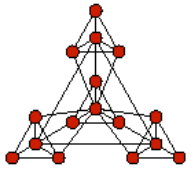
Hierarchical Organization of Modularity in Complex Networks

Erzsébet Ravasz

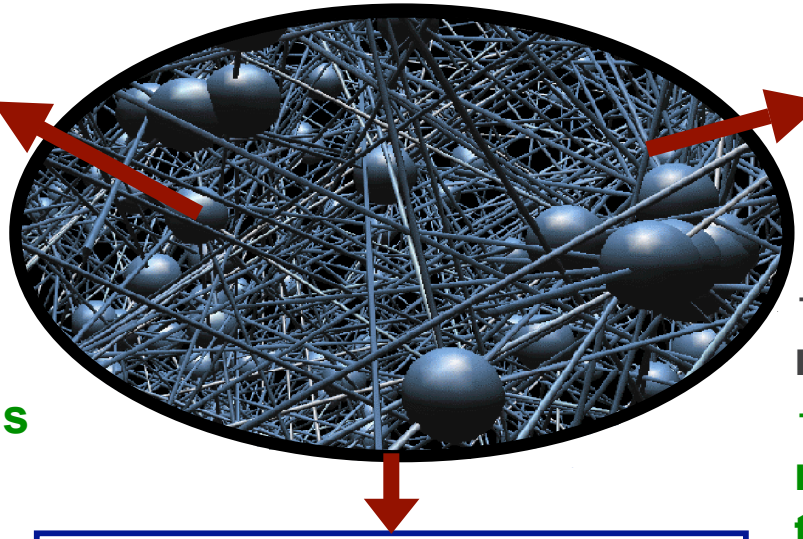
University of Notre Dame

Complexity



Many different components

→ atoms, particles
→ spins, oscillators
→ cells, DNA, proteins



A variety of interactions

→ basic forces of nature
→ genetic regulation, translation

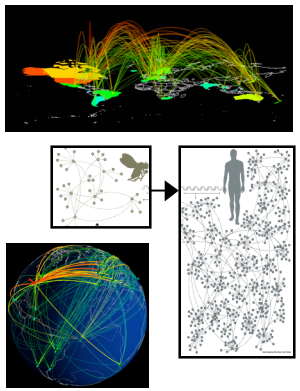
System as a whole

Complexity in:

→ topology of interactions

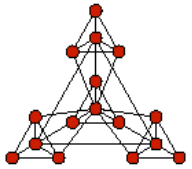
→ time evolution of the structure

→ dynamics on the structure



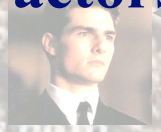
Network

Networks in life



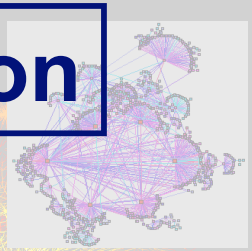
Society

→ Friendships, sexual contacts
→ Co-authorship, citations
→ Movie actors, business



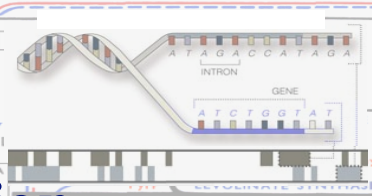
Days of Thunder
Far and Away
Eyes Wide Shut

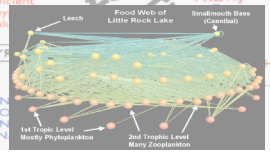
Communication



→ Internet
→ World Wide Web
→ Phone call networks

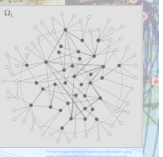
Biology

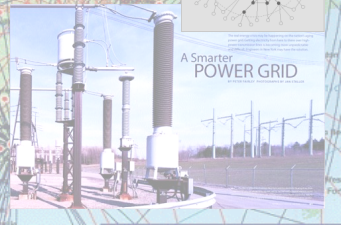




→ Genetic regulation
→ Protein-protein interactions
→ Metabolic pathways
→ Food webs
→ Neuron networks

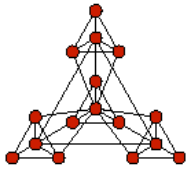
And more…





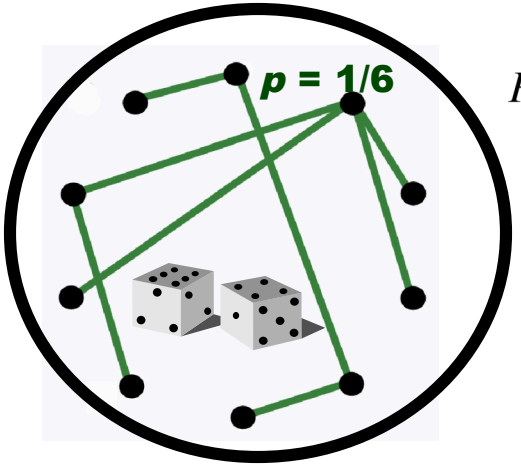
→ Airline routes
→ Word webs
→ Power grid

Random networks?

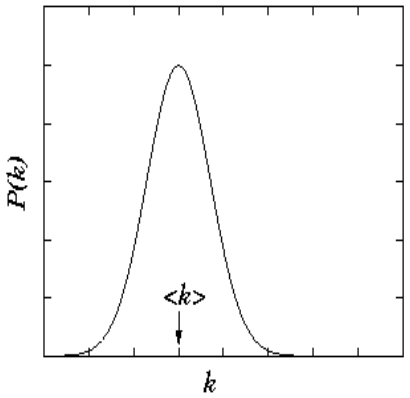


The Erdős-Rényi model

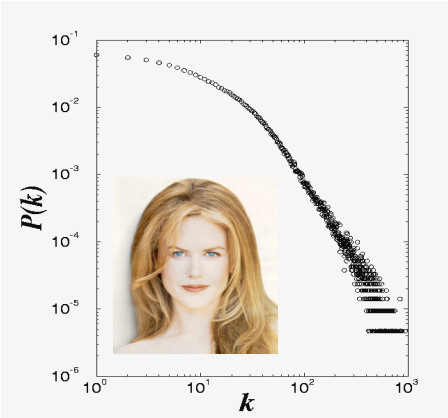
Is randomness a good approximation ?

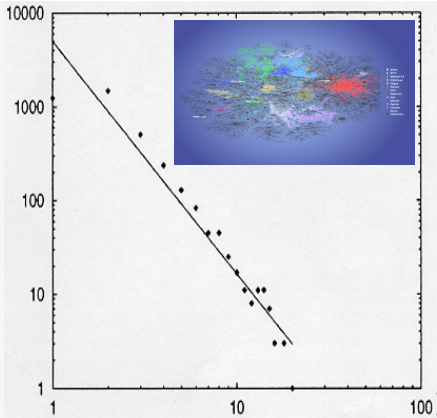


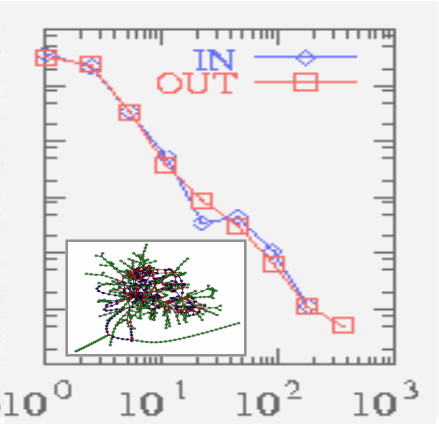
$$P(k) \simeq e^{-pN} \frac{(pN)^k}{k!}$$

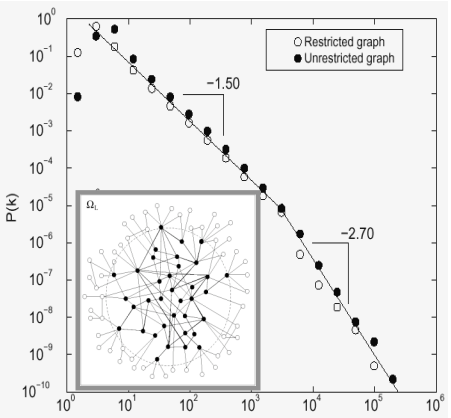


No characteristic connectivity!

