Statistical Mechanics of Networks

TROISIEME CYCLE DE LA SUISSE ROMANDE

Guido Caldarelli

Together with

C.Caretta, M. Catanzaro, F. Colaiori, D. Garlaschelli, L. Pietronero, V. Servedio

INFM and Dipartimento di Fisica, Università "La Sapienza" Roma, Italy

L. Laura, S. Leonardi, S. Millozzi, A. Marchetti-Spaccamela

Dipartimento di Sistemistica e Elettronica, Università "La Sapienza" Roma Italy

P. De Los Rios[a], G. Bianconi[b], A. Capocci[b]

[a]Université de Lausanne, [b]Université de Fribourg, Switzerland

S. Battiston, A. Vespignani

Ecole Normale Supériéure and Université de Paris Sud, Paris France

•Contents

Part 1 *20-11-2003*

BASICS

A. Networks as complex structures
B. Fractals, Self-similarity
C. Self-organization
D. Evidence of scale-free networks
E. Basic of Graphs

Part 2 *27-11-2003*

REAL GRAPHS

A. Technological data: Internet, WWW
B. Social data: Finance and Board of Directors
C. Biological data: Proteins

Part 3 *4-12-2003*

REAL TREES

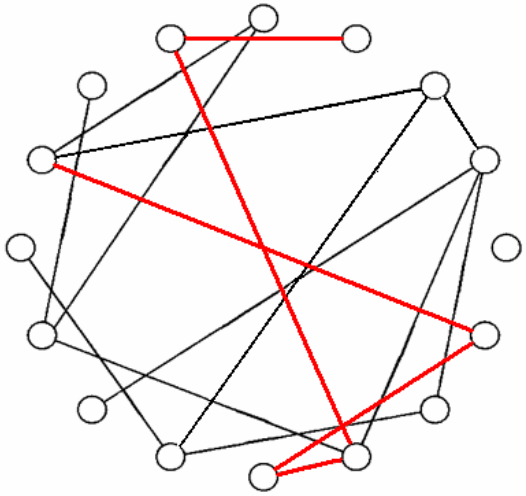
A. Food Webs
B. Geophysical data: the River Networks
C. Biological data: Taxonomy and Community Structures

Part 4 *11-12-2003*

MODELS

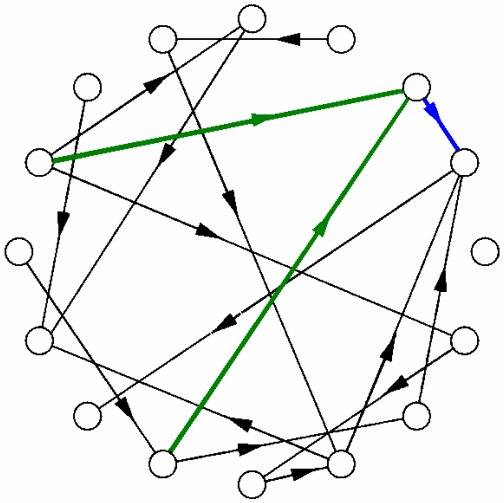
A. Random Graphs (Erdös-Renyi)
B. Small world
C. Preferential attachment
D. Fitness models

•4A Graph Definitions

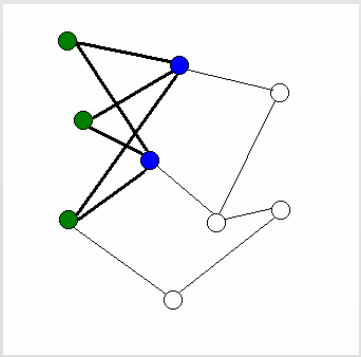


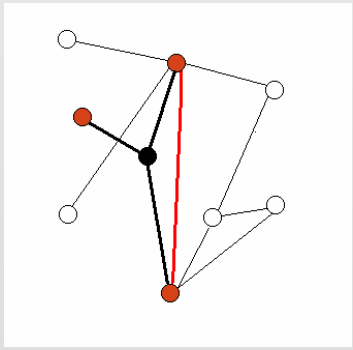
A **Graph** ***G(N,m)*** is an object composed by ***N*** *vertices* and ***m*** *edges*

Edges can be oriented →



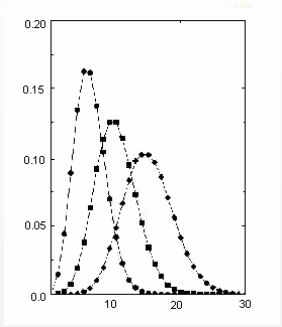
· Degree k (In-degree k_{in} and out-degree k_{out}) = number of edges (oriented) per vertex
· Distance d = number of edges amongst two vertices (in the connected region !)
· Diameter D = Maximum of the distances (in the connected region !)
· Clustering = cliques distribution, or clustering coefficient





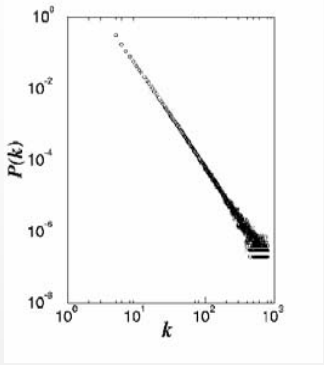
•4A Statistical measures

·1 Degree frequency density P(k) = how many times you find a vertex whose degree is k

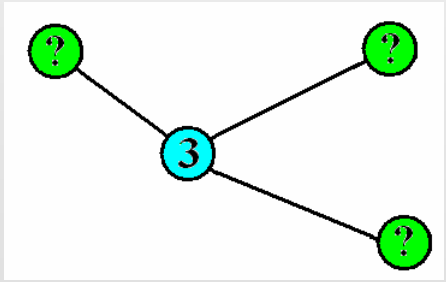


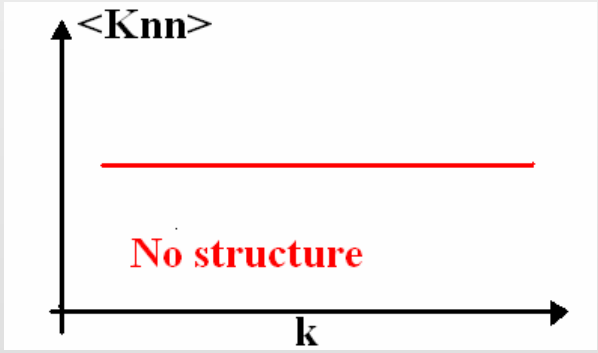
$$P(k) = e^{-pN}\frac{(pN)^k}{k!}$$

$$P(k) \propto k^{-\gamma}$$



·2 Degree Correlation Knn (k) = average degree of a neighbour of a vertex with degree k



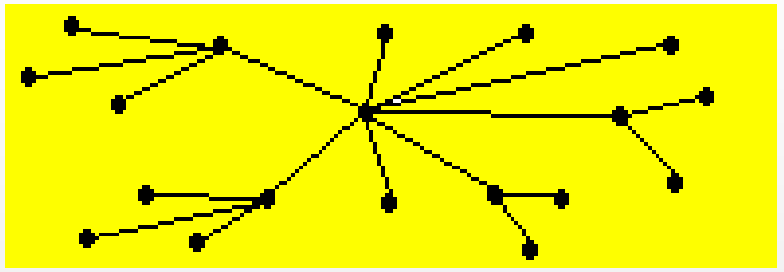


·3 Clustering Coefficient C(k) = the average value of c for a vertex whose degree is k

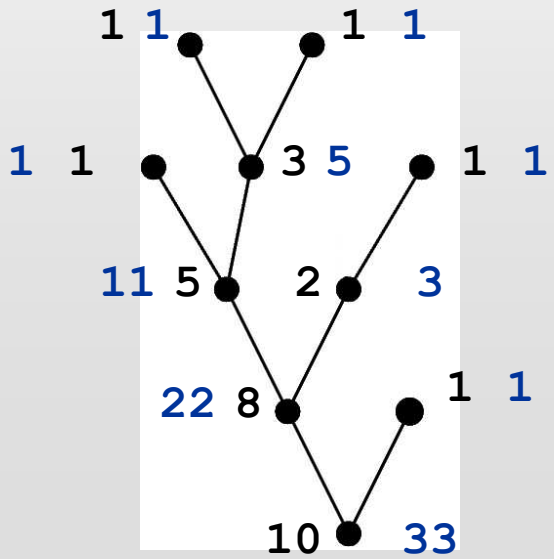
•4A Statistical measures

·4 Centrality betweenness b(k) = The probability that a vertex whose degree is k has betweenness b

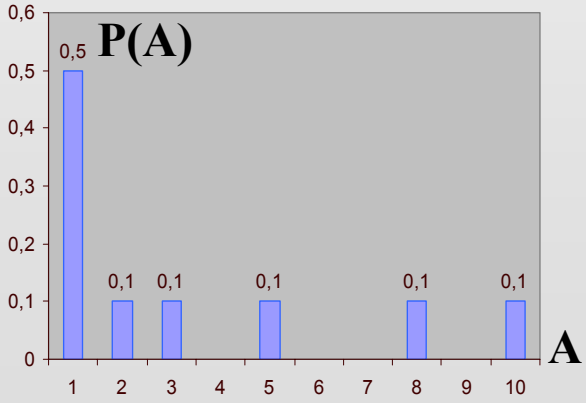
betweenness of I is the number of distances between any pair of vertices passing through I



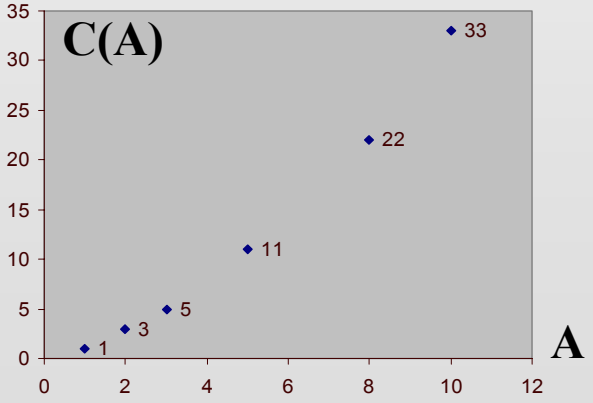
·5 TREES ONLY!!! P(A) = Probability Density for subbranches of size A



Size distribution:



Allometric relations:

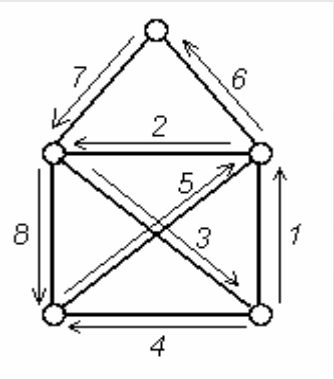


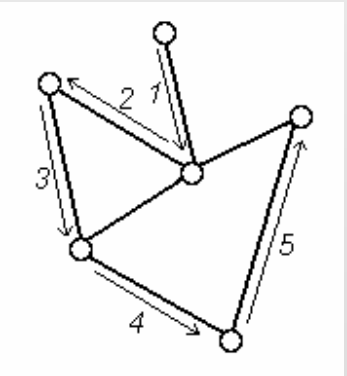
•4A Boring stuff (1/3)

- The graph **size** *n* is the number of its vertices.
- The graph **measure** *m* is the number of its edges.
- The **degree** of a vertex in a graph is the number of edges that connects it to other vertices.
- In the case of an oriented graph the degree can be distinguished in **in-degree** and **out-degree**.
- Whenever all the vertices share the same degree the graph is called **regular**.
- A series of consecutive edges forms a **path**.
  - The number of edges in a path is called the length of the path.
  - A Hamiltonian path is a path that passes once through all the vertices (not necessarily through all the edges) in the graph.
  - A Hamiltonian cycle is a Hamiltonian path which begins and ends in the same vertex.
  - An Eulerian path is a path that passes once through all the edges (not necessarily once through all the vertices) in the graph.
  - An Eulerian cycle is an Eulerian path which begins and ends in the same edge.

