# Systems Biology Across Scales: A Personal View
# XII. Intra-cellular Systems II: 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

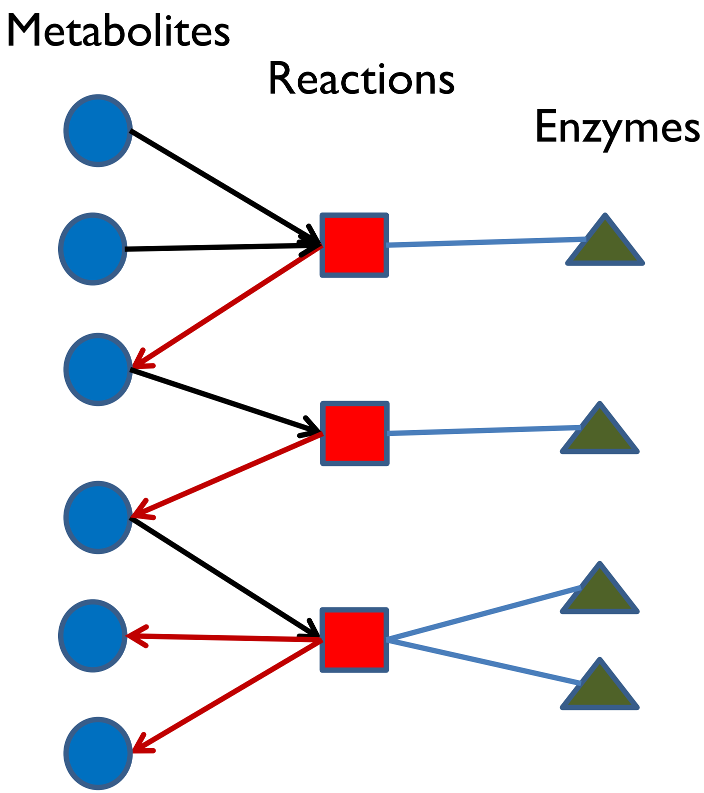
Chemical process through which cells break down nutrients to generate energy and/or into usable building blocks (*catabolic metabolism*) and then reassemble them using energy to form biological molecules necessary for the cell (*anabolic metabolism*).