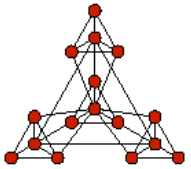






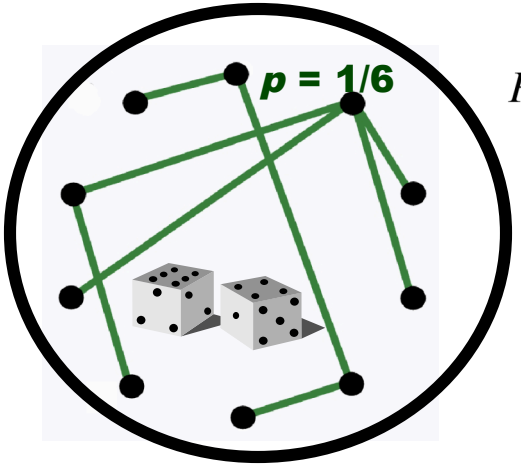


Random networks?

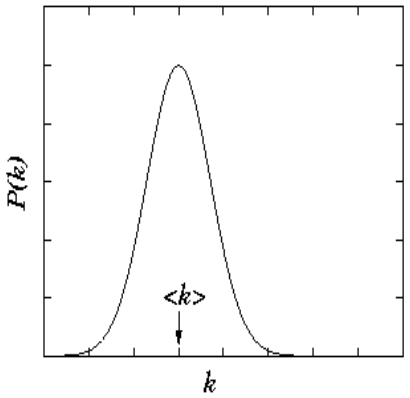


The Erdős-Rényi model

Is randomness a good approximation ?

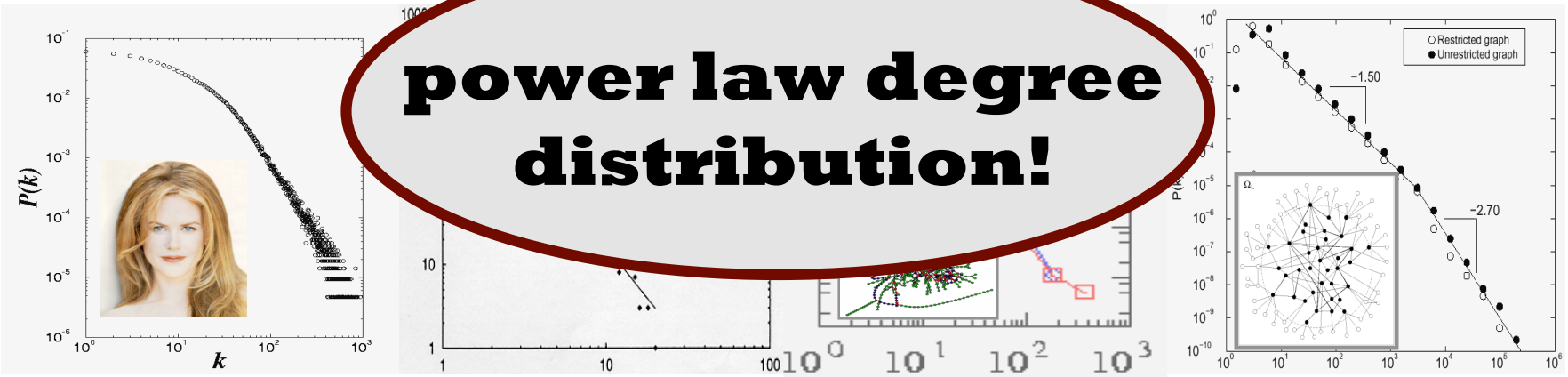


$$P(k) \simeq e^{-pN} \frac{(pN)^k}{k!}$$

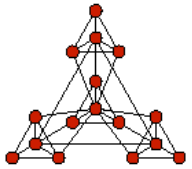


No characteristic connectivity!

power law degree distribution!