•4A Boring stuff (2/3)

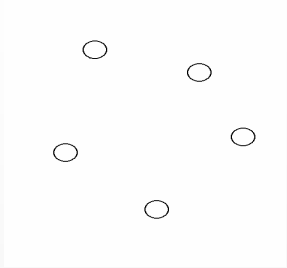
- Whenever all the vertices share the same degree the graph is called **regular**.
- A series of consecutive edges forms a **path**.
  - The number of edges in a path is called the length of the path.
  - A Hamiltonian path is a path that passes once through all the vertices (not necessarily through all the edges) in the graph.
  - A Hamiltonian cycle is a Hamiltonian path which begins and ends in the same vertex.
  - An Eulerian path is a path that passes once through all the edges (not necessarily once through all the vertices) in the graph.
  - An Eulerian cycle is an Eulerian path which begins and ends in the same edge.

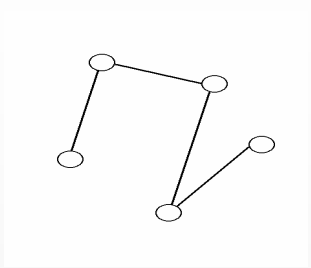


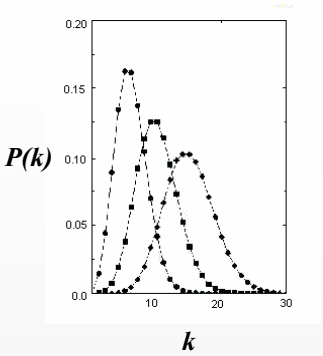


•4A Models (1)

Standard Theory of Random Graph
(Erdös and Rényi 1960)





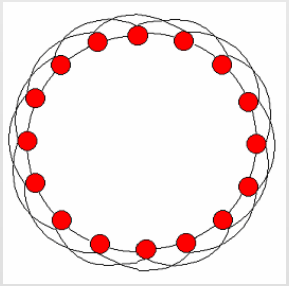


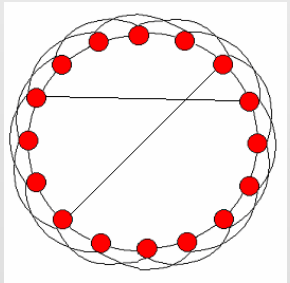
Random Graphs are composed by starting with n vertices.
With probability p two vertices are connected by an edge

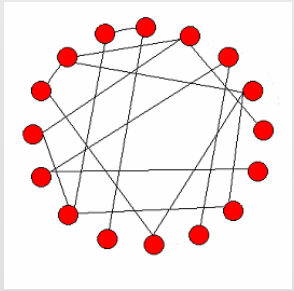
Degrees are Poisson distributed

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

Small World
(D.J. Watts and S.H. Strogatz 1998)







Small World Graph are composed by adding shortcuts to regular lattices

Degrees are peaked around mean value

•4A Models (2)

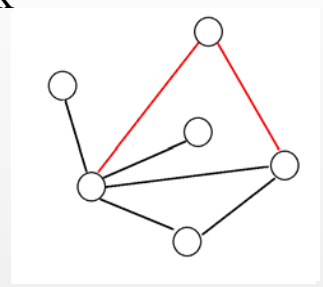
Model of Growing Networks
(A.-L. Barabási 1999)

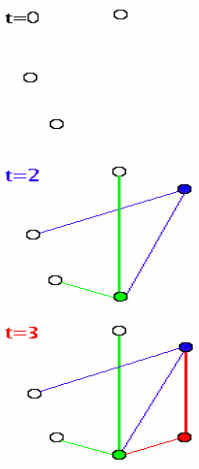
1) Growth

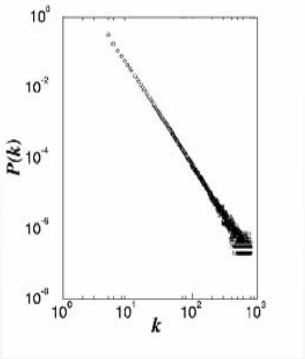
Every time step new nodes enter the system

2) Preferential Attachment

The probability to be connected depends on the degree P(k) ∝ k







Degrees are Power law distributed

$$P(k) \propto k^{-\gamma}$$

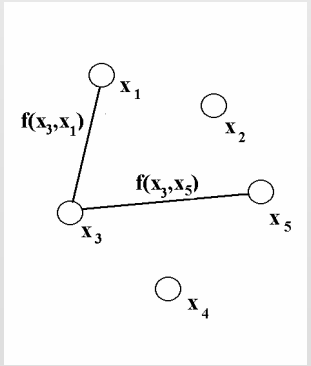
Intrinsic Fitness Model
(G.Caldarelli A. Capocci, P.De Los Rios, M.A. Munoz 2002)

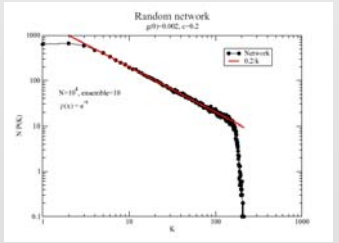
1) Growth or not

Nodes can be fixed at the beginning or be added

2) Attachment is related to intrinsic properties

The probability to be connected depends on the sites





Degrees are Power law distributed

$$P(k) \propto k^{-\gamma}$$

•4A Random Graphs (1)

A general derivation of interesting formulas for RG is present on
B. Bollobas *Graph Theory: an Introductory Course* (Springer-Verlag, New York, 1977)
Here we will present some results
One elegant approach is given by Generating Function approach

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

The P_k is the probability that a random vertex has a degree k $\longrightarrow$ $G_0(1) = 1$

$$P_k = \frac{1}{k!} \frac{\partial^k G_0(x)}{\partial x^k}\bigg|_{x=0}$$

$$\langle k \rangle = \sum k P_k = G_0'(1)$$

•4A Random Graphs (2)

•**The number m of links in a Random Graph is a random variable whose expectation value is**

$$E(m) = p\frac{N(N-1)}{2}$$

•**The probability to form a particular Graph *G(N,m)* is given by**

$$E(G(N,m)) = p^m (1-p)^{\frac{N(N-1)}{2}-m}$$

•**The degree has expectation value**

$$E(k) = 2m/N = p(N-1) \cong pN$$

Therefore the degree probability distribution is given by

$$P(k) = \binom{N-1}{k} p^k (1-p)^{(N-1)-k} \cong \frac{(pN)^k e^{-pN}}{k!}$$

•4A Random Graphs (3)

•We can give an estimate of the Clustering Coefficient for a complete graph it must be *1*.
If the graph is enough sparse then two points link each other with probability *p*

$$E(C) \cong p \cong \frac{< k >}{N}$$

•Same estimate can be given for the average distance l between two vertices.
If a graph has <*k*> average degree then
the first neighbours will be <*k*>
the second neighbours ≤ <*k*>^{2}
……..
the n-th neighbours ≤ <*k*>^{n}

•For the Diameter *D* → <*k*>^{D} *of order N*

$$l \leq D \cong \frac{\log(N)}{\log(k)}$$

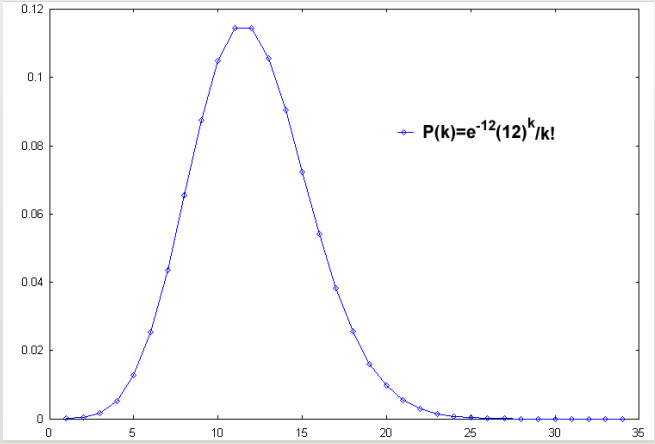
•4A Random Graphs (4)

$$G_0(x) \equiv \sum_{k=0}^{\infty} P_k x^k = \sum_{k=0}^{\infty} \binom{n-1}{k} p^k (1-p)^{(n-1)-k} x^k = (1-p+px)^{n-1} \cong e^{pn(x-1)}$$

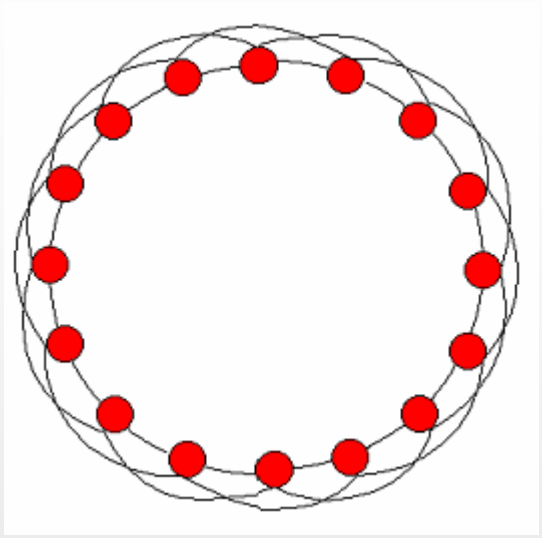
$$\langle k \rangle = \sum kP_k = G_0'(1) = pn$$

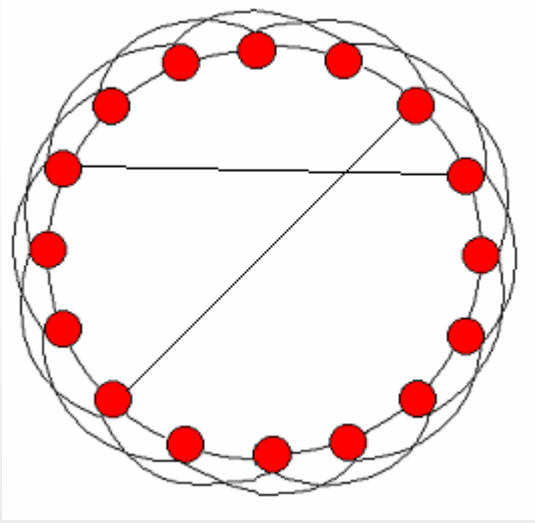
$$P_k = \frac{1}{k!} \frac{\partial^k G_0(x)}{\partial x^k}\bigg|_{x=0} = \frac{(pn)^k}{k!} e^{-pn}$$

As in the ordinary Poisson Distribution



•4B Small World (1)





Take a regular lattice and add *rewire* with probability ϕ some of the links
for analytical treatment, a slight modification is recommended.
Instead of rewiring add the new links proportional to the existing links

The total number of shortcuts is $L\chi\varphi \qquad (\chi = 2)$

Average degree is now $k = 2\chi(1+\varphi)$

Therefore for small ϕ the degree distribution is peaked around 2χ

•4B Small World (2)

Clustering Coefficient of the regular lattice
($\phi \to 0$ and k< 2/3N otherwise C=1)

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

For the average distance there is no result
but we can define a distance in the problem, given by the mean distance between two shortcuts endpoints.