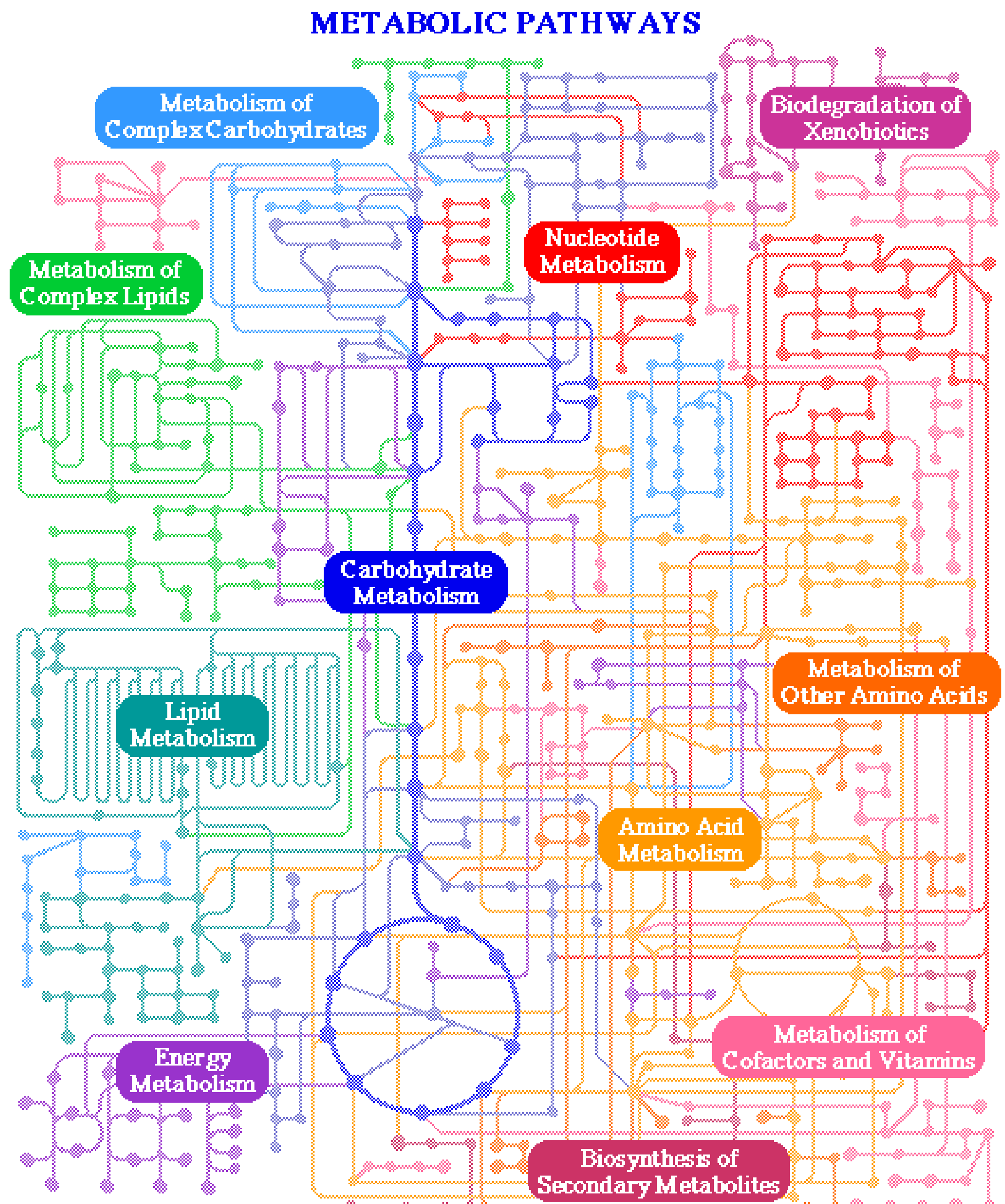
Uses sequence of chemical reactions (pathways) to convert *substrates* (initial inputs) successively into useful *products*. Reactions are aided by *enzymes*.

