Systems Biology: A Personal View
XI. Intra-cellular Systems III: Metabolism and Modularity

Sitabhra Sinha
IMSc Chennai

Intra-cellular biochemical networks

- ☐ **Metabolic networks**
  Nodes: metabolites (substrates & products of metabolism)
  Links: chemical reactions (directed)
- ☐ Genetic regulatory networks
  Nodes: Genes & Proteins
  Links: regulatory interactions (directed)
- ☐ Protein-Protein interaction network
  Nodes: Proteins
  Links: physical binding and formation of protein complex (undirected)
- ☐ Signaling network
  Nodes: Signaling molecules e.g., kinase, cyclicAMP, Ca
  Links: chemical reactions (directed)

Metabolism

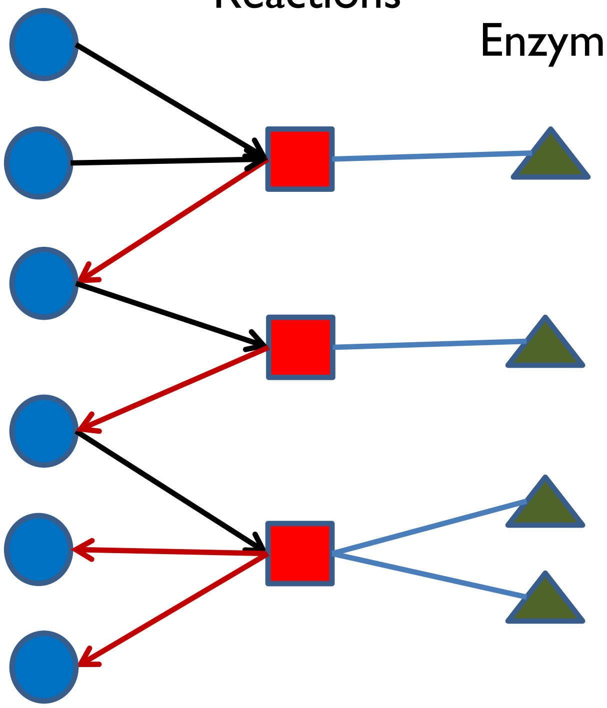
Chemical process through which cells break down nutrients to generate energy and/or into usable building blocks (*catabolic metabolism*) and then reassemble them using energy to form biological molecules necessary for the cell (*anabolic metabolism*).

Uses sequence of chemical reactions (pathways) to convert *substrates* (initial inputs) successively into useful *products*. Reactions are aided by *enzymes*.