We have that in the regular lattice *(start with χ=1 and generalize)*

$$l_i = \frac{1}{N}\sum_{j=1,L} d_{i,j} = \frac{1}{N} 2(0+1+2+...+N/2) = \frac{2}{N}\frac{N^2}{8} = \frac{N}{4} \rightarrow \quad l_i = \frac{N}{4\chi}$$

We have that in the Random Graph

$$l \cong \frac{\log(N)}{\log(k)}$$

•4B Small World (3)

Several conjectures, made but neither the actual distribution of path lengths nor the <*l*> has been found

Now in Small World graphs, the behaviour must be intermediate between the regular lattice and Random Graph.
If we define a characteristic length in the system as for example
ξ = average distance between two endpoints of shortcuts (not the same!)

$$\xi = \frac{L}{2(L\chi\varphi)} = \frac{1}{2\chi\varphi}$$ **ξ diverges when $\phi \to 0$**

ξ is characteristic distance we can define in the model so that we make the ansatz

$$l = \xi G(L/\xi) = \frac{L}{2\varphi\chi} G(x)$$

$$G(x) = \begin{cases} 1 & x << 1 \\ \dfrac{\log(x)}{x} & x >> 1 \end{cases}$$

•4C BA model (1)

The basic approach is through *continuum theory*, degree is now a continuum variable:

Start with m_0 vertices and add $\forall t$ m new links

$$\frac{\partial k_i}{\partial t} = \frac{mk_i}{\sum_{j=1,N} k_j} = \frac{mk_i}{2tm} = \frac{k_i}{2t} \rightarrow k_i(t) = m\left(\frac{t}{t_i}\right)^{\beta}, \beta = \frac{1}{2}$$

As for the degree distribution we can compute the $P(k_i<k)$

$$P(k_i(t) < k) = P(t_i > t\frac{m^{1/\beta}}{k^{1/\beta}})$$

•4C BA model (2)

The distribution of incoming vertices is uniform in time

$$P(t_i) = \frac{1}{m_0 + t}$$

$$P(k_i(t) < k) = P(t_i > t\frac{m^{1/\beta}}{k^{1/\beta}}) = 1 - t\frac{m^{1/\beta}}{k^{1/\beta}}\frac{1}{(t+m_0)}$$

From which we obtain

$$P(k) = \frac{\partial P(k_i(t) < k)}{\partial k} = \frac{2tm^{1/\beta}}{k^{1/\beta+1}}\frac{1}{(t+m_0)} \xrightarrow[t\to\infty]{} 2m^{1/\beta}k^{-\gamma}$$

$$\boxed{\gamma = \frac{1}{\beta} + 1 = 3}$$

•4C BA model (3)

Same result can be obtained from Rate equation approach where $N_k(t)$ is the number of nodes whose degree is k

$$P(k) = \frac{dN_k}{dt} = m\frac{(k-1)N_{k-1}(t) - kN_k(t)}{\sum_{k=0,N} kN_k(t)} + \delta_{k,m}$$

Asymptotically one obtains the same result since

$$\begin{cases} N_k(t) = tP(k) \\ \sum_{k=0,N} N_k(t) = 2mt \end{cases}$$

We can now check the robustness of preferential attachment with respect to different choice of function as for example $P(k) \propto k^{\alpha}$

•4C BA model (4)

The rate equation now is *(m=1)*

$$P(k) = \frac{dN_k}{dt} = \frac{(k-1)^\alpha N_{k-1}(t) - k^\alpha N_k(t)}{\sum_{k=0,N} k^\alpha N_k(t)} = \frac{(k-1)^\alpha N_{k-1}(t) - k^\alpha N_k(t)}{M_\alpha(t)} + \delta_{k,1}$$

$\boxed{\alpha << 1}$ **Sublinear case**

$$M_\alpha(t) = \mu t \rightarrow 1 \leq \mu = \mu(\alpha) \leq 2$$

$$P(k) = \frac{\mu}{k^\alpha} \prod_{j=1,k} (1 + \frac{\mu}{j^\alpha})^{-1}$$

This product can be expanded in series. The result is a stretched exponential

•4C BA model (5)

$\alpha >> 1$ **Superlinear case**

No analytical solution of

$$P(k) = \frac{dN_k}{dt} = \frac{(k-1)^\alpha N_{k-1}(t) - k^\alpha N_k(t)}{\sum_{k=0,N} k^\alpha N_k(t)} = \frac{(k-1)^\alpha N_{k-1}(t) - k^\alpha N_k(t)}{M_\alpha(t)} + \delta_{k,1}$$

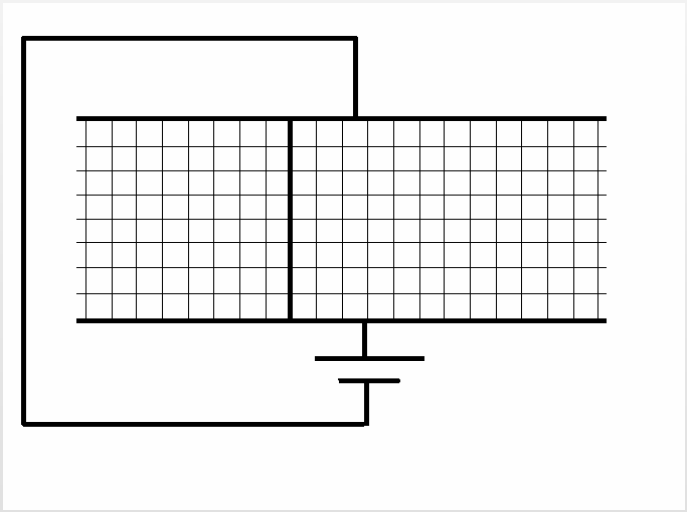
From recursion procedure some indication of the behaviour.
For α >2 there is one large hub + leaves
In general the number of nodes with degree larger than value *j* is finite
NO MORE SCALE FREE BEHAVIOUR

•4D Fitness Model

Without introducing growth or preferential attachment we can have power-laws
We consider "disorder" in the Random Graph model
(i.e. vertices differ one from the other).

This mechanism is responsible of self-similarity in Laplacian Fractals

•Dielectric Breakdown



•In a perfect dielectric

•In reality

•4D Fitness Model: Undirected Graphs

1. **Assign to every vertex one real positive number x that we call fitness. fitnesses are drawn from probablity distribution $\rho(x)$**

2. **Link two vertices with fitnesses x and y according to a probability function f(x,y)=f(y,x) (choice function).**

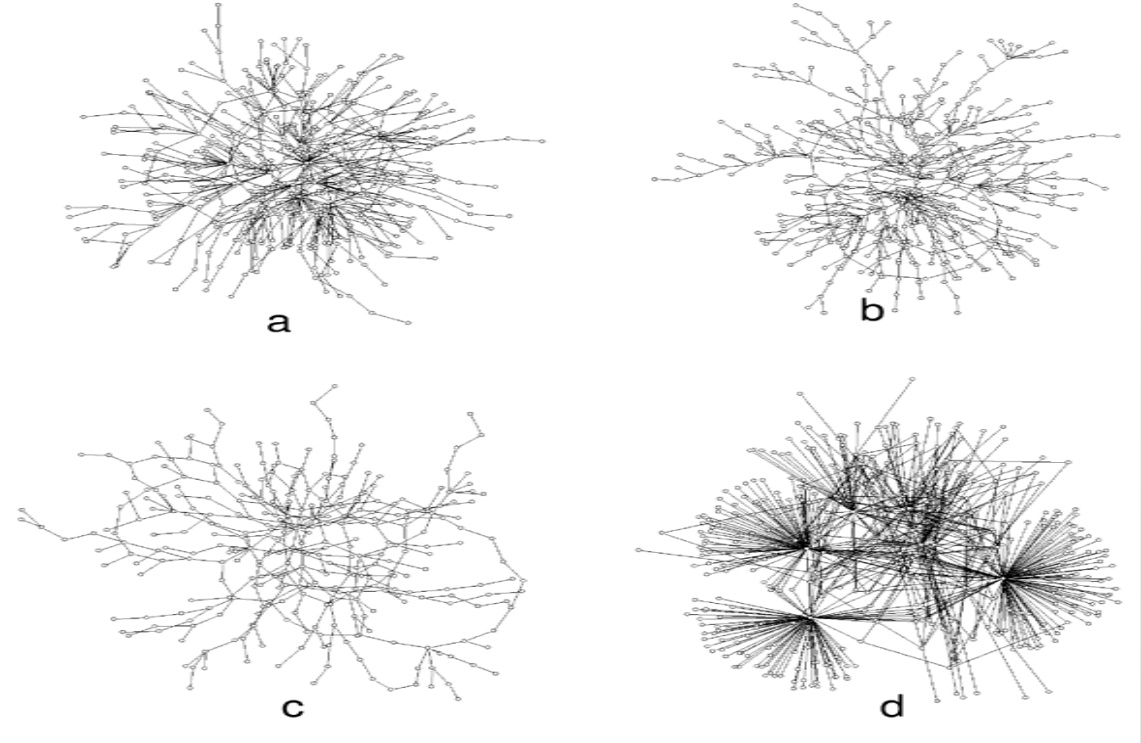
The model can be considered

- **STATIC** — **if N is kept fixed**
- **DYNAMIC** — **if N is growing**

This is a GOOD GETS RICHER model
No preferential attachment is present.

G. Caldarelli, A. Capocci, P. De Los Rios, M. A. Muñoz *Phys. Rev. Lett.* **89** 258702 (2002).
V.D.P. Servedio, P. Buttà, G. Caldarelli ArXiv:cond-mat/0309659 (2003).

