Network Biology: Understanding the Cell's Functional Organization

Albert-Laszlo Barabasi

& Zoltan N.Oltvai

Nature 2004

Introduction

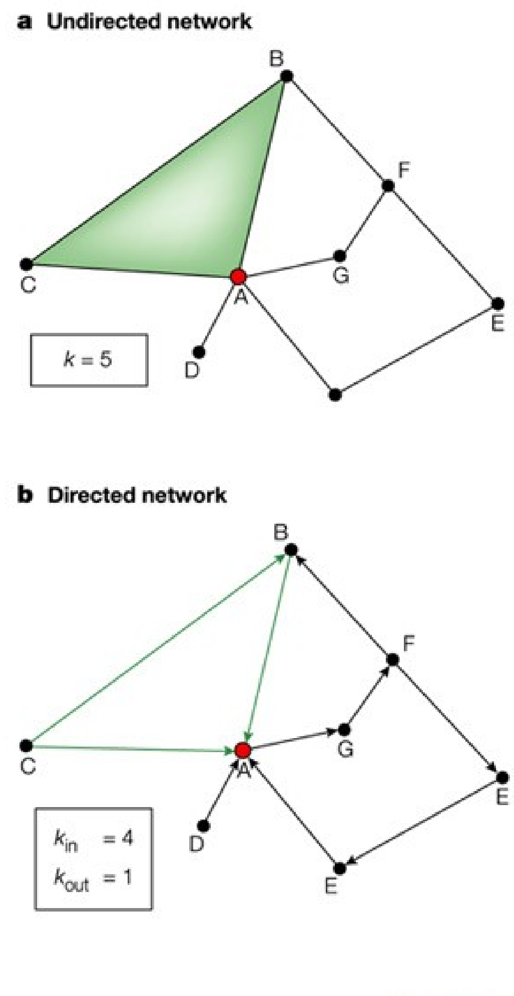
- Discrete biological functions involve complex interactions
- New technologies allow us to collect interaction data
- Graphs are a natural way to model interactions
- We can then use graph theory to analyze the data

Universality

- We know a lot about other complex real world networks
  - Internet
  - Computer Chips
  - Society
- Research has shown that these networks are governed by a few universal laws
- Do these same laws apply to biological networks?

Network Measures

- The **degree** of a node is the number of links it has to other nodes, denoted by k.
- The average degree of a graph <k> = 2L/N



Network Measures

- The **degree distribution**, P(k), gives the probability that a selected node has exactly k links.
- P(k) is obtained by counting the number of nodes with k links and dividing by N
- The degree distribution is used to classify networks

Network Measures

- The **shortest path** between two nodes is the minimum number of links that must be traversed to travel from one node to the other
- The **mean path length** <l> is the average of the shortest paths between all pairs of nodes

Network Measures

- The **clustering coefficient** measures the tendency of nodes to form clusters
- If a is connected to b and b is connected to c, then a is connected to c ?
- $C_i = 2n_i / (k_i (k_i - 1))$, where n_i is the number of links connecting the k_i neighbors of node i to each other
- What is the maximum value of C_i ?

Network Measures

Clustering Coefficient Example:

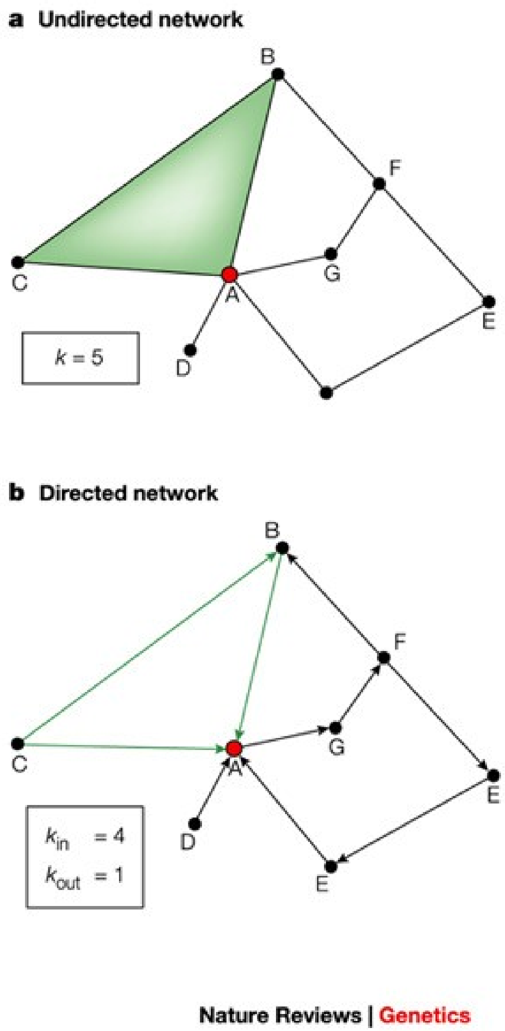
$C_i = 2n_i / (k_i (k_i - 1))$

$C_A = 2(1) / (5*4)$

$= 2/20 = 0.1$

$C_F = C_E = C_D$

$= C_G = 0$



Network Measures

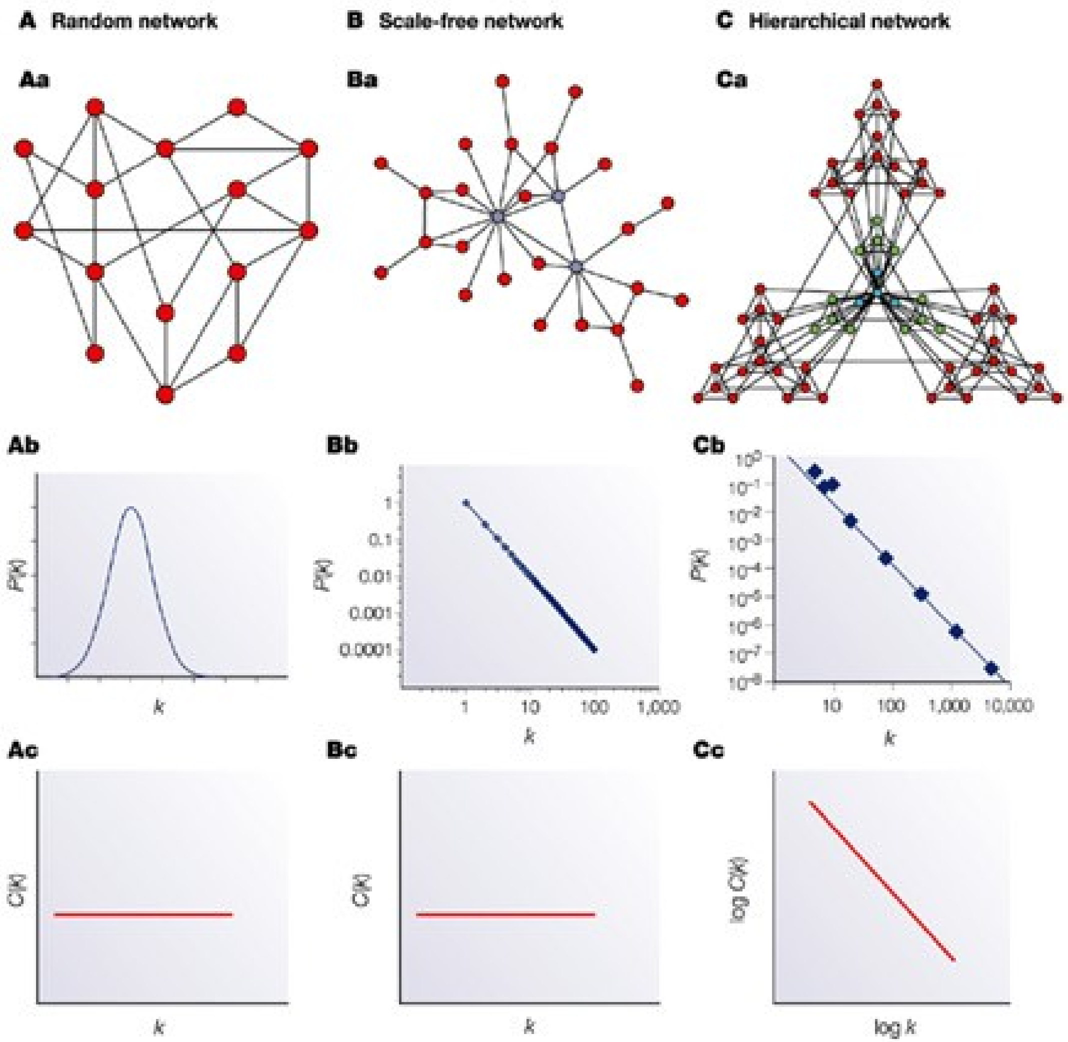
- <C> is the average clustering coefficient. It measures the overall tendency of nodes to form clusters.
- C(k) is the average clustering coefficient of all nodes with k links.

