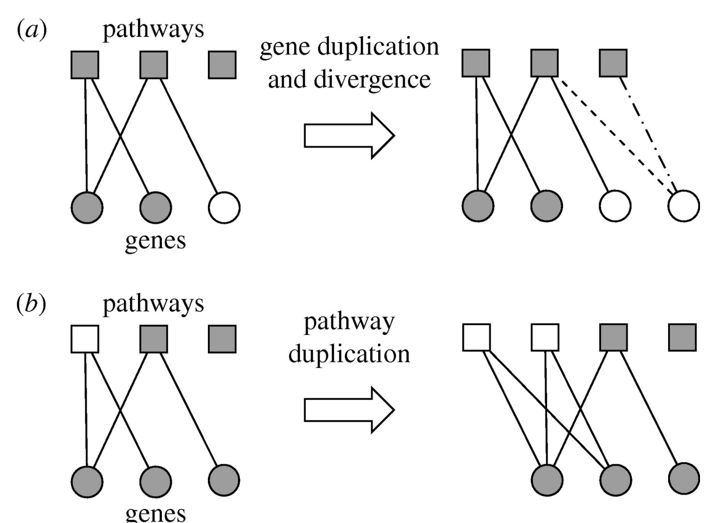
Minimal or link-selection model

$\Pi \sim k \rightarrow \gamma = 3$

Duplication models

$$\Pi_s = \frac{R_s^{-\alpha}}{\sum_j R_j^{-\alpha}}$$

$$P(k) \sim k^{-\left(1+\frac{1}{\alpha}\right)}$$



Conclusion: Understanding topological variety

Models with Pref. Attachment

Barabási-Albert Model
$\Pi \sim k$ ➔ $\gamma=3$

(...)

(...)

(...)

(...)

(...)

Models without Pref. Attachment

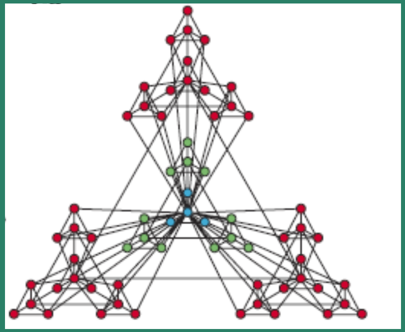
Ranking models

$$\Pi_s = \frac{R_s^{-\alpha}}{\sum_j R_j^{-\alpha}}$$

$$P(k) \sim k^{-\left(1+\frac{1}{\alpha}\right)}$$

Minimal or link-selection model
$\Pi \sim k$ ➔ $\gamma=3$

Hierarchical nets



Duplication models

$$\Pi_s = \frac{R_s^{-\alpha}}{\sum_j R_j^{-\alpha}}$$

$$P(k) \sim k^{-\left(1+\frac{1}{\alpha}\right)}$$

Conclusion: Understanding topological variety

Power-laws: BA-model, minimal model, etc, etc.

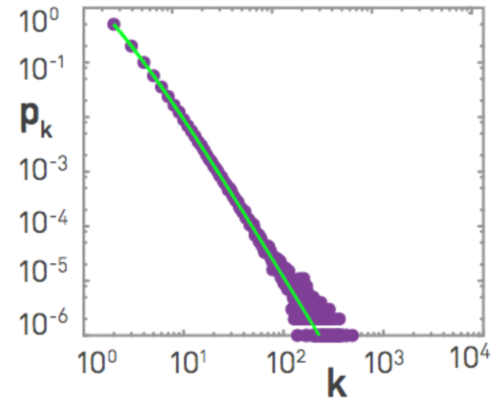
Stretched exponentials: If preferential attachment is sublinear you get a stretched exponential.

Fitness-induced corrections: Ranking models, Fitness models, initial attractiveness model.

Small-degree saturations: Initial attractiveness adds a random component to preferential attachment, particularly for low degrees.

High degree cutoffs: Node and link removal, costs and cutoffs, and node ageing, can induce high-degree cutoffs.

Hierarchical structure & power-law dep. in clustering: Minimal/link-selection model, duplication models and hierarchical networks model.



Conclusion: Understanding topological variety

Power-laws: BA-model, minimal model, etc, etc.

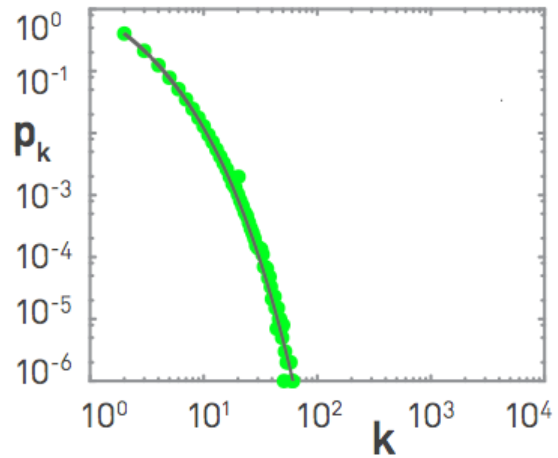
Stretched exponentials: If preferential attachment is sublinear you get a stretched exponential.

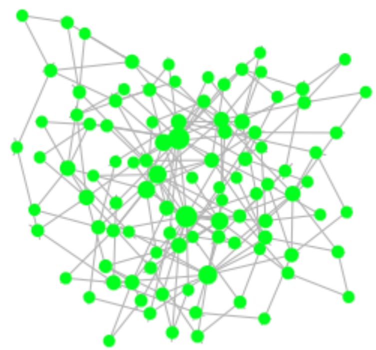
Fitness-induced corrections: Ranking models, Fitness models, initial attractiveness model.

Small-degree saturations: Initial attractiveness adds a random component to preferential attachment, particularly for low degrees.

High degree cutoffs: Node and link removal, costs and cutoffs, and node ageing, can induce high-degree cutoffs.

Hierarchical structure & power-law dep. in clustering: Minimal/link-selection model, duplication models and hierarchical networks model.





Conclusion: Understanding topological variety

Power-laws: BA-model, minimal model, etc, etc.

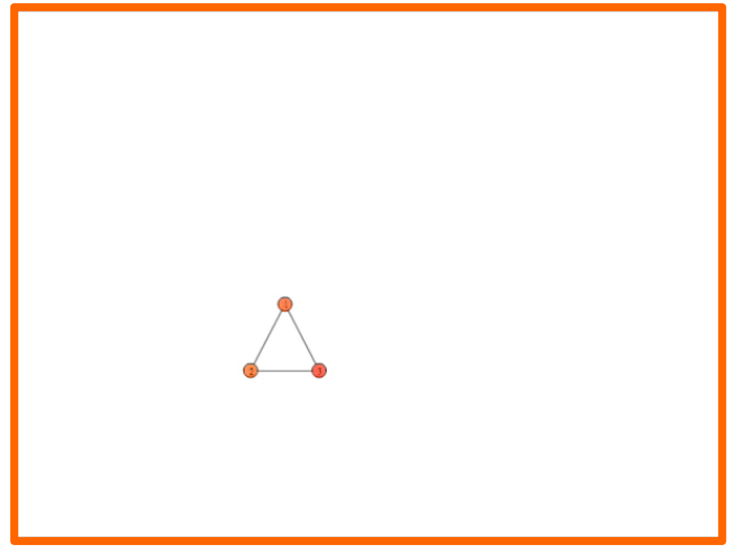
Stretched exponentials: If preferential attachment is sublinear you get a stretched exponential.