•4D Fitness Model

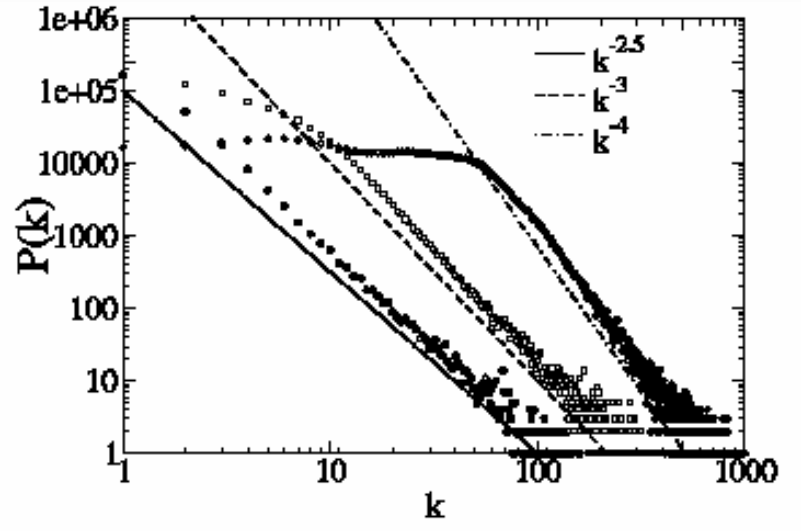


Different realizations of the model

a) b) c) have ρ(x) power law with exponent 2.5 ,3 ,4 respectively.

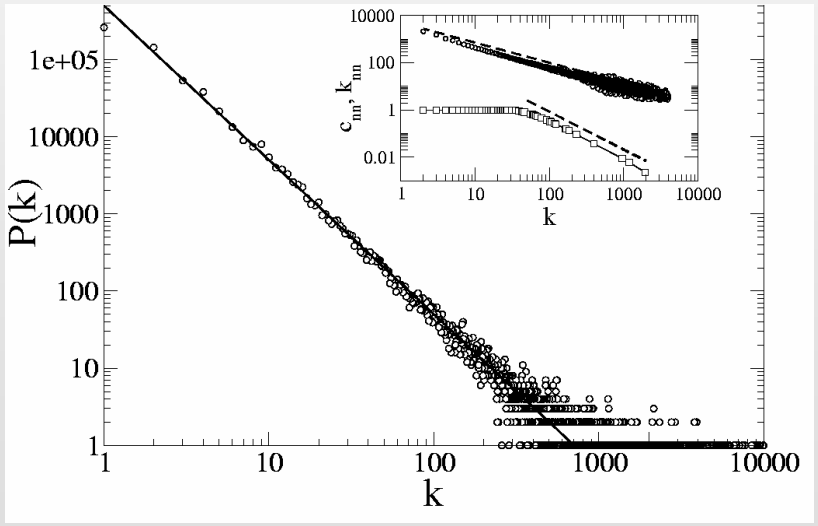
d) has ρ(x)=exp(-x) and a threshold rule.

•4D Fitness Model



Degree distribution for cases
a) b) c) with ρ(x) power law with exponent 2.5 ,3 ,4 respectively.

Degree distribution for the case
d) with ρ(x)=exp(-x) and a threshold rule.



•4D Fitness Model: Formulation of the Problem

The Degree probability distribution $P(k)$ is a functional of $\rho(x)$ and $f(x,y)$.

DIRECT PROBLEM

Given a fitness $\rho(x) \rightarrow$ **which choice function $f(x,y)$ produces scale free graphs? i.e. $P(k) = ck^{\alpha}$**

INVERSE PROBLEM

Given a **choice function $f(x,y)$ $\rightarrow$ which** fitness $\rho(x)$ **produces scale free graphs? i.e. $P(k) = ck^{\alpha}$**

•4D Fitness Model: Useful formulas

- Fitness probability distribution

$$R(x) = \int_o^x \rho(y)dy \qquad \rho(y) \geq 0 \rightarrow \begin{cases} R(x) \\ R(\infty) = 1 \end{cases}$$ Non decreasing

---

- Vertex degree

$$k(x) \equiv \frac{K(x)}{N} \equiv \int_o^\infty \rho(y) f(x,y)dy \rightarrow 0 \leq k(x) \leq 1$$

- Vertex degree Probability Distribution

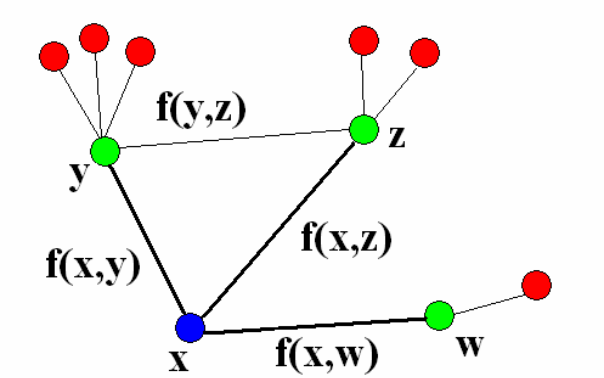
$$P(k(x))dk = \rho(x)dx \rightarrow P(k(x)) = \rho(x)x'(k)$$

$$\boxed{P(k(x)) = \frac{\rho(x)}{k'(x)}}$$

•4D Fitness Model: Useful formulas

- Degree Correlation

$$K_{nn}(x) = \frac{N\int_0^\infty f(x,y)k(y)\rho(y)dy}{k(x)}$$



- Vertex Clustering Coefficient

$$C(x) = \frac{\int\int_0^\infty f(x,z)f(y,z)f(x,y)\rho(y)\rho(z)dydz}{k^2(x)}$$

•4D Fitness Model: Form of P(k)

$$P(k) = \frac{\rho(x)}{k'(x)}$$

We impose $P(k)=c(k(x))^{\alpha} \rightarrow$

$$\frac{\rho(x)}{k'(x)} = c(k(x))^{\alpha}$$

Multipling both sides of the equation for *k'(x)* and integrating from *0* to *x*

$$k(x) = \begin{cases} \left(k_0^{-|\alpha+1|} - \frac{|\alpha+1|}{c} R(x) \right)^{-\frac{1}{\alpha+1}} & \alpha < 1 \\ k_0 e^{\frac{R(x)}{c}} & \alpha = 1 \\ \left(k_0^{\alpha+1} + \frac{\alpha+1}{c} R(x) \right)^{\frac{1}{\alpha+1}} & \alpha > 1 \end{cases}$$

$$k(x) \equiv \frac{K(x)}{N} \equiv \int_o^{\infty} \rho(y) f(x,y) dy \rightarrow k_0 \equiv k(0)$$

$$R(x) = \int_o^x \rho(y) dy$$

•4D Fitness Model: Form of P(k)

We now have a constraint on the fitness distribution $\rho(x)$ and choice function $f(x,y)$

Some exact results

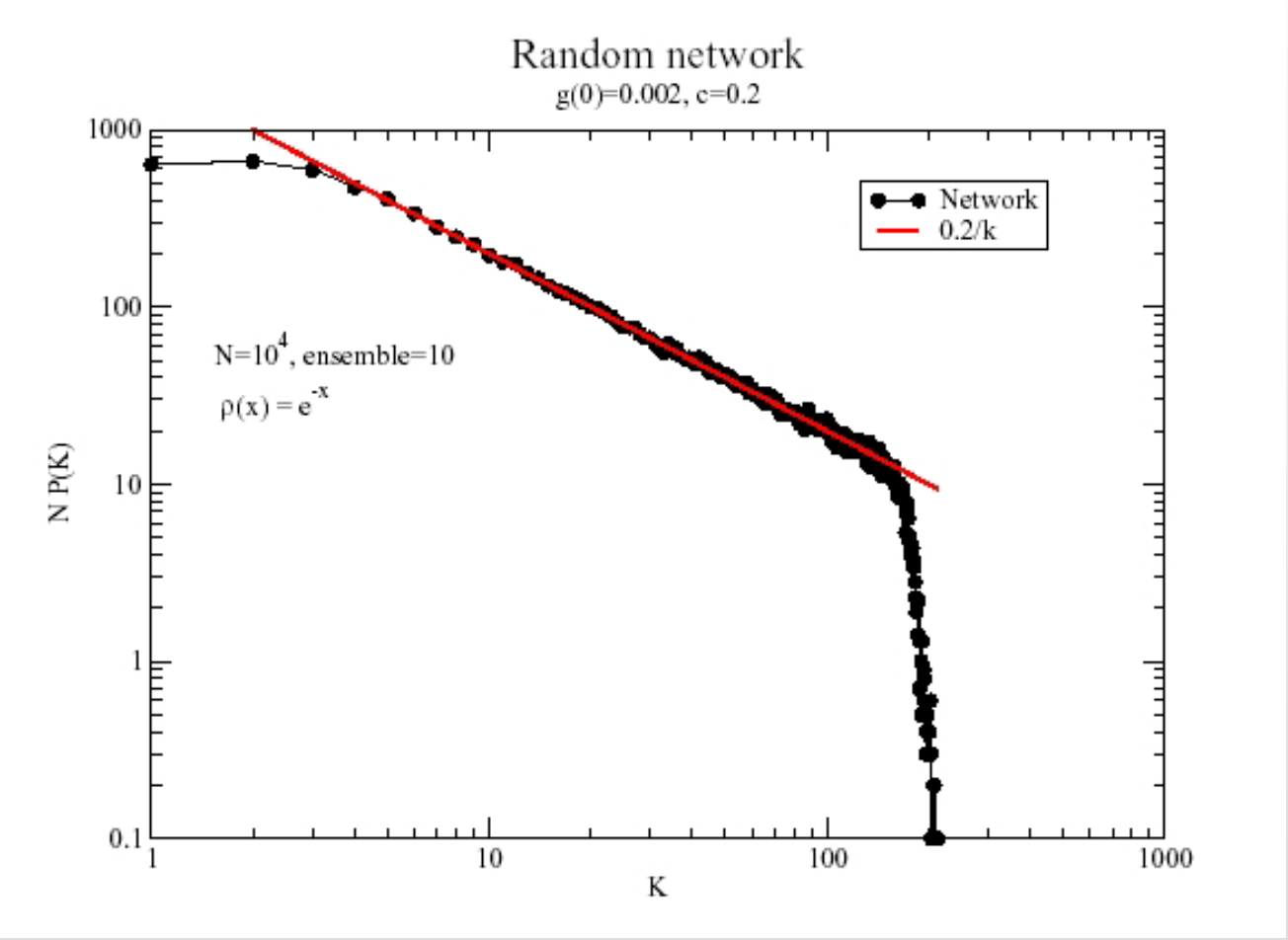
$$f(x,y) = g(x)g(y) = \frac{1}{< k >^2} k(x)k(y) \rightarrow \forall \rho(x)$$

$$f(x,y) = f(x \pm y) = k(x \pm y) \mp k'(x \pm y) \rightarrow \rho(x) = e^{-x}$$

$$k(x) = \begin{cases} \left(k_0^{-|\alpha+1|} - \frac{|\alpha+1|}{c} R(x) \right)^{-\frac{1}{\alpha+1}} & \alpha < 1 \\ k_0 e^{\frac{R(x)}{c}} & \alpha = 1 \\ \left(k_0^{\alpha+1} + \frac{\alpha+1}{c} R(x) \right)^{\frac{1}{\alpha+1}} & \alpha > 1 \end{cases}$$

