Lecture III
Introduction to complex networks

Santo Fortunato

INSTITUTE
FOR SCIENTIFIC INTERCHANGE
FOUNDATION

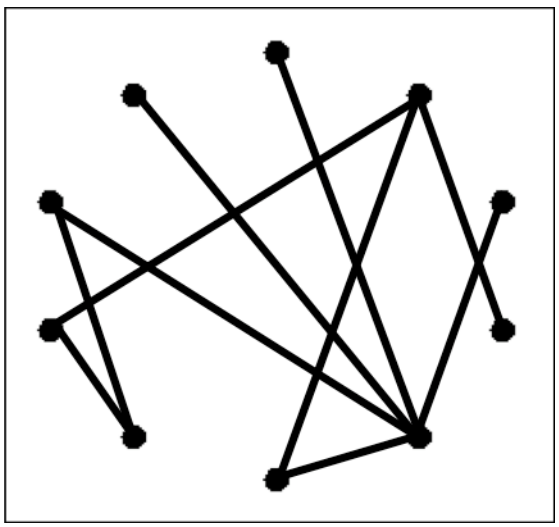
Plan of the course

I. Networks: definitions, characteristics, basic concepts in graph theory
II. Real world networks: basic properties
III. Models
IV. Community structure I
V. Community structure II
VI. Dynamic processes in networks

Classical random graphs

Solomonoff & Rapoport (1951), Erdös-Rényi (1959)

n vertices, edges with probability p: static random graphs



Average number of edges: $|E| = pn(n-1)/2$

Average degree: $< k > = p(n-1)$

$p=c/n$ to have finite average degree

Related formulation

n vertices, m edges: each configuration is equally probable

Classical random graphs

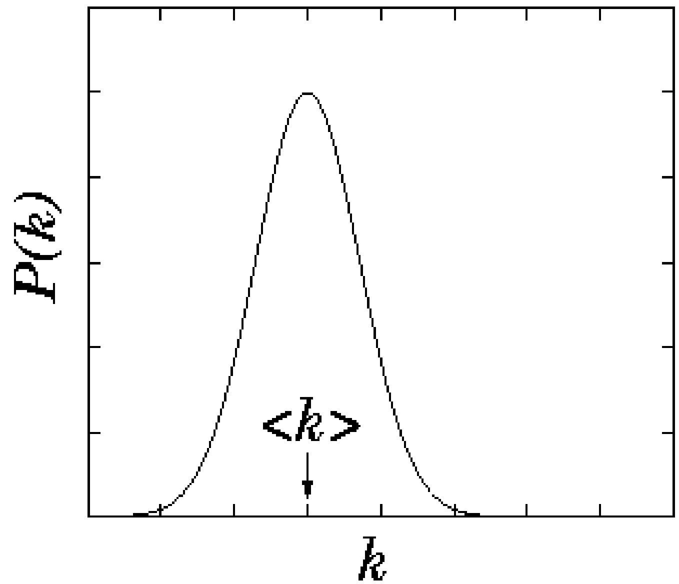
Probability to have a vertex of degree k

- connected to k vertices,
- not connected to the other n-k-1

$$P(k) = \binom{n-1}{k} p^k (1-p)^{n-k-1}$$

Large n, fixed p(n-1)=< k **>** : Poisson distribution

$$P(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$$



Exponential decay at large k

Classical random graphs

Properties

I. Poisson degree distribution (on large graphs)

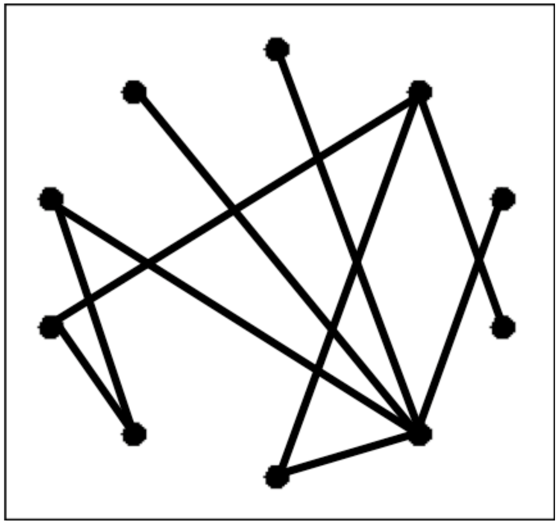
II. Small clustering coefficient: $\langle C\rangle = p = \langle k\rangle / n$ (goes to zero in the limit of infinite graph size for sparse graphs)

III. Short distances: diameter $l \sim \log(n)/\log(\langle k\rangle)$ (number of neighbors at distance d: $\langle k\rangle^d$)

Classical random graphs

<k> < 1: many small subgraphs

< k > > 1: giant component + small subgraphs



Classical random graphs

Self-consistent equation for relative size of giant component

u — probability that a vertex, taken at random, does NOT belong to the giant component

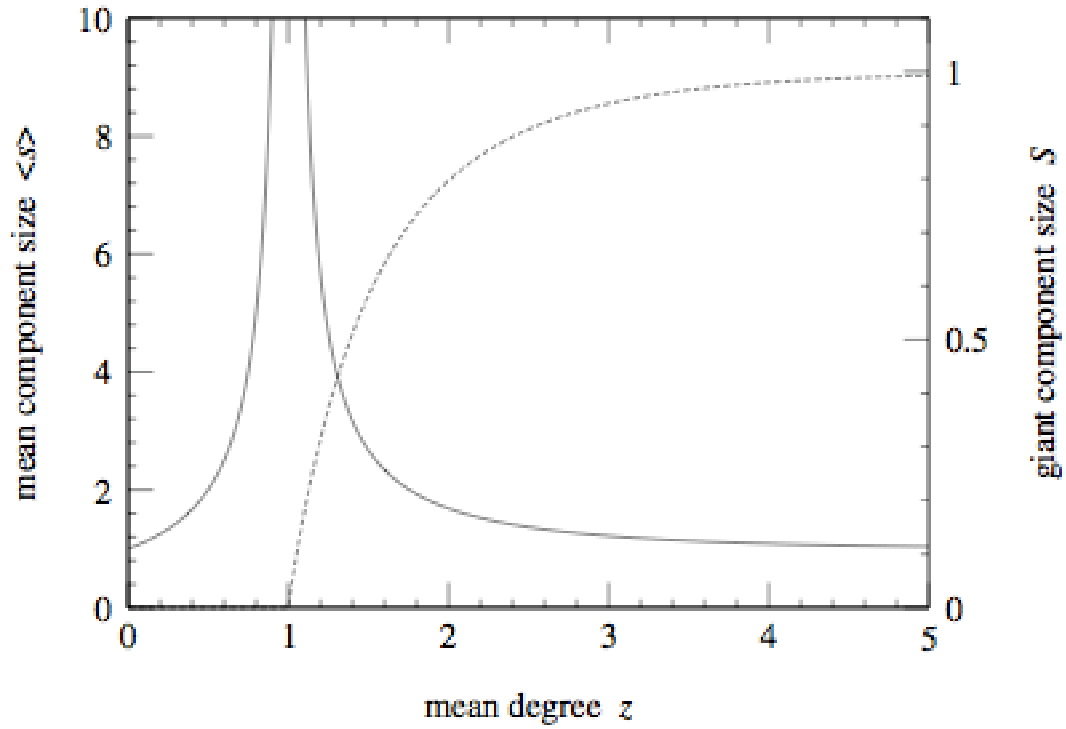
$$u = \sum_{k=0}^{\infty} P(k) u^k = e^{-<k>} \sum_{k=0}^{\infty} \frac{(<k> u)^k}{k!} = e^{<k>(u-1)}$$