Metabolic network: The set of all reactions in all pathways

# Representing Metabolic Networks

Tripartite directed network




Projection to only metabolites

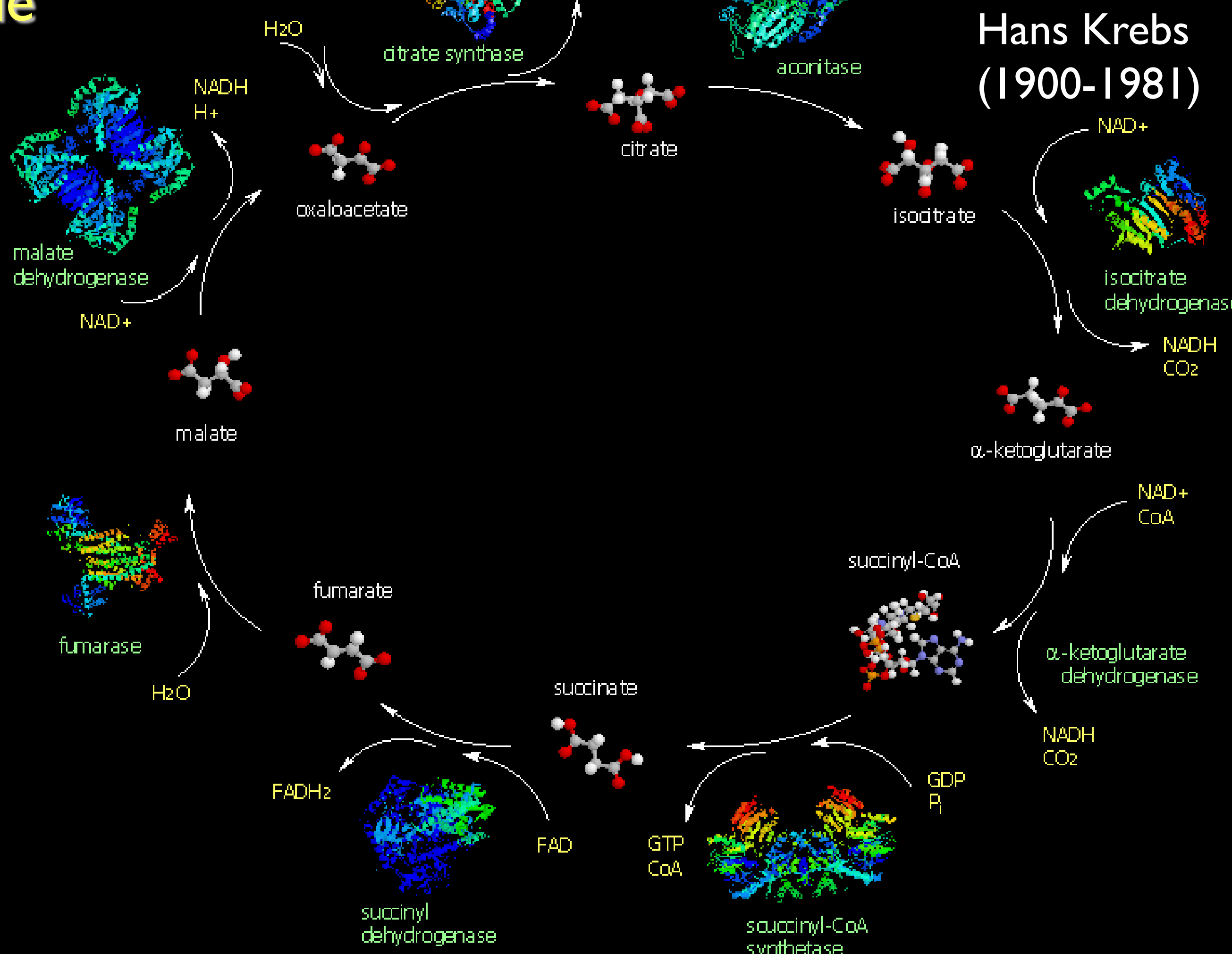
KEGG

# 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.





Hans Krebs (1900-1981)

# Graphical representation of metabolic networks

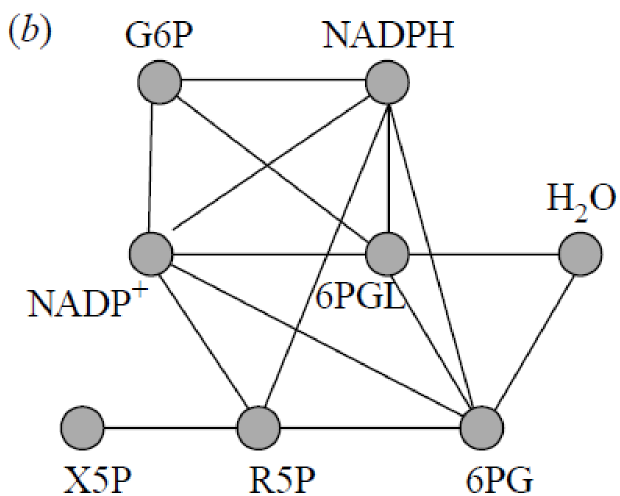
(*a*)

1 glucose 6-phosphate (G6P) + 1 $NADP^+$ $\xRightarrow{zwf}$ 1 6-phosphoglucono δ-lactone (6PGL) + 1 NADPH