Fitness-induced corrections: Ranking models, Fitness models, initial attractiveness model.

Small-degree saturations: Initial attractiveness adds a random component to preferential attachment, particularly for low degrees.

High degree cutoffs: Node and link removal, costs and cutoffs, and node ageing, can induce high-degree cutoffs.

Hierarchical structure & power-law dep. in clustering: Minimal/link-selection model, duplication models and hierarchical networks model.



Conclusion: Understanding topological variety

Power-laws: BA-model, minimal model, etc, etc.

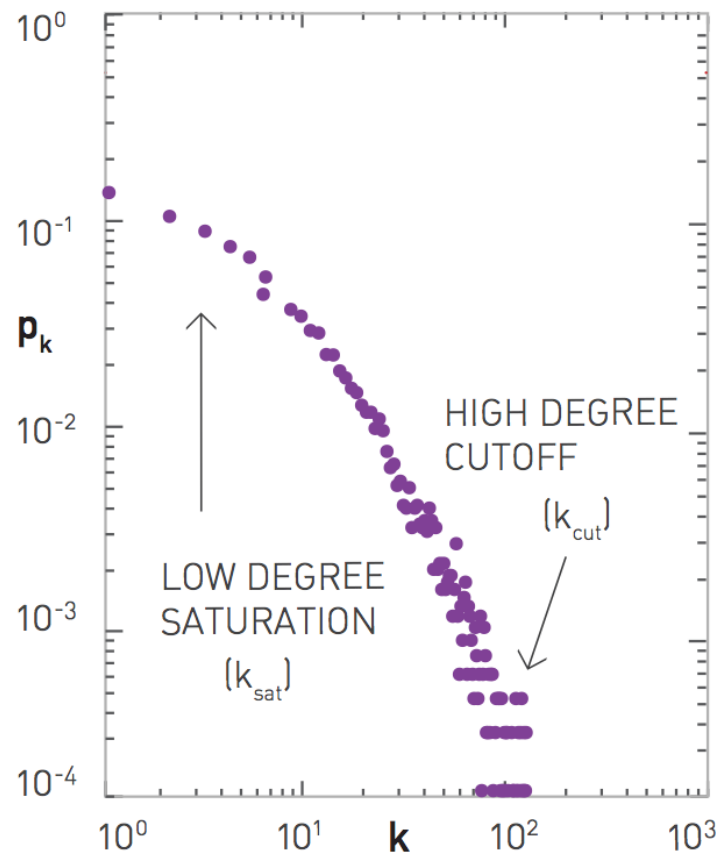
Stretched exponentials: If preferential attachment is sublinear you get a stretched exponential.

Fitness-induced corrections: Ranking models, Fitness models, initial attractiveness model.

Small-degree saturations: Initial attractiveness adds a random component to preferential attachment, particularly for low degrees.

High degree cutoffs: Node and link removal, costs and cutoffs, and node ageing, can induce high-degree cutoffs.

Hierarchical structure & power-law dep. in clustering: Minimal/link-selection model, duplication models and hierarchical networks model.



Conclusion: Understanding topological variety

Power-laws: BA-model, minimal model, etc, etc.

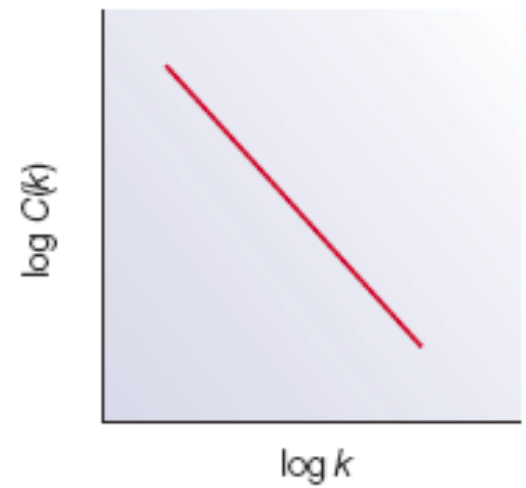
Stretched exponentials: If preferential attachment is sublinear you get a stretched exponential.

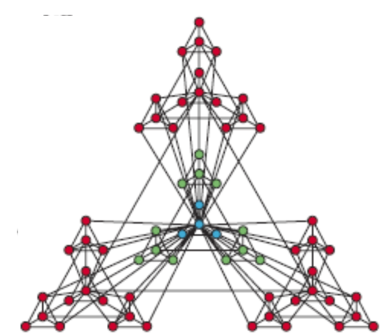
Fitness-induced corrections: Ranking models, Fitness models, initial attractiveness model.

Small-degree saturations: Initial attractiveness adds a random component to preferential attachment, particularly for low degrees.

High degree cutoffs: Node and link removal, costs and cutoffs, and node ageing, can induce high-degree cutoffs.

Hierarchical structure & power-law dep. in clustering: Minimal/link-selection model, duplication models and hierarchical networks model.





Classes of models in network science

- ***Static & generative models.*** ER model, Watts-Strogratz model, Configuration Model, etc.

- ***Evolving network models.*** BA model, Initial attractiveness model, fitness model, internal links model, node deletion model, accelerated model, aging model, costs model, minimal model, ranking model, duplication model, hierarchical networks model, etc.

Which models or principles should I consider to justify each case:

1) Network with a power-law degree distribution with $\gamma=3$.
2) Network with a power-law degree distribution with $\gamma=3$ with a significant exponential cutoff for large $k \sim k_{max}$.
3) Network with a power-law degree distribution with $\gamma=3$ and large clustering coefficient.
4) Power-law degree distribution with $\gamma<3.0$.
5) Power-law degree distribution with $\gamma>3.0$.
6) A network with exponential degree distribution.
7) A power-law degree distribution with saturation for low k.

Solutions:

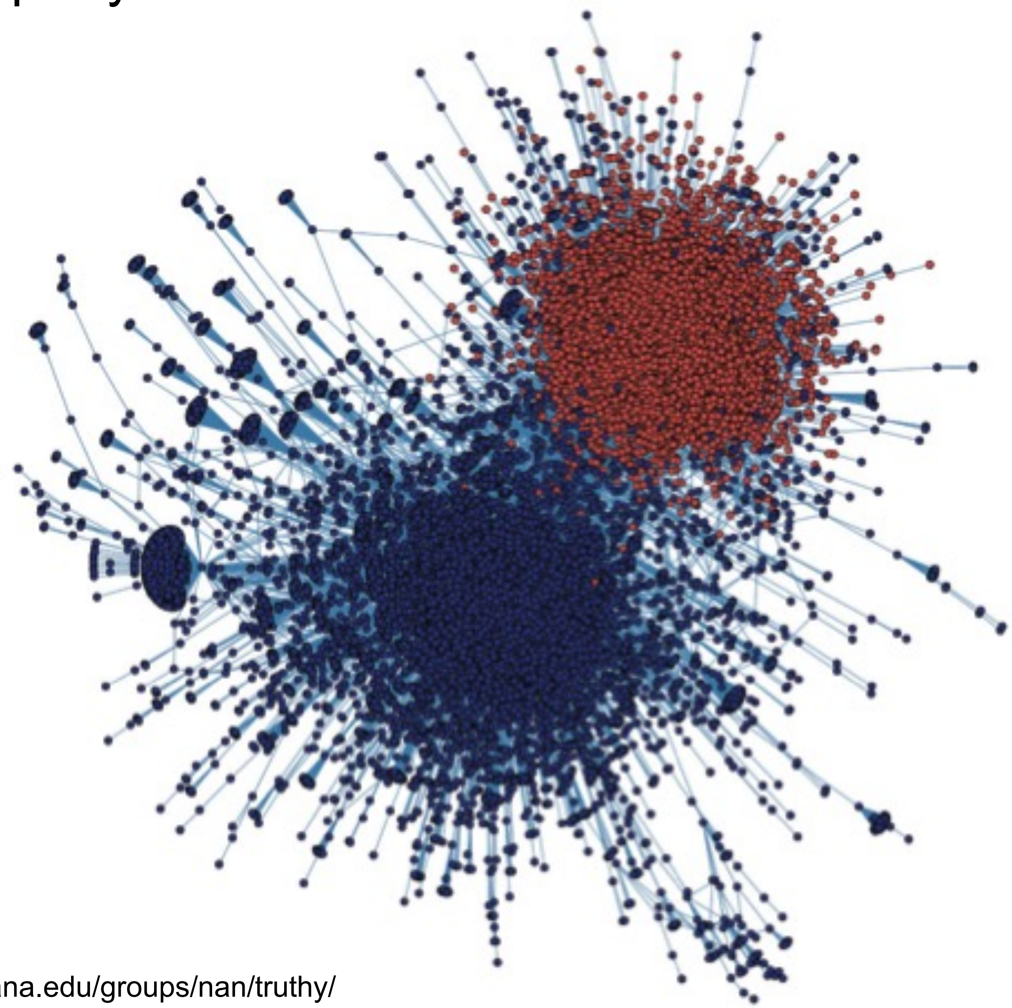
1) BA model
2) BA model with costs or cutoffs
3) DMS minimal model
4) BA model with internal links chosen through preferential attachment
5) BA model with internal links chosen randomly (preferential attachment with initial attractiveness will also work)