Scale-free models

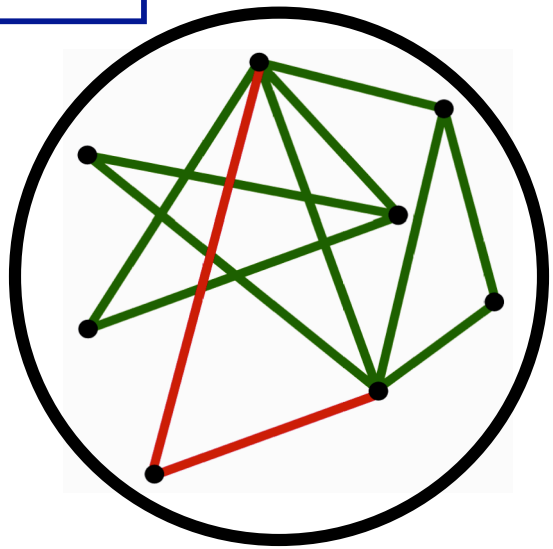


The Barabási-Albert model

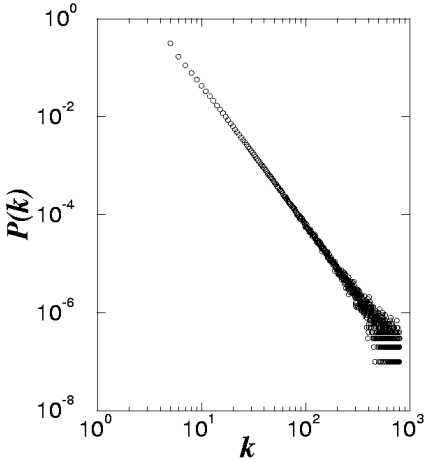
→ Growth process

→ Preferential attachment

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



$$P(k) = \frac{2m^2}{k^3}$$

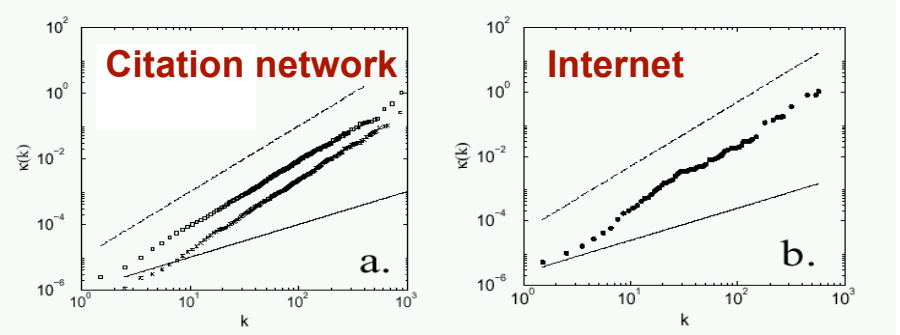


Preferential attachment in disguise

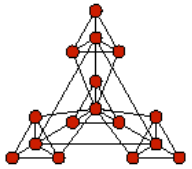
→ Protein duplication

→ Triad formation

→ Finite memory of nodes



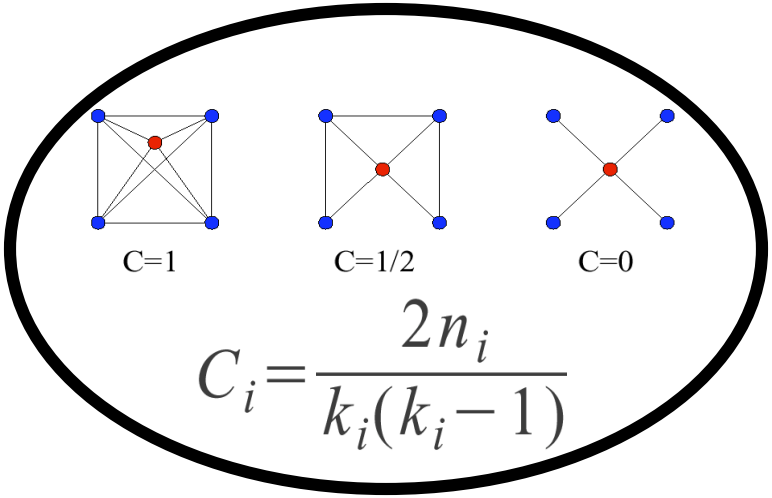
Clustering

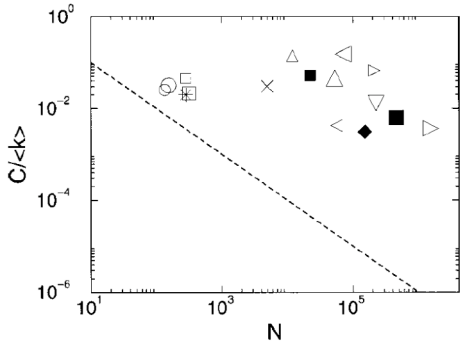


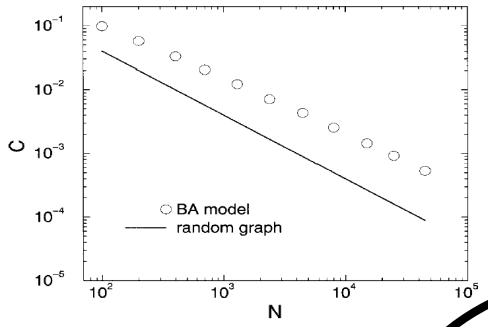
Clustering coefficient

→ High average in real networks

→ Scale-free model: $C \sim (\ln N)^2/N$

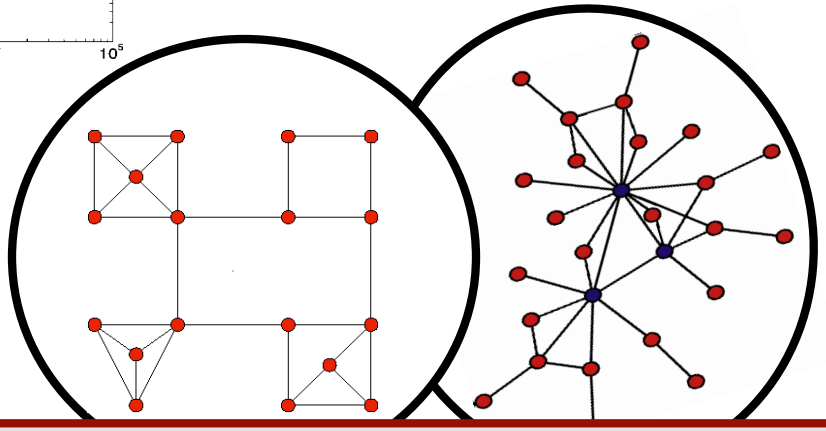






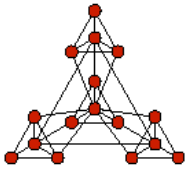
Modular organization

→ WWW communities
→ Scientific groups
→ Research areas
→ Circles of friends
→ Protein complexes



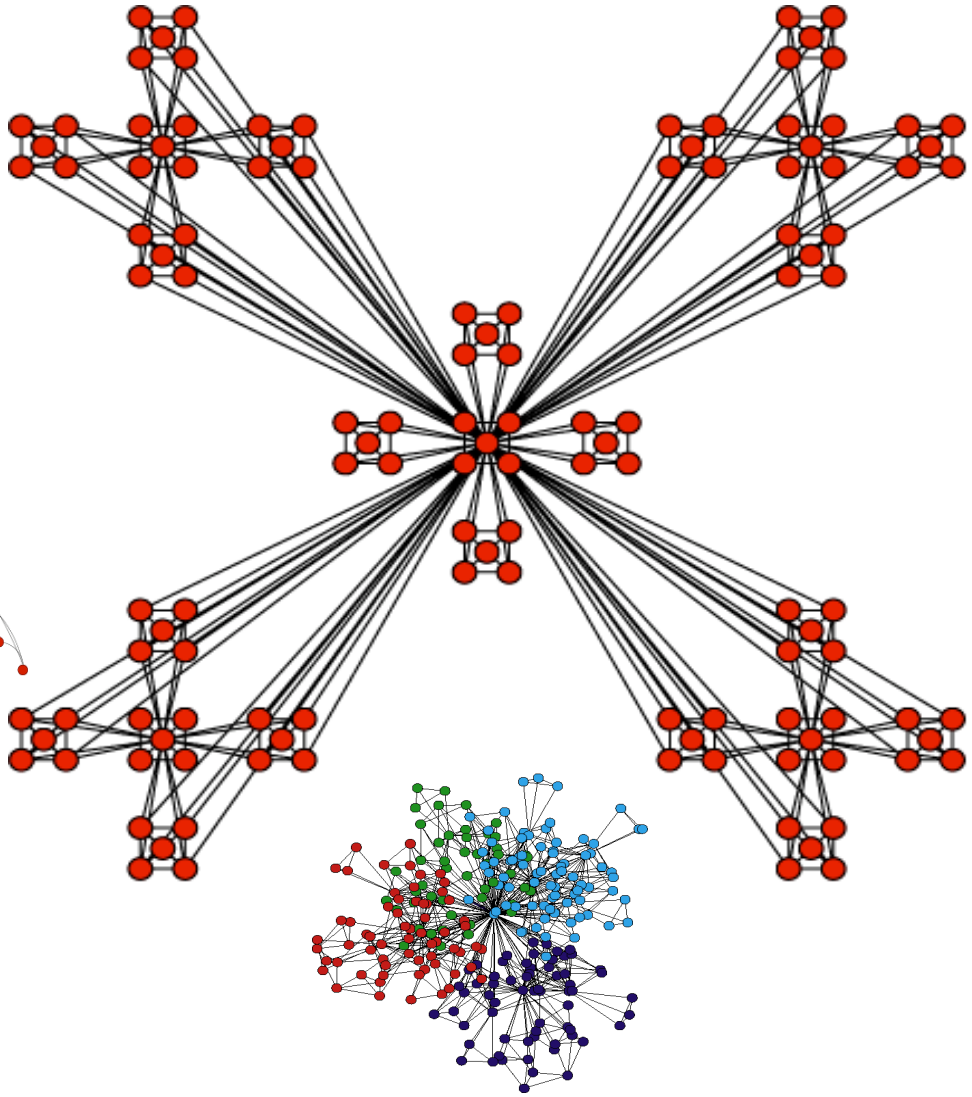
MODULAR AND SCALE-FREE?

Module hierarchy

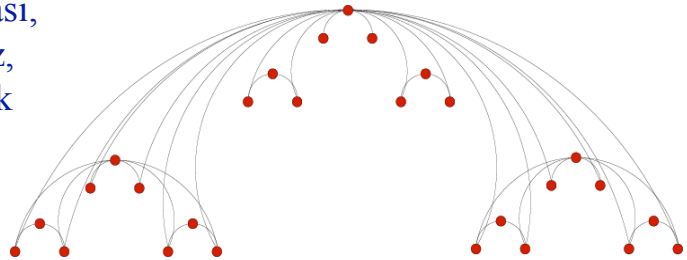


Hierarchical models

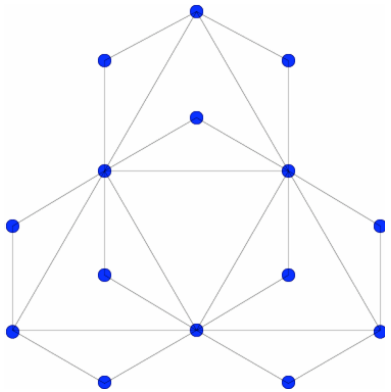
→ Iterative construction
→ Hierarchically nested modules