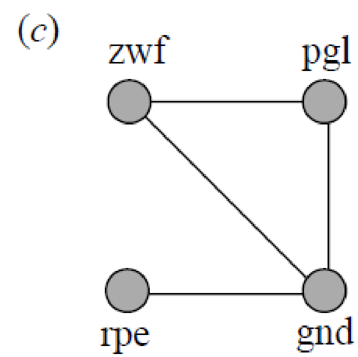
1 6-phosphoglucono δ-lactone + 1 $H_2O$ $\xRightarrow{pgl}$ 1 6-phosphogluconate (6PG)

1 6-phosphogluconate + 1 $NADP^+$ $\xRightarrow{gnd}$ 1 ribulose 5-phosphate (R5P) + 1 NADPH

1 ribulose 5-phosphate $\xLeftrightarrow{rpe}$ 1 xylulose 5-phosphate (X5P)




Substrate graph

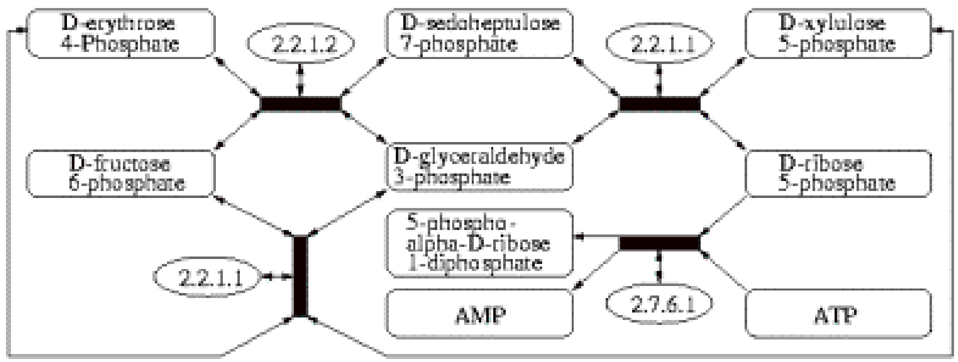


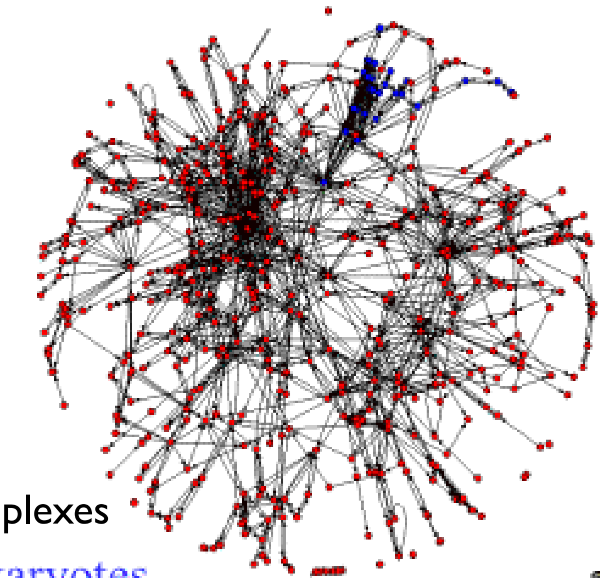

Reaction graph

Wagner & Fell (2001) The small world inside large metabolic networks, *Proc Roy Soc Lond B*

