Structure of biological networks

Presentation by
Atanas Kamburov

Overview

- Motivation
- Definitions
- Large-scale properties of cellular networks
    - the scale-free property
    - the small-world effect
    - disassortativity
    - modularity
- Bottom-up characterization of cellular networks: network motifs
- Summary
- Literature

Motivation

- Biological functions can rarely be related to *individual molecules*
- Instead, they arise from the **complex molecular interactions** that take place in the cell
- The numerous interactions form **interaction networks** where nodes represent molecules and edges represent functional interactions
- understanding the **structure** and dynamics of such networks is a central challenge for biology
 ➔ network biology, systems biology

Network shapes
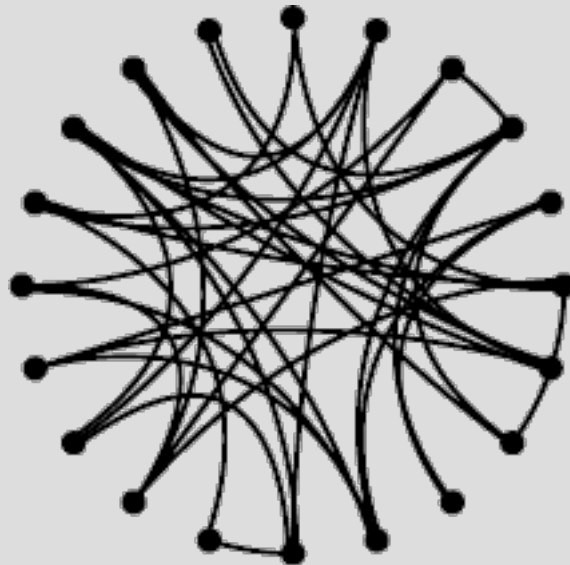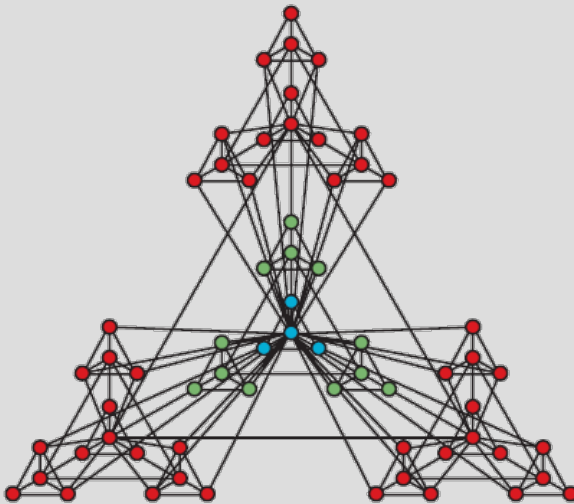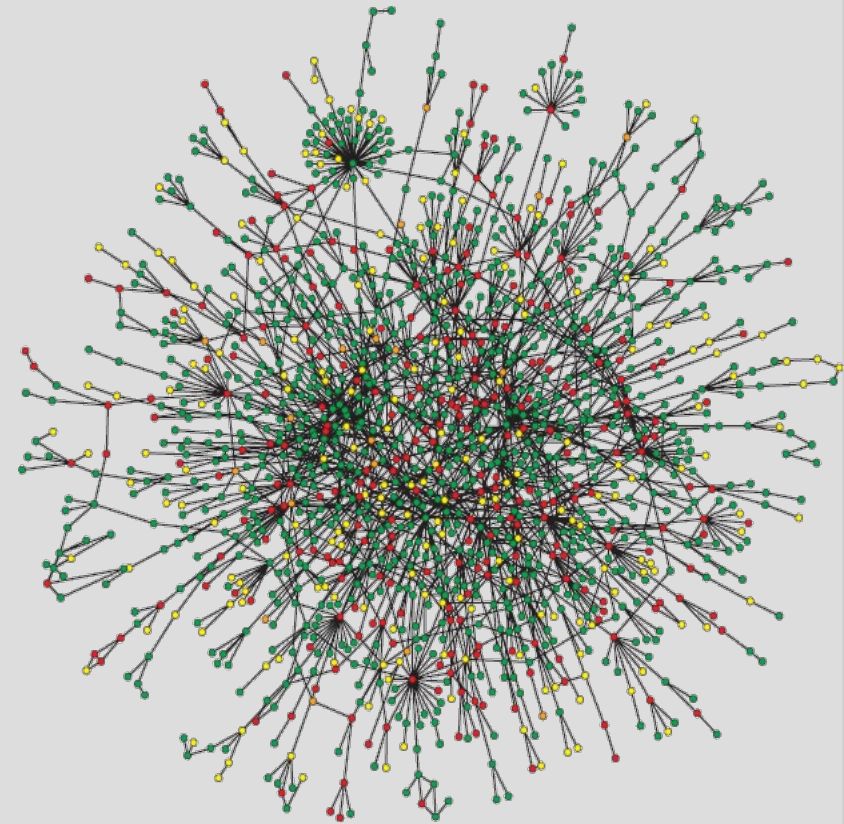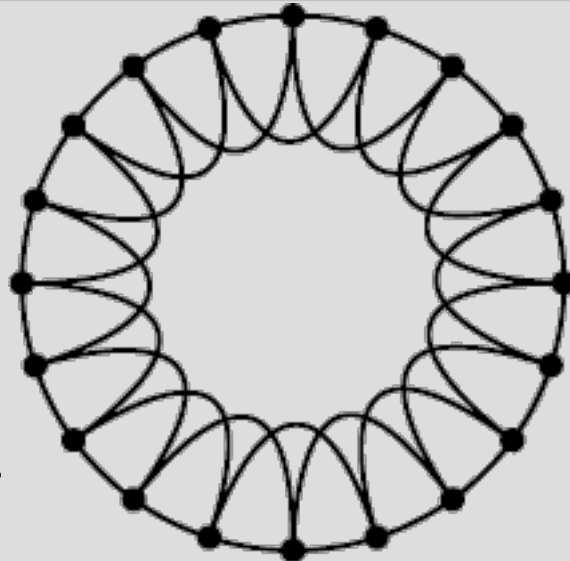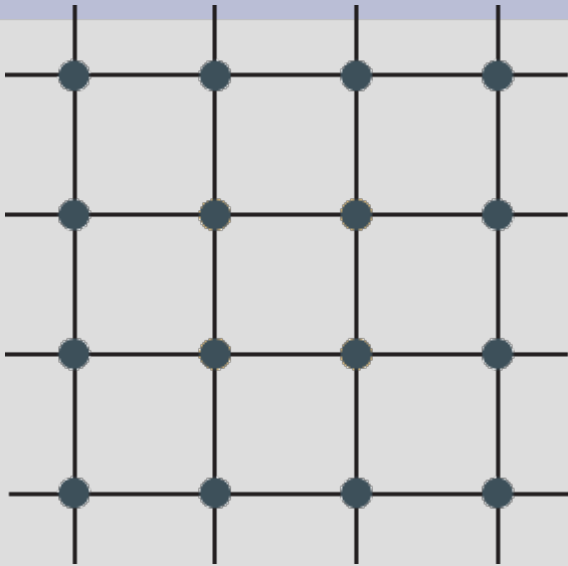
Network characterization