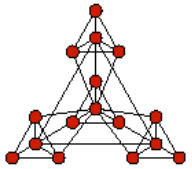
• Barabási, Ravasz, Vicsek



• Dorogovtsev, Goltsev, Mendes



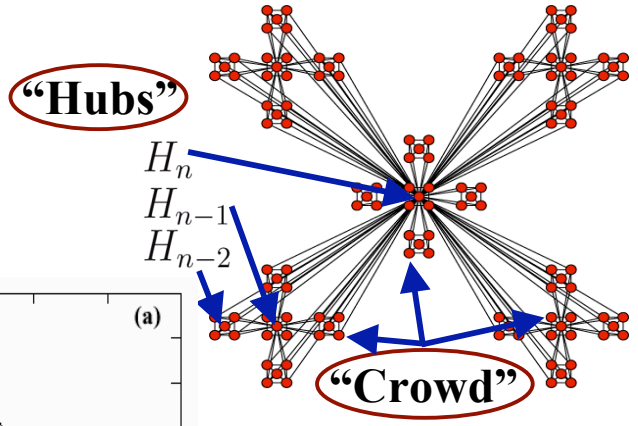
Properties of the model

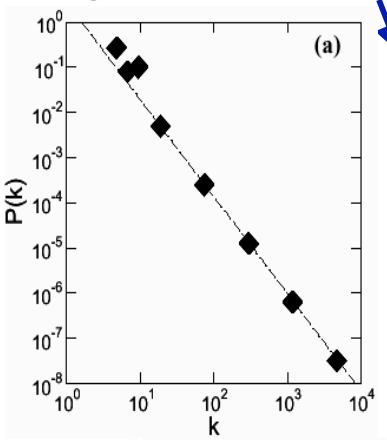


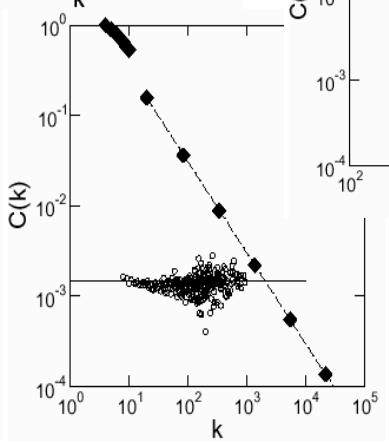
Power law degree distribution

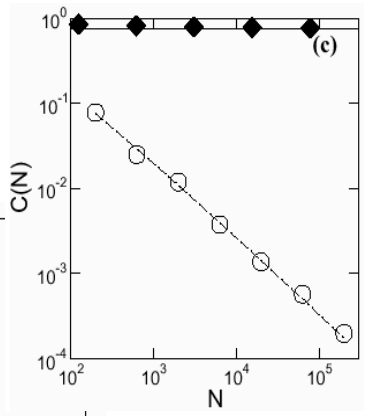
- $k_i(H_i) = \sum_{l=1}^{i} 4^l = \frac{4}{3}\left(4^i - 1\right) \implies \ln k_i(H_i) \simeq i \cdot \ln 4 + \ln\frac{4}{3}$
- $N(H_i) = 4 \cdot 5^{n-i-1} \implies \ln N(H_i) = c_n - i \cdot \ln 5$
- $\ln N(H_i) \simeq c'_n - \ln k_i \frac{\ln 5}{\ln 4} \implies N(H_i) \sim k_i^{-\frac{\ln 5}{\ln 4}}$
- $P(k_i) \sim \frac{N(H_i)}{(k_{i+1} - k_i)} \sim k_i^{-\gamma}$

$$\gamma = 1 + \frac{\ln 5}{\ln 4} \simeq 2.16$$









Large average clustering

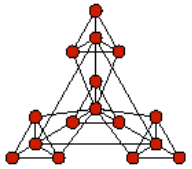
$$C \simeq 0.74$$

Hierarchical clustering

- $n_i \sim k_i \implies C(H_i) \sim \frac{2\,k_i}{k_i(k_i - 1)}$

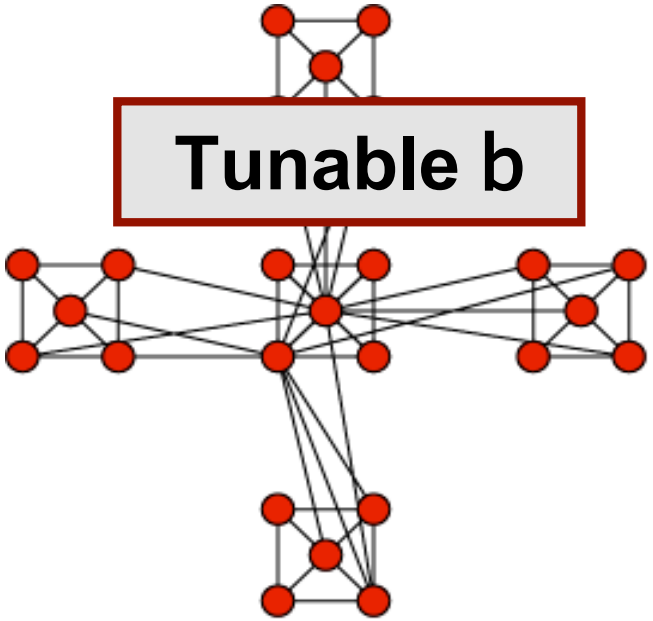
$$C(k) \sim \frac{1}{k}$$

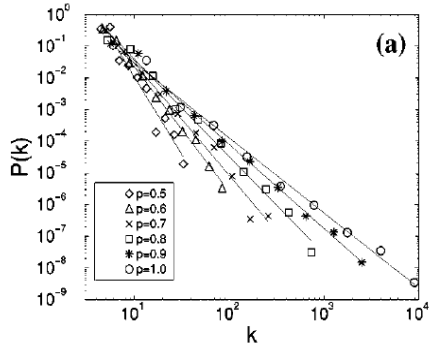
Stochastic version

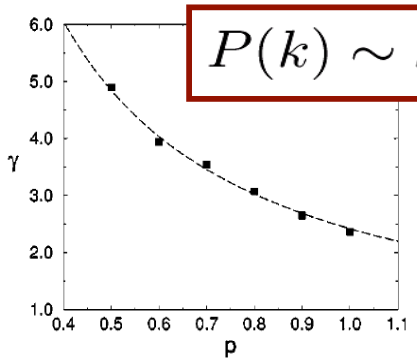


Definition of the stochastic model

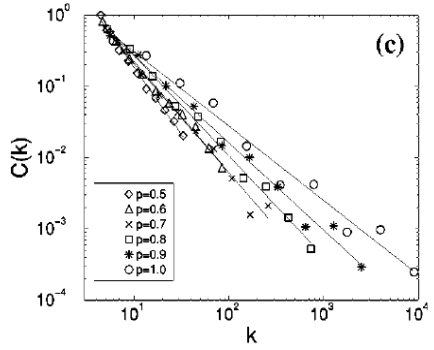
→ p^i fraction of new nodes
→ links back to old nodes
→ preferential attachment

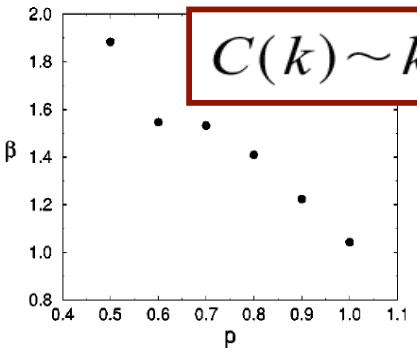






$$P(k) \sim k^{-\gamma}$$

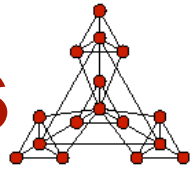




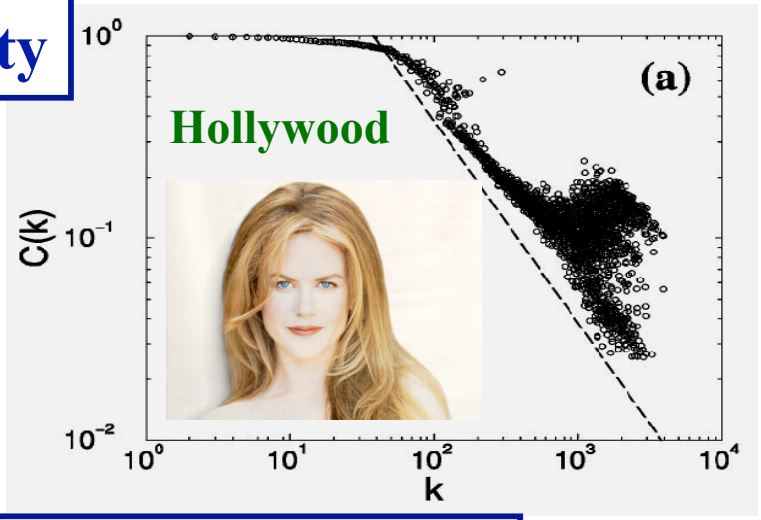
$$C(k) \sim k^{-\beta}$$

Scaling of *C(k)* indicates hierarchy!