Next step? Beyond degree distributions

Assortativity in complex networks

Political Homophily in Twitter

Assortativity = homophiliy



Network of Retweets
Political retweet network:
red: right
blue: left

Conover, Ratkiewicz, et al. 2011
Data available from: http://cnets.indiana.edu/groups/nan/truthy/

Mixing patterns

- ***Assortative mixing***: “likes link with likes”
- ***Disassortative mixing***: “likes link with dislikes”

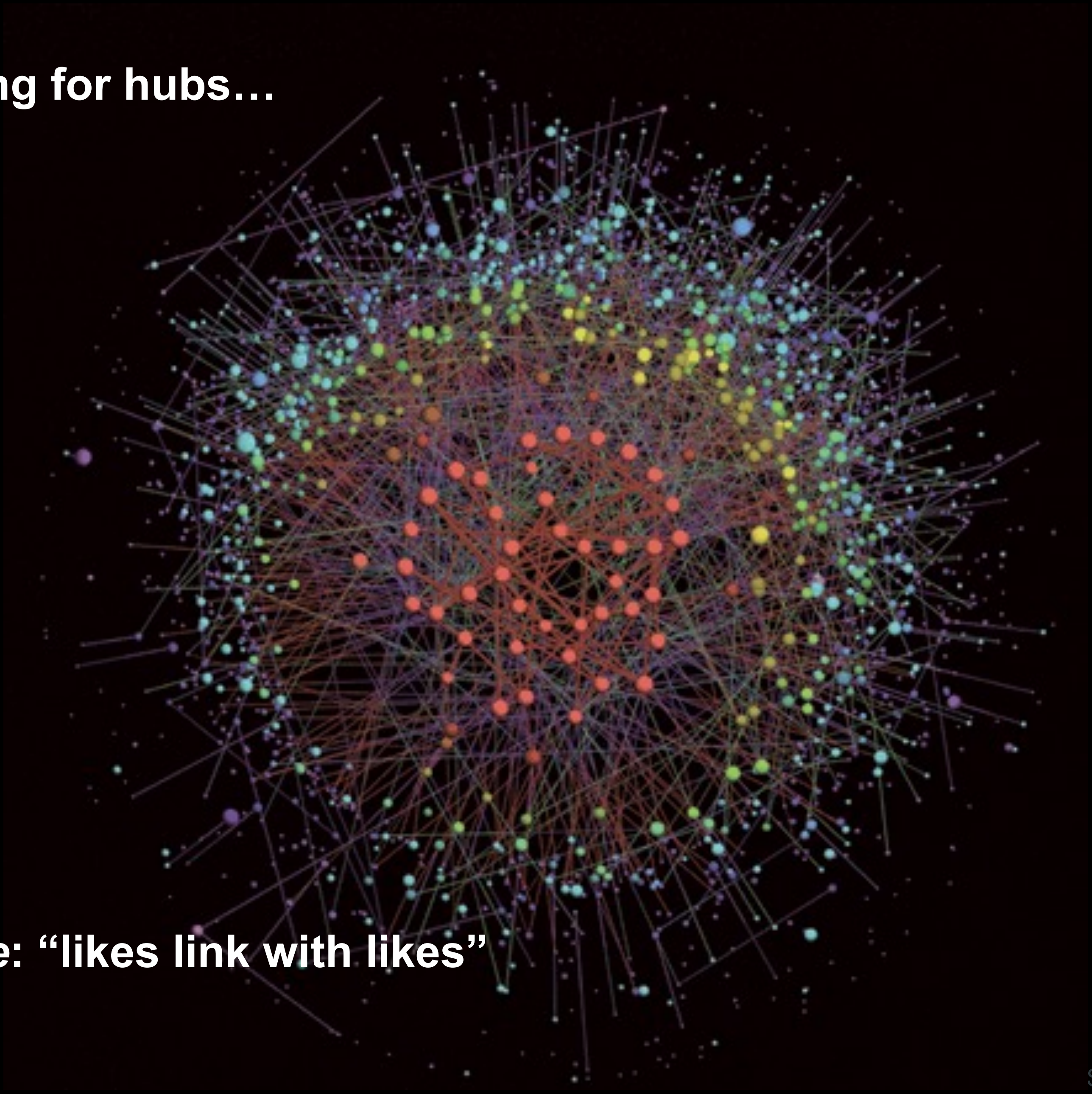
Nodes can be (dis)assortative on various attributes (age, sex, geography), topological attributes $\boxed{\text{(degree,}}$ clustering, etc.).

Examples:

Assortative mixing (in social nets): political beliefs, race, obesity, etc.

Disassortative mixing: food webs (predator/prey), dating networks (gender), economic networks (producers / consumers)