Metabolic network: The set of all reactions in all pathways

Representing Metabolic Networks

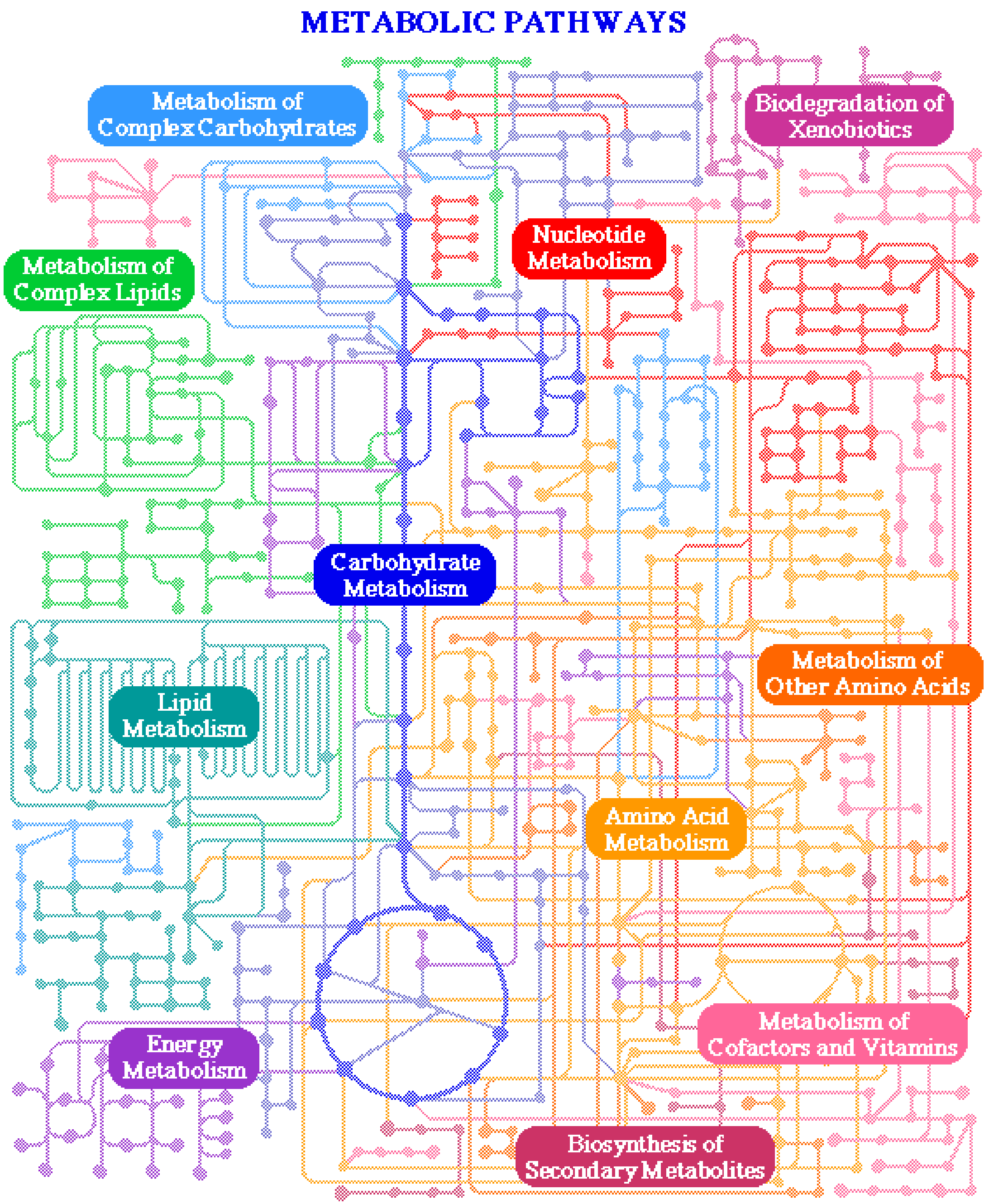
Tripartite directed network



Metabolites

Reactions

Enzymes



KEGG

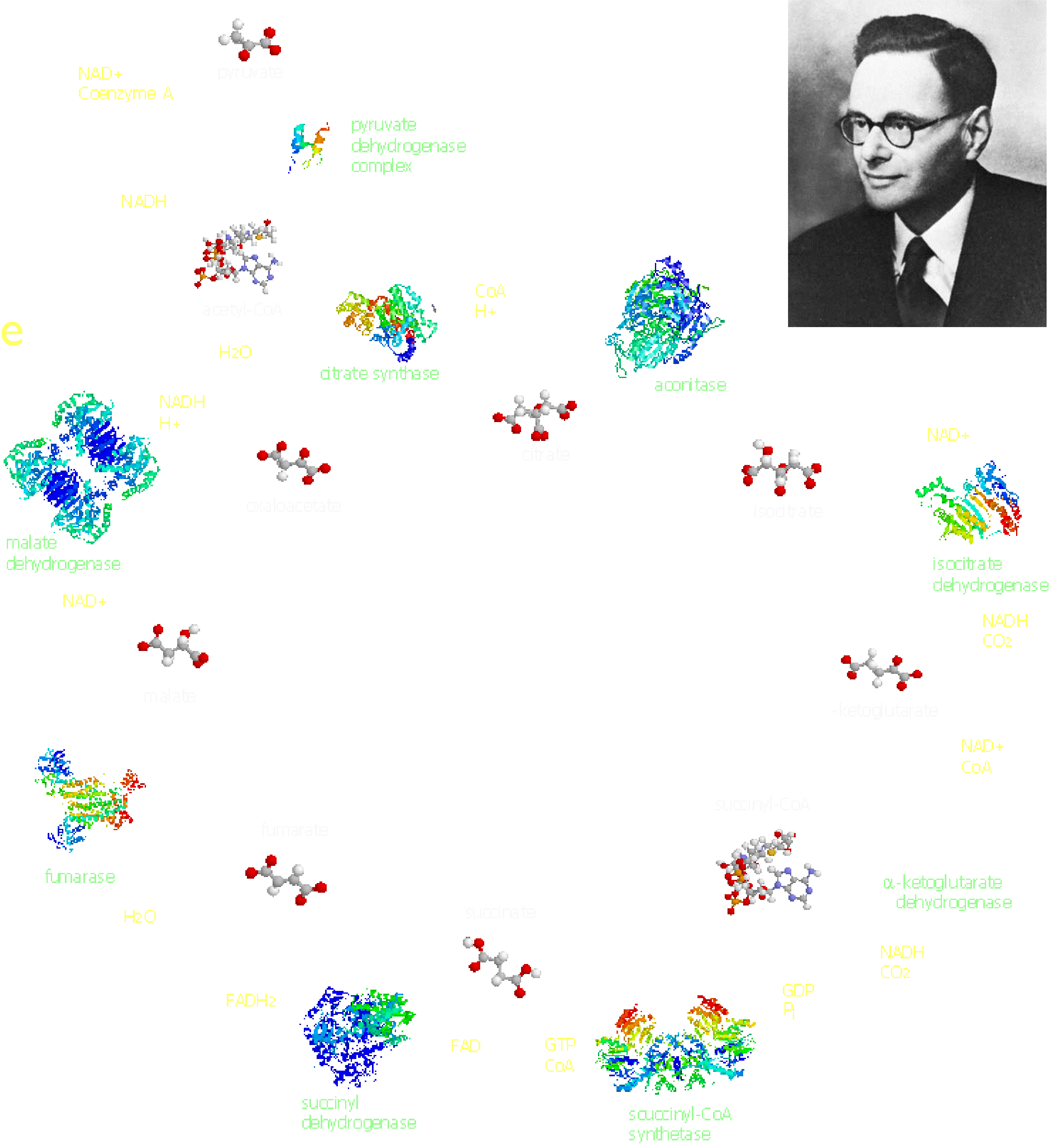
Projection to only metabolites

Citric acid cycle