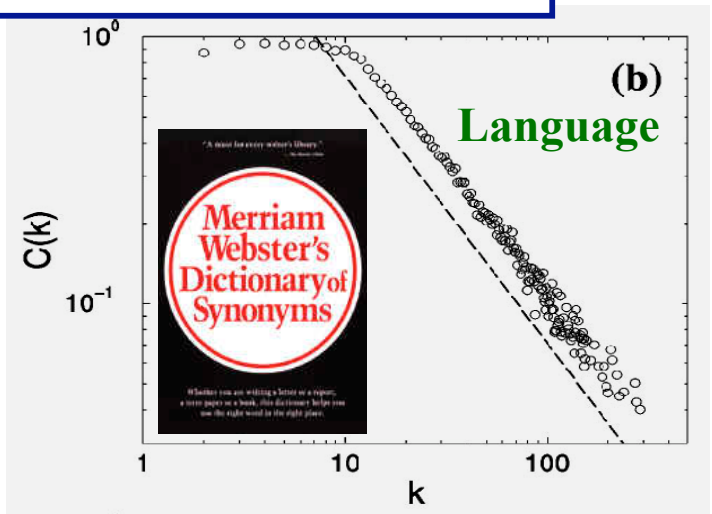
Hierarchical networks



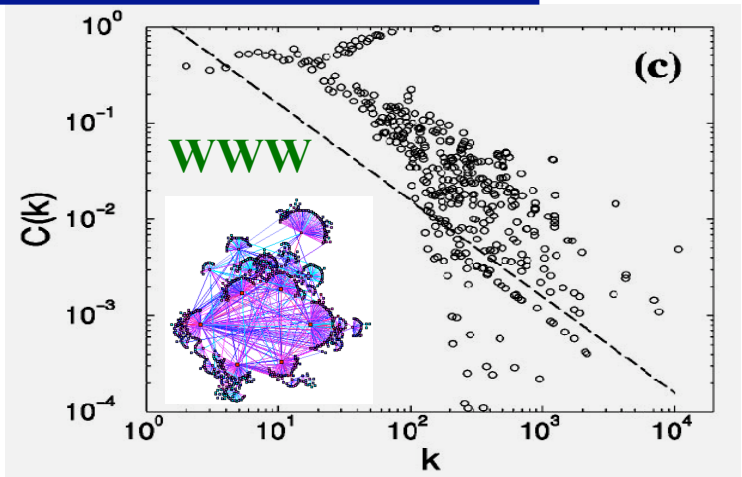
Society

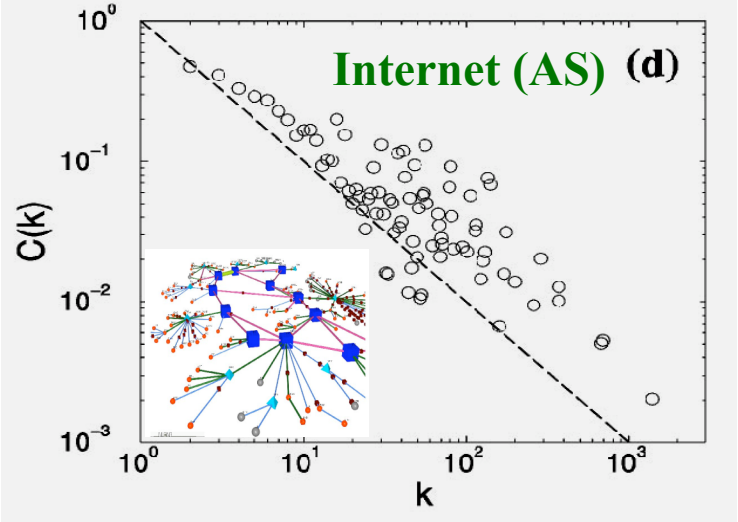


Human communication

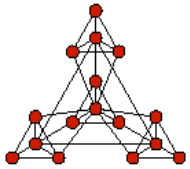


The electronic skin

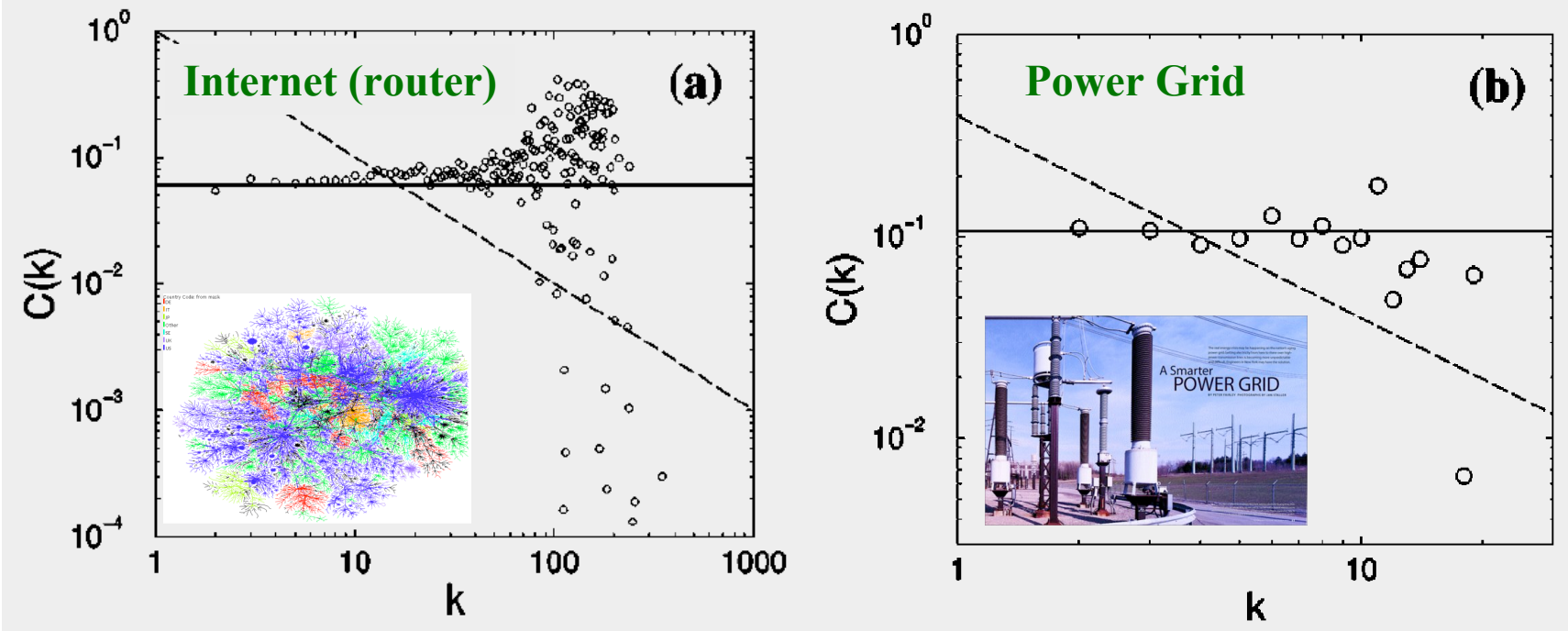




Absence of hierarchy

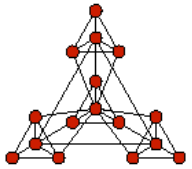


Geographically localized networks

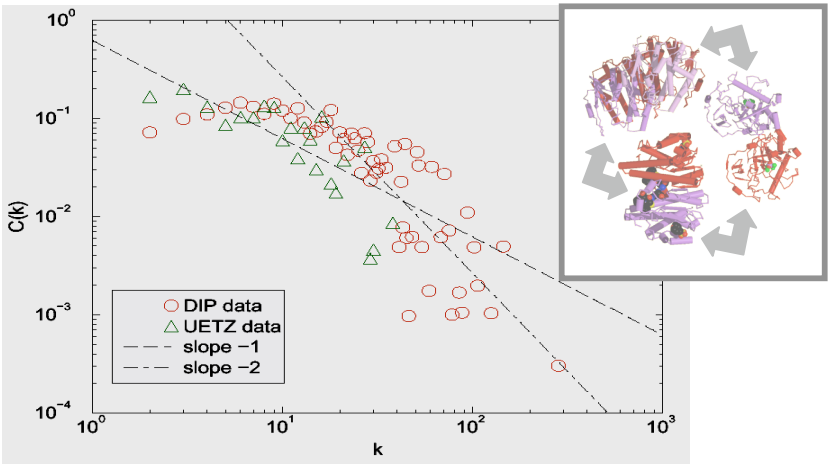


Economic pressure to minimize link lengths

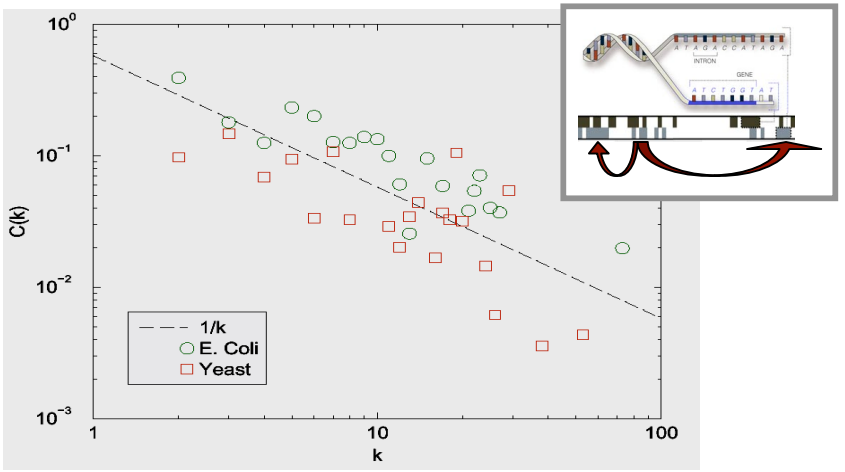