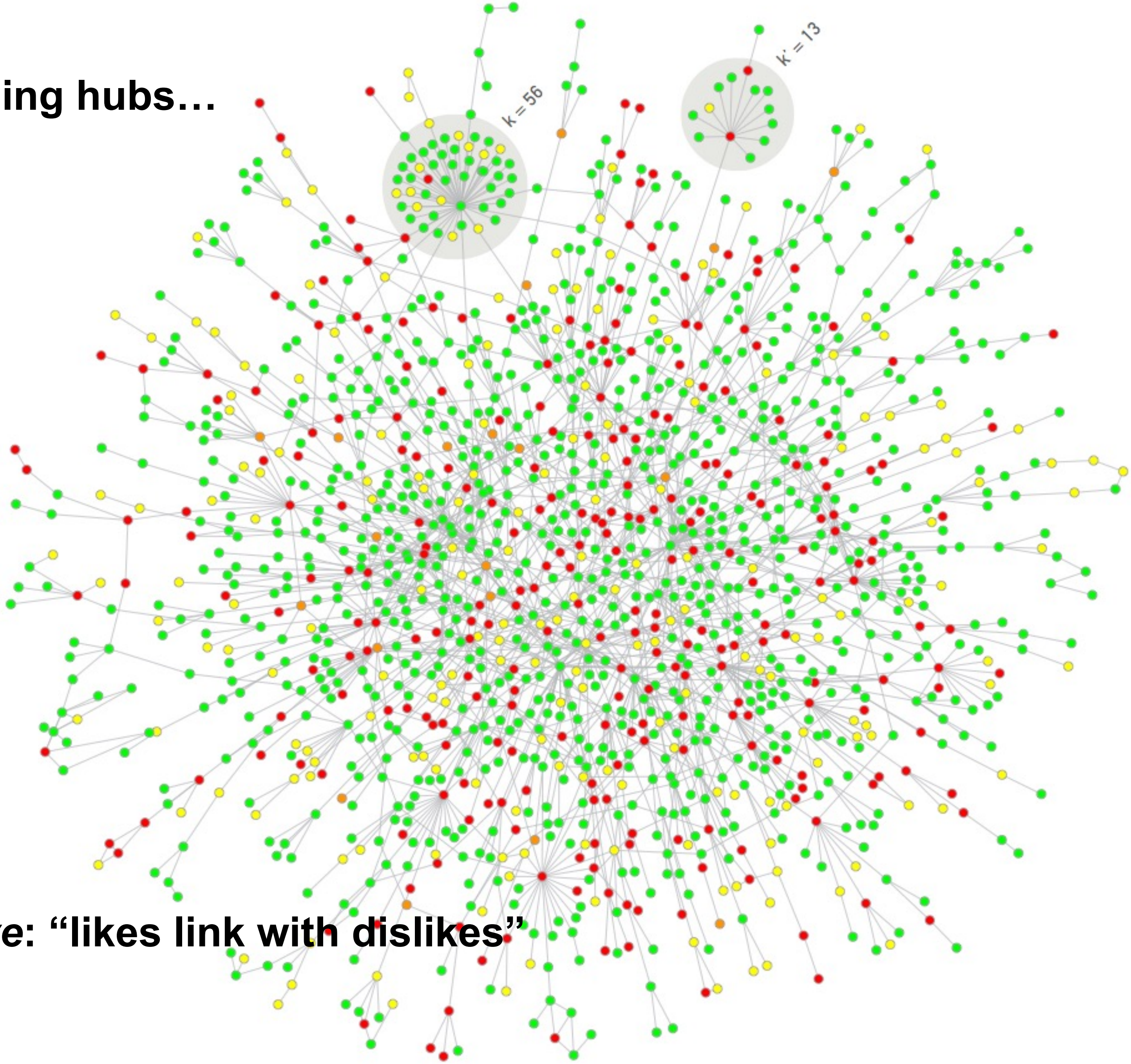
Hubs looking for hubs…



***Assortative*: “likes link with likes”**

Science, 2023/24

Hubs avoiding hubs…



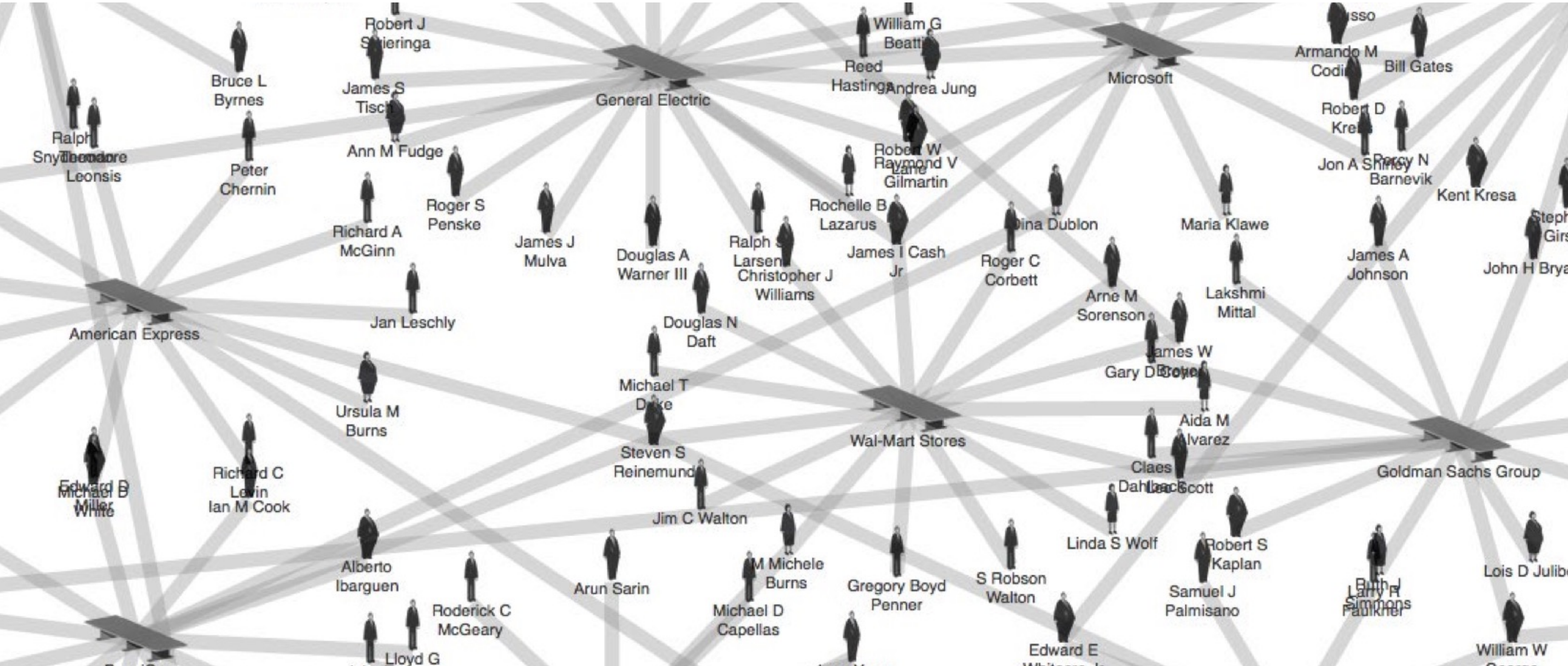
***Disassortative*: “likes link with dislikes”**

Degree assortativity

- ***Social networks***: Hubs tend to date each other. If popularity is attractive, this is even clearer among the most popular.

Degree assortativity

- ***Social networks***: Hubs tend to date each other. If popularity is attractive, this is even clearer among the most popular.



Degree assortativity

- ***Social networks***: Hubs tend to date each other. If popularity is attractive, this is even clearer among the most popular.
- ***Biological & technological networks*** (e.g., PPI nets): hubs avoid linking to other hubs, connecting instead to many small degree nodes.