$$S = 1 - u \quad \rightarrow \quad S = 1 - e^{-<k>S}$$

Average cluster size

$$\langle s \rangle = \frac{1}{1 - <k> + <k> S}$$

Classical random graphs



Near the transition (<k>~1)

$$S \sim (<k> - 1)^{\beta}$$

$$<s> \sim |<k> - 1|^{-\gamma}$$

$$\beta = 1, \gamma = 1$$

Generalized random graphs

The configuration model (Molloy & Reed, 1995, 1998)

Basic idea: building random graphs with arbitrary degree distributions

How it works

I. Choose a degree sequence compatible with some given distribution
II. Assign to each vertex his degree, taken from the sequence, in that each vertex has as many outgoing stubs as its degree
III. Join the stubs in random pairs, until all stubs are joined

Generalized random graphs

Simple model: easy to handle analytically

Important point: the probability that the degree of vertex reached by following a randomly chosen edge is k is not given by P(k)!

Reason: vertices with large degree have more edges and can be reached more easily than vertices with low degree

Conclusion: distribution of degree of vertices at the end of randomly chosen edge is proportional to kP(k)

Generalized random graphs

Excess degree: number of edges leaving a vertex reached from a randomly selected edge, other than the edge followed

Distribution of excess degree: $Q(k)$

$$Q(k) = \frac{(k+1)P(k+1)}{\sum_k kP(k)} = \frac{(k+1)P(k+1)}{<k>}$$

Generalized random graphs

Chance of finding loops within small (i.e. non-giant) components goes as $O(n^{-1})$

Consequence: graphs of the configuration model are essentially loopless, tree-like, unlike real-world networks

Generating functions

$$G_0(x) = \sum_{k=0}^{\infty} P(k) x^k$$

$$G_1(x) = \sum_{k=0}^{\infty} Q(k) x^k$$

Generalized random graphs

$$G_1(x) = G_0'(x)/ < k >$$

$$< k >= G_0'(1)$$

$$< k^2 > - < k >= G_0'(1)G_1'(1)$$

$$< s >= 1 + \frac{< k >^2}{2 < k > - < k^2 >}$$

Condition for existence of giant component:

$$\sum_k k(k-2)P(k) = 0 \quad \leftrightarrow \quad G_1'(1) = 1$$

Generalized random graphs

$u =$ probability that a randomly chosen edge is not in the giant component

$$u = \sum_{k=0}^{\infty} Q(k) u^k = G_1(u)$$

$$1 - S = \sum_{k=0}^{\infty} P(k) u^k = G_0(u)$$

Self-consistent equations for the relative size S of the giant component

Generalized random graphs

Example: power-law degree distribution

$$P(k) = \begin{cases} 0 & \text{for } k = 0 \\ k^{-\alpha}/\zeta(\alpha) & \text{for } k \geq 1 \end{cases}$$

$$G_0(x) = \frac{Li_\alpha(x)}{\zeta(\alpha)}, \quad G_1(x) = \frac{Li_{\alpha-1}(x)}{x\zeta(\alpha-1)} \qquad \mathrm{Li}_s(z) = \sum_{k=1}^{\infty} \frac{z^k}{k^s}$$

$$G_1'(1) = 1 \;\; \rightarrow \zeta(\alpha-2) = 2\zeta(\alpha-1)$$

Critical exponent value: $\alpha_c = 3.4788$

For $\alpha < \alpha_c$ there is always a giant connected component!

For $\alpha > \alpha_c$ there is no giant connected component!

Generalized random graphs

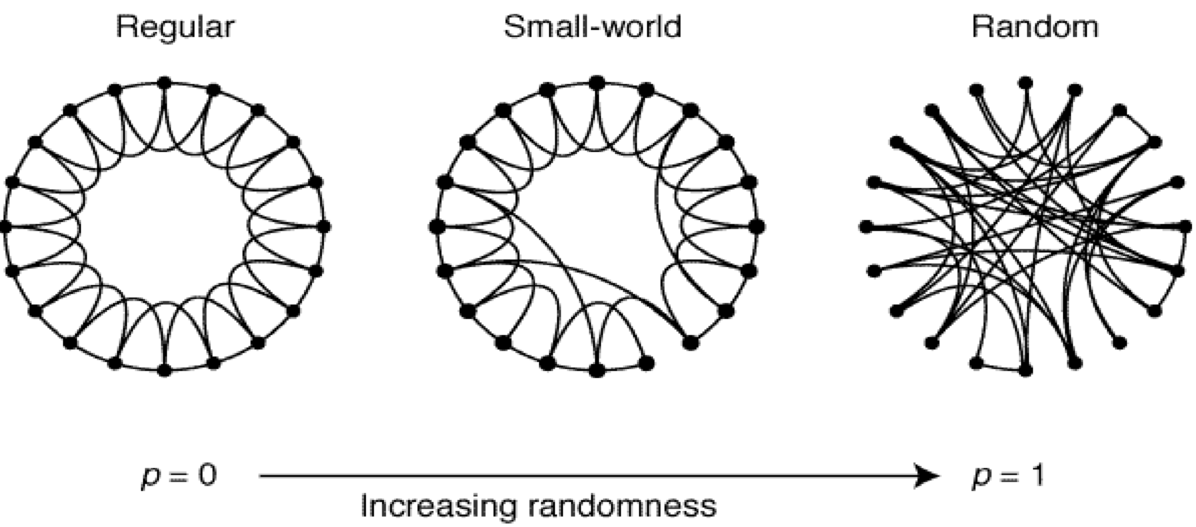
$$S = 1 - G_0(u), \quad u = G_1(u)$$

$$u = \frac{\mathrm{Li}_{\alpha-1}(u)}{u\zeta(\alpha-1)}$$

For $\alpha < 2$, u=0 and S=1 ➔ All vertices are in the giant component!