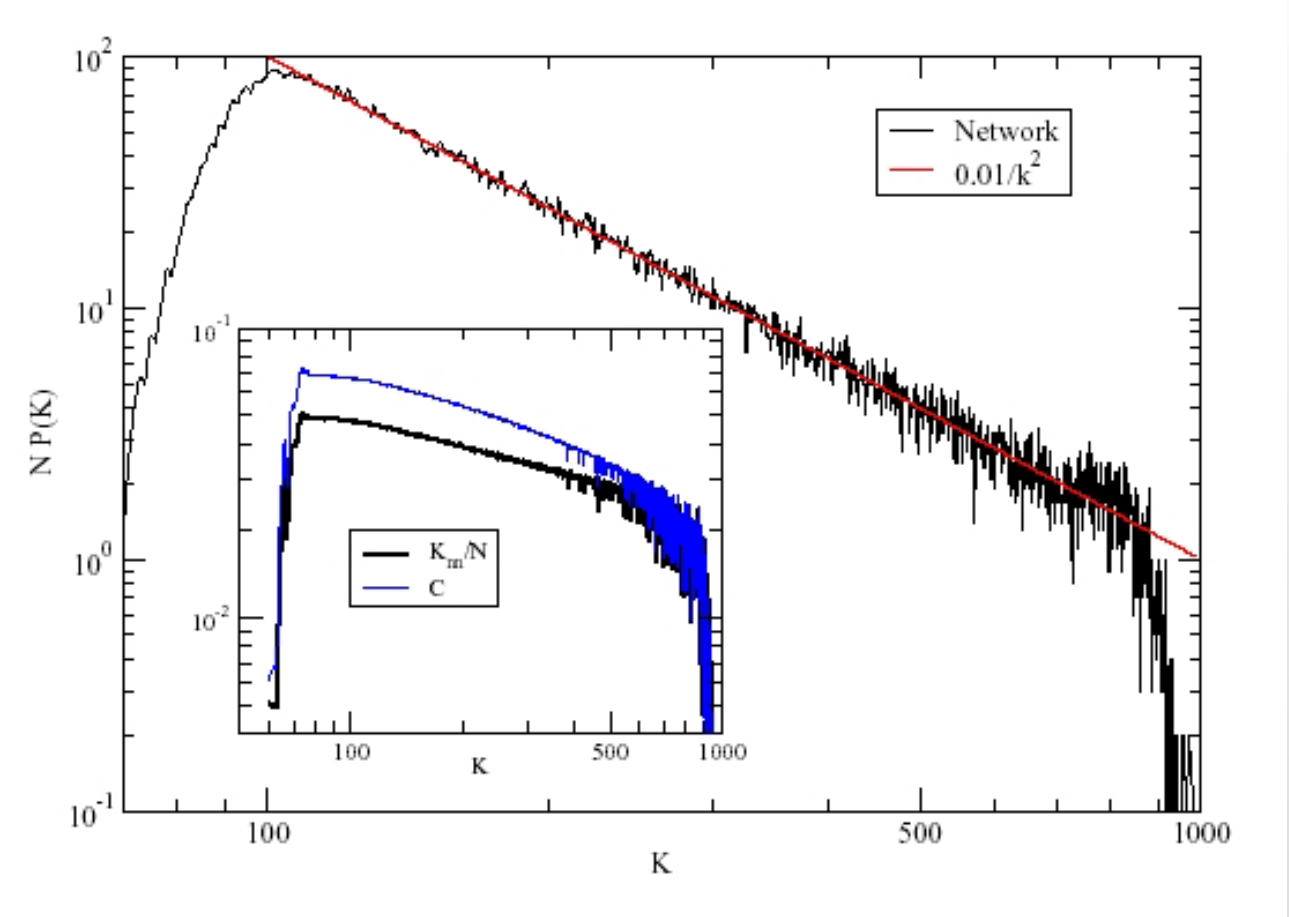
•4D Fitness Model: Exact cases

Special case $f(x,y)=g(x)g(y)$



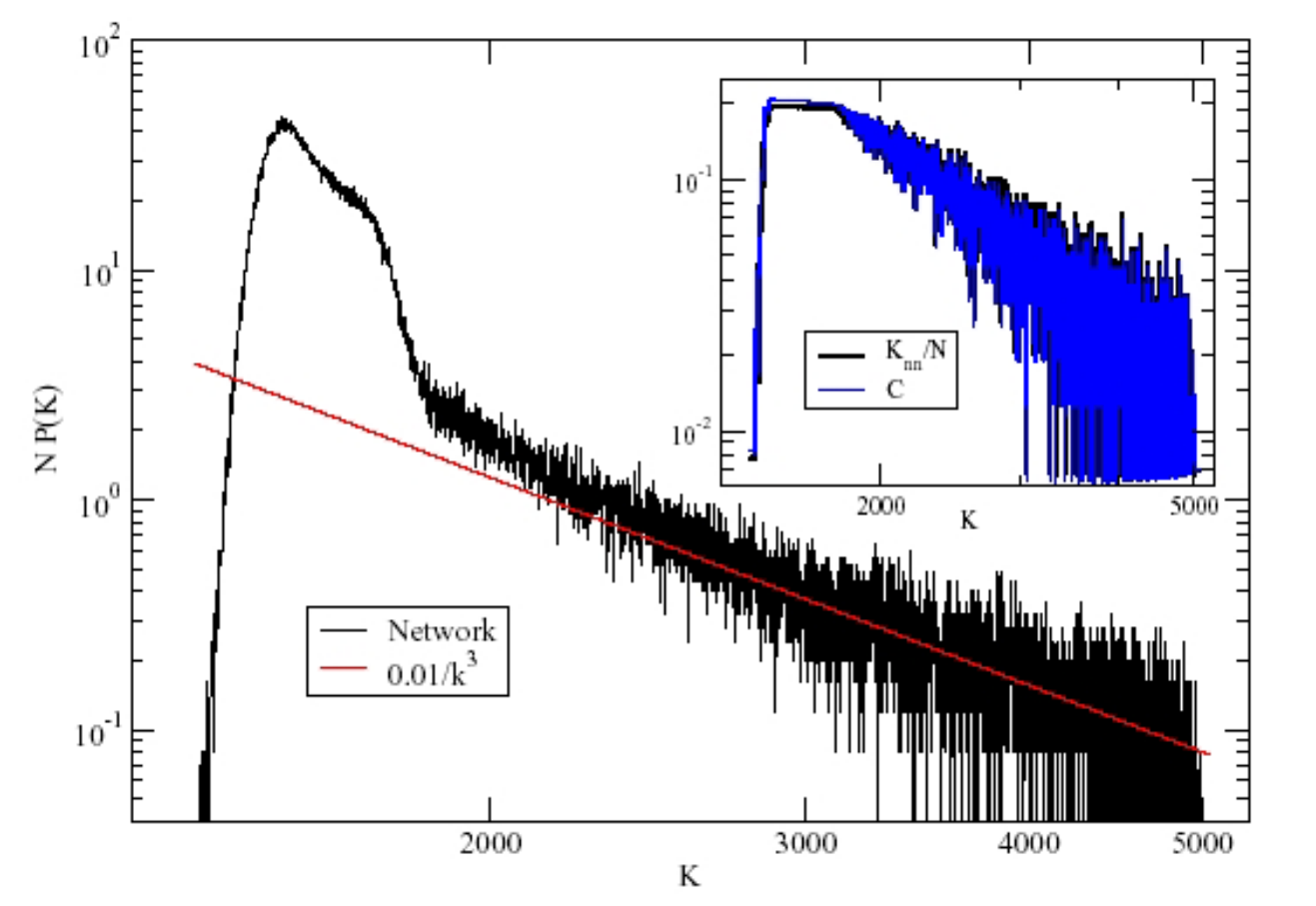
•4D Fitness Model: Exact cases

Special case $f(x,y)=f(x+y)$



•4D Fitness Model: Form of P(k)

Special case $f(x,y)=f(x-y)$



•4D Fitness Model: Conclusions

- **Using the intrinsic fitness model it is possible to create scale-free networks with any desired power-law exponent**
- **This is possible for any fitness probability distribution $\rho(x)$, it does not matter if they are *(e.g.)* exponential, power-law or Gaussian.**
- **We found analytic expressions for the choice function f(x,y) in three cases:**
  - **$f(x,y)=f(x)f(y)\ \forall\ \rho(x)$,**
  - **$f(x,y)=f(x \pm y)$ $\rho(x)=e^{-x}$**
- **If f(x,y)=f(x)f(y) both vertex degree correlation and clustering coefficient are constant**

•4E Data and Models

TABLE I. The general characteristics of several real networks. For each network we have indicated the number of nodes, the average degree $\langle k \rangle$, the average path length ℓ, and the clustering coefficient C. For a comparison we have included the average path length ℓ_{rand} and clustering coefficient C_{rand} of a random graph of the same size and average degree. The numbers in the last column are keyed to the symbols in Figs. 8 and 9.

Network	Size	$\langle k \rangle$	ℓ	ℓ_{rand}	C	C_{rand}	Reference	Nr.
WWW, site level, undir.	153 127	35.21	3.1	3.35	0.1078	0.00023	Adamic, 1999	1
Internet, domain level	3015–6209	3.52–4.11	3.7–3.76	6.36–6.18	0.18–0.3	0.001	Yook *et al.*, 2001a, Pastor-Satorras *et al.*, 2001	2
Movie actors	225 226	61	3.65	2.99	0.79	0.00027	Watts and Strogatz, 1998	3
LANL co-authorship	52 909	9.7	5.9	4.79	0.43	1.8×10^{-4}	Newman, 2001a, 2001b, 2001c	4
MEDLINE co-authorship	1 520 251	18.1	4.6	4.91	0.066	1.1×10^{-5}	Newman, 2001a, 2001b, 2001c	5
SPIRES co-authorship	56 627	173	4.0	2.12	0.726	0.003	Newman, 2001a, 2001b, 2001c	6
NCSTRL co-authorship	11 994	3.59	9.7	7.34	0.496	3×10^{-4}	Newman, 2001a, 2001b, 2001c	7
Math. co-authorship	70 975	3.9	9.5	8.2	0.59	5.4×10^{-5}	Barabási *et al.*, 2001	8
Neurosci. co-authorship	209 293	11.5	6	5.01	0.76	5.5×10^{-5}	Barabási *et al.*, 2001	9
E. coli, substrate graph	282	7.35	2.9	3.04	0.32	0.026	Wagner and Fell, 2000	10
E. coli, reaction graph	315	28.3	2.62	1.98	0.59	0.09	Wagner and Fell, 2000	11
Ythan estuary food web	134	8.7	2.43	2.26	0.22	0.06	Montoya and Solé, 2000	12
Silwood Park food web	154	4.75	3.40	3.23	0.15	0.03	Montoya and Solé, 2000	13
Words, co-occurrence	460.902	70.13	2.67	3.03	0.437	0.0001	Ferrer i Cancho and Solé, 2001	14
Words, synonyms	22 311	13.48	4.5	3.84	0.7	0.0006	Yook *et al.*, 2001b	15
Power grid	4941	2.67	18.7	12.4	0.08	0.005	Watts and Strogatz, 1998	16
C. Elegans	282	14	2.65	2.25	0.28	0.05	Watts and Strogatz, 1998	17

R.Albert A.-L. Barabási Statistical Mechanics of Complex Networks
Review of Modern Physics **74** 47 (2002).