- Networks can be characterized using global and/or local measures
- global properties
  - node connectivity distribution
  - path lengths
- local properties
  - node clustering ratio
 - frequent patterns (motifs)

Some definitions

- Networks can be **directed** or **undirected** depending on the nature of the interactions
  - *directed*: material flow, information flow
  - *undirected*: physical interactions
- **Degree (connectivity) k_i of a node** is the number of links that the node has to other nodes
  - directed networks have two types of degree: **in**degree and **out**degree
- **Average degree <k> = sum(k_i) / #nodes**
- **Degree distribution P(k)** is the probability that a randomly picked node has k links

**a** Undirected network

$k = 5$
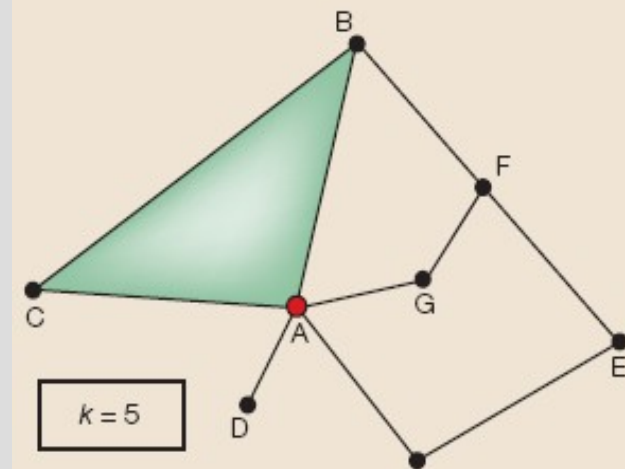
b Directed network

$k_{in} = 4$

$k_{out} = 1$
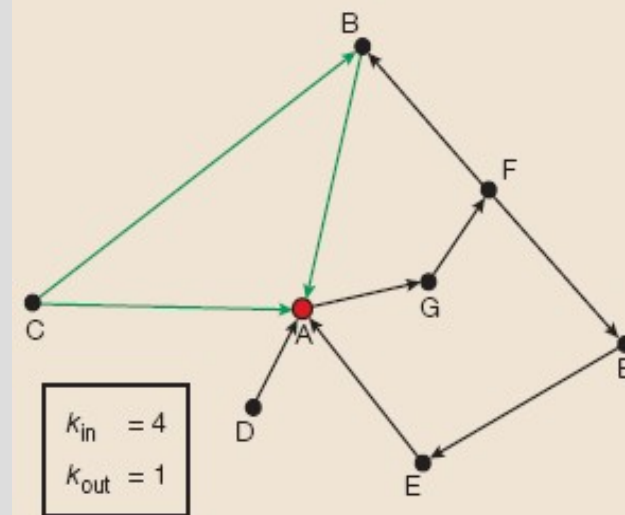
The scale-free property

- For decades, cellular networks were considered to be either completely regular or completely random
- Both models are unable to explain the existence of nodes with high connectivity