Also known as Tricarboxylic Acid (TCA) or Krebs cycle

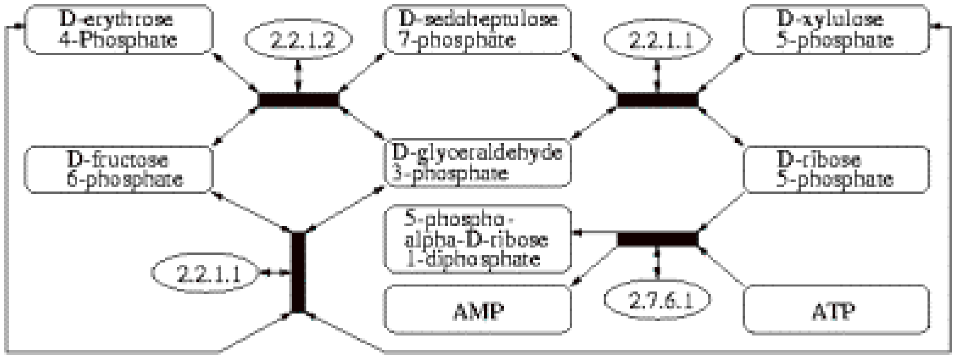
is a series of enzyme-catalysed chemical reactions lying at the heart of aerobic metabolism.

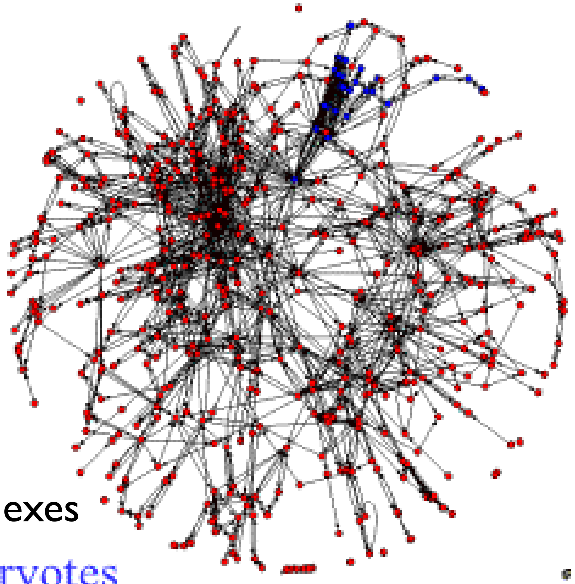
Involved in the breakdown of all 3 major food groups: carbohydrates, lipids and proteins.