Biological systems



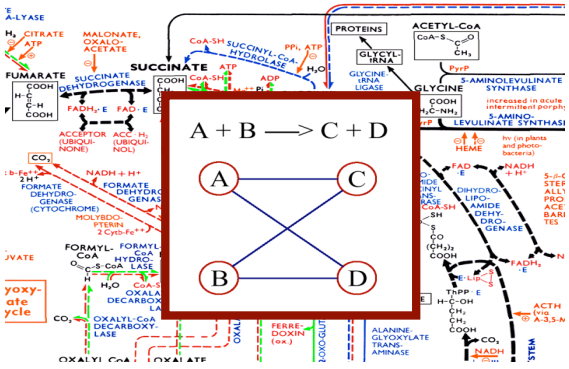
Protein-protein interaction

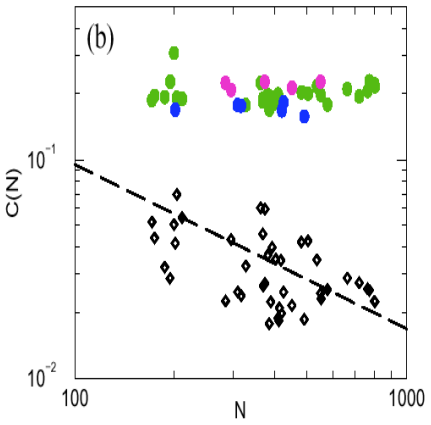


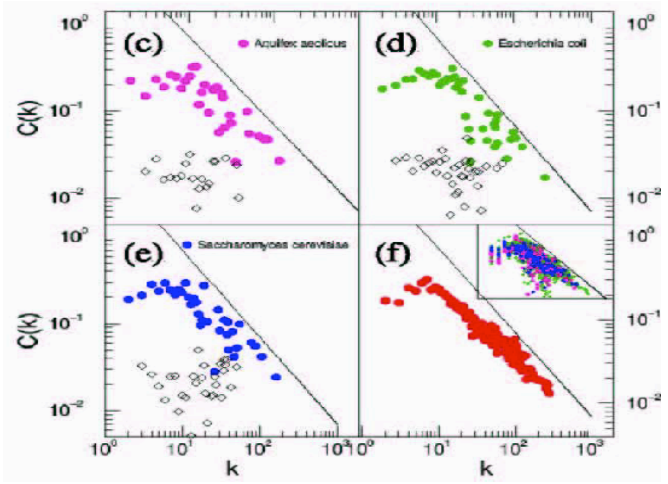
Regulatory networks



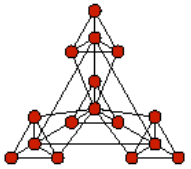
Metabolic networks



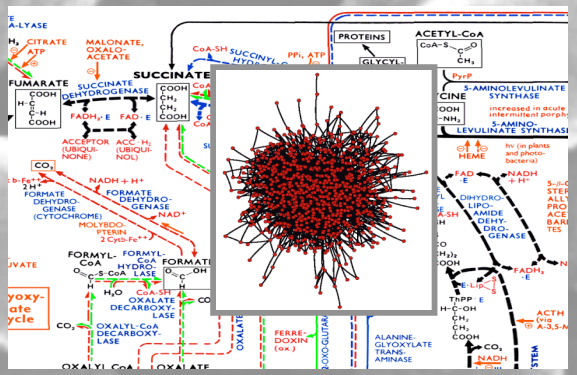


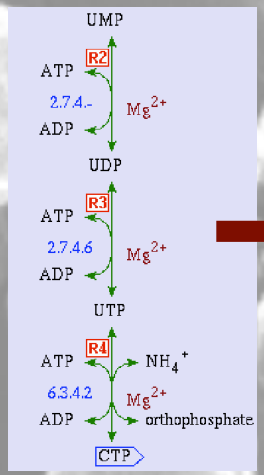


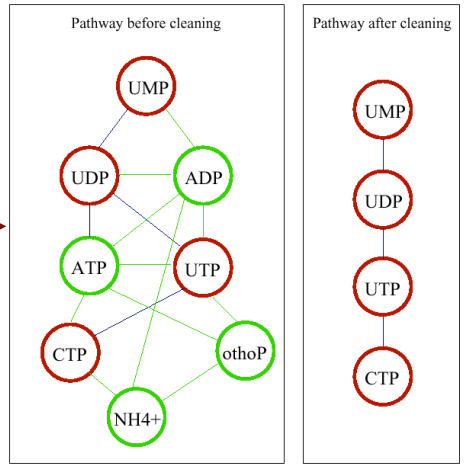
The metabolism of *E. Coli*



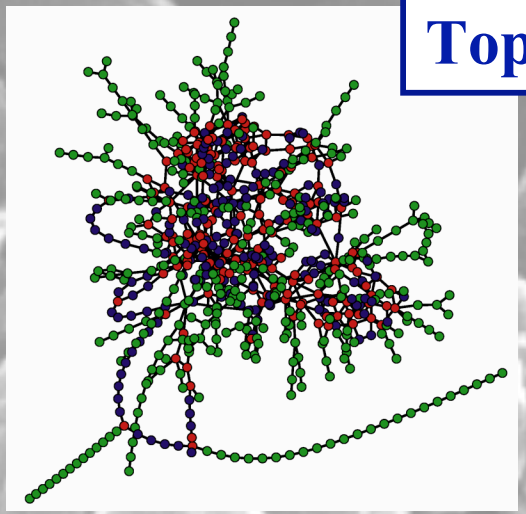
Biochemical reduction

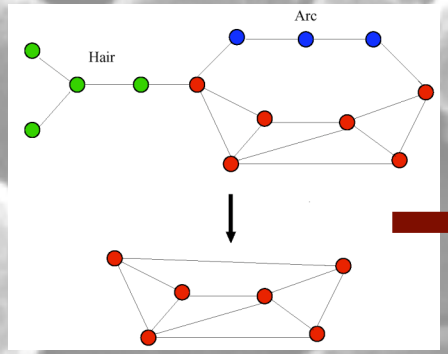