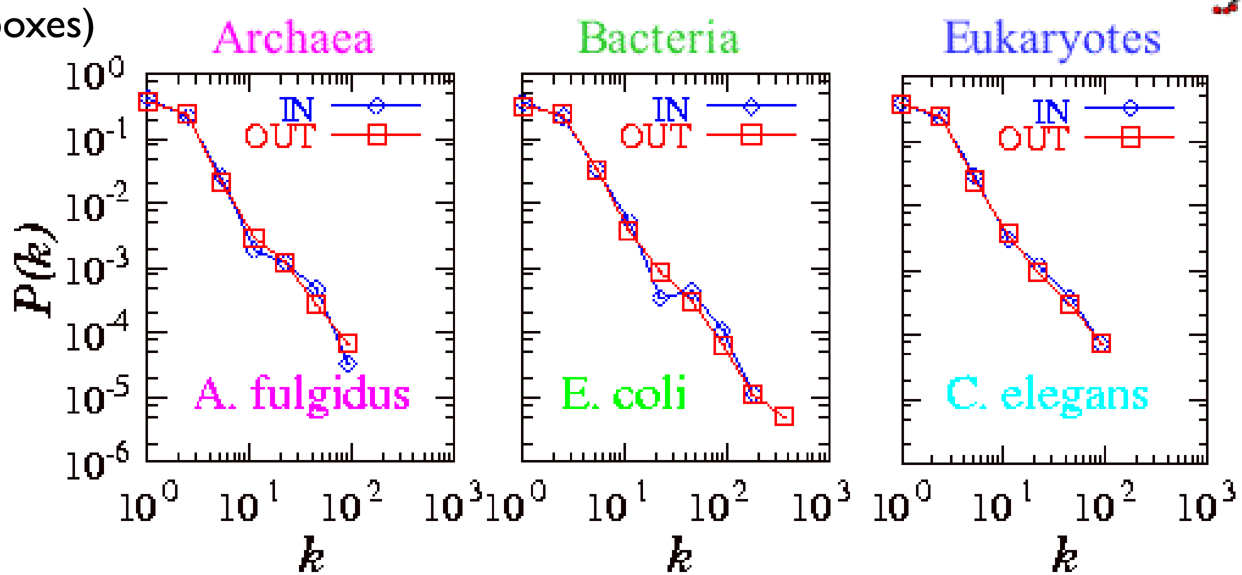
# Scale-free nature of degree distribution of metabolic networks