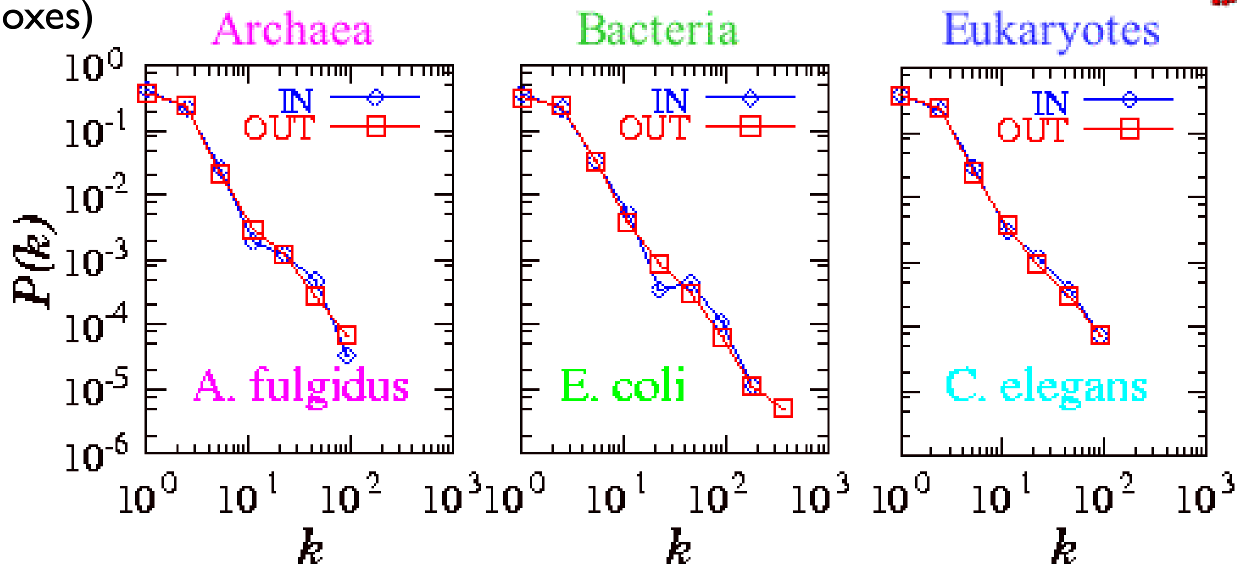
Scale-free nature of degree distribution of metabolic networks

A portion of the WIT database for *E. coli.*





Nodes are substrates & products, linked by enzyme-substrate complexes (black boxes)