- Recent studies show that in many biological networks, P(k) follows a power law distribution $P(k) \sim k^{-\gamma}$ with $2 < \gamma < 3$
- **Scale-free** network
  - many nodes with few links
  - few nodes with many links **(hubs)**
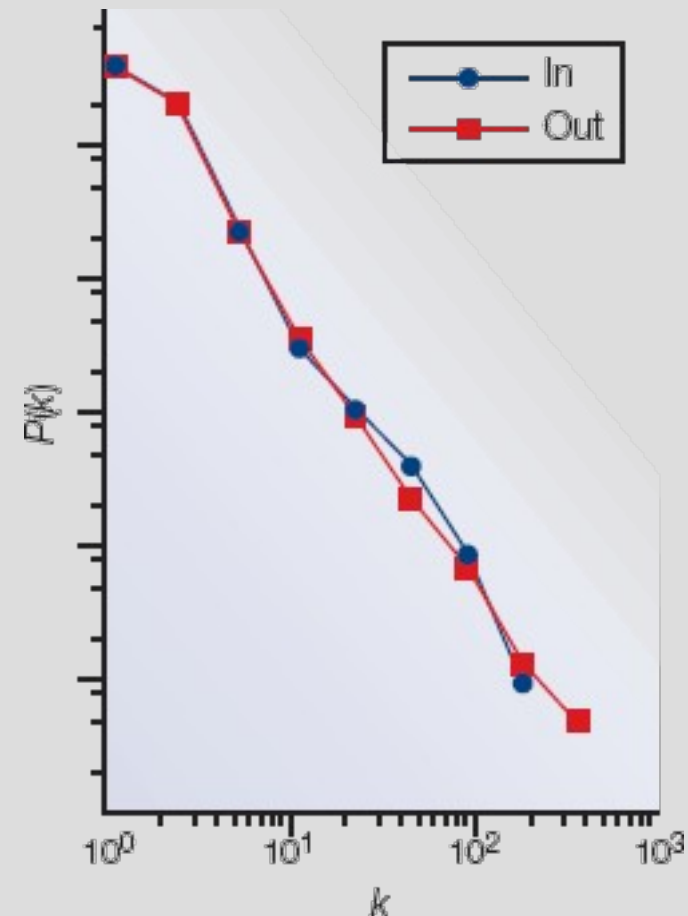
The scale-free property

- Metabolic networks have scale-free nature

- Most protein interaction networks and gene regulatory networks also have scale-free nature

- **Not every** cellular network is scale-free
 - e.g. transcription regulatory networks of S.cer. and E.coli
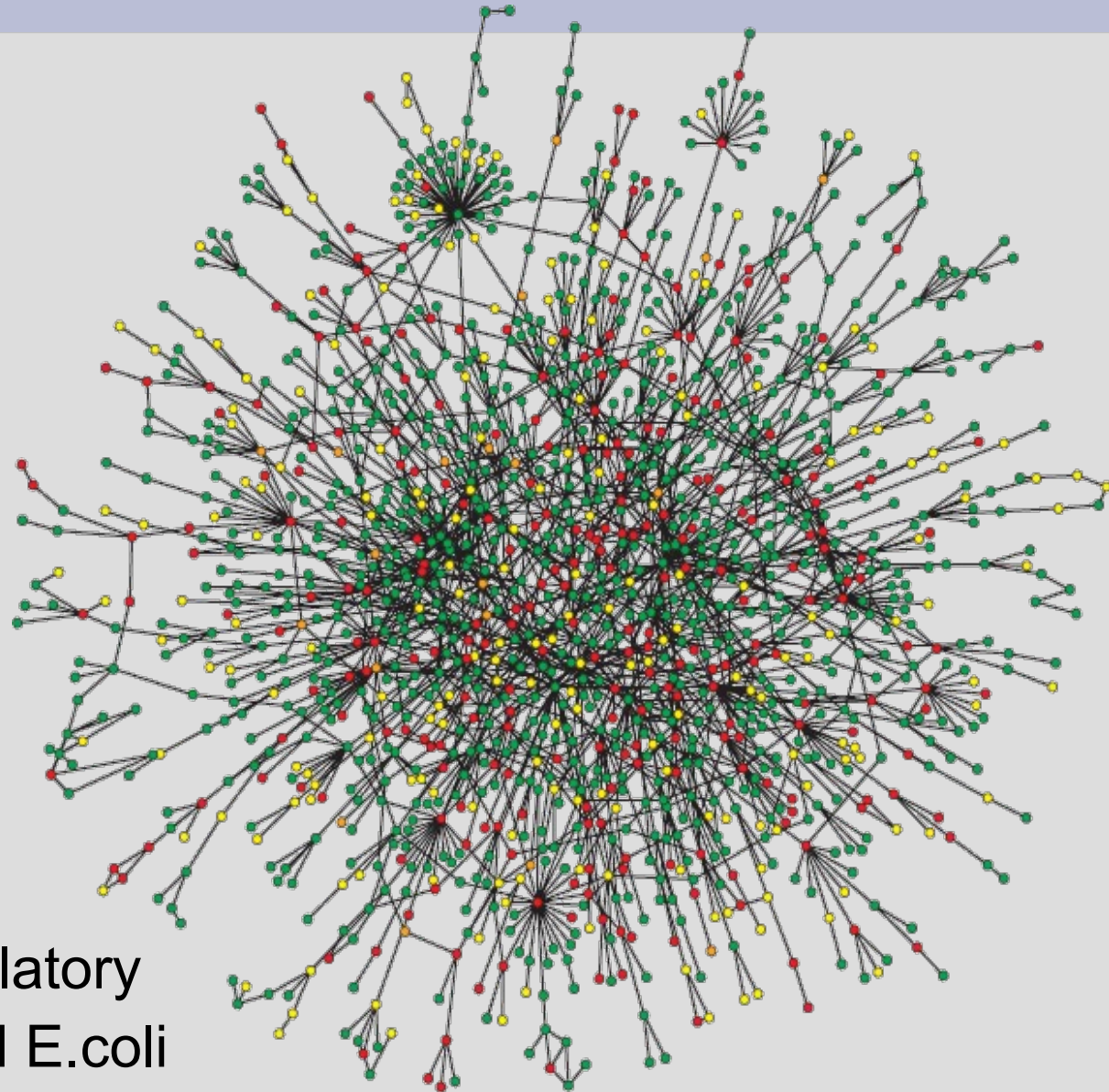
The scale-free property