***Assortativity* is a preference for nodes to attach to others that are similar in some way. Though the specific measure of similarity may vary, network theorists often examine assortativity in terms of a node's degree**

We will discuss 3 key measures

- ***Social networks***: Hubs tend to date each other. If
  po
  m
- ***Bi***
  hu
  ma

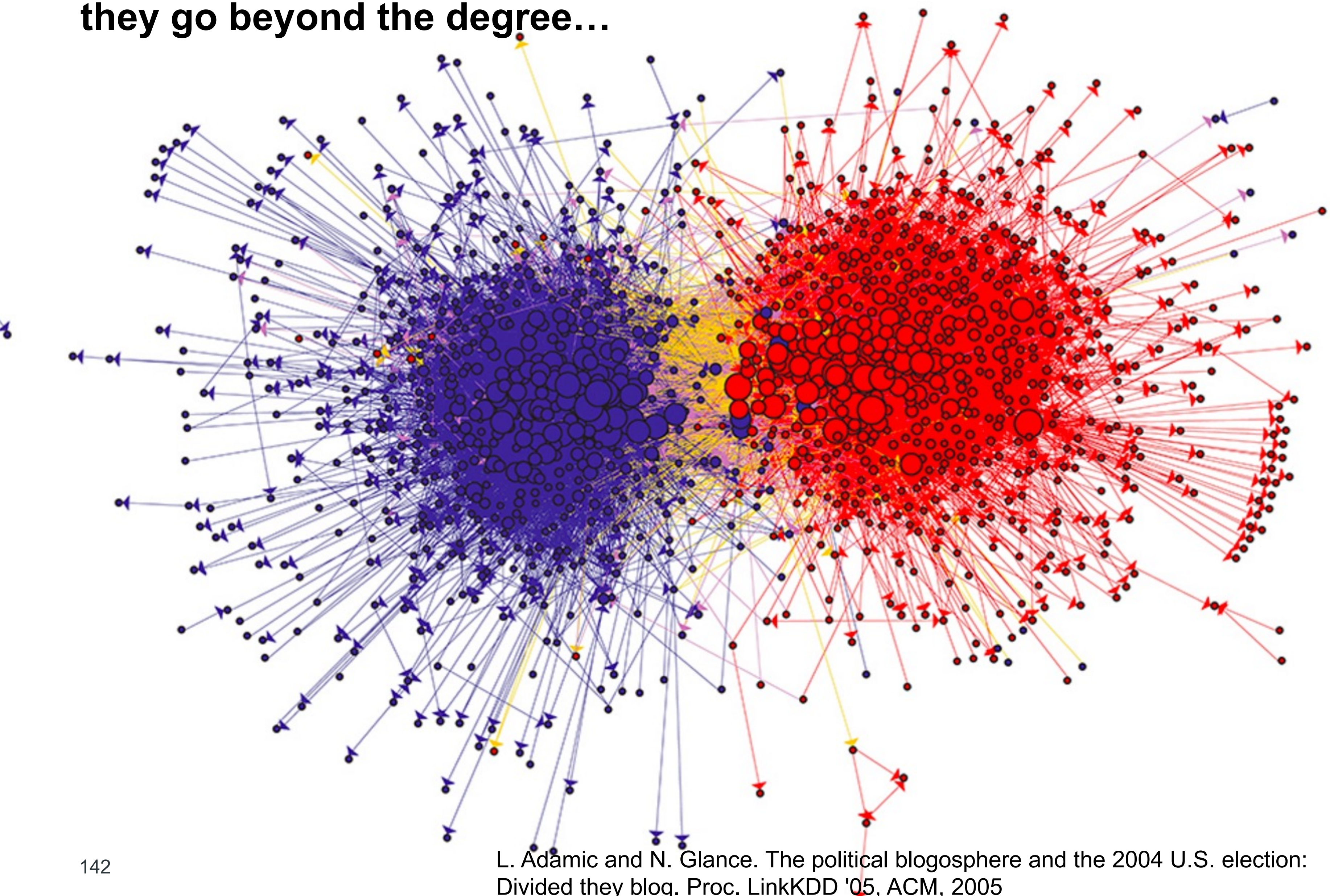
Three assortativity measures:

1. Pearson coefficient
2. Degree correlation matrix
3. Degree correlation function

***Assortativity* is a preference for nodes to attach to others that are similar in some way. Though the specific measure of similarity may vary, network theorists often examine assortativity in terms of a node's degree**

Mixing patterns?

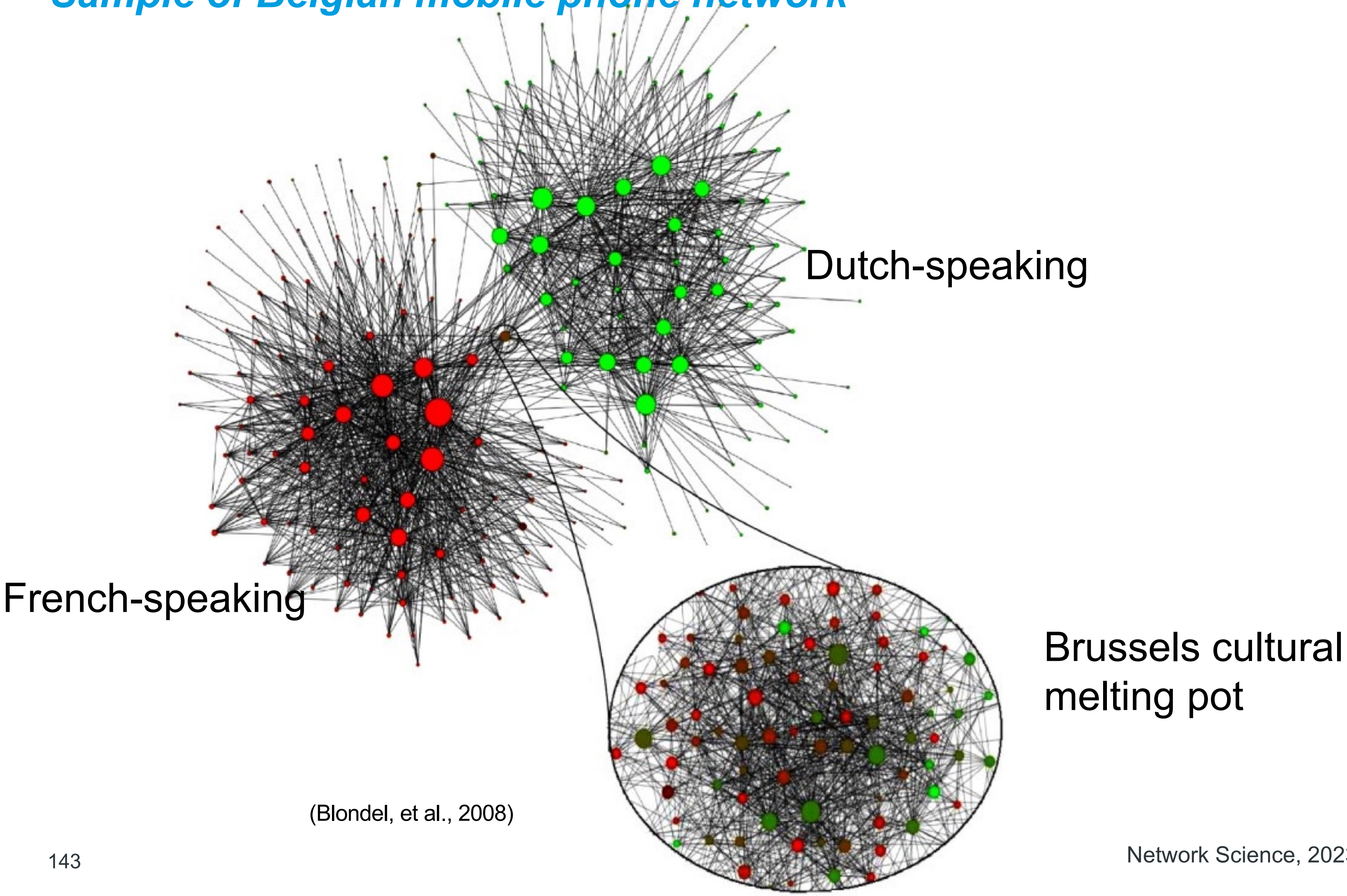
Our understanding of correlations remain incomplete as often they go beyond the degree…