•4E Data and Models

TABLE II. The scaling exponents characterizing the degree distribution of several scale-free networks, for which $P(k)$ follows a power law (2). We indicate the size of the network, its average degree $\langle k \rangle$, and the cutoff κ for the power-law scaling. For directed networks we list separately the indegree (γ_{in}) and outdegree (γ_{out}) exponents, while for the undirected networks, marked with an asterisk (*), these values are identical. The columns l_{real}, l_{rand}, and l_{pow} compare the average path lengths of real networks with power-law degree distribution and the predictions of random-graph theory (17) and of Newman, Strogatz, and Watts (2001) [also see Eq. (63) above], as discussed in Sec. V. The numbers in the last column are keyed to the symbols in Figs. 8 and 9.

Network	Size	$\langle k \rangle$	κ	γ_{out}	γ_{in}	ℓ_{real}	ℓ_{rand}	ℓ_{pow}	Reference	Nr.
WWW	325 729	4.51	900	2.45	2.1	11.2	8.32	4.77	Albert, Jeong, and Barabási 1999	1
WWW	4×10^7	7		2.38	2.1				Kumar *et al.*, 1999	2
WWW	2×10^8	7.5	4000	2.72	2.1	16	8.85	7.61	Broder *et al.*, 2000	3
WWW, site	260 000				1.94				Huberman and Adamic, 2000	4
Internet, domain*	3015–4389	3.42–3.76	30–40	2.1–2.2	2.1–2.2	4	6.3	5.2	Faloutsos, 1999	5
Internet, router*	3888	2.57	30	2.48	2.48	12.15	8.75	7.67	Faloutsos, 1999	6
Internet, router*	150 000	2.66	60	2.4	2.4	11	12.8	7.47	Govindan, 2000	7
Movie actors*	212 250	28.78	900	2.3	2.3	4.54	3.65	4.01	Barabási and Albert, 1999	8
Co-authors, SPIRES*	56 627	173	1100	1.2	1.2	4	2.12	1.95	Newman, 2001b	9
Co-authors, neuro.*	209 293	11.54	400	2.1	2.1	6	5.01	3.86	Barabási *et al.*, 2001	10
Co-authors, math.*	70 975	3.9	120	2.5	2.5	9.5	8.2	6.53	Barabási *et al.*, 2001	11
Sexual contacts*	2810			3.4	3.4				Liljeros *et al.*, 2001	12
Metabolic, *E. coli*	778	7.4	110	2.2	2.2	3.2	3.32	2.89	Jeong *et al.*, 2000	13
Protein, *S. cerev.**	1870	2.39		2.4	2.4				Jeong, Mason, *et al.*, 2001	14
Ythan estuary*	134	8.7	35	1.05	1.05	2.43	2.26	1.71	Montoya and Solé, 2000	14
Silwood Park*	154	4.75	27	1.13	1.13	3.4	3.23	2	Montoya and Solé, 2000	16
Citation	783 339	8.57			3				Redner, 1998	17
Phone call	53×10^6	3.16		2.1	2.1				Aiello *et al.*, 2000	18
Words, co-occurrence*	460 902	70.13		2.7	2.7				Ferrer i Cancho and Solé, 2001	19
Words, synonyms*	22 311	13.48		2.8	2.8				Yook *et al.*, 2001b	20

R.Albert A.-L. Barabási Statistical Mechanics of Complex Networks
Review of Modern Physics **74** 47 (2002).

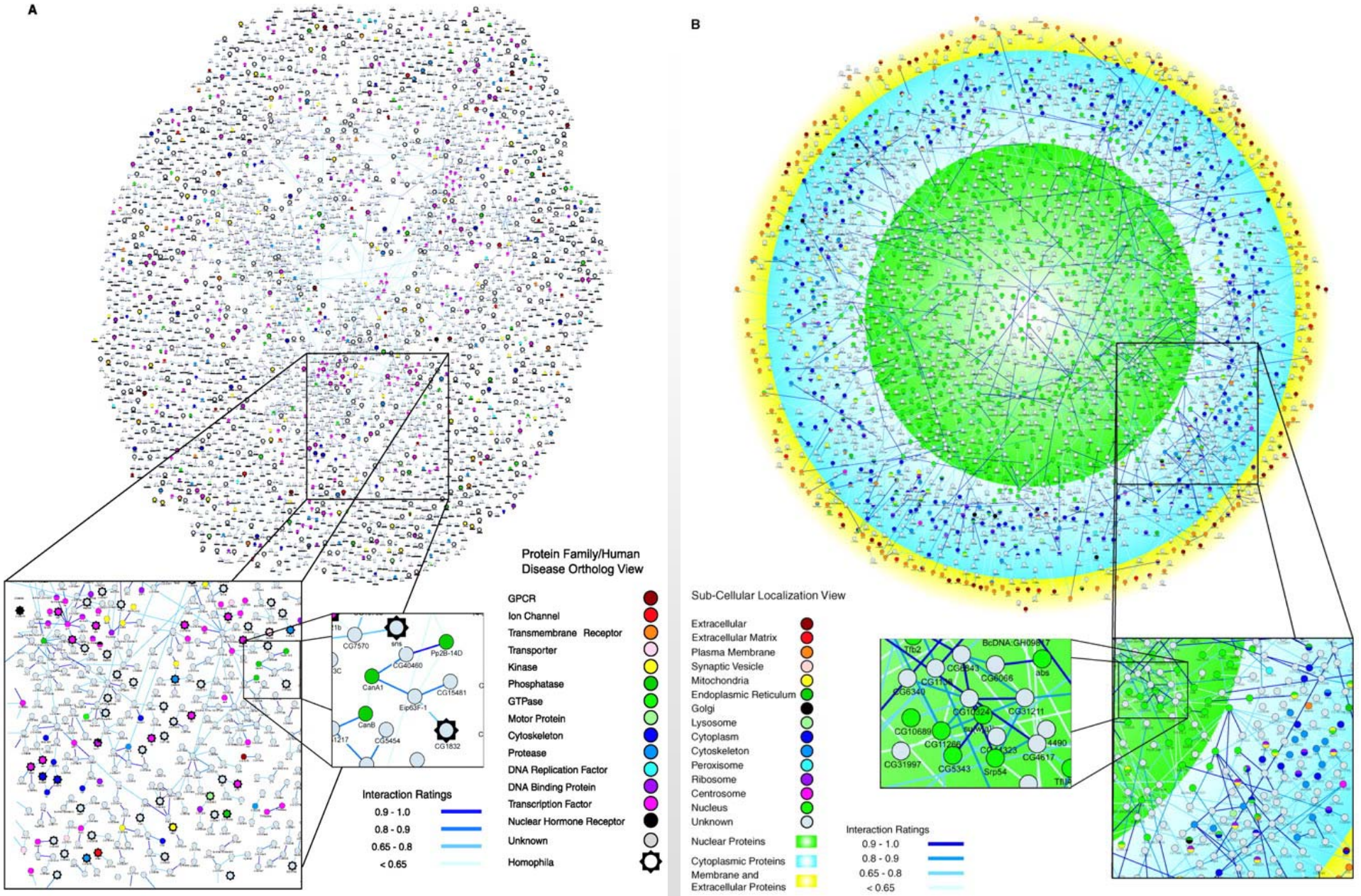
•4E Future?

Science, Vol. 302, Issue 5651, 1727-1736, December 5, 2003

A Protein Interaction Map of *Drosophila melanogaster*

L. Giot,[1*] J. S. Bader,[1*†] C. Brouwer,[1*] A. Chaudhuri,[1*] B. Kuang,[1] Y. Li,[1] Y. L. Hao,[1] C. E. Ooi,[1] B. Godwin,[1] E. Vitols,[1] G. Vijayadamodar,[1] P. Pochart,[1] H. Machineni,[1] M. Welsh,[1] Y. Kong,[1] B. Zerhusen,[1] R. Malcolm,[1] Z. Varrone,[1] A. Collis,[1] M. Minto,[1] S. Burgess,[1] L. McDaniel,[1] E. Stimpson,[1] F. Spriggs,[1] J. Williams,[1] K. Neurath,[1] N. Ioime,[1] M. Agee,[1] E. Voss,[1] K. Furtak,[1] R. Renzulli,[1] N. Aanensen,[1] S. Carrolla,[1] E. Bickelhaupt,[1] Y. Lazovatsky,[1] A. DaSilva,[1] J. Zhong,[2] C. A. Stanyon,[2] R. L. Finley, Jr.,[2] K. P. White,[3] M. Braverman,[1] T. Jarvie,[1] S. Gold,[1] M. Leach,[1] J. Knight,[1] R. A. Shimkets,[1] M. P. McKenna,[1] J. Chant,[1‡] J. M. Rothberg[1]

Drosophila melanogaster is a proven model system for many aspects of human biology. Here we present a two-hybrid–based protein-interaction map of the fly proteome. A total of 10,623 predicted transcripts were isolated and screened against standard and normalized complementary DNA libraries to produce a draft map of 7048 proteins and 20,405 interactions. A computational method of rating two-hybrid interaction confidence was developed to refine this draft map to a higher confidence map of 4679 proteins and 4780 interactions. Statistical modeling of the network showed two levels of organization: a short-range organization, presumably corresponding to multiprotein complexes, and a more global organization, presumably corresponding to intercomplex connections. The network recapitulated known pathways, extended pathways, and uncovered previously unknown pathway components. This map serves as a starting point for a systems biology modeling of multicellular organisms, including humans.


A
B
Protein Family/Human Disease Ortholog View
GPCR
Ion Channel
Transmembrane Receptor
Transporter
Kinase
Phosphatase
GTPase
Motor Protein
Cytoskeleton
Protease
DNA Replication Factor
DNA Binding Protein
Transcription Factor
Nuclear Hormone Receptor
Unknown
Homophila
Interaction Ratings
0.9 - 1.0
0.8 - 0.9
0.65 - 0.8
< 0.65
Sub-Cellular Localization View
Extracellular
Extracellular Matrix
Plasma Membrane
Synaptic Vesicle
Mitochondria
Endoplasmic Reticulum
Golgi
Lysosome
Cytoplasm
Cytoskeleton
Peroxisome
Ribosome
Centrosome
Nucleus
Unknown
Nuclear Proteins
Cytoplasmic Proteins
Membrane and Extracellular Proteins
Interaction Ratings
0.9 - 1.0
0.8 - 0.9
0.65 - 0.8
< 0.65

•4E Future?