- Effects of the scale-free nature of cellular networks
  - Robustness: random removal of nodes often does not affect the network integrity and function
  - Attack vulnerability of hubs: if any hubs are removed, the networks is disintegrated
    - In yeast, only 10% of the proteins with less than 5 links are essential; from proteins with more than 15 links, over 60% are essential

The scale-free property

- Evolutionary origin of the scale-free nature of cellular networks:
  two fundamental processes
  - growth process: new nodes join the system
    over an extended time period
  - preferential attachment: new nodes prefer
    to connect to nodes with many connections
- The scale-free property of protein interaction
  networks is probably rooted in gene duplication
  - products of duplicated genes have same interaction partners
  - the duplication of a hub's interaction partner is more probable


After duplication

Proteins

Before duplication

Proteins
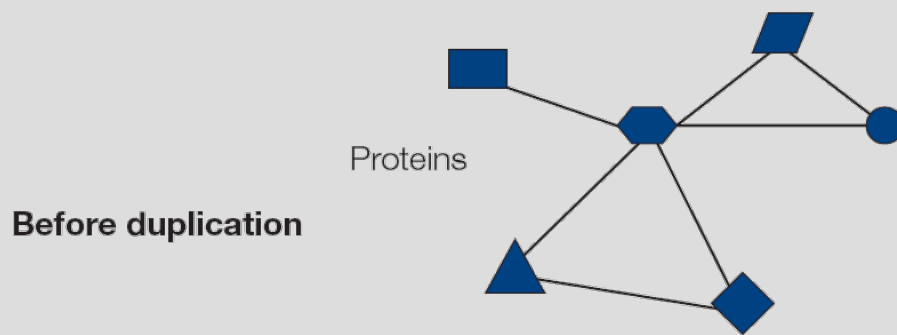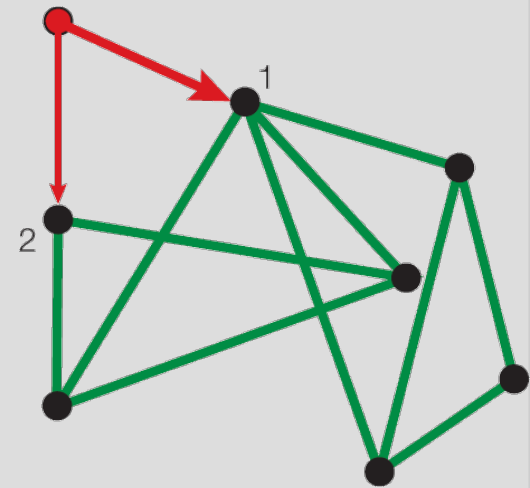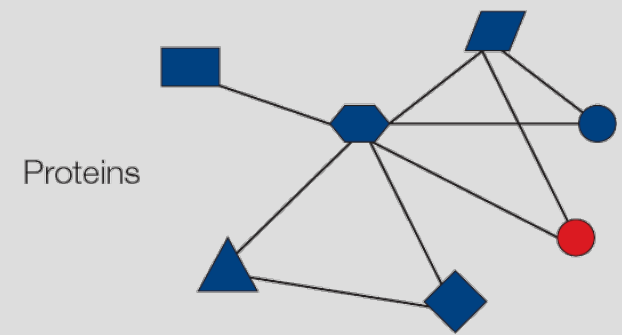
Disassortativity of biological networks