Network Measures

- Depend on the size of the network:
  - <k> average degree
  - <l> average length
  - <C> average clustering coefficient

- Independent of the size of the network:
  - P(k) degree distribution
  - C(k) average clustering coefficient function

Network Models



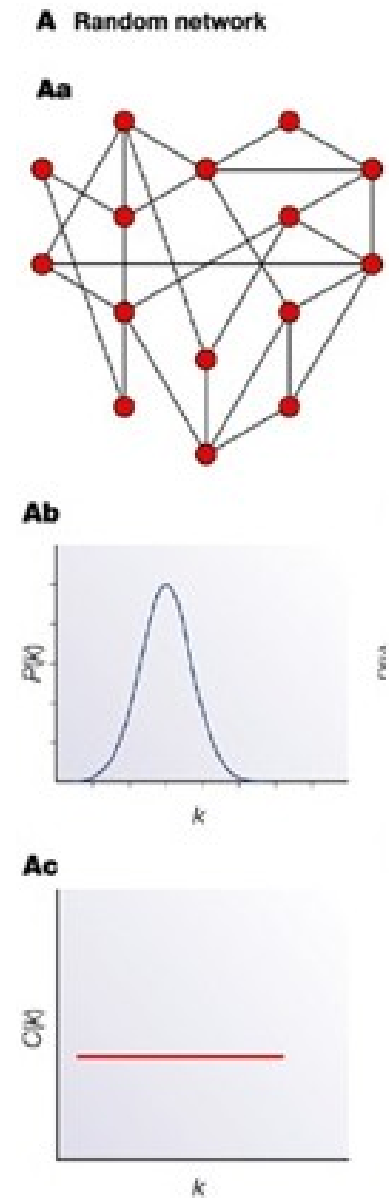
Nature Reviews | Genetics

Random

- Paul Erdos and Alfred Renyi initiated the study of random networks in 1960
- Erdos-Renyi (ER) model of a random network:
  - Start with N nodes
  - Connect each pair of nodes with probability p
  - Results in a graph with pN(N-1)/2 expected links

Random

- P(k) follows a Poisson distribution
- Most nodes have a degree that is close to <k>
- C(k) is constant
- The mean path length <l> ~ log N, which indicates that the network has the small-world property (which we will discuss soon)

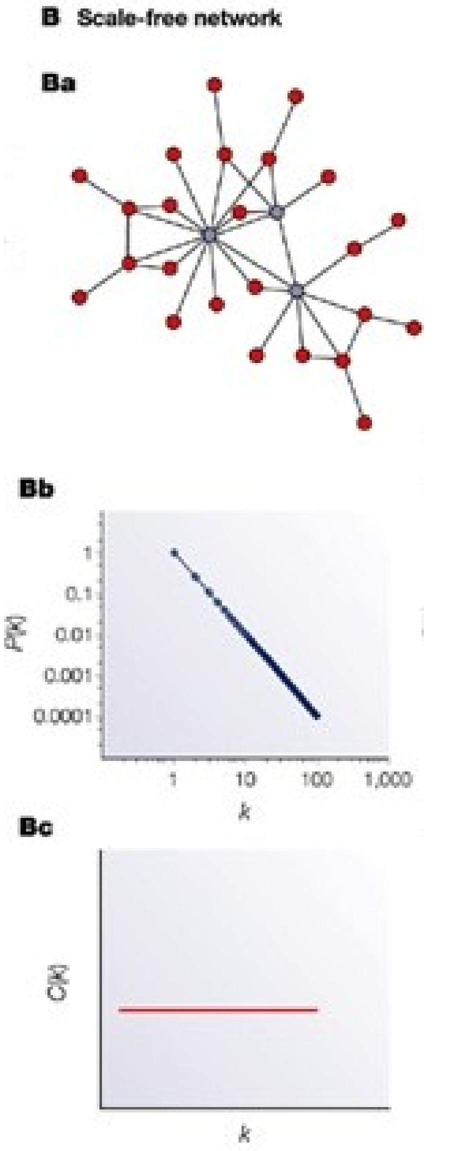


Scale-free

- Networks with a power-law degree distribution are called scale-free
- $P(k) \sim k^{-\gamma}$ where γ is the degree exponent
- For most networks $2 > \gamma > 3$
- The smaller γ is, the more important the role of hubs is
- When $\gamma > 3$, scale-free features disappear and network behaves like a random one