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



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

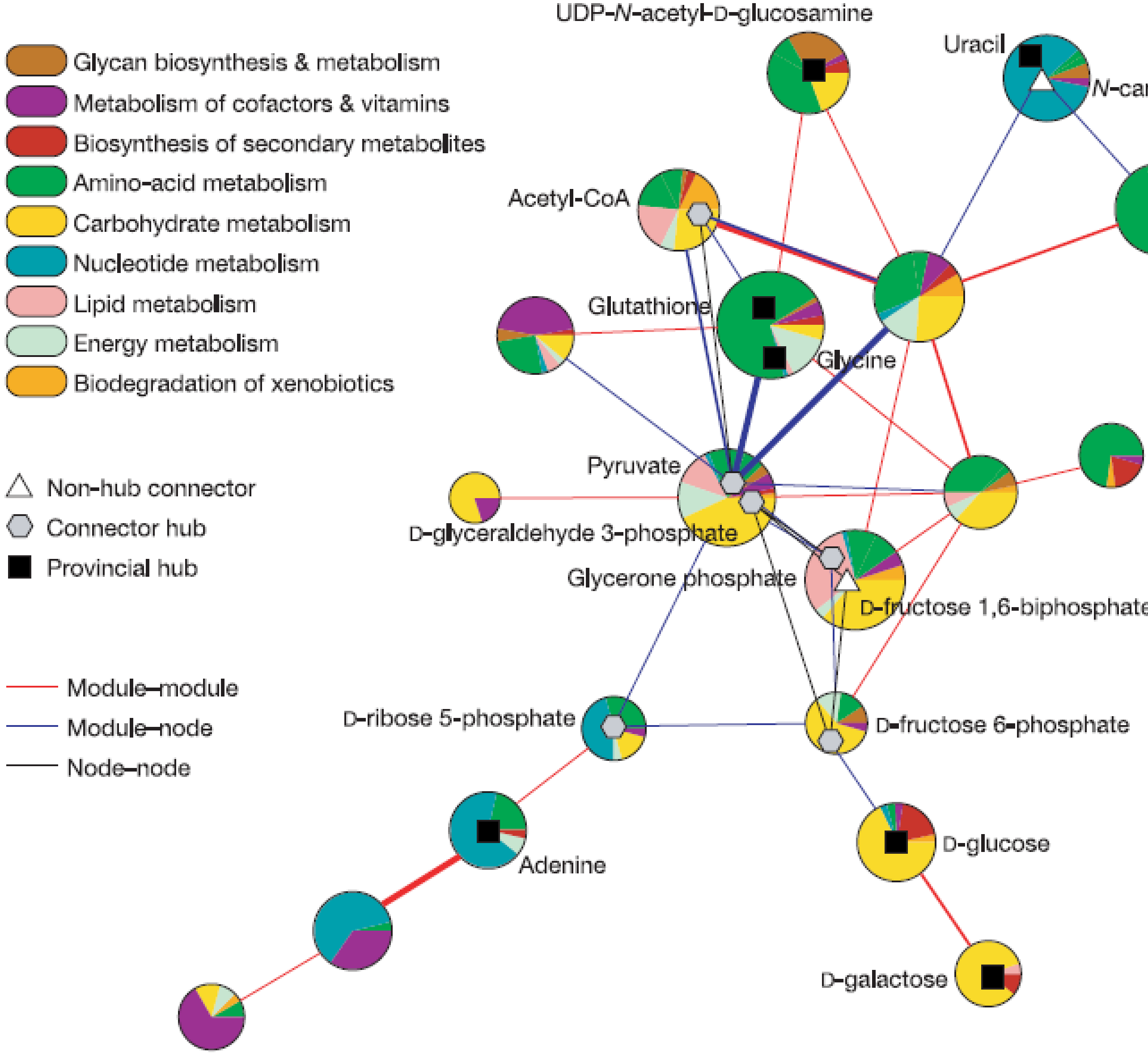



Organisms from all three domains of life are **scale-free** networks!

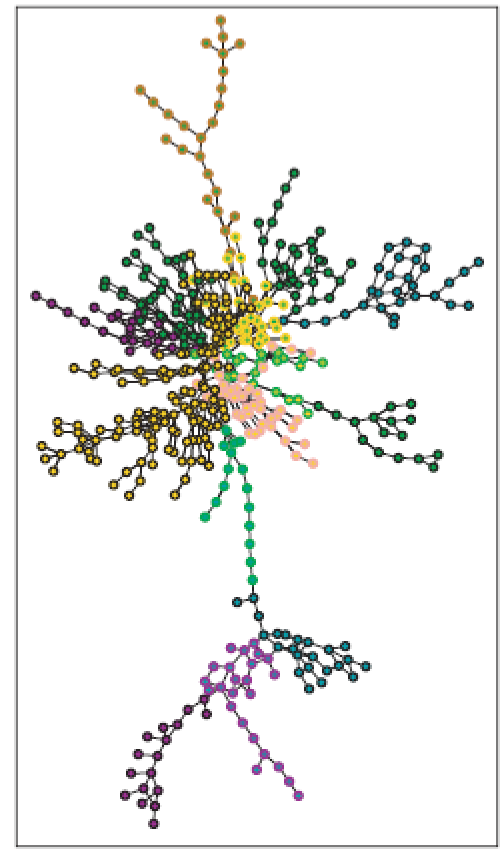
H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, and A.L. Barabasi, Nature, **407** 651 (2000)

# 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.



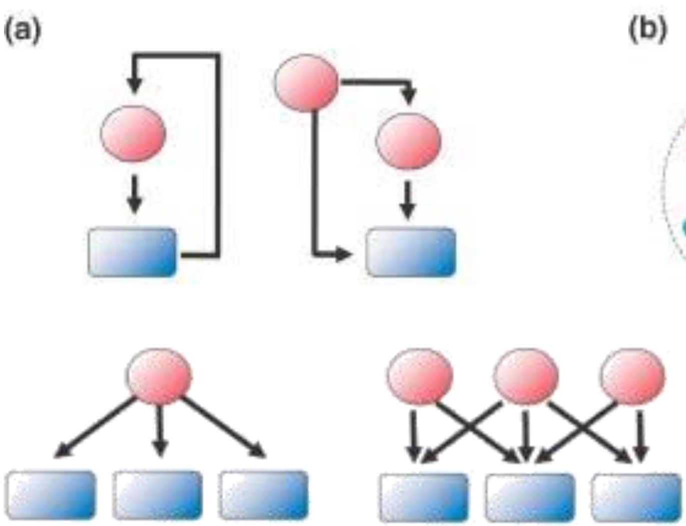
Metabolic network of *E. coli* (N=473, L =674).