- An interesting observation: hubs avoid connecting to each other **(disassortativity)**

- To this point, dissasortativity cannot be explained by network biologists

The small-world property

- In regular networks, the average path length L between couples of nodes is relatively long
- L is shorter for random networks: the **small-world** property
- Even a few rewiring procedures of a regular network suffice to achieve a small-world effect
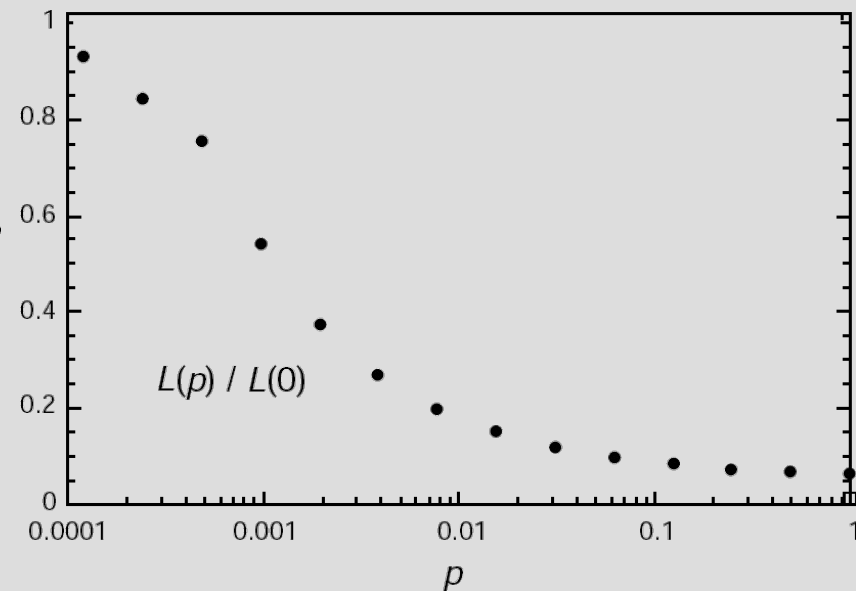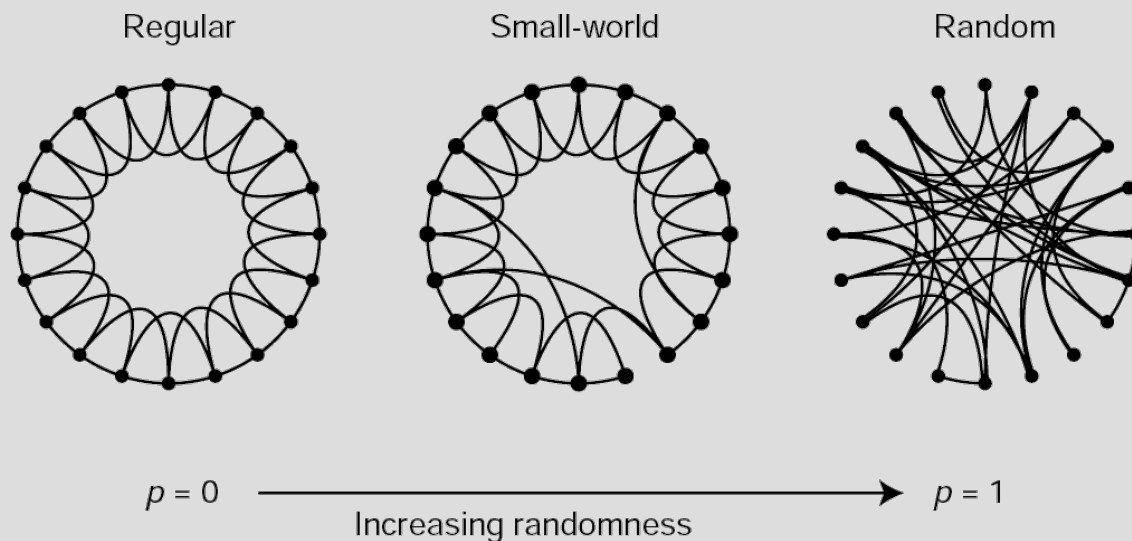
The small-world property