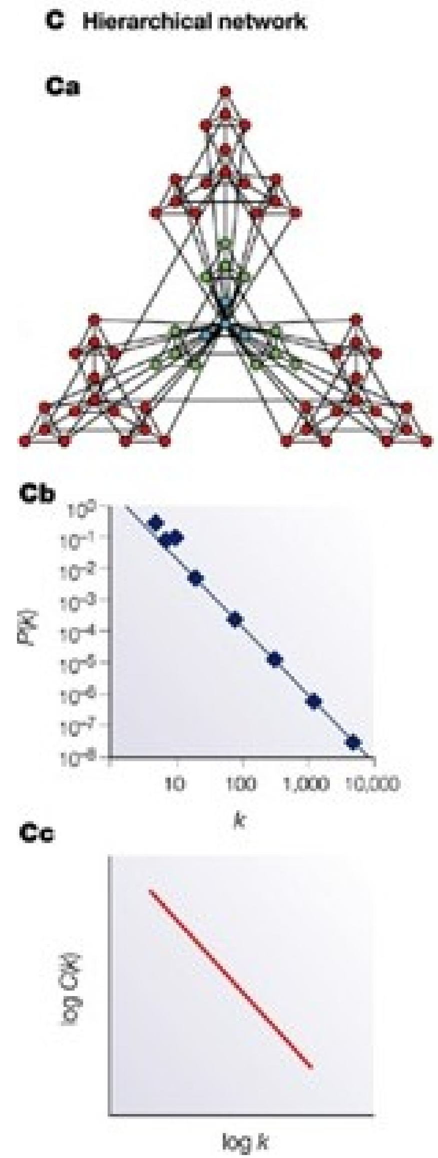
Scale-free

- Scale-free networks are characterized by a few hub nodes of high degree, and many nodes of low degree
- C(k) is constant, like random networks
- <l> ~ log log N, which means it has ultra-small-world property



Hierarchical

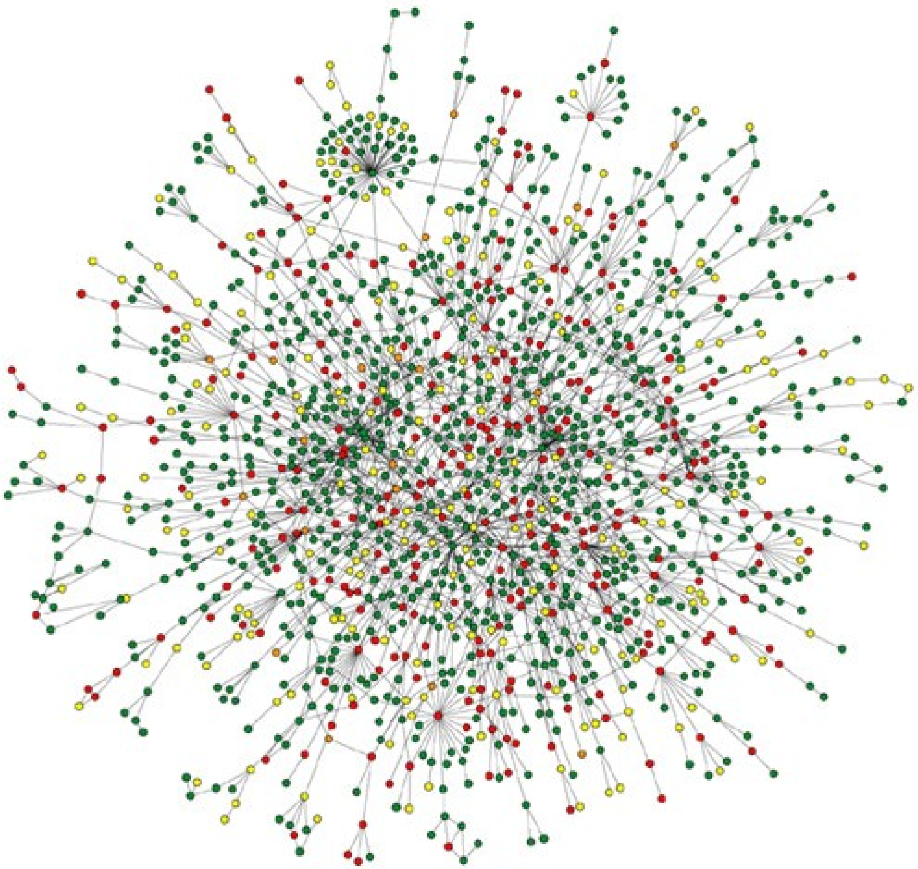
- A hierarchical architecture implies that sparsely connected nodes are part of highly clustered areas, with communication between the clusters being maintained by a few hubs
- P(k) follows a power-law degree distribution
- $C(k) \sim k^{-1}$, unlike random and scale-free networks