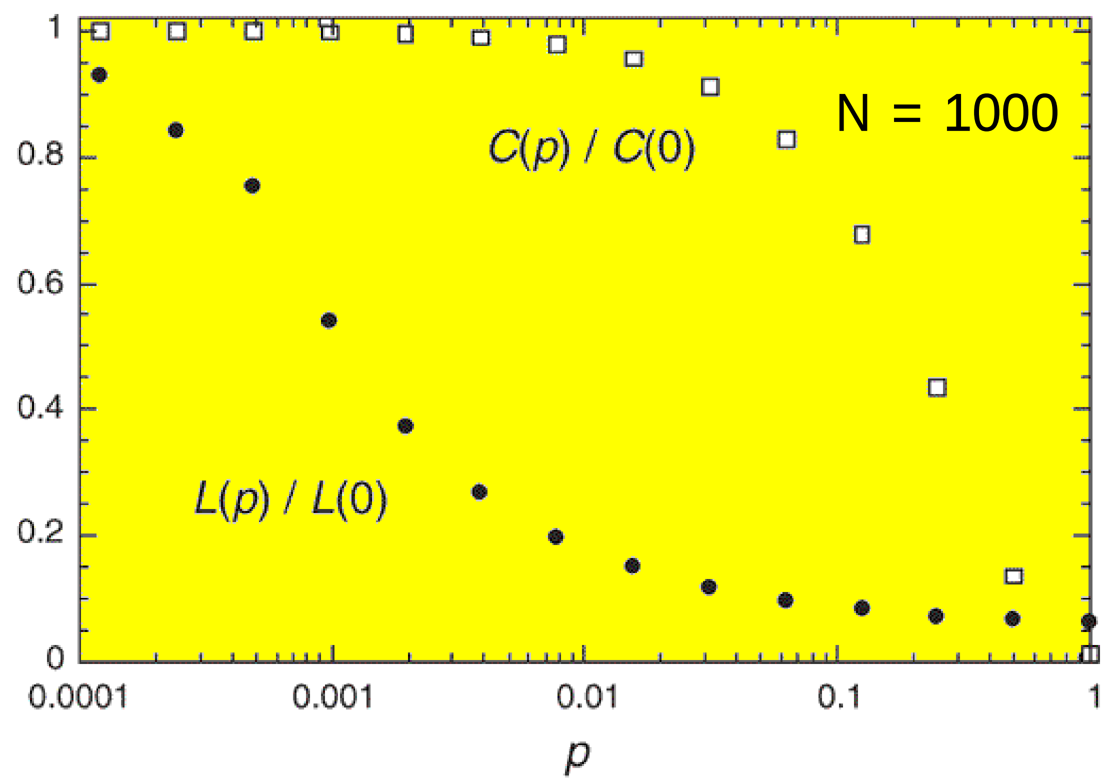
W. Aiello, F. Chung, L. Lu, *Proc. 32th ACM Symposium on Theory of Computing* 171-180 (2000)

Small-world networks



L vertices form a regular lattice, with degree 2k. With probability p, each edge is rewired randomly among those in the clockwise sense

=>Shortcuts



- Large clustering coeff.
- Short typical path

Watts & Strogatz,

Nature **393**, 440 (1998)

Small-world networks

Problems of original formulation:

1) Only one extreme of an edge is rewired
2) No vertex has to be connected to itself
3) No multiple edges
4) The graph may become disconnected ➔ average distance between vertices may get ill-defined

Formulation by Monasson and Newman-Watts: edges are added to the system, not rewired!

p = probability per edge that there is a shortcut anywhere in the graph

Mean number of shortcuts = Lkp, mean degree = 2Lk(1+p)

R. Monasson, *European Physical Journal B* **12**, 555 (1999)

M. E. J. Newman & D. J. Watts, *Physics Letters A* **263**, 341 (1999)

Small-world networks

Clustering coefficient

$$C(0) = \frac{3(k-1)}{2(2k-1)}$$

WS model

$$\tilde{C}(p) = \frac{3(k-1)}{2(2k-1)}(1-p)^3$$

MNW model

$$C(p) = \frac{3(k-1)}{2(2k-1) + 4kp(p+2)}$$

Small-world networks

Degree distribution

WS model

Two contributions:

1) Probability that n^1_i neighbors remain such

$$P_1(n_i^1) = \binom{k}{n_i^1}(1-p)^{n_i^1} p^{k-n_i^1}$$

2) Probability that n^2_i vertices have become neighbors because of rewiring (for large L)

$$P_2(n_i^2) = \frac{(kp)^{n_i^2}}{n_i^2!}\exp(-pk)$$

Small-world networks

Degree distribution

WS model

$$P(j) = \sum_{n=0}^{\min(j-k,k)} \binom{k}{n} (1-p)^n p^{k-n} \frac{(pk)^{j-k-n}}{(j-k-n)!} e^{-pk}$$

MNW model

Each vertex has degree at least 2k plus a contribution which has a binomial distribution

$$P(j) = \binom{L}{j-2k} \left[\frac{2kp}{L}\right]^{j-2k} \left[1 - \frac{2kp}{L}\right]^{L-j+2k}$$

Small-world networks

Average shortest path

Scaling form

$$\ell(L,p) = \frac{L}{k} f(Lkp)$$

$$f(x) \sim \begin{cases} 1 & \text{for } x \ll 1 \\ (\log x)/x & \text{for } x \gg 1. \end{cases}$$

M. Barthélémy, L. A. N. Amaral, *Physical Review Letters* **82**, 3180 (1999)

A. Barrat, M. Weigt, *European Physical Journal B* **13**, 547 (2000)

M. E. J. Newman & D. J. Watts, *Physics Letters A* **263**, 341 (1999)

Small-world networks

Average shortest path

$$\text{if } Lkp \ll 1 \quad \rightarrow \quad \ell(L,p) \sim \frac{L}{k}$$

$$\text{if } Lkp \gg 1 \quad \rightarrow \quad \ell(L,p) \sim \log L$$

$L \rightarrow \infty$ for any non-zero value of p the average shortest path is “small”!

Phase transition at p=0!

Statistical physics approach

Microscopic processes of the
many component units

↓

Macroscopic statistical and dynamical
properties of the system

Cooperative phenomena
Complex topology

⟷

Natural outcome of
the dynamical evolution

⇒ Find microscopic mechanisms

Modelling growing networks

(1) The number of vertices (N) is **NOT** fixed.

Networks continuously expand by the addition of new vertices

Examples:
WWW: addition of new documents
Citation: publication of new papers

(2) The attachment is NOT uniform.

A vertex is linked with higher probability to a vertex that already has a large number of links.

Examples :
WWW : new documents link to well known sites (CNN, YAHOO, New York Times, etc)
Citation : well cited papers are more likely to be cited again

Price's model

D. de Solla Price, *Networks of scientific papers*, Science **149**, 510 (1965)

D. de Solla Price, *A general theory of bibliometric and other cumulative advantage processes*, J. Amer. Soc. Inform. Sci. **27**, 292 (1976)

Citation networks have a broad distribution of in-degree!