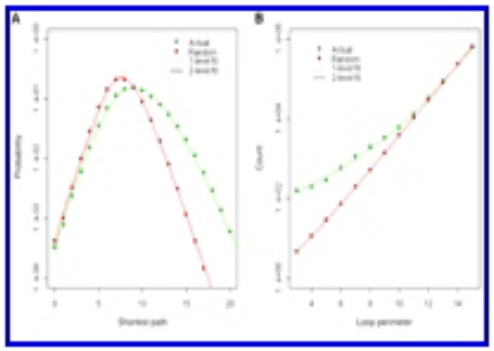
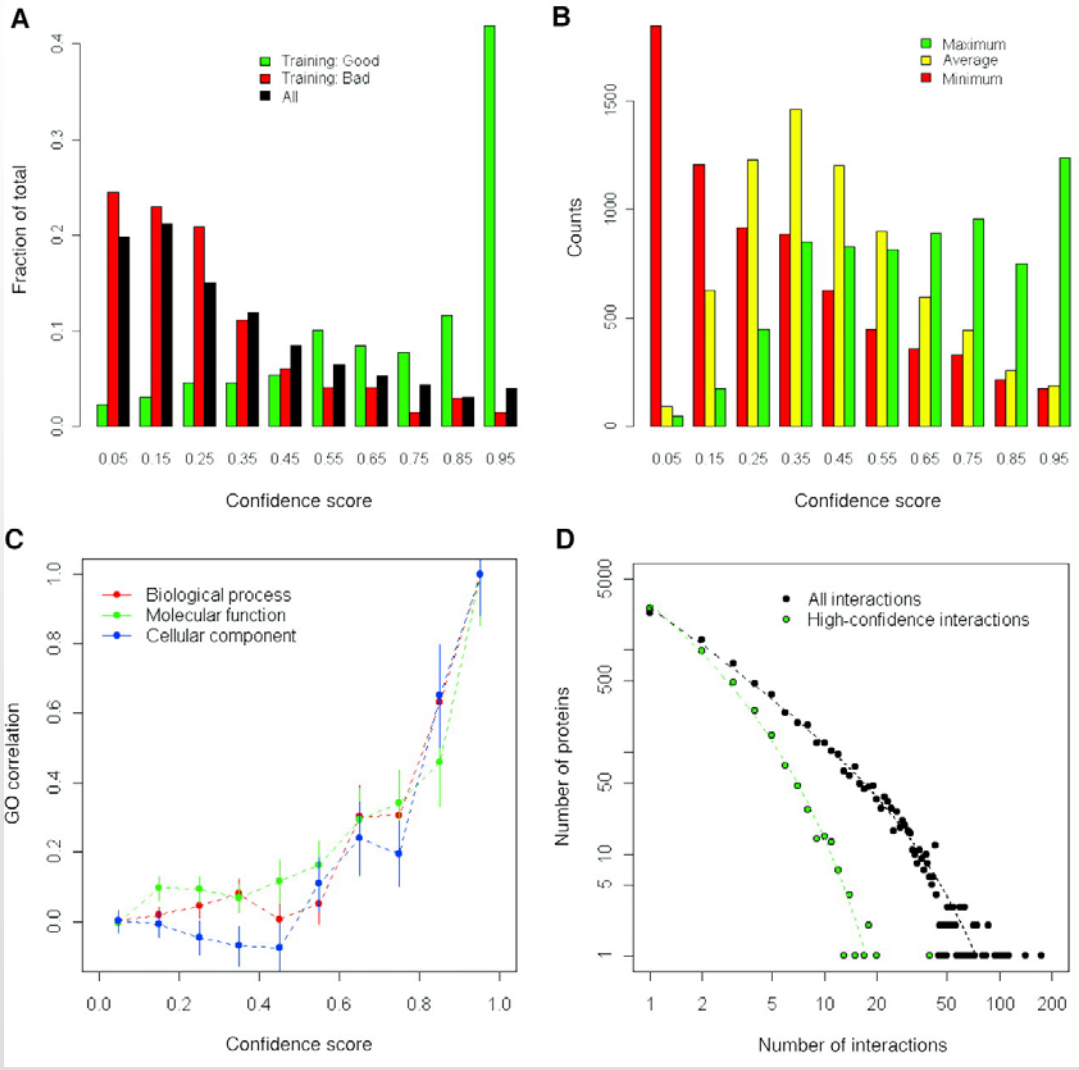


Fig. 3. Statistical properties of the refined *Drosophila* interaction map. The high-confidence *Drosophila* protein-protein interactions form a small-world network with evidence for a hierarchy of organization. Network properties are presented for the giant connected component, in which 3659 pairwise interactions connect 3039 proteins into a single cluster (see text for details). (**A**) The probability distribution for the shortest path between a pair of proteins in the actual network (green points) peaks at 9 to 11 links, with a mean of 9.4 links. In contrast, an ensemble of randomly rewired networks shows a mean separation of 7.7 links between proteins. Biological organization may be responsible for flattening the actual network by enhancing links between proteins that are already close. (**B**) Clustering, or enhancement of connections between proteins that are already close, is analyzed quantitatively by counting the number of closed loops (triangles, squares, pentagons, etc.) in which the perimeter is formed by a series of proteins connected head-to-tail, with no protein repeated. The actual network (green points) shows an enhancement of loops with perimeter up to 10 to 11 relative to the random network (red points). In both (A) and (B), the one-level and two-level models produce nearly indistinguishable fits for the random networks, indicating the absence of structured clustering. [View Larger Version of this Image (14K GIF file)]

•4E Future?


A
Training: Good
Training: Bad
All
Fraction of total
Confidence score
B
Maximum
Average
Minimum
Counts
Confidence score
C
Biological process
Molecular function
Cellular component
GO correlation
Confidence score
D
All interactions
High-confidence interactions
Number of proteins
Number of interactions

•4E Future?

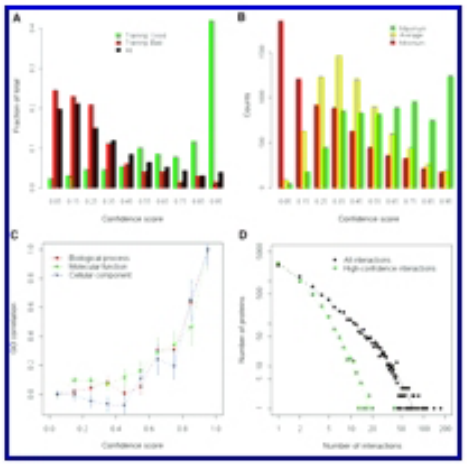
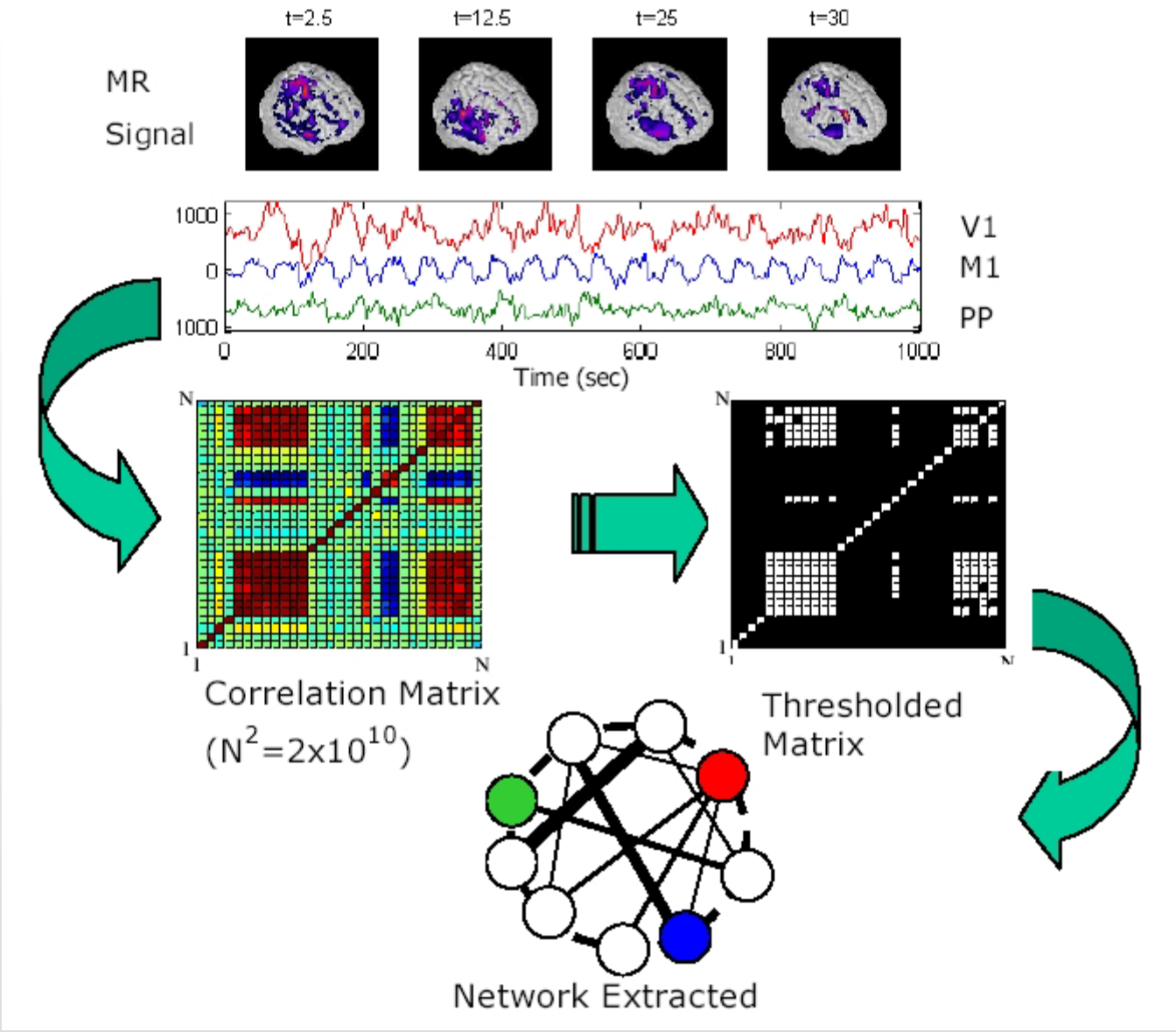
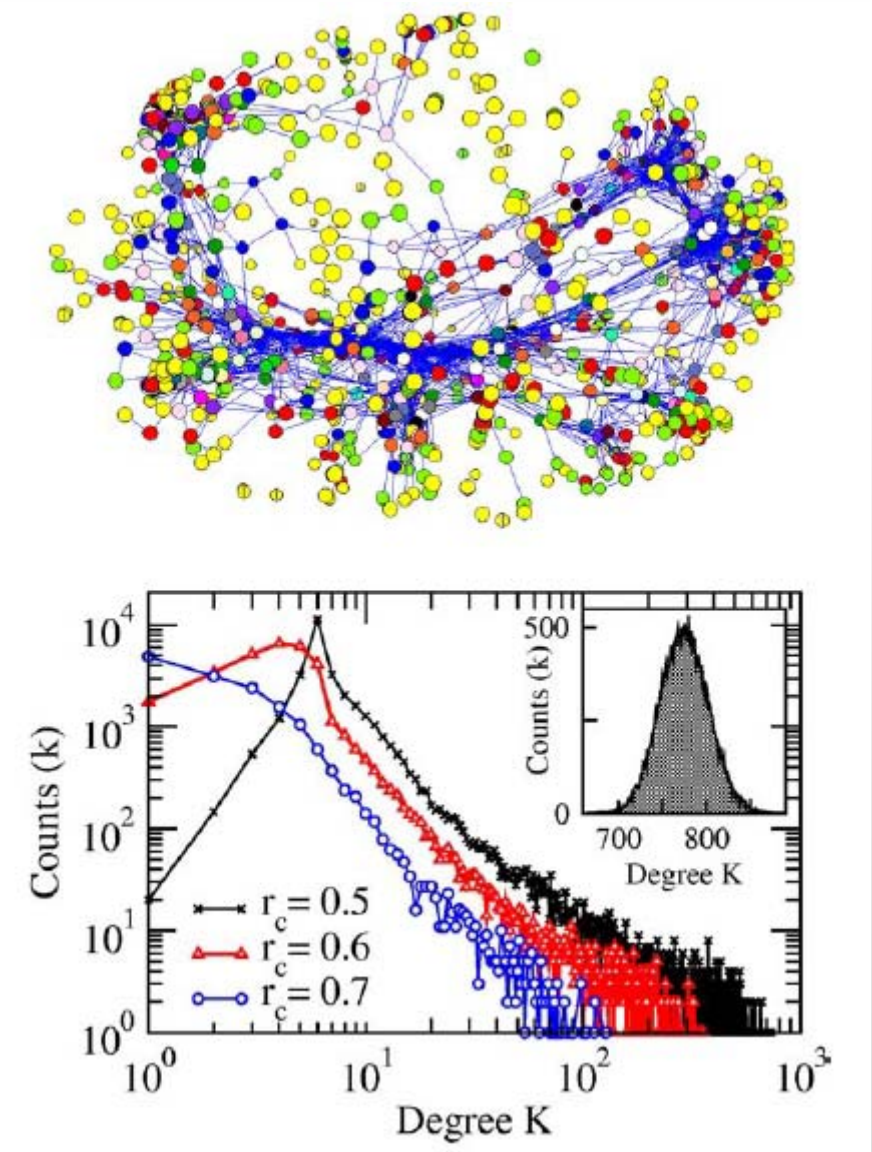