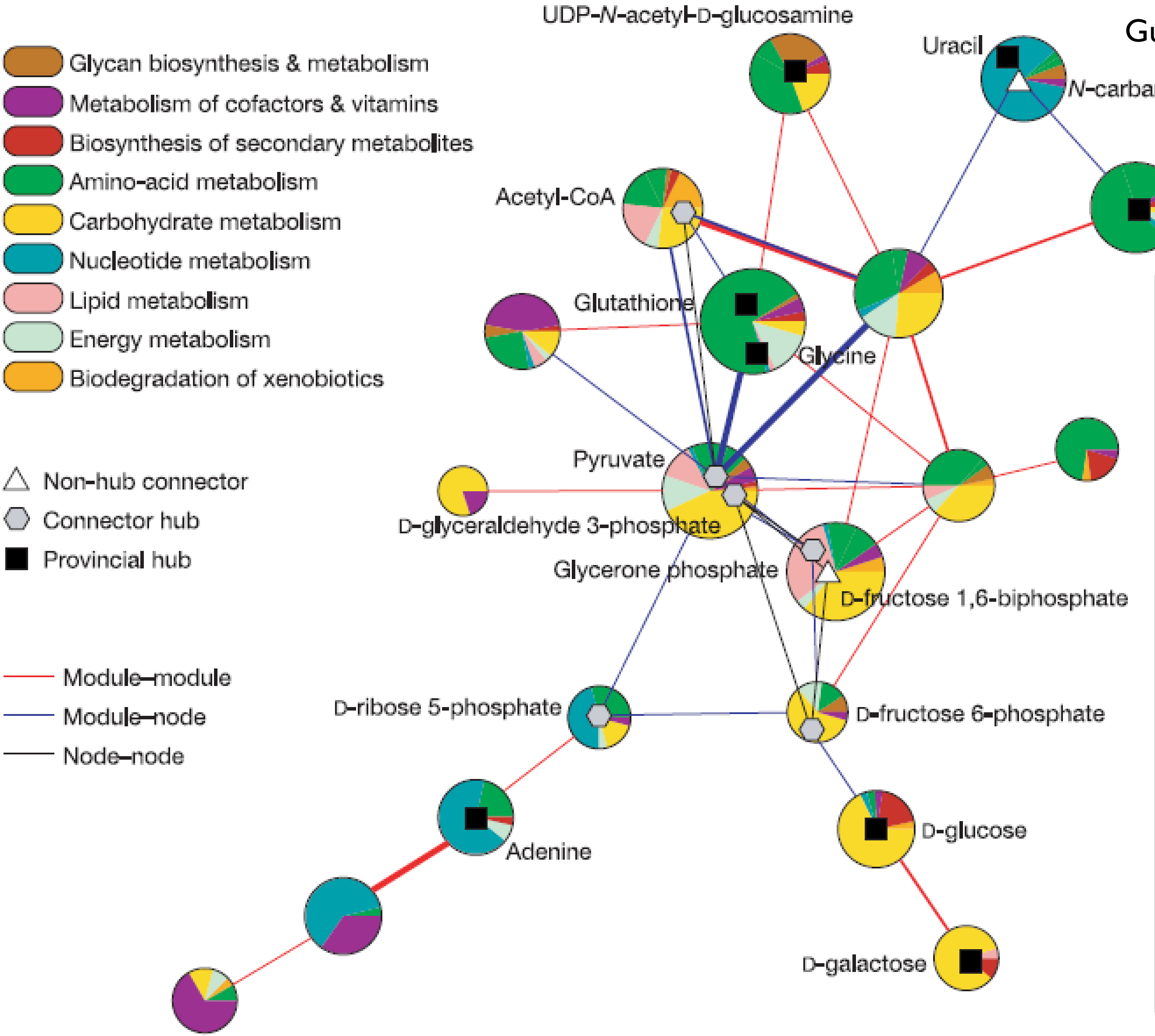


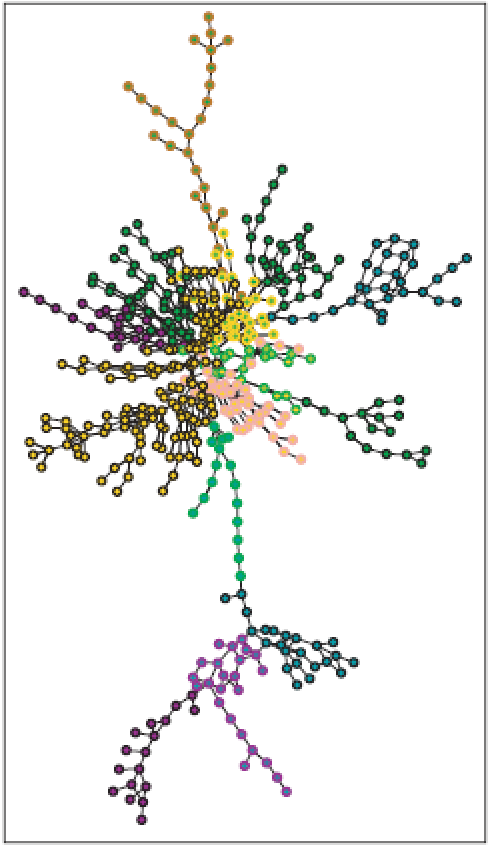
Organisms from all three domains of life are **scale-free** networks!

H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, and A.L. Barabasi, Nature, **407** 651 (2000)

Modular nature of metabolic networks

Guimera and Amaral, *Nature* (2005)





Each circle represents a module and is colored according to the KEGG pathway classification of the metabolites it contains.

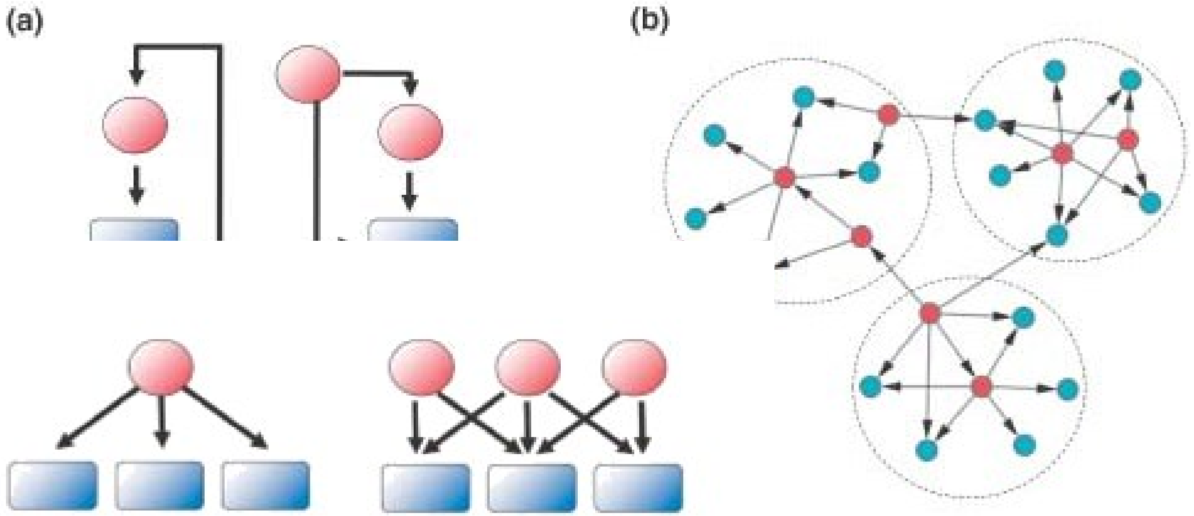
Metabolic network of *E. coli* (N=473, L =674).