Idea: popular papers become more popular (cumulative advantage)

H. A. Simon, *On a class of skew distribution functions*, Biometrika **42**, 425 (1955)

Price's model

Directed graph of n vertices

Mean in-degree/out-degree m

$$\sum_k kP(k) = m$$

Principle: probability that a newly appearing paper cites a previous paper is proportional to the number k of citations of the paper

Problem: what happens if a paper has no citations?

Solution: proportionality to $k+k_0$, where k_0 is constant

Price took $k_0 = 1$ (publication of the paper is a sort of first citation)

Price's model

Probability that a new edge attaches to any of the vertices with in-degree k:

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

Mean number of new edges (per vertex added) attached to vertices of in-degree k:

$$(k+1)P(k)\frac{m}{m+1}$$

P(k,n)= probability distribution of in-degree k when there are n vertices in the graph

Price's model

Master equation

$$(n+1)P(k,n+1) - nP(k,n) = [kP(k-1,n) - (k+1)P(k,n)]\frac{m}{m+1} \qquad \text{for } k \geq 1$$

$$(n+1)P(0,n+1) - nP(0,n) = 1 - P(0,n)\frac{m}{m+1} \qquad \text{for } k = 0$$

Stationary solution

$$P(k,n+1) = P(k,n) = P(k)$$

$$P(k) = \begin{cases} [kP(k-1) - (k+1)P(k)]m/(m+1) & \text{for } k \geq 1, \\ 1 - P(0)m/(m+1) & \text{for } k = 0. \end{cases}$$

Price's model

Stationary solution

$$P(k) = \begin{cases} [kP(k-1) - (k+1)P(k)]m/(m+1) & \text{for } k \geq 1, \\ 1 - P(0)m/(m+1) & \text{for } k = 0. \end{cases}$$

$$P(0) = \frac{m+1}{2m+1}$$

$$P(k) = \frac{P(k-1)k}{k+2+1/m}$$

$$P(k) = \frac{k(k-1)\ldots 1}{(k+2+1/m)\ldots(3+1/m)} P(0) = (1+1/m)B(k+1, 2+1/m)$$

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$ Beta function!

Price's model

For large a and fixed b: $B(a,b) \to a^{-b}$

$$P(k) = (1+1/m)B(k+1, 2+1/m) \to k^{-(2+1/m)} \text{ for large } k$$

For generic k_0 :

$$P(k) = \frac{m+1}{m(k_0+1)+1} \frac{B(k+k_0, 2+1/m)}{B(k_0, 2+1/m)} \to k^{-(2+1/m)}$$

Exponent of the power law tail is independent of k_0

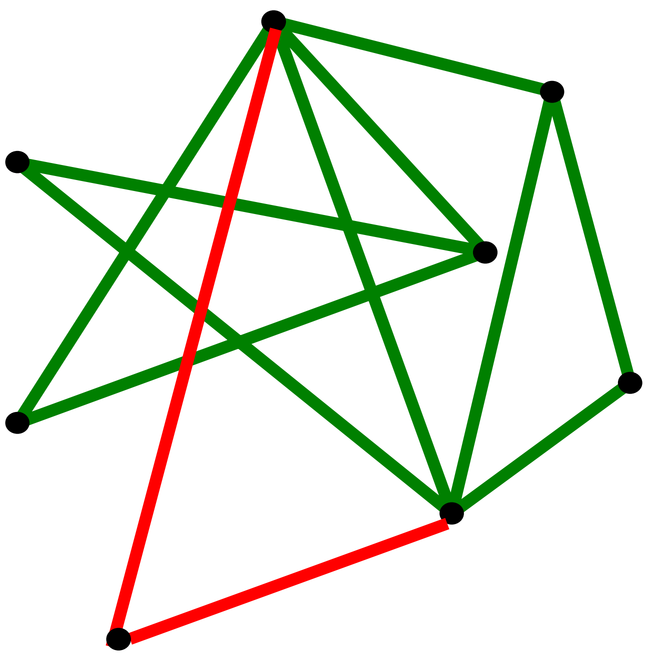
The exponent only depends on m and is always larger than 2!

Barabási-Albert model

Analogies with Price's model

1) Cumulative advantage $\Leftrightarrow$ preferential attachment
2) Average degree of vertices is m at all stages of the evolution

$$\prod_{j\to i} = \frac{k_i}{\sum_l k_l}$$



Differences from Price's model

1) Networks are undirected (which solves Price's problem with uncited papers)
2) Number of edges coming with new vertex is fixed to m (so m is integer ≥ 1!)

A.-L. Barabási, R. Albert, Science **286,** 509 (1999)

Barabási-Albert model

Probability that a new edge attaches to any of the vertices with in-degree k:

$$\frac{kP(k)}{\sum_k kP(k)} = \frac{kP(k)}{2m}$$

Mean number of vertices of degree k gaining an edge when a single new vertex with m edges is added:

$$m \times kP(k)/2m = \frac{kP(k)}{2}$$

P(k,n)= probability distribution of in-degree k when there are n vertices in the graph

Barabási-Albert model

Master equation

$$(n+1)P(k,n+1) - nP(k,n) = \frac{1}{2}(k-1)P(k-1,n) - \frac{1}{2}kP(k,n) \quad \text{for } k > m$$

$$(n+1)P(m,n+1) - nP(m,n) = 1 - \frac{1}{2}mP(m,n) \quad \text{for } k = m$$

Stationary solution

$$P(k,n+1) = P(k,n) = P(k)$$

$$P(k) = \begin{cases} \frac{1}{2}(k-1)P(k-1) - \frac{1}{2}kp_k & \text{for } k > m, \\ 1 - \frac{1}{2}mp_m & \text{for } k = m. \end{cases}$$

Barabási-Albert model

Stationary solution

$$P(k) = \begin{cases} \frac{1}{2}(k-1)P(k-1) - \frac{1}{2}kp_k & \text{for } k > m, \\ 1 - \frac{1}{2}mp_m & \text{for } k = m. \end{cases}$$

$$P(m) = \frac{2}{m+2}$$

$$P(k) = \frac{P(k-1)(k-1)}{k+2}$$

$$P(k) = \frac{(k-1)(k-2)\ldots m}{(k+2)(k+1)\ldots(m+3)} P(m) = \frac{2m(m+1)}{(k+2)(k+1)k}$$

Barabási-Albert model

$$P(k) = \frac{(k-1)(k-2)\dots m}{(k+2)(k+1)\dots(m+3)} P(m) = \frac{2m(m+1)}{(k+2)(k+1)k}$$