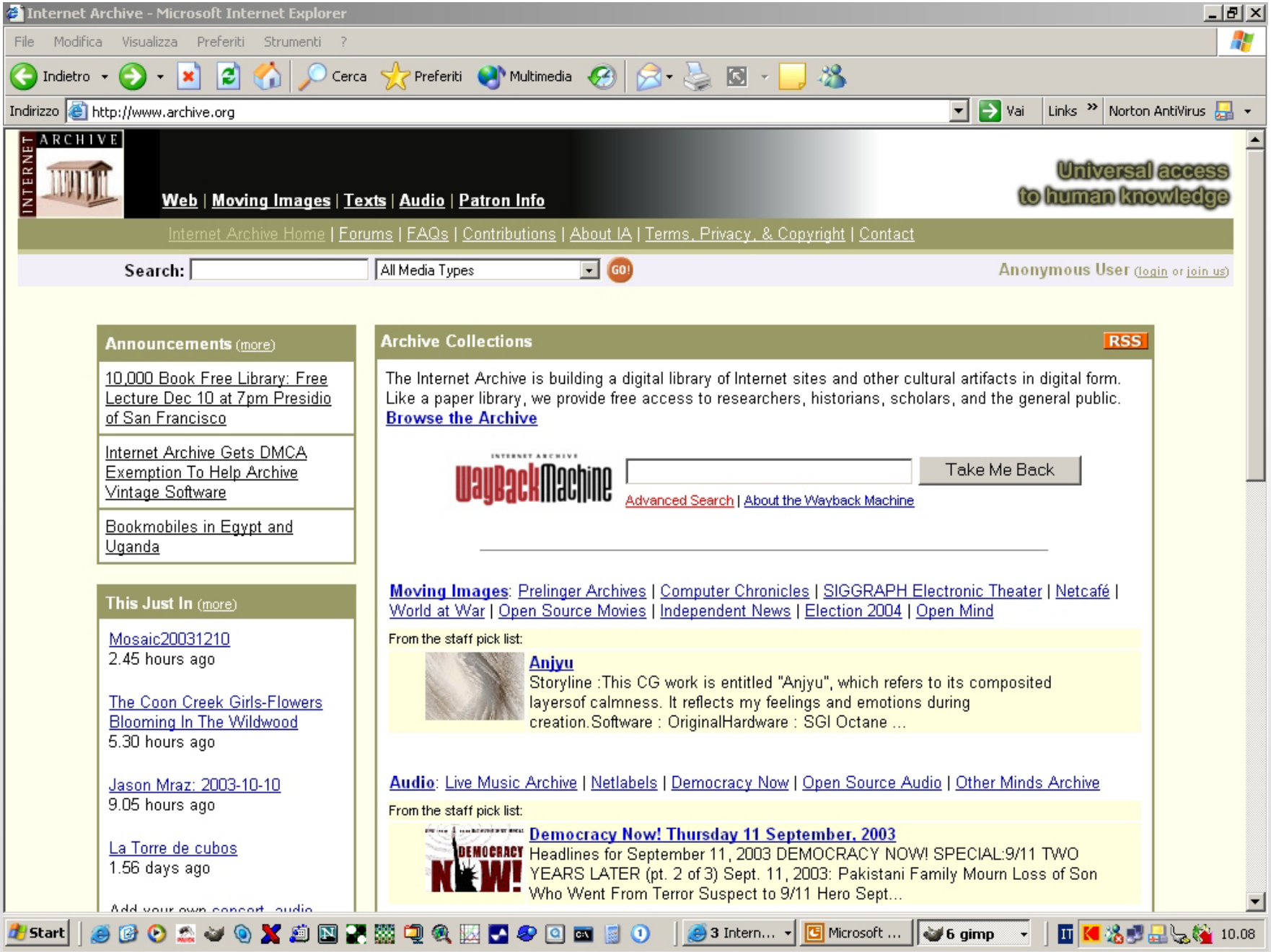
Fig. 2. Confidence scores for protein-protein interactions (**A**) *Drosophila* protein-protein interactions have been binned according to confidence score for the entire set of 20,405 interactions (black), the 129 positive training set examples (green), and the 196 negative training set examples (red). (**B**) The 7048 proteins identified as participating in protein-protein interactions have been binned according to the minimum, average, and maximum confidence score of their interactions. Proteins with high-confidence interactions total 4679 (66% of the proteins in the network, and 34% of the protein-coding genes in the *Drosophila* genome). (**C**) The correlation between GO annotations for interacting protein pairs decays sharply as confidence falls from 1 to 0.5, then exhibits a weaker decay. Correlations were obtained by first calculating the deepest level in the GO hierarchy at which a pair of interacting proteins shared an annotation (interactions involving unannotated proteins were discarded). The average depth was calculated for interactions binned according to confidence score, with bin centers at 0.05, 0.1,..., 0.95. Finally, the correlation for the bin centered at x was defined as $[Depth(x) - Depth(0)]/[Depth(0.95) - Depth(0)]$. This procedure effectively controls for the depth of each hierarchy and for the probability that a pair of random proteins shares an annotation. (**D**) The number of interactions per protein is shown for all interactions (black circles) and for the high-confidence interactions (green circles). Linear behavior in this log-log plot would indicate a power-law distribution. Although regions of each distribution appear linear, neither distribution may be adequately fit by a single power-law. Both may be fit, however, by a combination of power-law and exponential decay, $\text{Prob}(n) \sim n^{-\alpha}\exp^{-\beta n}$, indicated by the dashed lines, with r^2 for the fit greater than 0.98 in either case (all interactions: $\alpha = 1.20 \pm 0.08$, $\beta = 0.038 \pm 0.006$; high-confidence interactions: $\alpha = 1.26 \pm 0.25$, $\beta = 0.27 \pm 0.05$). Note that the power-law exponents are within 1σ for the two interaction sets. [View Larger Version of this Image (29K GIF file)]


t=2.5
t=12.5
t=25
t=30
MR
Signal
1000
0
1000
0
200
400
600
800
1000
Time (sec)
V1
M1
PP
N
1
1
N
Correlation Matrix
(N^2=2x10^{10})
N
1
1
N
Thresholded
Matrix
Network Extracted



Example of a network extracte dusing the methods described before. Top panel shows a pictorial representation of the network (1/8 of all nodes are shown, colored according to its degree: yellow = 1, green = 2, red =3, blue=4, etc).

The bottom panel shows the distribution of links plotted for three values of the correlation threshold. The inset depicts the link distribution
for an equivalent randomly connected network.


Internet Archive - Microsoft Internet Explorer
File Modifica Visualizza Preferiti Strumenti ?
Indietro Cerca Preferiti Multimedia
Indirizzo http://www.archive.org
Vai Links Norton AntiVirus
INTERNET ARCHIVE
Web | Moving Images | Texts | Audio | Patron Info
Universal access to human knowledge
Internet Archive Home | Forums | FAQs | Contributions | About IA | Terms, Privacy, & Copyright | Contact
Search: All Media Types GO!
Anonymous User (login or join us)
Announcements (more)
10,000 Book Free Library: Free Lecture Dec 10 at 7pm Presidio of San Francisco
Internet Archive Gets DMCA Exemption To Help Archive Vintage Software
Bookmobiles in Egypt and Uganda
This Just In (more)
Mosaic20031210
2.45 hours ago
The Coon Creek Girls-Flowers Blooming In The Wildwood
5.30 hours ago
Jason Mraz: 2003-10-10
9.05 hours ago
La Torre de cubos
1.56 days ago
Archive Collections
RSS
The Internet Archive is building a digital library of Internet sites and other cultural artifacts in digital form. Like a paper library, we provide free access to researchers, historians, scholars, and the general public.
Browse the Archive
WaybackMachine
Take Me Back
Advanced Search | About the Wayback Machine
Moving Images: Prelinger Archives | Computer Chronicles | SIGGRAPH Electronic Theater | Netcafé | World at War | Open Source Movies | Independent News | Election 2004 | Open Mind
From the staff pick list:
Anjyu
Storyline :This CG work is entitled "Anjyu", which refers to its composited layersof calmness. It reflects my feelings and emotions during creation.Software : OriginalHardware : SGI Octane ...
Audio: Live Music Archive | Netlabels | Democracy Now | Open Source Audio | Other Minds Archive
From the staff pick list:
Democracy Now! Thursday 11 September, 2003
Headlines for September 11, 2003 DEMOCRACY NOW! SPECIAL:9/11 TWO YEARS LATER (pt. 2 of 3) Sept. 11, 2003: Pakistani Family Mourn Loss of Son Who Went From Terror Suspect to 9/11 Hero Sept...
Start
3 Intern...
Microsoft ...
6 gimp
10.08