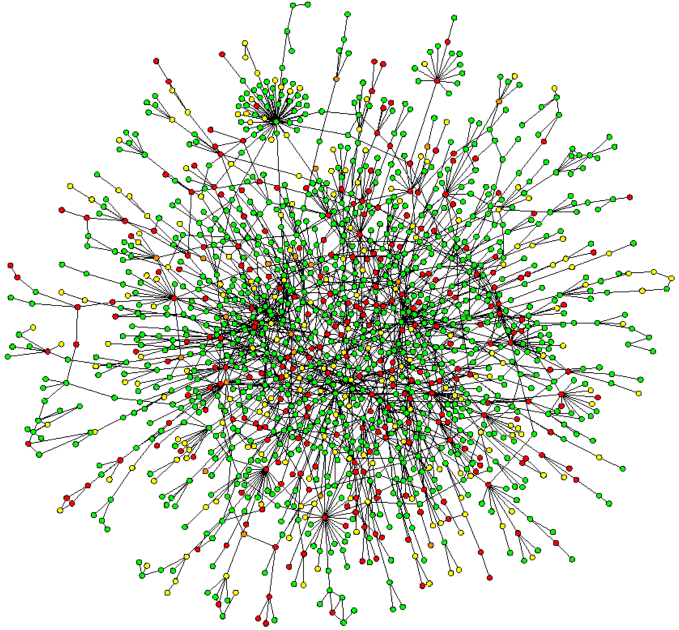
Modular Networks: dense connections *within* certain sub-networks (modules) & relatively few connections *between* modules

Modules: A *mesoscopic* organizational principle of networks

Going beyond ***motifs*** but more detailed than ***global*** description (L, C etc.)

Micro ***Meso*** ***Macro***





Kim & Park, WIREs Syst Biol & Med, 2010

Modular Biology (Hartwell et al, Nature 1999)
Functional modules as a critical level of biological organization

Modules in biological networks are often associated with specific functions

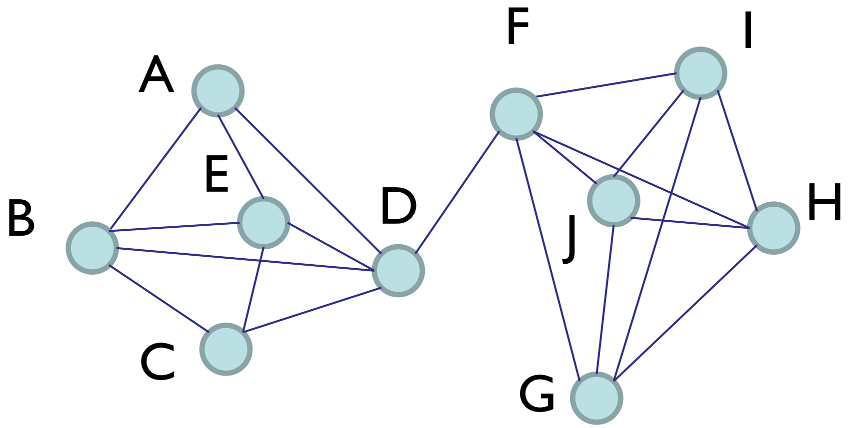
Problem:

Given a network,
how do we find the modules (communities)
into which it can be divided ?

Community Detection in Networks

Also referred to as Graph Partitioning or Module Determination

How to divide the nodes of a network into several groups such that nodes in each group are densely or strongly inter-connected



E.g., it is clear that node clusters I: {A,B,C,D,E} and II: {F,G,H,I,J} constitute two separate groups that are highly intra-connected but has only a single connecting the two groups

The corresponding adjacency matrix will have an almost block-diagonal form – the two blocks corresponding to node clusters I & II

However for large networks the modular character may not be visually apparent – and adjacency matrices need to be partitioned

Graph partitioning

A classic problem in computer science from 1960s

How to divide the nodes of a network into a given number of non-overlapping groups of given sizes such that the number of edges between groups is minimized ?

A generalization of this problem,
How to divide the nodes into several groups such that most links are within groups and few links are between groups

referred to as

Community detection

How we define "most" and "few" can vary from one algorithm to another

Spectral partitioning

Fiedler 1973

Consider a network of N nodes and L links

Aim: to divide the N nodes into 2 groups (Groups A and B, say) to reduce the cut size (number of links between the two groups)

$R = (1/2)\ \Sigma_{ij} A_{ij}$ such that i and j belong to different groups

Partitioning into more than 2 groups can be done by repeated bisection

For each node, a label $s = \{-1,+1\}$ is defined

$s_i = +1$ if node i belongs to group A, $= -1$ if i belongs to group B

Thus

$(1/2)\ (1 - s_i s_j) = 1$ if i & j are in different groups,

$= 0$ if i & j are in same group

$\Rightarrow R = (1/4)\ \Sigma_{ij} A_{ij}(1 - s_i s_j) = (1/4)\ \Sigma_{ij} (k_i \delta_{ij} - A_{ij})\ s_i s_j$