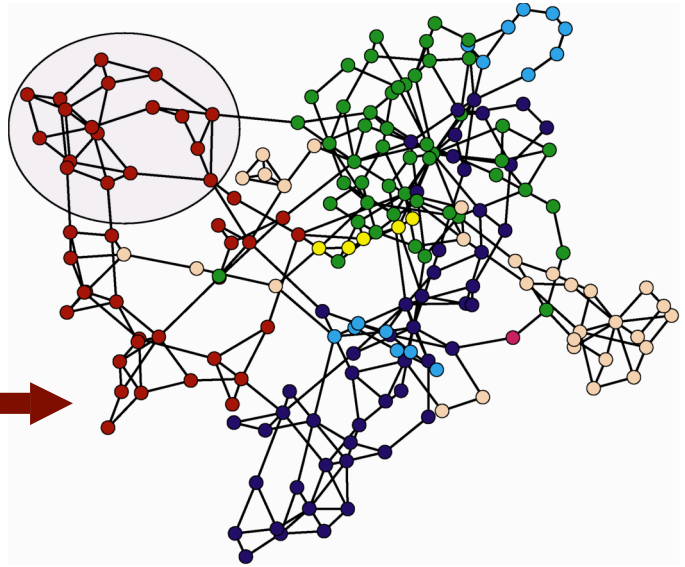




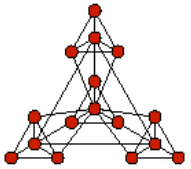
Topological reduction





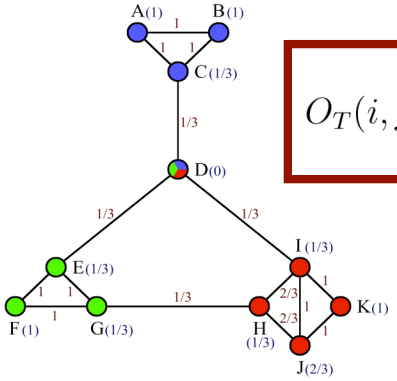


Finding the modules



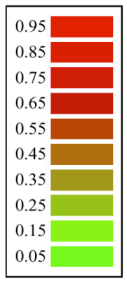
Hierarchical clustering

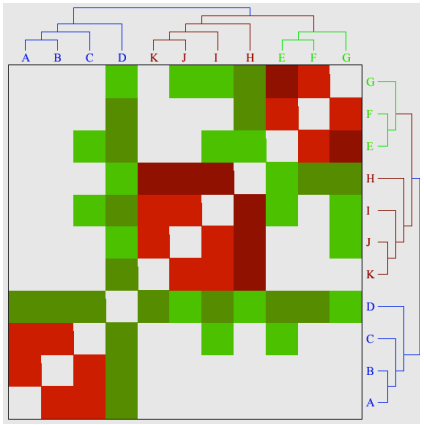
→ Similarity matrix

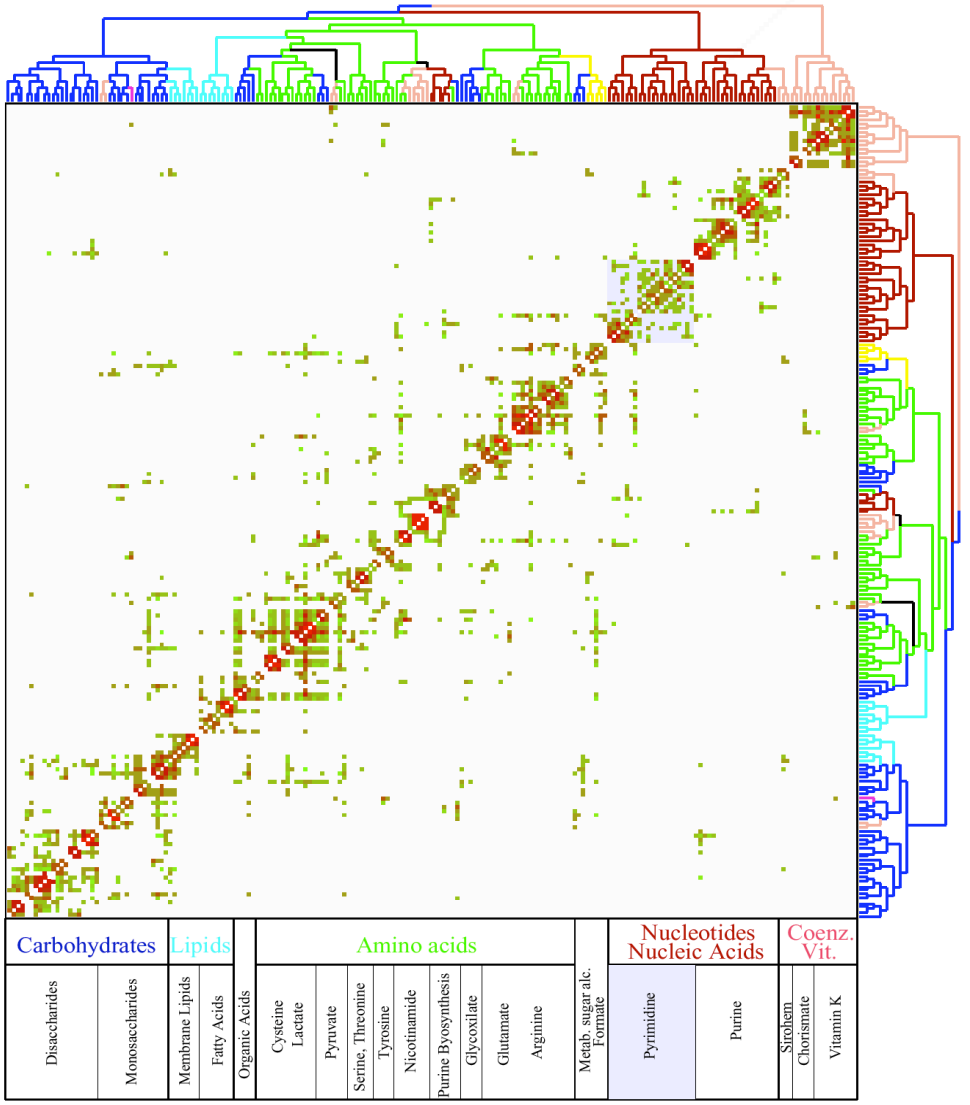


$$O_T(i,j) = \frac{\sum_{l=1}^{N}{}' l_{i,l} \cdot l_{j,l} + l_{i,j}}{\min(k_i, k_j) + 1 - l_{i,j}}$$

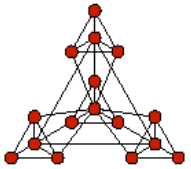
→ Average linkage clustering (UPGMA)