- Biological networks have even smaller average path lengths than random networks: **ultra-small-world** effect
  - first documented for metabolism
    - → often, 3-4 reactions suffice to convert one metabolite to another
    - → local concentration perturbations quickly reach the whole system
    - → ultra-small-world effect underlies selection: mean path length of metabolic networks is the same in bacteria and complex multicellular species
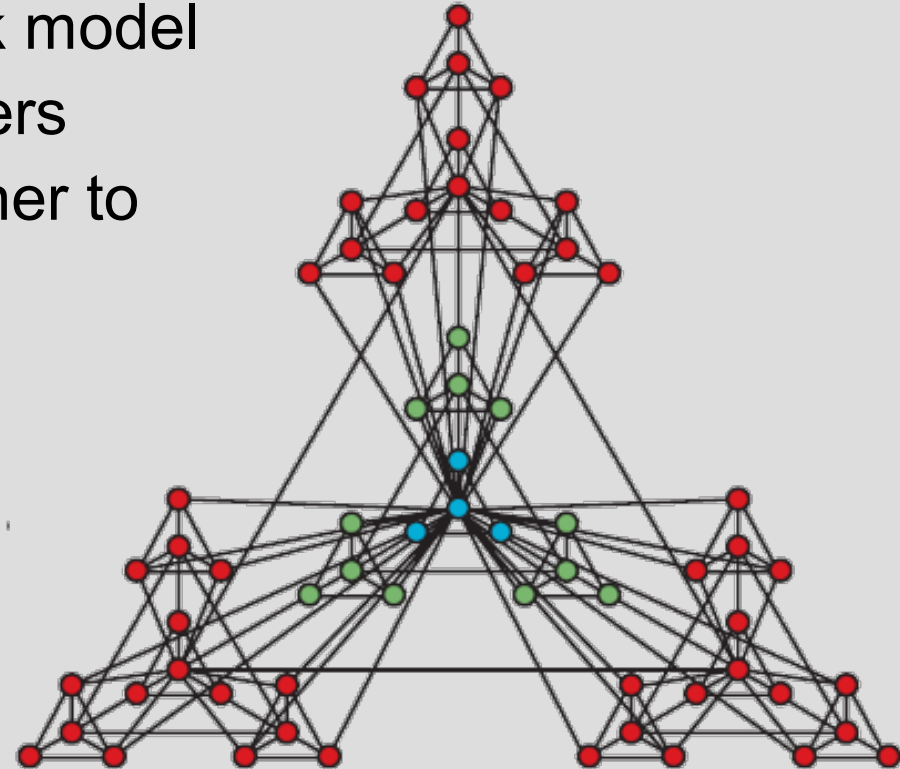
Modularity
of cellular networks

- Cellular functions often carried out in a highly **modular** manner
  - protein complexes
  - temporally coregulated molecules that control cell cycle
- A sign for this is the significantly high **average clustering coefficient <C>** of most cellular networks
  - ➜ **The clustering coefficient $C_I = 2n_I / k_I(k_I-1)$** shows the fraction of neighbours of node I that are directly linked to each other (n_I is the number of links between neighbours of I)
  - ➜ <C>=sum(C_I) / number of nodes

$k = 5$
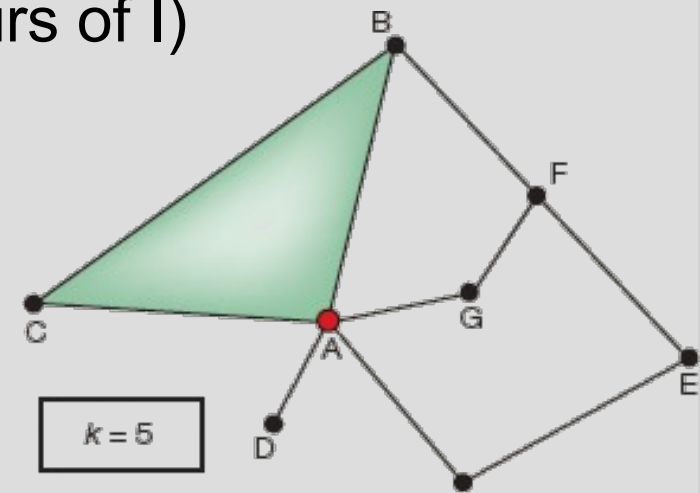
Modularity
of cellular networks

- Functional modules are interconnected in the cell to form a "Network of networks"

- The scale-free property and modularity coexist in the hierarchical network model (Ravasz et al.,2002), where clusters are combined in an iterative manner to generate a hierarchical network
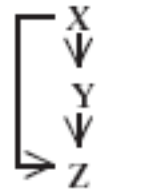
Architectural features of cellular networks
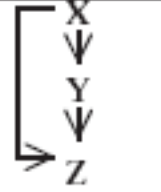
- Topological properties of most biological networks
    - **scale-free**
        - **robustness**
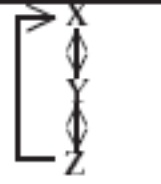   - **attack vulnerability** of hubs
    - **disassortativity**
    - **ultra-small-world** property
   - **modularity**