$\Rightarrow R = (1/4)\ \Sigma_{ij} L_{ij} s_i s_j$ where $\mathbf{L} = \mathbf{D} - \mathbf{A}$ is the Laplacian matrix

In matrix notation $R = (1/4)\ \mathbf{s}^T \mathbf{L}\ \mathbf{s}$ where $\mathbf{s} = \{ s_1 s_2 \ldots s_N \}$

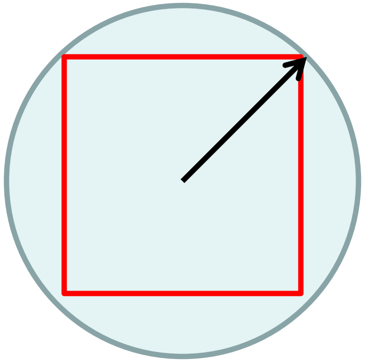
Goal of Partitioning: To find **s** that minimizes **R** given **L**

Partitioning as minimization

Fiedler 1973

If s_i were allowed to take any possible value, then differentiation gives the optimum

But s_i are restricted to $\{-1,+1\} \Rightarrow$ difficult problem



s can be seen as a vector that points to any one of the 2^N vertices of N-dimensional hypercube

Possible approximate solution:
Allow s_i to take any value subject to the constraints that
(i) $\Sigma_i s_i^2 = N \Rightarrow$ **s** is a vector in N-dimensional unit hypersphere
(ii) $\Sigma_i s_i = N_A - N_B$ where N_A, N_B are the sizes of the two groups

In matrix notation $\mathbf{I}^T \mathbf{s} = N_A - N_B$

The minimization problem can now be solved as

$$\partial/\partial s_i \left[\Sigma_{jk} L_{jk} s_j s_k + \lambda \left(N - \Sigma_j s_j^2 \right) + 2\mu \left([N_A - N_B] - \Sigma_j s_j\right)\right] = 0$$

where λ, μ are Lagrange multipliers for enforcing the constraints

$\Rightarrow \Sigma_j L_{ij} s_j = \lambda s_i + \mu$ In matrix notation, $\mathbf{L}\,\mathbf{s} = \lambda\,\mathbf{s} + \mu\,\mathbf{I}$

Partitioning using the Laplacian spectrum

Multiplying $\mathbf{L}\,\mathbf{s} = \lambda\,\mathbf{s} + \mu\,\mathbf{1}$ by $\mathbf{1}^T$ on the left

we get $\lambda\,[N_A - N_B] + \mu\,N = 0 \quad \Rightarrow \quad \mu/\lambda = -[N_A - N_B]/N$

Using $\mathbf{1}^T\,\mathbf{s} = N_A - N_B$ and $\mathbf{1}^T\,\mathbf{L} = 0$ (1 is eigenvector of L with eigenvalue 0)

Defining a new vector $\mathbf{x} = \mathbf{s} + (\mu/\lambda)\,\mathbf{1} = \mathbf{s} - \mathbf{1}\,[N_A - N_B]/N$

$\Rightarrow \mathbf{L}\,\mathbf{x} = \mathbf{L}\,(\mathbf{s} + (\mu/\lambda)\,\mathbf{1}) = \mathbf{L}\,\mathbf{s} = \lambda\,\mathbf{s} + \mu\,\mathbf{1} = \lambda\,\mathbf{x}$

Thus **x** is an eigenvector of the Laplacian

But which eigenvector ?

The one that gives the smallest value of cut size R

We can't choose $\mathbf{1} = \{1,1,\ldots,1\}$ as it is orthogonal to x because $\mathbf{1}^T\,\mathbf{x} = 0$

Note that cut size is proportional to the eigenvalue λ

as $R = (1/4)\,\mathbf{s}^T\,\mathbf{L}\,\mathbf{s} = (1/4)\,\mathbf{x}^T\,\mathbf{L}\,\mathbf{x} = (1/4)\,\lambda\,\mathbf{x}^T\mathbf{x} = \lambda[N_A N_B]/N$

Thus we have to choose the eigenvector corresponding to the lowest non-zero eigevalue (smallest eigenvalue of L is 0 with eigenvector **1**)

Finally, optimal partition **s** is obtained from $\mathbf{s} = \mathbf{x} + \mathbf{1}\,[N_A - N_B]/N$

From Relaxation Approxn to the Network

For the actual network, the optimal partition **s** is subject to the additional constraint that (i) $s_i = +1$ or -1, and (ii) exactly N_A of the components are $+1$ and N_B are -1