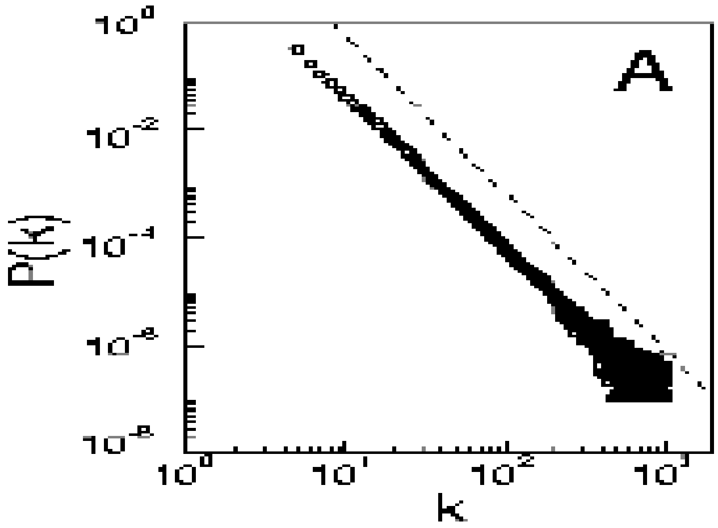
For large k:

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

Important difference from Price's model: the exponent of the degree distribution does not depend on m!

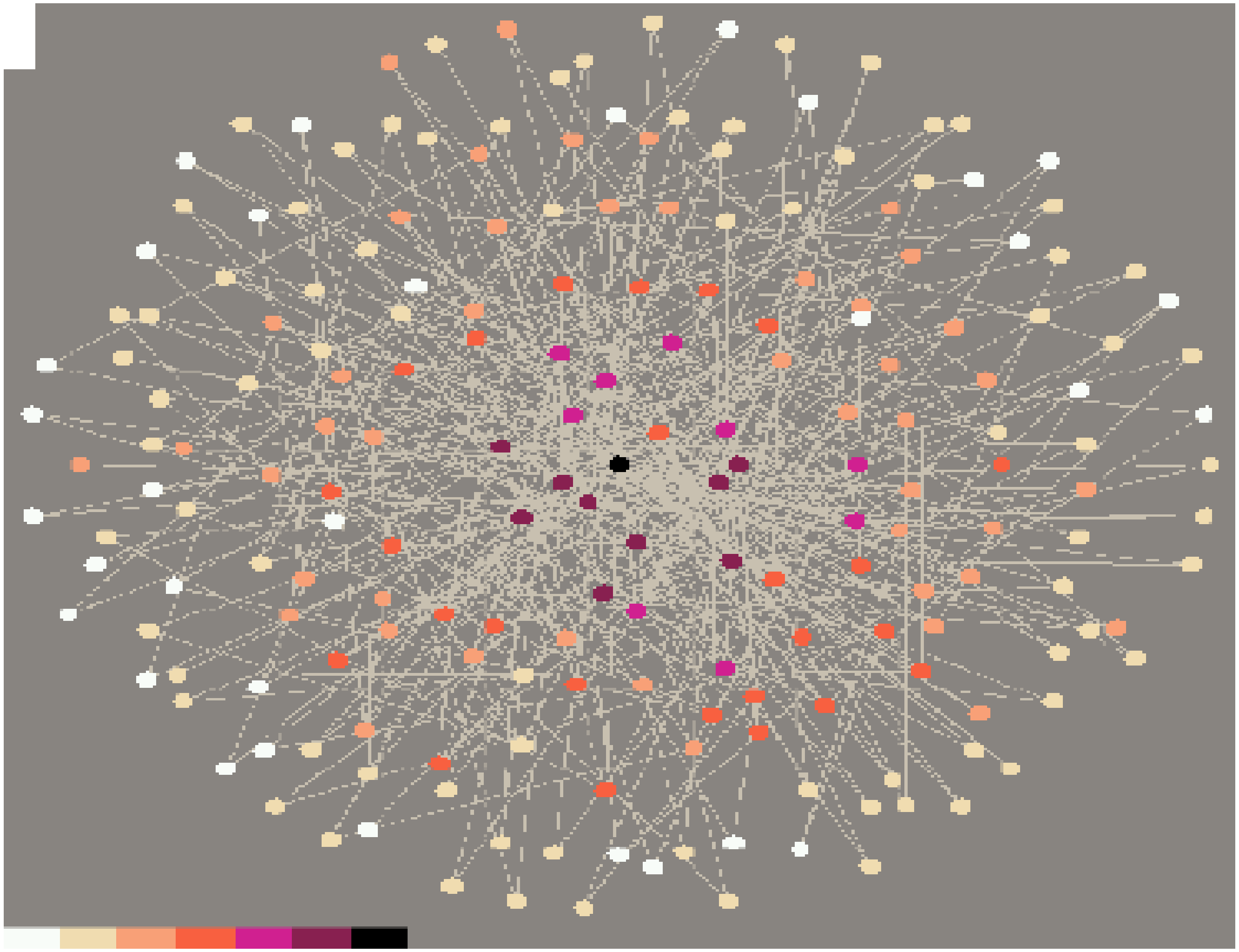
P. L. Krapivsky, S. Redner, F. Leyvraz, Phys. Rev. Lett. **85,** 4629 (2000)

S. N. Dorogovtsev, J. F. F. Mendes, A. N. Samukhin, Phys. Rev. Lett. **85,** 4633 (2000)



Connectivity distribution

BA network



Barabási-Albert model

Features

1) Correlation between the degree of vertices and their age: early vertices have a much higher probability to reach high degree values!

2) The average shortest path scales as log n/[log (log n)]

3) The clustering coefficient scales as $n^{-3/4}$ (slower than random graphs)

Limits

1) The correlation between degrees of neighboring vertices goes to zero in the infinite size limit

2) The clustering coefficient goes to zero in the infinite size limit

3) All vertices belong to the same connected component

4) Only one exponent value for the degree distribution

Real networks are characterized by relevant degree correlations, high clustering coefficients, different values for the degree exponent and may have several components => refined versions of the BA model account for these features

Simulations of the BA model

Typical starting point: complete graph with m+1 vertices, where m is the average degree of the graph

Naïve procedure too costly: O(n) for a loop over all vertices (to check their degree and compute the linking probability) and $O(n^2)$ to build the graph

Faster procedure [complexity O(n)]:

1)One has to maintain a list, in which the label of each vertex is repeated as many times as its degree

2)When a new vertex joins the system, one picks labels at random from the list, all with the same probability, and the new vertex gets linked to the vertex corresponding to the selected label

Microscopic mechanism:

Non-linear preferential attachment

$$\prod_{j \to i} = \frac{k_i^\alpha}{\sum_l k_l^\alpha}$$

(1) $\alpha<1$: P(k) has exponential decay (power law times stretched exponential)! $f(x) = e^{x^\alpha}$

(2) $\alpha>1$: one or more vertices is attached to a macroscopic fraction of vertices (condensation); the degree distribution of the other vertices is exponential

(3) $\alpha=1$: $P(k) \sim k^{-3}$

P.L. Krapivsky, S. Redner, F. Leyvraz, Phys. Rev. Lett. 85, 4629 (2000)

Preferential attachment: is it justified?

$$\prod_{j\to i} = \frac{k_i^{\alpha}}{\sum_l k_l^{\alpha}}$$

One could try to test the hypothesis on real networks

Method: measure the fraction of edges that get attached to vertices with degree k in a (short) time window

Problem: data could be noisy if one focuses on single degree classes!