Network motifs

- Scale-free and hierarchical features determine the large-scale feature of networks
- An alternative approach for network analysis: finding **subgraphs that are significantly overrepresented (motifs)**
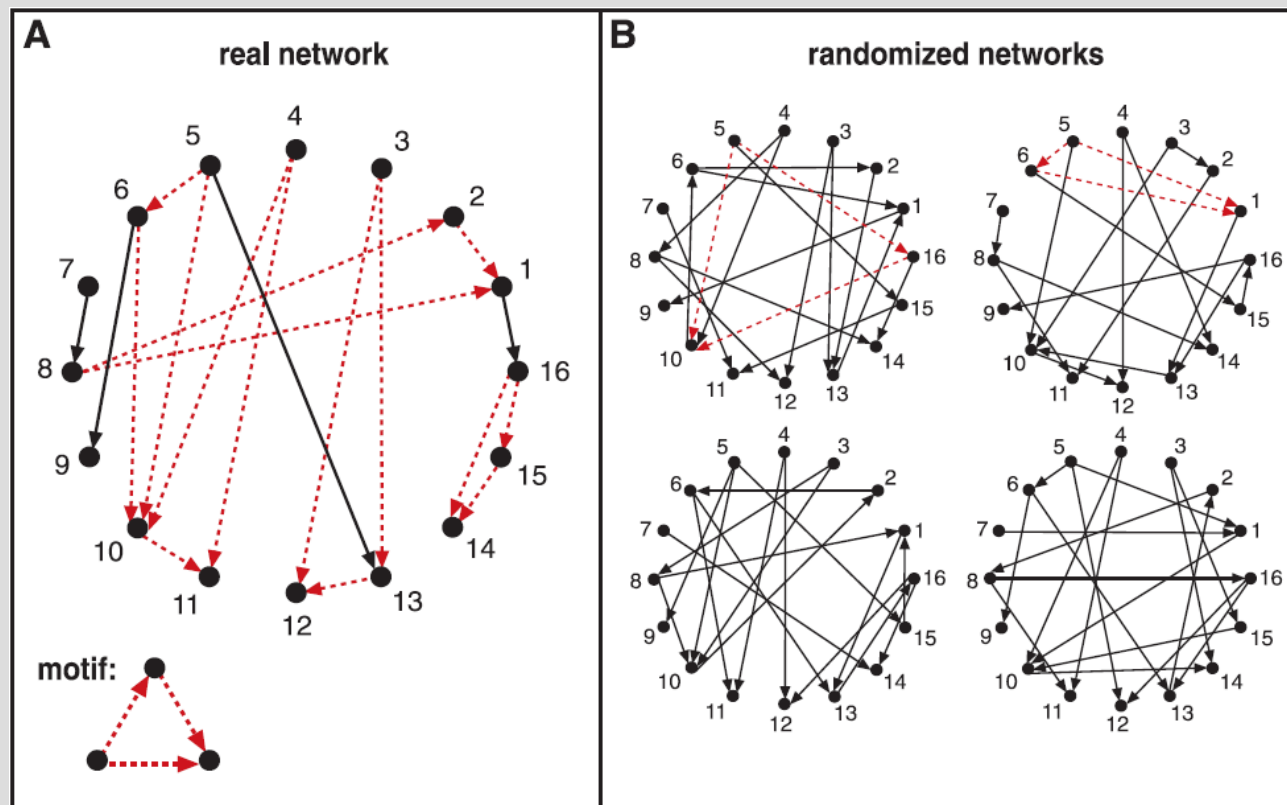
Network	Nodes	Edges	N_{real}	$N_{rand} \pm SD$	Z score	N_{real}	$N_{rand} \pm SD$	Z score	N_{real}	$N_{rand} \pm SD$	Z score
Gene regulation (transcription)				Feed-forward loop			Bi-fan				
E. coli	424	519	40	7 ± 3	10	203	47 ± 12	13			
S. cerevisiae*	685	1,052	70	11 ± 4	14	1812	300 ± 40	41			
Neurons				Feed-forward loop			Bi-fan			Bi-parallel	
C. elegans†	252	509	125	90 ± 10	3.7	127	55 ± 13	5.3	227	35 ± 10	20
Electronic circuits (forward logic chips)				Feed-forward loop			Bi-fan			Bi-parallel	
s15850	10,383	14,240	424	2 ± 2	285	1040	1 ± 1	1200	480	2 ± 1	335
s38584	20,717	34,204	413	10 ± 3	120	1739	6 ± 2	800	711	9 ± 2	320
s38417	23,843	33,661	612	3 ± 2	400	2404	1 ± 1	2550	531	2 ± 2	340
s9234	5,844	8,197	211	2 ± 1	140	754	1 ± 1	1050	209	1 ± 1	200
s13207	8,651	11,831	403	2 ± 1	225	4445	1 ± 1	4950	264	2 ± 1	200
Electronic circuits (digital fractional multipliers)				Three-node feedback loop			Bi-fan			Four-node feedback loop	
s208	122	189	10	1 ± 1	9	4	1 ± 1	3.8	5	1 ± 1	5
s420	252	399	20	1 ± 1	18	10	1 ± 1	10	11	1 ± 1	11
s838‡	512	819	40	1 ± 1	38	22	1 ± 1	20	23	1 ± 1	25
World Wide Web				Feedback with two mutual dyads			Fully connected triad			Uplinked mutual dyad	
nd.edu§	325,729	1.46e6	1.1e5	$2e3 \pm 1e2$	800	6.8e6	$5e4 \pm 4e2$	15,000	1.2e6	$1e4 \pm 2e2$	5000