L. Adamic and N. Glance. The political blogosphere and the 2004 U.S. election: Divided they blog. Proc. LinkKDD '05, ACM, 2005

Modules and communities

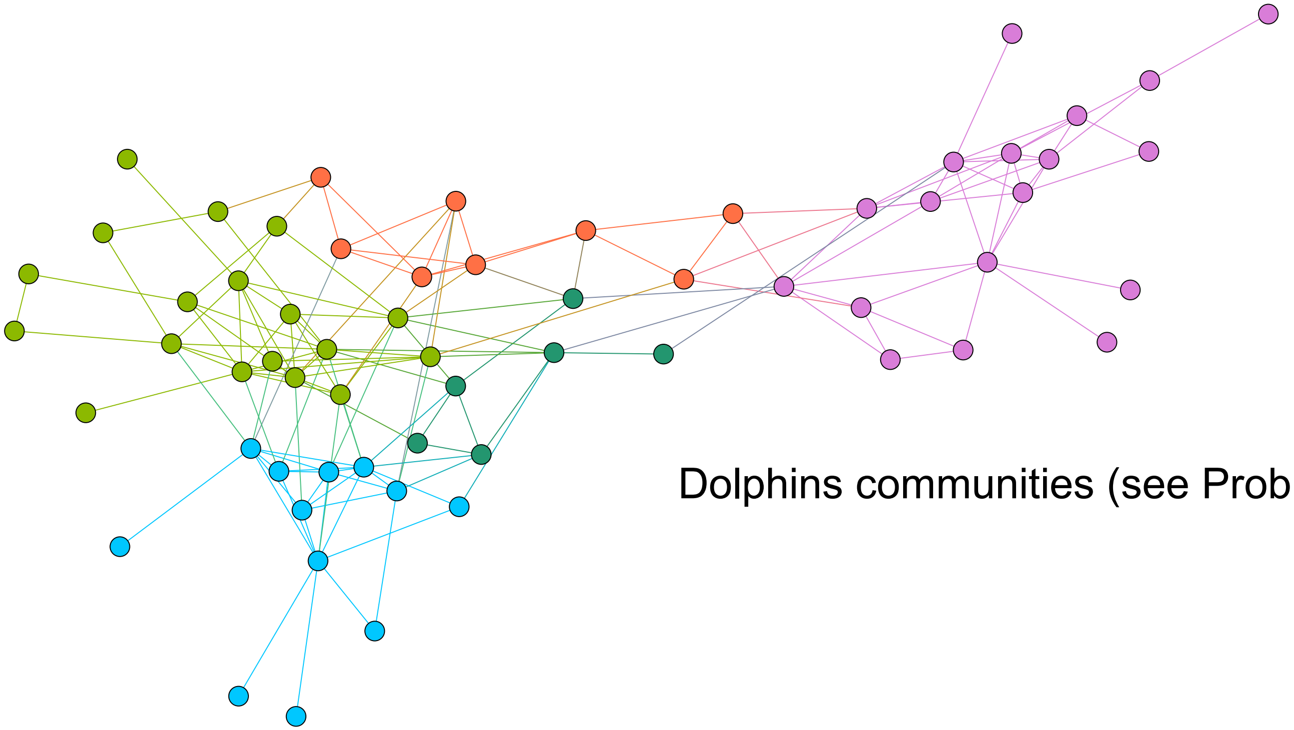
Sample of Belgian mobile phone network



(Blondel, et al., 2008)

Modules and communities

In network science we call a ***community*** a group of nodes that have a higher likelihood of connecting to each other than to nodes from other communities.

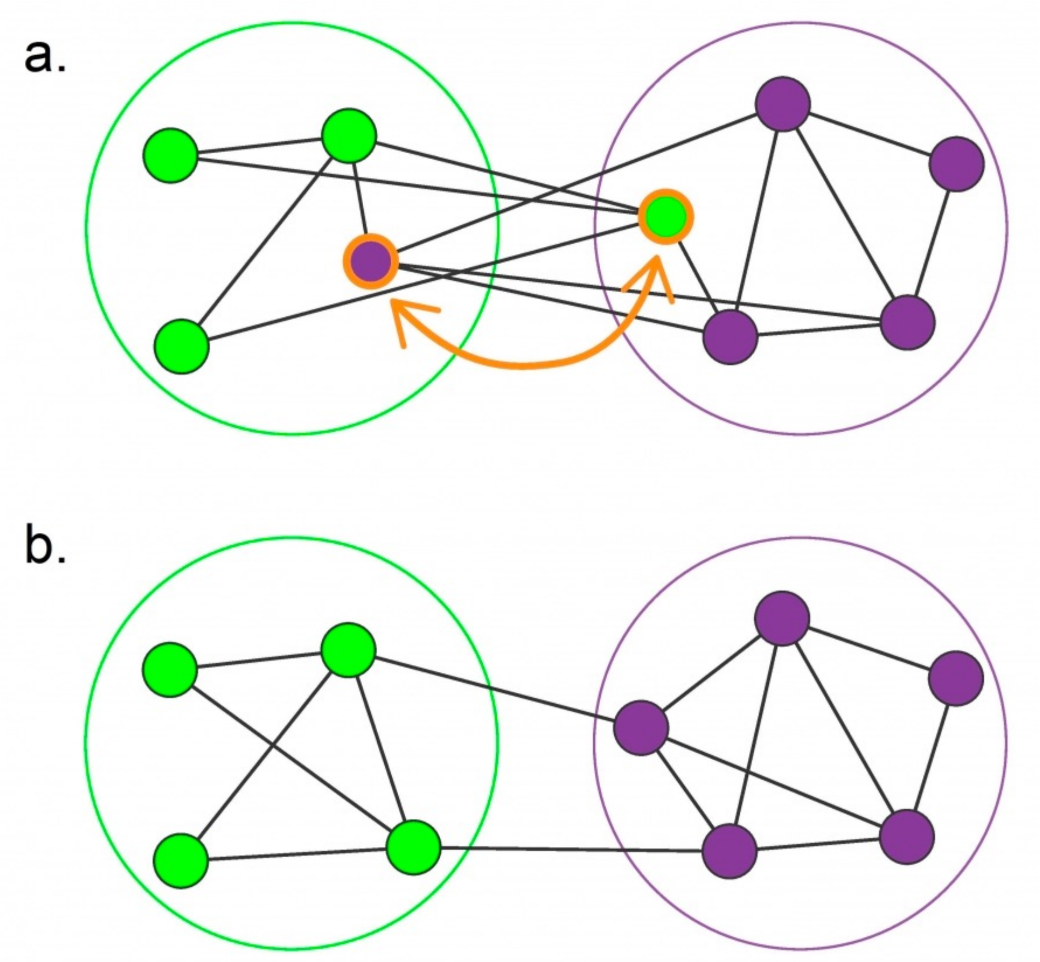


Dolphins communities (see Problem set 1)

General idea

If we have a quantity (let’s call it modularity)we can decide if a particular community partition is better than some other one.

Moreover, if we optimize this quantity, we can find the “best” partitions.