Cellular networks are scale-free

- Computer and social networks are scale-free
- So cellular networks must be scale-free too

Cellular networks are scale-free



Nature Reviews | Genetics

Cellular networks are scale-free

- Metabolic networks are scale free (Jeong et al. *Nature* 2000; Wagner & Fell *Proc. R. Soc. Lond.* 2001)
  - Most metabolic substrates participate in only one or two reactions
  - A few participate in dozens (hubs)
- Genetic regulatory networks (Featherstone & Brodie *Bioessays* 2002; Agrawal *Phys. Rev. Lett*. 2002)
  - Nodes are genes
  - Links are derived from expression data

Cellular networks are scale-free

- Protein domain networks (Wuchty *Mol. Biol. Evol.* 2001; Apic et al. *Bioinformatics* 2001)
  - Constructed based on protein domain interactions

Small-world effect

- Everyone in the world knows everyone else through an average chain of relatively few people
- From a famous experiment by Stanley Milgram in 1967 (*Psychology Today*)
- In 1998, Watts and Strogatz showed that networks such as the neural network of C. elegans and power grids exhibit the small world property (*Nature*)

Small-world effect

- Random networks have the small-world property
- Scale-free networks are ultra-small (Chung & Lu *Proc. Natl Acad. Sci*. 2002; Cohen & Havlin *Phys. Rev. Lett*. 2003)
- Metabolic network of of parasitic bacterium has same mean path length as the network of a large multi-cellular organism (Jeong et al. *Nature* 2000)
  - Indicates that mechanisms have maintained the average path length during evolution

Assortativity

- Social networks are assortative
  - Well connected people tend to know each other
- Cellular networks are dissassortative (Maslov & Sneppen *Science* 2002)
  - Highly connected nodes don't link directly to each other
  - Hubs tend to link to nodes with small degree

Subnets of scale-free networks are not scale-free: Sampling properties of networks

Michael P.H. Stump, Carsten Wiuf, and Robert M. May

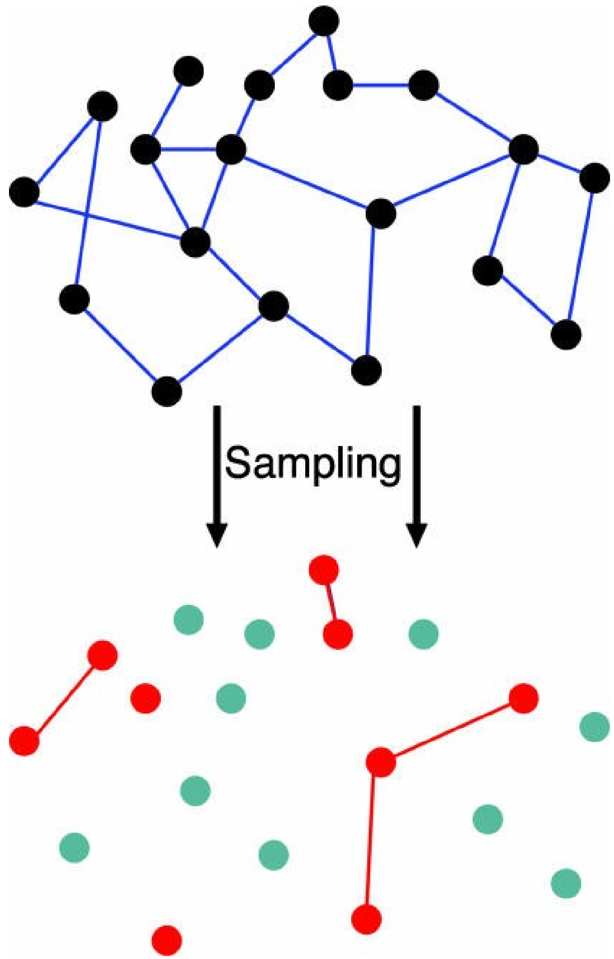
PNAS 2005

Introduction

- Over the last few years, many networks have been characterized as scale-free
- But many of the networks have been subnets of much larger networks
  - Protein interaction networks
  - Gene regulation networks
  - Metabolic networks
- For some model organisms, protein interaction data covers < 20% of the proteins known to exist
- How well does a subnet represent its network?
- Claim: random subnets of scale-free networks are not scale-free themselves

Random Sampling

- Start with a complete network of size N
- Include each node in the subnet with probability p



Results

- Deviation from scale-free behavior is more pronounced as γ increases.
- Subnets have more nodes with few connections.
- As k increases, subnets follow power-law.