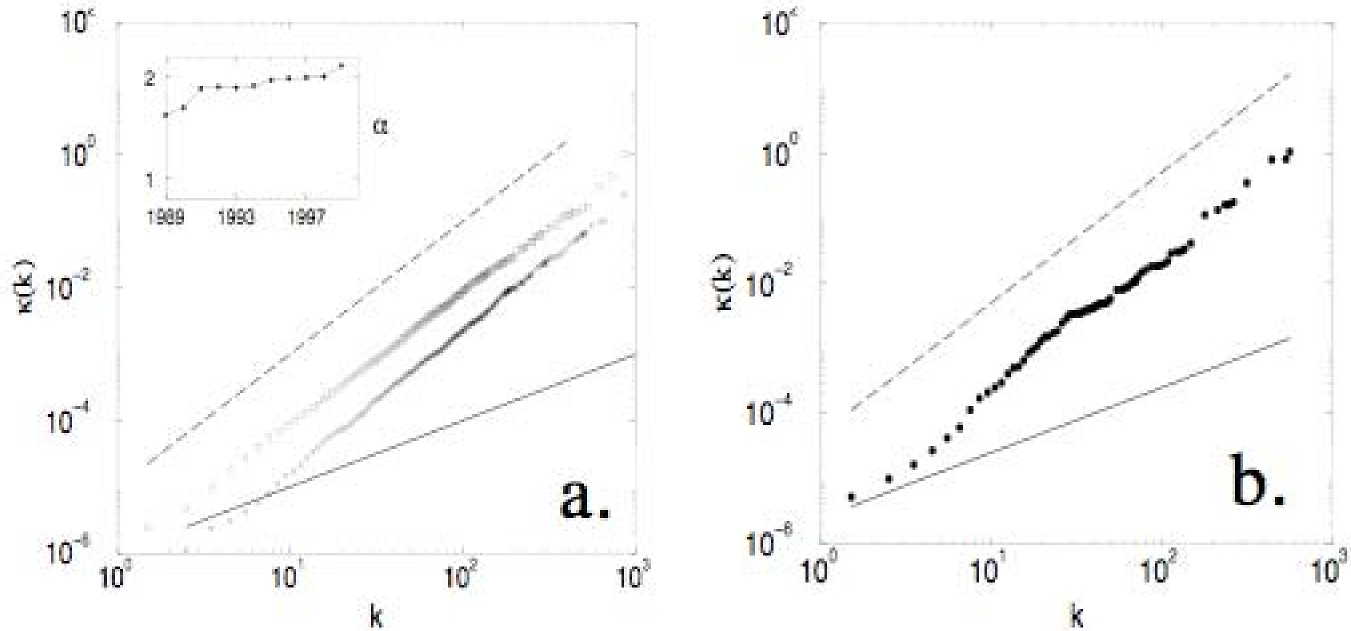
Solution:

$$\kappa(k) = \int_0^k \Pi(k)dk$$

H. Jeong, Z. Néda, A.-L.Barabási, Europhys. Lett. 61, 567 (2003)

Preferential attachment: is it justified?

$$\prod_{j \to i} = \frac{k_i^\alpha}{\sum_l k_l^\alpha} \quad => \quad \kappa(k) \propto k^{\alpha+1}$$



Linear preferential attachment

(1) GROWTH :

At every timestep we add a new vertex with *m* edge(connected to the vertices already present in the system).

(2) PREFERENTIAL ATTACHMENT :

The probability Π that a new vertex will be connected to vertex *i* depends on the connectivity k_i of that vertex and a constant k_0 (**attractivity**), with $-m < k_0 < \infty$

$$\prod_{j \rightarrow i} = \frac{k_i + k_0}{\sum_l (k_l + k_0)}$$

S. N. Dorogovtev, J. F. F. Mendes, A. N. Samukhin, Phys. Rev. Lett. **85,** 4633 (2000)

Stationary solution

$$P(k) = \begin{cases} [(k-1)P(k-1) - kP(k)]m/(2m+k_0) & \text{for } k > m, \\ 1 - P(m)m^2/(2m+k_0) & \text{for } k = m, \end{cases}$$

$$P(m) = \frac{2m+k_0}{m^2+2m+k_0}$$

$$P(k) = \frac{(k-1)\ldots m}{(k+2+k_0/m)\ldots(m+3+k_0/m)} P(m) = \frac{B(k, 3+k_0/m)}{B(m, 2+k_0/m)}$$

$$P(k) \sim k^{-(3+k_0/m)} \quad \text{for large } k$$

Extension to directed graphs:

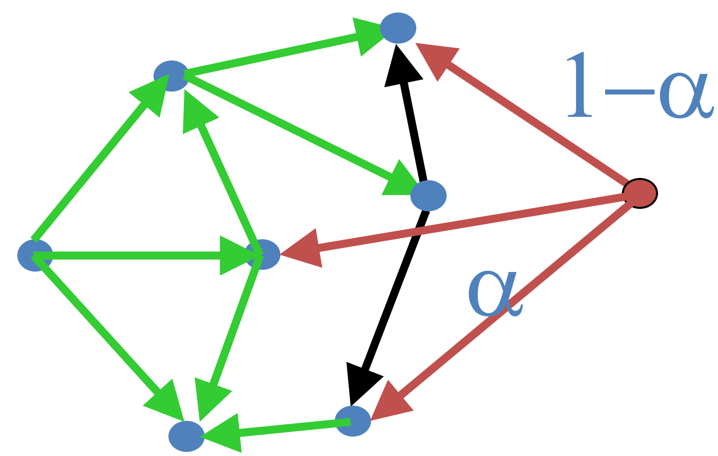
$$\prod_{j \to i} = \frac{k_i^{in} + k_0}{\sum_j (k_j^{in} + k_0)}$$

$$P(k_{in}) \sim (k_{in})^{-(2+k_0/m)}$$

No problems with vertices with zero indegree!