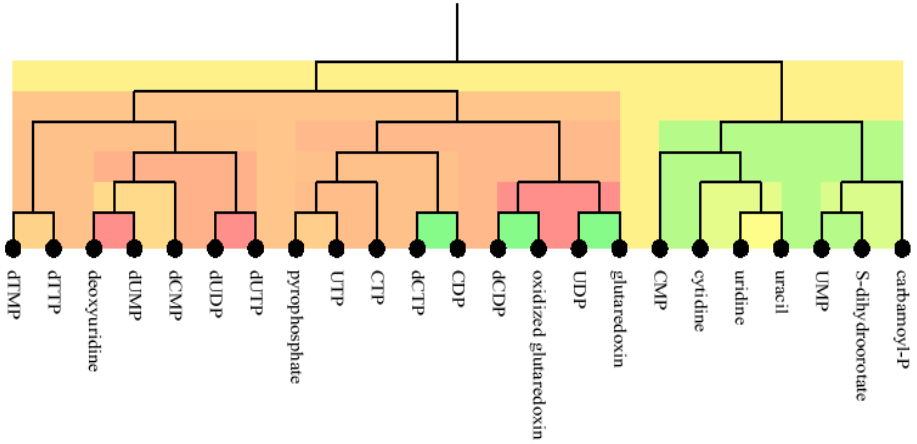


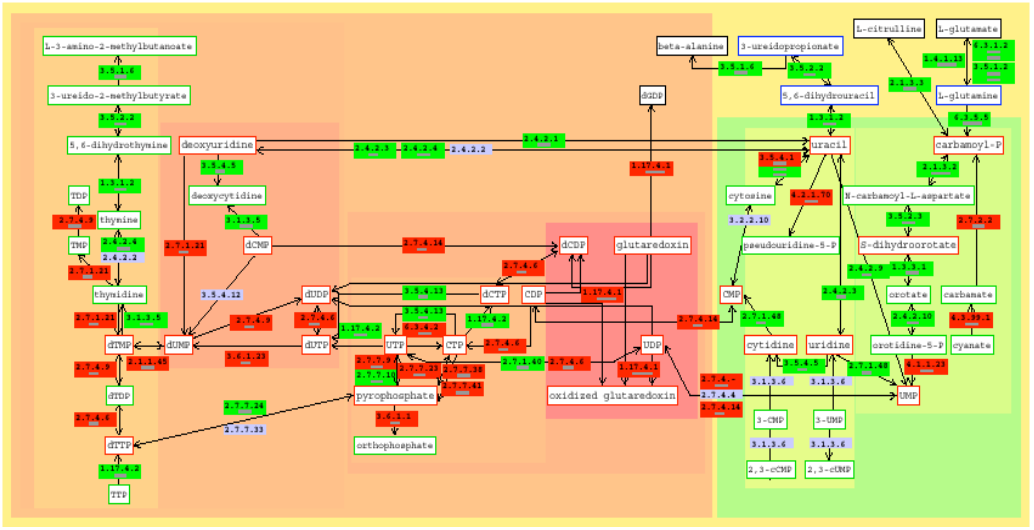


Module Lethality

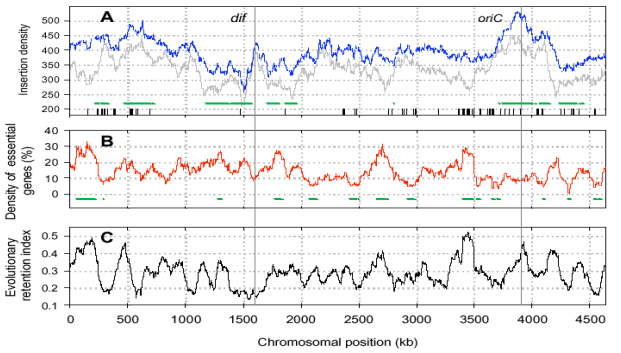


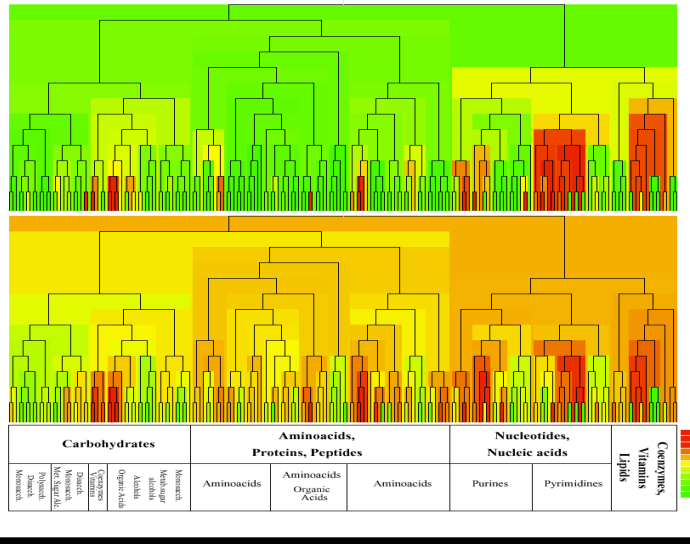
Pyrimidine metabolism



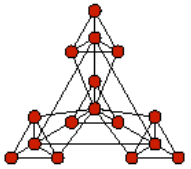


Genome-wide lethality measurement

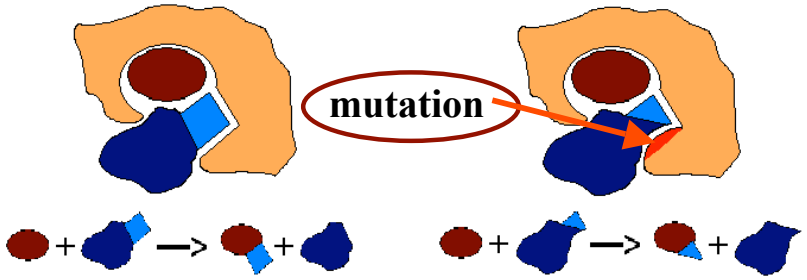




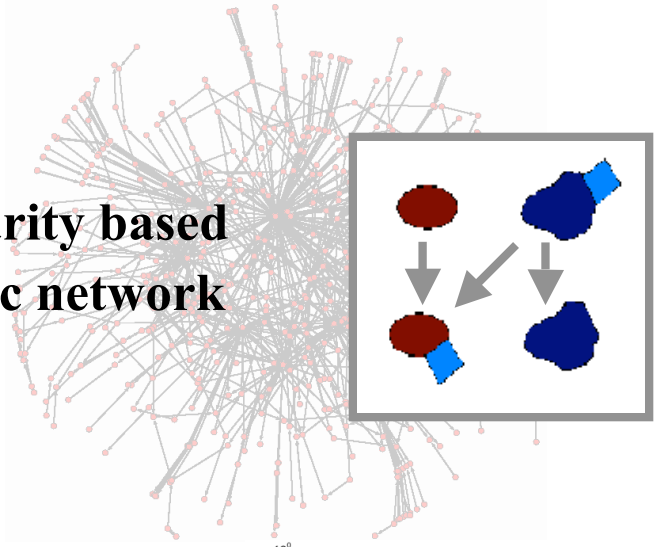
Modeling metabolism



Evolution via mutation of a catalytic site



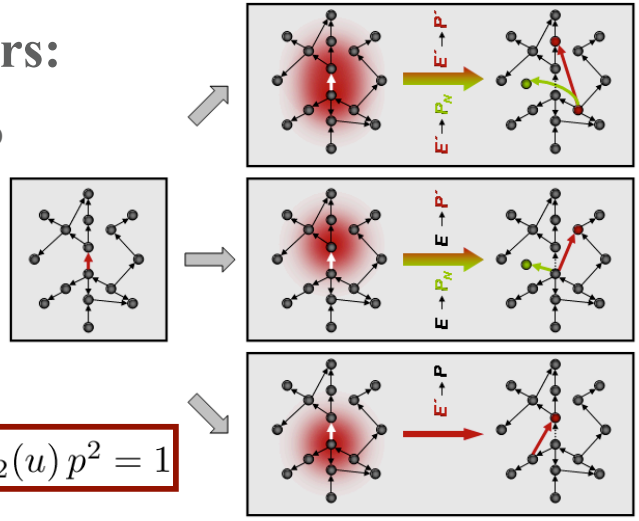
→ similarity based metabolic network



Algorithm of the model

→ parameters:
P_{new} , P_{dup}

→ 3 cases:



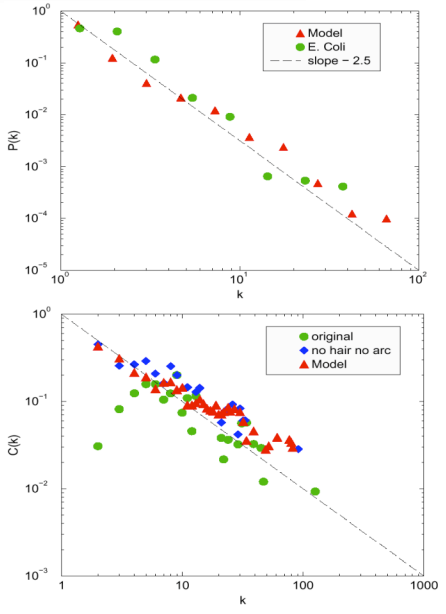
→ neighbor:

$$N_1(u)\,p + N_2(u)\,p^2 = 1$$

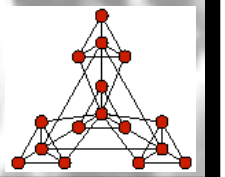
Results

→ P_{new}= 0.60

→ P_{dup}= 0.85



Thank you!



http://www.nd.edu/~networks

Thanks to:

- Albert László Barabási
- Zoltán Oltvai
- Stefan Wuchty & Nina Lehmann
- My group