Modularity

Compare two networks: our given network and its zero modularity (i.e., randomized) counterpart with *n* modules.

$$M = \sum_{r=1}^{n} \left[\frac{E_{r\,|\,\mathrm{net}}}{E} - \left(\frac{k_r}{2E} \right)^2 \right]$$

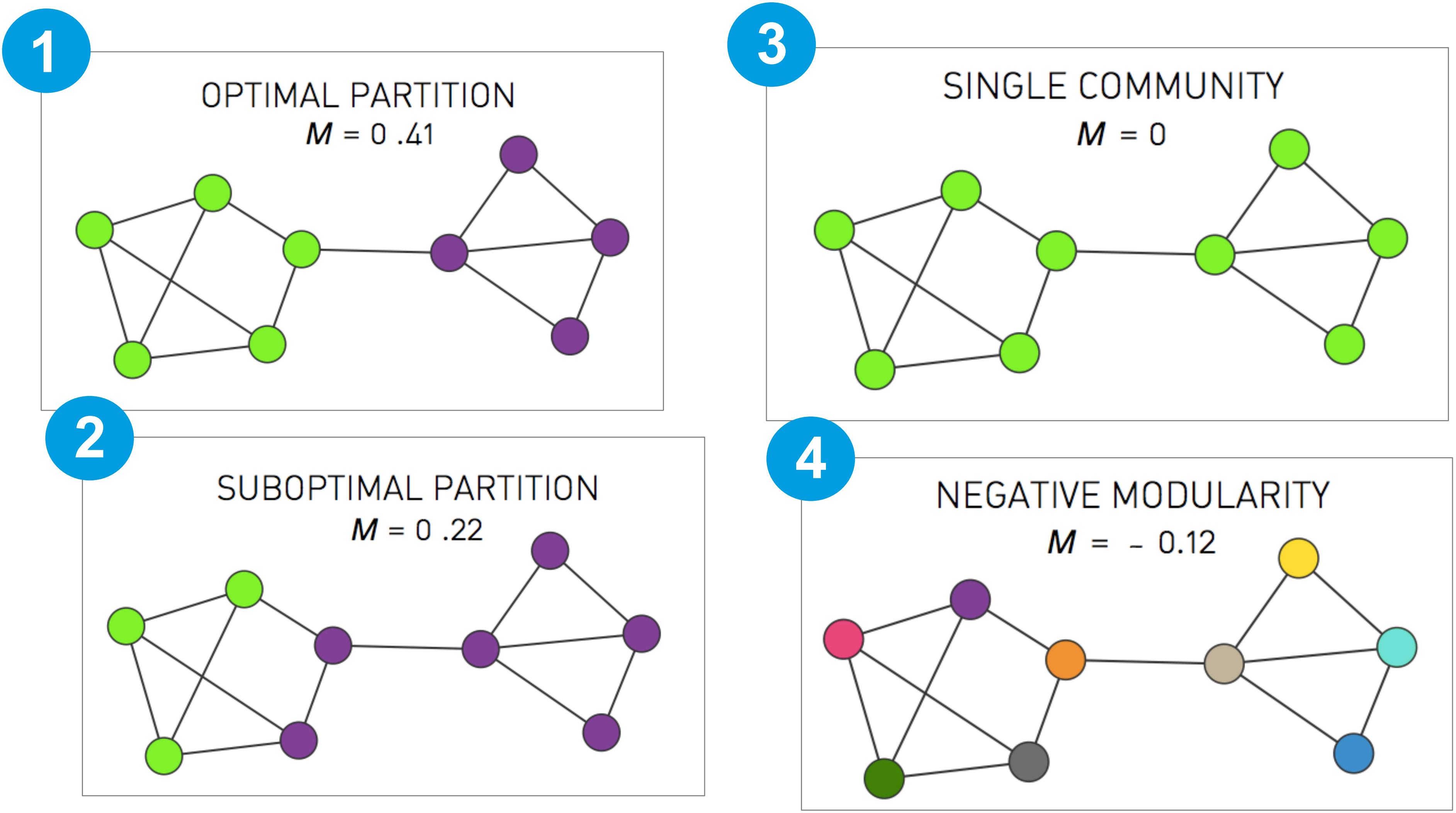
k_r = total degree of nodes in r

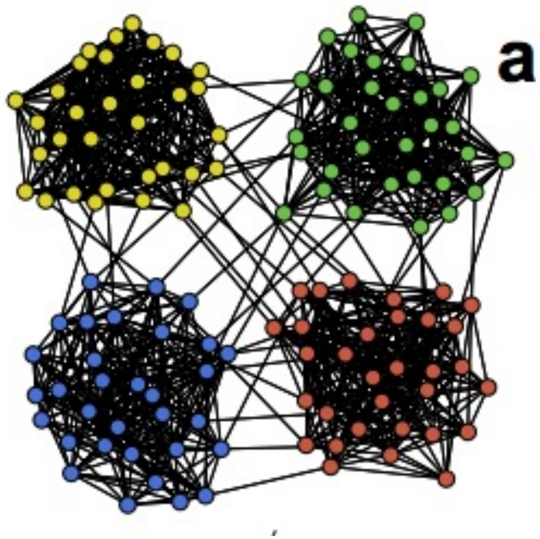
Internal links within module r

Internal links in module r for a randomized network

Modularity

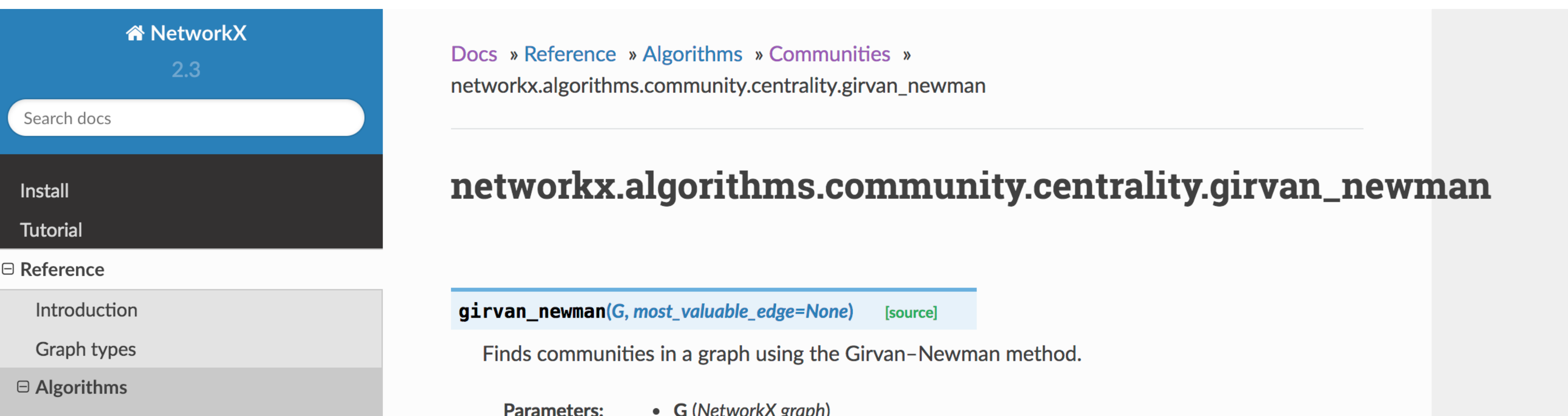
$$M = \sum_{r=1}^{n}\left[\frac{E_r}{E} - \left(\frac{k_r}{2E}\right)^2\right]$$





We will analyze a few algorithms capable of doing this efficiently.