The copying model

Originally proposed as a model for the Web graph



a. Selection of a vertex

b. Introduction of a new vertex

c. The new vertex copies m edges of the selected one

d. Each new edge is kept with probability 1-α, rewired at random with probability α

J. M. Kleinberg, S. R. Kumar, P. Raghavan, S. Rajagopalan, A. Tomkins, Proc. Int. Conf. Combinatorics & Computing, LNCS **1627,** 1 (1999)

The copying model

Probability for a vertex to receive a new edge after t vertices have been added to the system:

• Due to random rewiring: α/t

• Because it is neighbour of the selected vertex: $(1-\alpha)k_{in}/(mt)$ (it equals the probability that a randomly chosen edge is attached to a vertex with in-degree k_{in})

$$\prod_{j\to i} = \frac{\alpha}{t} + (1-\alpha)\frac{k_{in}^{i}}{mt}$$

effective preferential attachment, without a priori knowledge of degrees!

The copying model

Degree distribution:

$$P(k_{in}) \sim (k_0 + k_{in})^{-\frac{2-\alpha}{1-\alpha}}$$

=> Heavy-tails

In addition, bipartite motifs are formed (as in the Web graph)!

=> model for WWW and evolution of genetic networks

Fitness models

- In models based on preferential attachment, the older a vertex, the higher its degree (degree is correlated with age)

- In real systems, this is not true! Young vertices may become more important/connected than older vertices (Ex. Web)

Fitness model: each vertex has a fitness η_i with distribution $\rho(\eta)$

$$\prod_{j \to i} = \frac{\eta_i k_i}{\sum_l \eta_l k_l}$$

G. Bianconi, A.-L. Barabási, Phys. Rev. Lett. 86, 5632 (2001)

Fitness models

- If η is the same for all vertices ➔ preferential attachment
- If $\rho(\eta)$ is uniform in [0:1]

$$P(k) \sim \frac{k^{-2.26}}{\ln(k)}$$

Ranking model

Criteria of standard models: complete knowledge of prestige of vertices (degree, fitness)

$$\prod_{t\rightarrow i} = f(k_i, \eta_i, \ldots)$$

Problem: in real systems usually only a partial knowledge is possible !

Ranking is easier!

Ranking model

1) Prestige measure is chosen.
2) New node t joined to j with probability

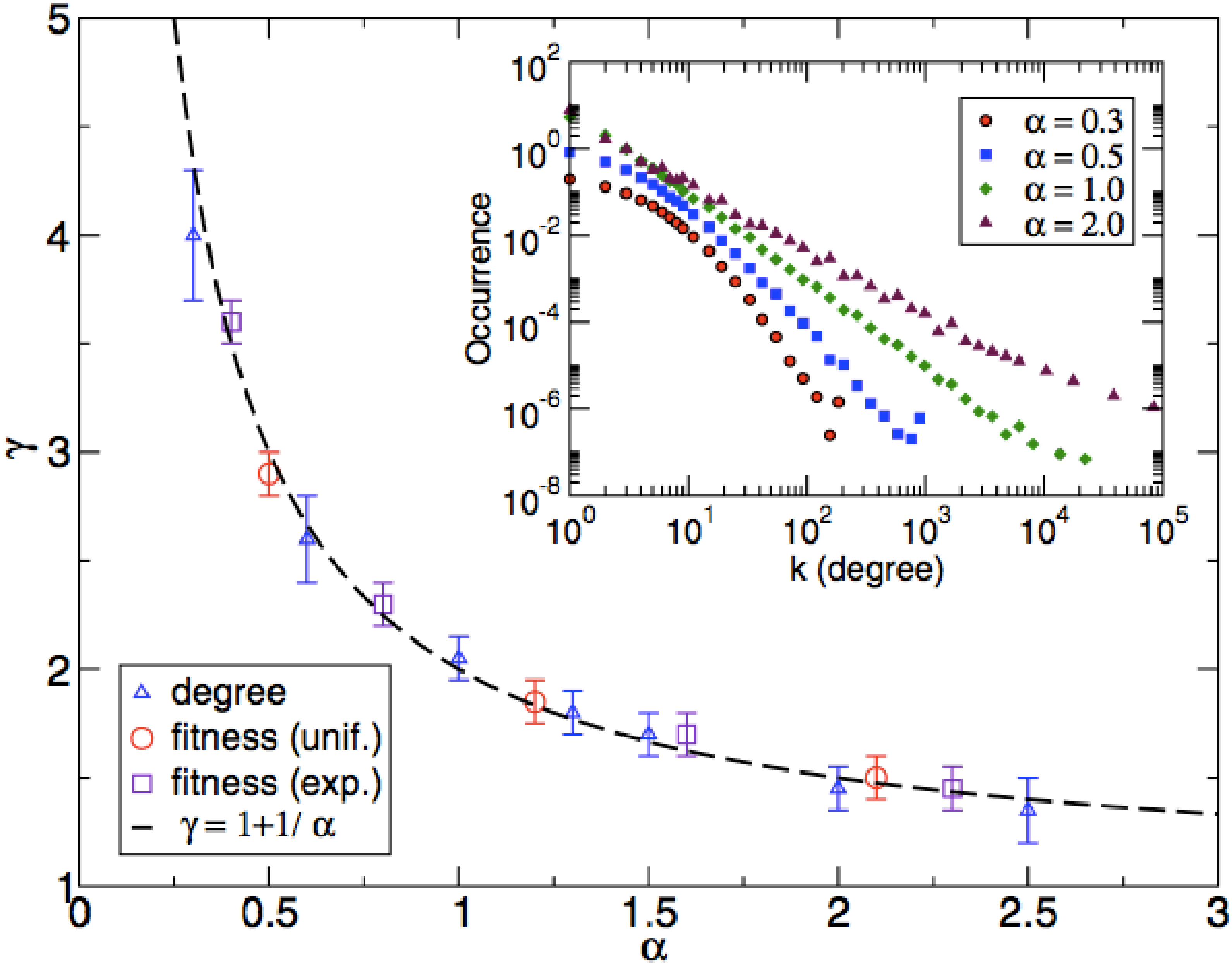
$$\prod_{t \to j} = \frac{R_j^{-\alpha}}{\sum_{k=1}^{t-1} R_k^{-\alpha}}$$

Result:

$$P(k) \sim k^{-1-1/\alpha}$$

S. Fortunato, A. Flammini, F. Menczer, Phys. Rev. Lett. **96,** 218701 (2006)

Ranking model



Weighted growing networks

- Growth: at each time step a new vertex is added with m edges to be connected with previous vertices

- Preferential attachment: the probability that a new edge is connected to a given vertex is proportional to the vertex strength

The preferential attachment follows the probability distribution:

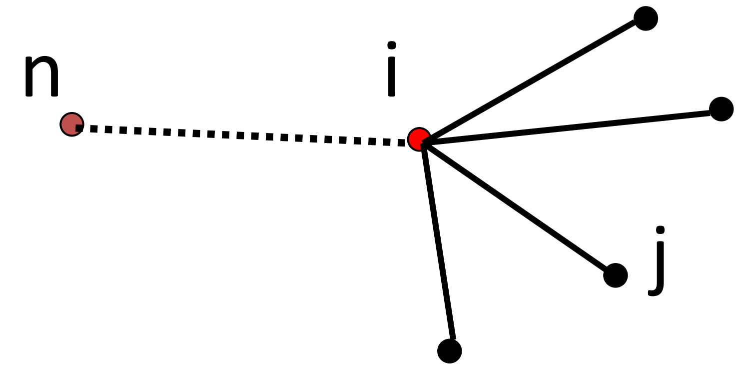
$$\prod_{k \to i} = \frac{s_i}{\sum_j s_j}$$

Preferential attachment driven by weights

A. Barrat, M. Barthélemy, A. Vespignani, Phys. Rev. Lett. 92, 228791 (2004)

Redistribution of weights

New vertex: n, attached to i
New weight $w_{ni}=w_0=1$
Weights between i and its other neighbours:



$$w_{ij} \, ! \, w_i + \Delta w_{ij}$$

$$\Delta w_{ij} = \delta \frac{w_{ij}}{s_i}$$

$$s_i \longrightarrow s_i + w_0 + \delta$$

Only parameter

The new traffic n-i increases the traffic i-j

Analytical results

Power law distributions for k, s and w:

$P(k) \sim k^{-\gamma}$; $P(s) \sim s^{-\gamma}$

$$\gamma = \frac{4\delta + 3}{2\delta + 1}$$

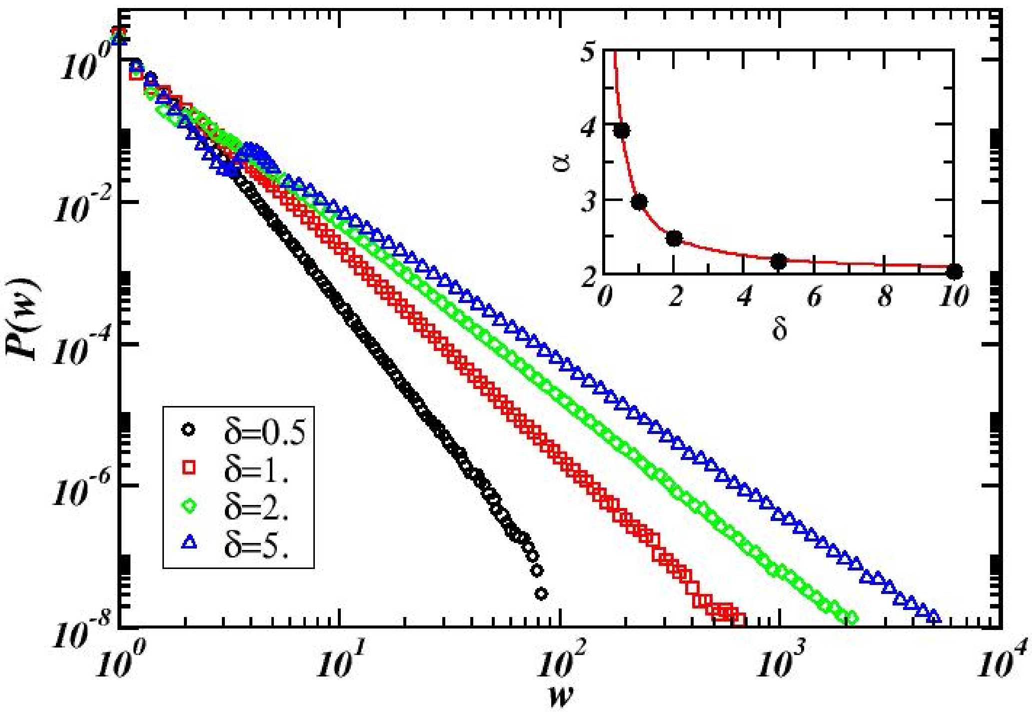
$$\mathbf{P(w) \sim w^{-\alpha}}, \quad \alpha = \mathbf{2} + \frac{\mathbf{1}}{\delta}$$

Correlations topology/weights:

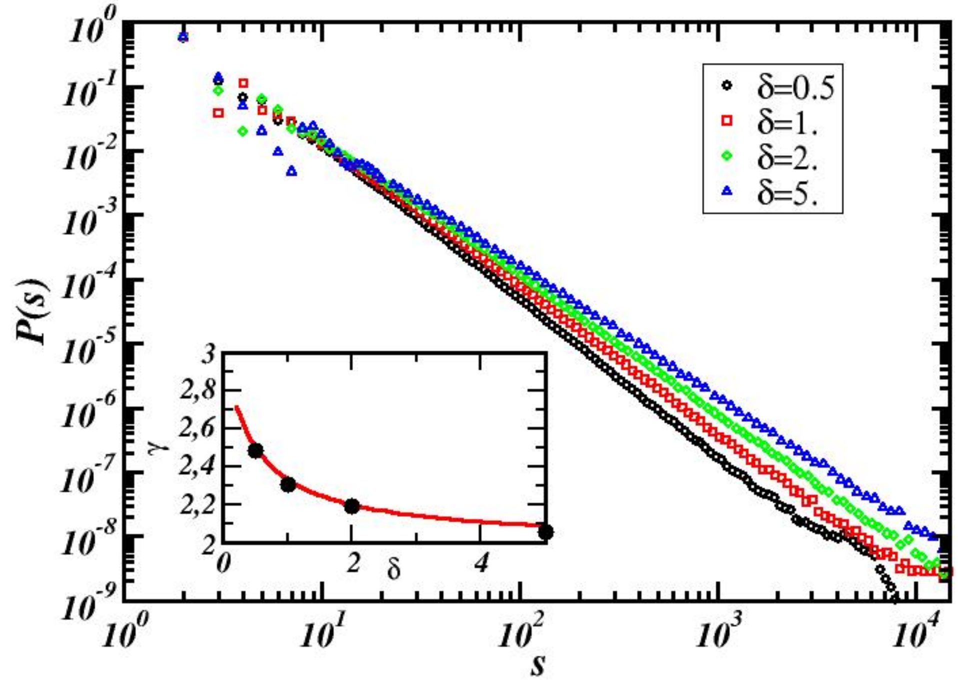
$$\mathbf{s_i = (2\delta + 1)k_i - 2m\delta}$$

$w_{ij} \sim \min(k_i, k_j)^{\alpha}$, $\alpha = 2\delta / (2\delta + 1)$

Numerical results: P(w), P(s)



(N=10^5)



Models: other ingredients/features

- Vertex/edge deletion
- Edge rewiring
- Clustering
- Assortatitivy
- More connected components
- ...

Model validation: comparison with large scale datasets!

Plan of the course

I. Networks: definitions, characteristics, basic concepts in graph theory
II. Real world networks: basic properties
III. Models
IV. Community structure I
V. Community structure II
VI. Dynamic processes in networks