Thus, we need to choose **s** as close as possible to ideal value subject to the constraints ⇒ maximize the vector length, i.e.,

$\mathbf{s}^T \mathbf{s} = \mathbf{s}^T (\mathbf{x} + \mathbf{1}[N_A - N_B]/N) = \Sigma_i \, s_i \, (x_i + [N_A - N_B]/N)$

by assigning $s_i = +1$ for the nodes corresponding to the N_A largest (most positive) values of **x,** i.e., the components of the eigenvector of the lowest non-zero eigenvalue of **L**, and, $s_i = -1$ to the remaining N_B nodes

Note: If $N_A \neq N_B$, we can either choose (i) N_A elements to be $+1$ (N_B elements -1) or (ii) N_A elements to be -1 (N_B elements $+1$)
The one having lower cut size is the optimal partition

Community detection

How to quantify the degree of modularity for a given partitioning of a network into communities ?
Is there a distinction between links within a module and that between a module and the rest ?

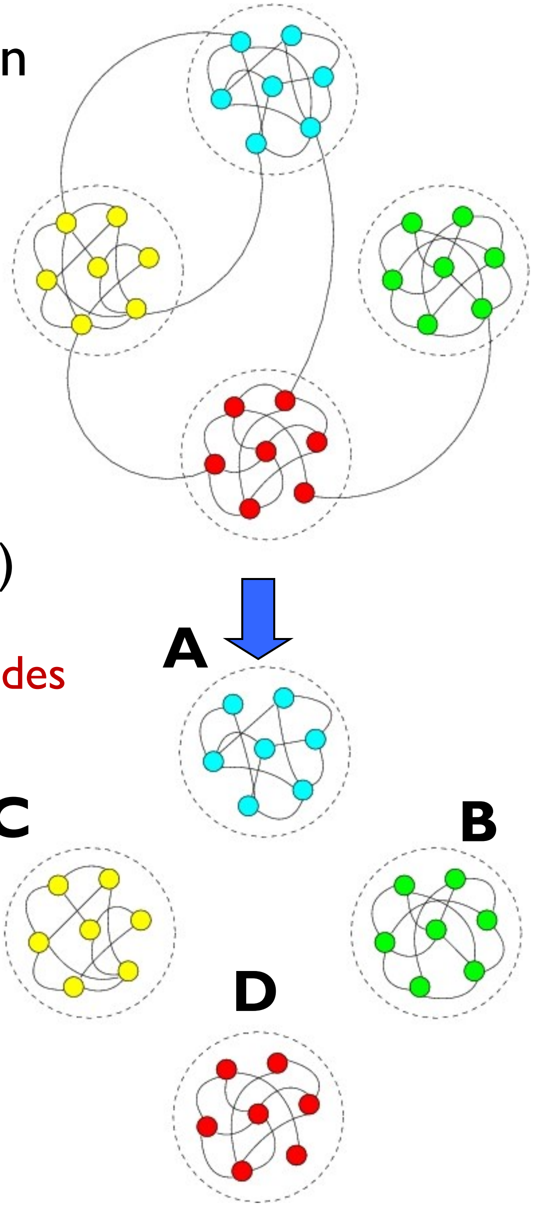
One suggested measure:

$$Q \equiv \frac{1}{2L}\sum_{i,j}\left[A_{ij} - \frac{k_i k_j}{2L}\right]\delta_{c_i c_j}$$

=1 if nodes are in same community

(Newman, EPJB, 2004)

probability of an edge betn 2 nodes proportional to their degrees



A: Adjacency matrix
L : Total number of links
k_i : degree of *i*-th node
c_i : label of module to which *i*-th node belongs

For a random network, Q = 0
i.e., the connection density within a module is no different from that anywhere else in the network

Community detection

For directed & weighted networks:

$$Q^W \equiv \frac{1}{L^W}\sum_{i,j}\left[W_{ij} - \frac{s_i^{\mathrm{in}} s_j^{\mathrm{out}}}{L^W}\right]\delta_{c_i c_j} \qquad \left(L^W = \sum_{i,j} W_{ij}\right)$$

W: Weight matrix

s_i : strength of *i*-th node

Modules determined through a generalization of the spectral method (Leicht & Newman, 2008)

Calculate eigenvector corresponding to largest +ve eigenvalue of symmetrized modularity matrix **B + B^T** where

$$B_{ij} = W_{ij} - [s_i^{\mathrm{in}}\, s_j^{\mathrm{out}} / L^W]$$

and then assign communities based on the signs of the elements of the eigenvector.

Simplest generalization of the method to more than 2 communities is to use repeated bisection