Network motifs

Motifs are rarely independent:
motif clusters
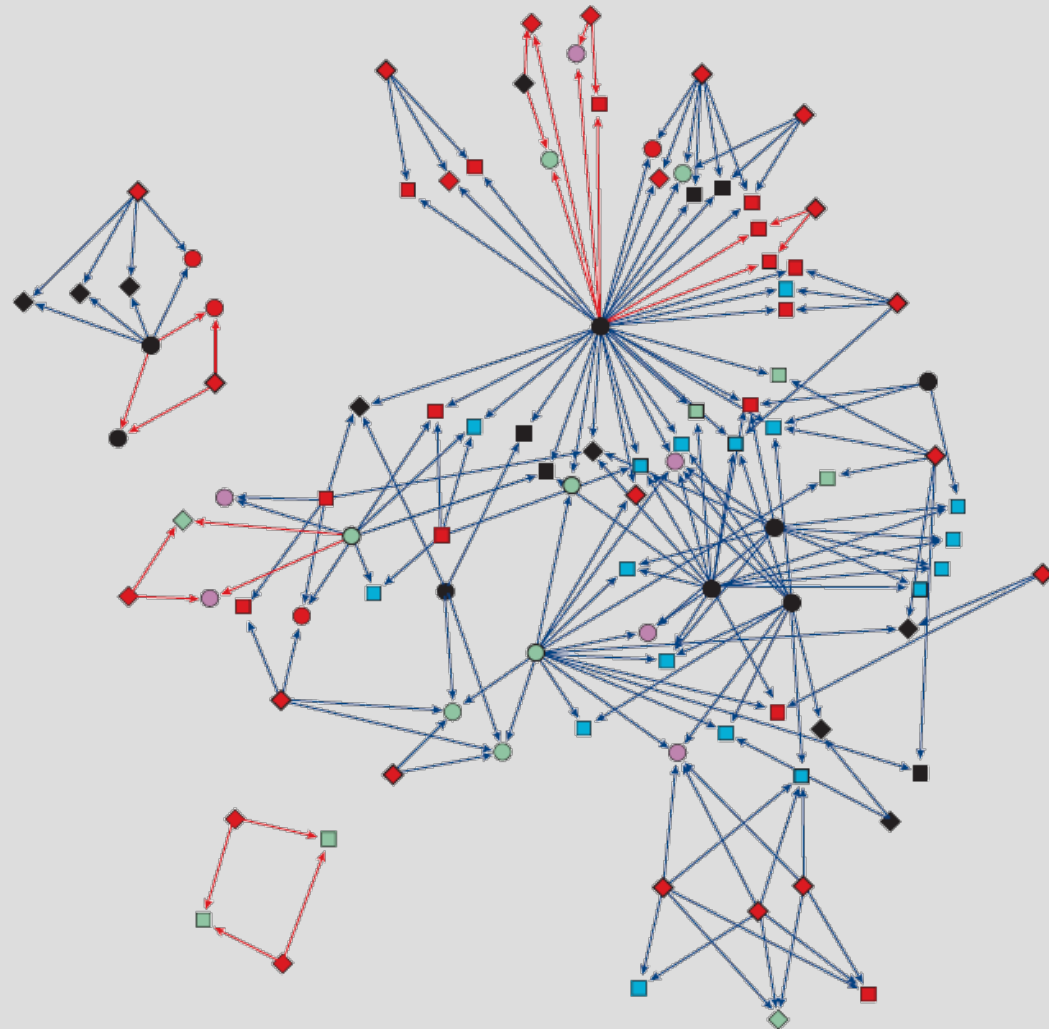
Summary

- Most **(but not all!)** cellular networks are scale-free, ultra-small-world networks organized most probably in a hierarchical manner
- To characterize the large-scale properties of the network, measures like node connectivity, path length, clustering coefficient are used
- Motifs are subgraphs that are significantly overrepresented in real networks. Each class of complex networks has own characteristic motifs

Discussion

- Network biology still at its infancy

- Topology does not determine the function of cellular networks
  - System dynamics

Thanks!

Sources

- Barabasi A, Oltvai Z. **Network biology: Understanding the cell's functional organization** (2004) *Nature Genetics 5 (101-113)*
- Watts D, Strogatz S. **Collective dynamics of "small-world" networks** (1998) *Nature 393 (440-442)*
- Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U. **Network motifs: Simple building blocks of complex networks** (2002) *Science 298 (824-27)*
- Ravasz E, Somera A, Mongru D, Oltvai Z, Barabasi A. **Hierarchical organization of modularity in metabolic networks** (2002) *Science 297 (1551-1555)*

- The above publications are sources for the pictures used in this presentation.