In the meantime, we suggest that you use, for example:



NetworkX
2.3

Search docs

Install
Tutorial
⊟ Reference
Introduction
Graph types
⊟ Algorithms

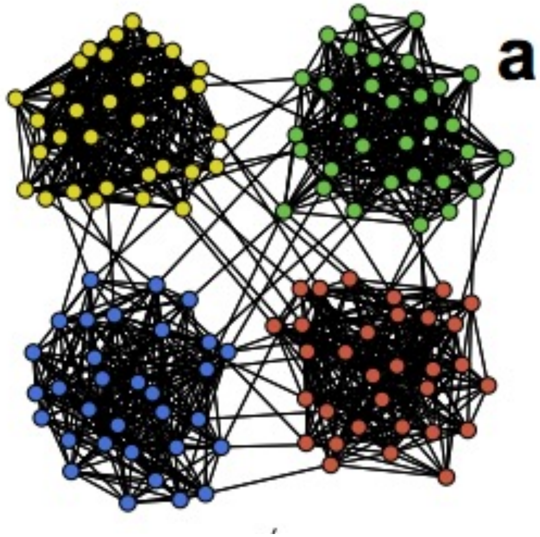
Docs » Reference » Algorithms » Communities » networkx.algorithms.community.centrality.girvan_newman

networkx.algorithms.community.centrality.girvan_newman

`girvan_newman(G, most_valuable_edge=None)` [source]

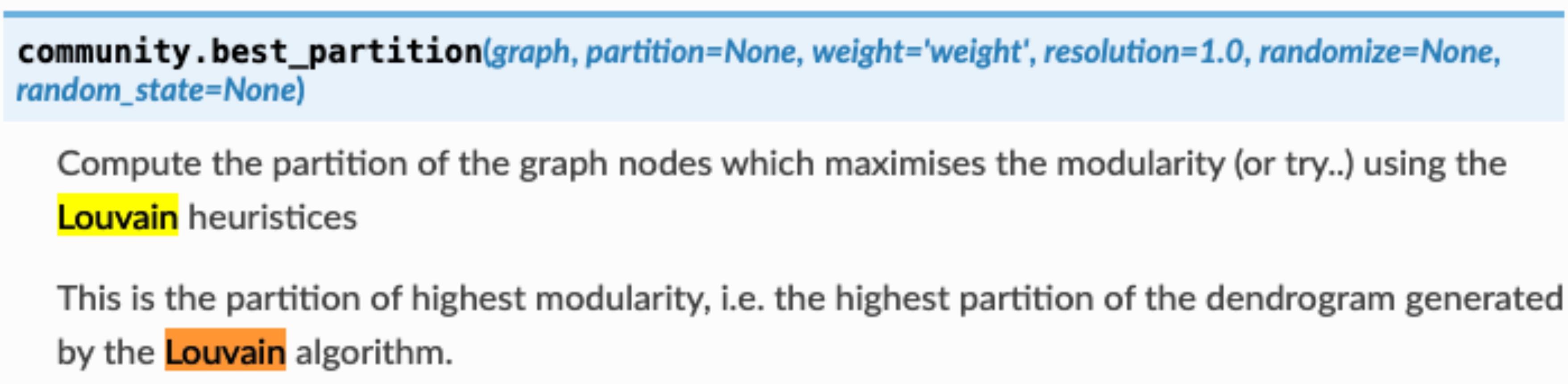
Finds communities in a graph using the Girvan–Newman method.

Parameters: • G (NetworkX graph)



We will analyze a few algorithms capable of doing this efficiently.

In the meantime, we suggest that you use, for example:



community.best_partition(*graph, partition=None, weight='weight', resolution=1.0, randomize=None, random_state=None*)

Compute the partition of the graph nodes which maximises the modularity (or try..) using the Louvain heuristices

This is the partition of highest modularity, i.e. the highest partition of the dendrogram generated by the Louvain algorithm.

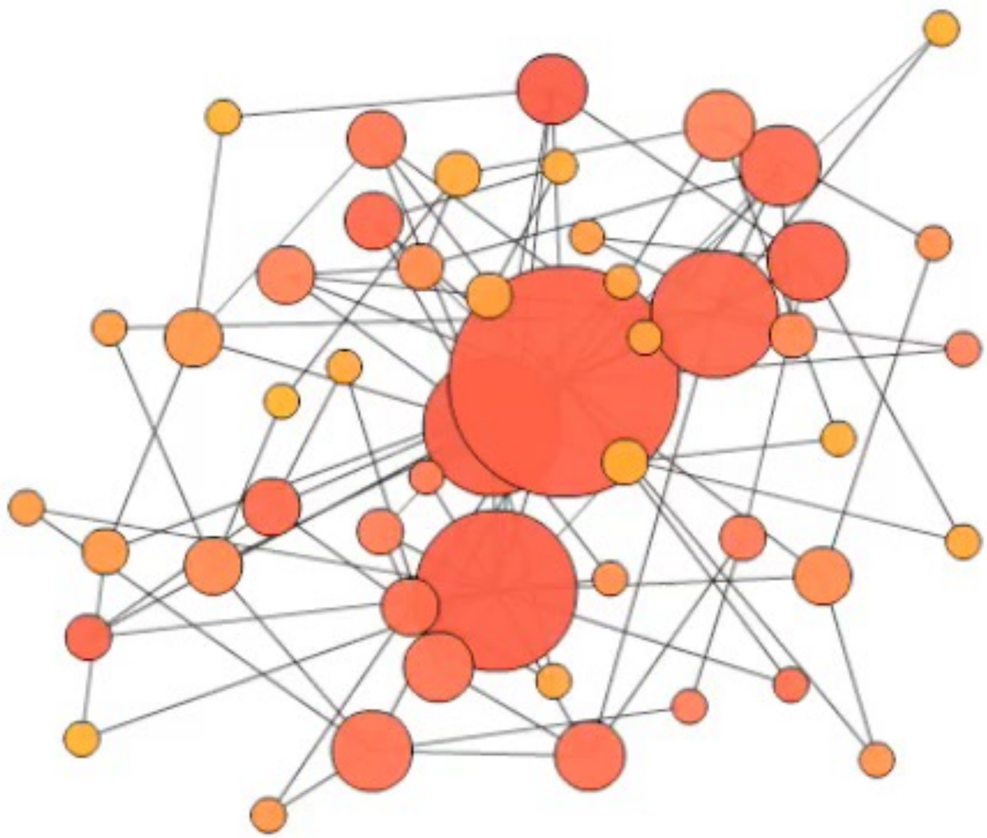
What’s the average degree of your friends?

Friendship paradox:
On average my friends are more popular than I am

What's next?

Resilience of complex networks and cascading effects

Next step: Network robustness



How many nodes do we have to delete to fragment the network into isolated components?

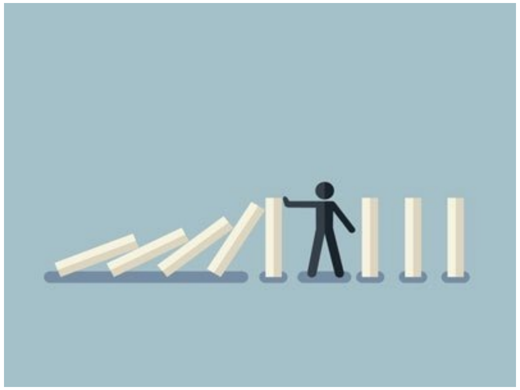
How is resilience connected with the network measures we have been discussing?

Next step: Network robustness

robustness

Random failures & attacks

Cascading effects



Building robustness

Modeling cascading failures