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

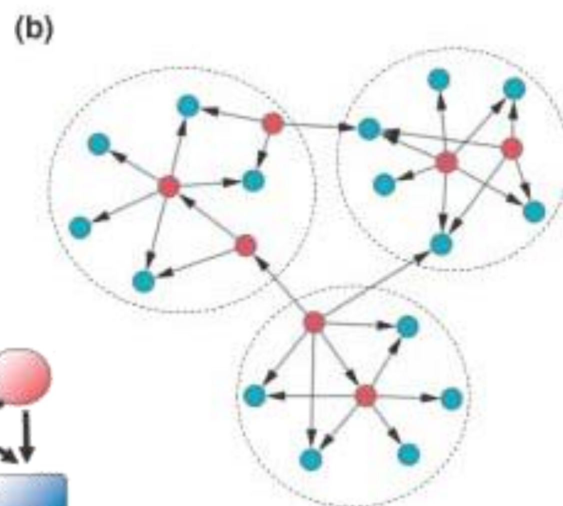
# Modules: A *mesoscopic* organizational principle of networks

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

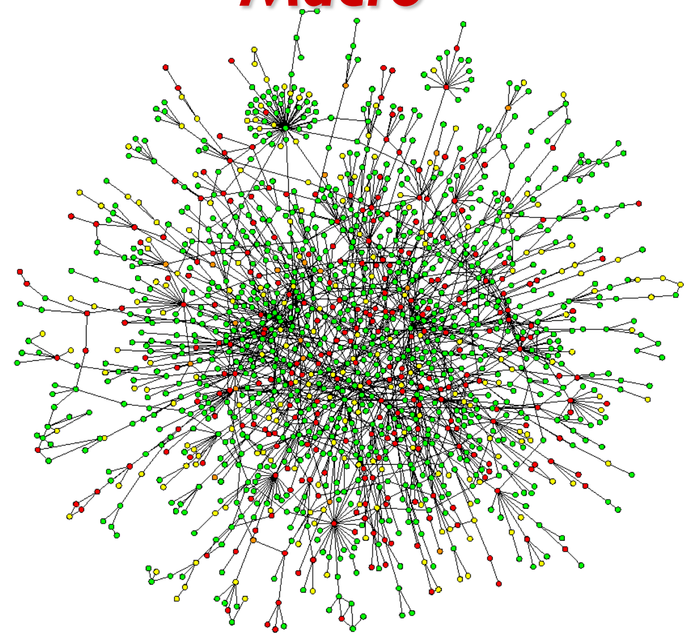
***Micro***




***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

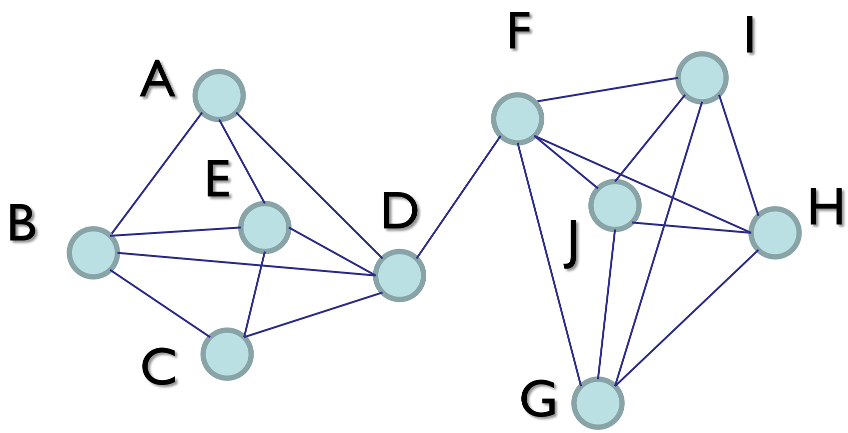
Problem:

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 link 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 $\mathbf{s} = \{ s_1 s_2 \ldots s_N \}$

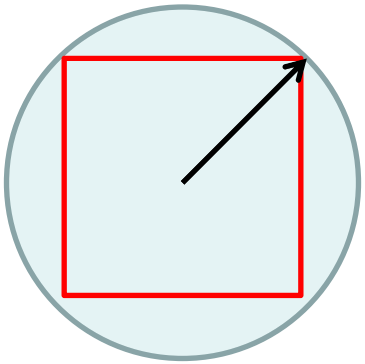
Goal of Partitioning: To find **s** that minimizes **R** given **L**

# 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

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

where $\lambda$, $\mu$ are Lagrange multipliers for enforcing the constraints

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

# 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 $\lambda\,[N_A - N_B] + \mu\,N = 0 \;\Rightarrow\; \mu/\lambda = -[N_A - N_B]/N$

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 $\lambda$

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:

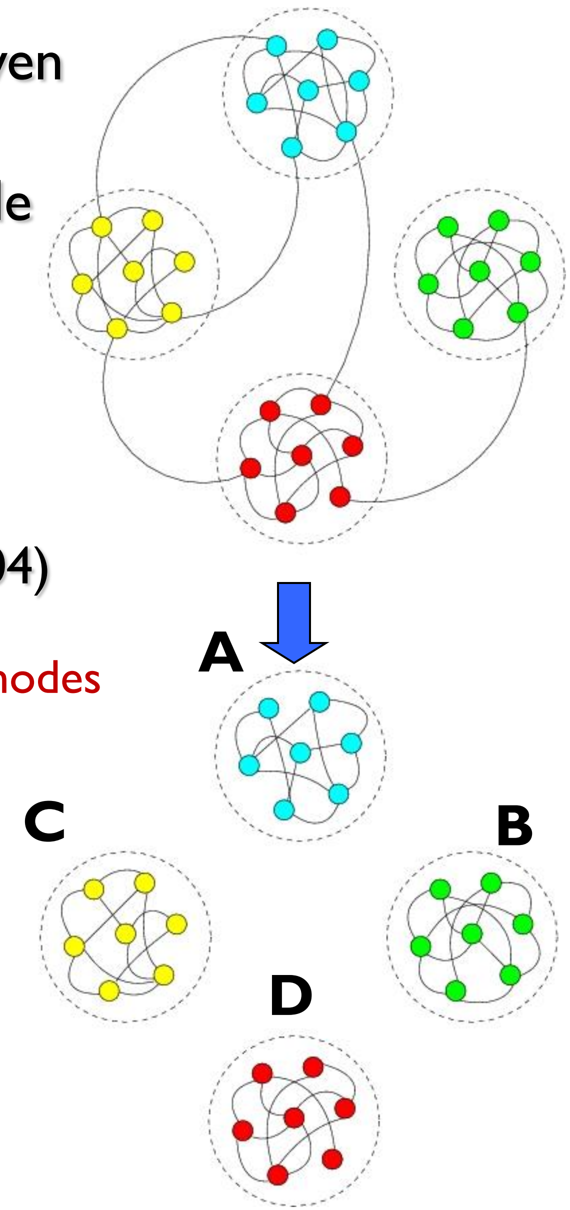
$$Q \equiv \frac{1}{2L}\sum_{i,j}\left[A_{ij} - \frac{k_i k_j}{2L}\right]\delta_{c_i c_j}$$

=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

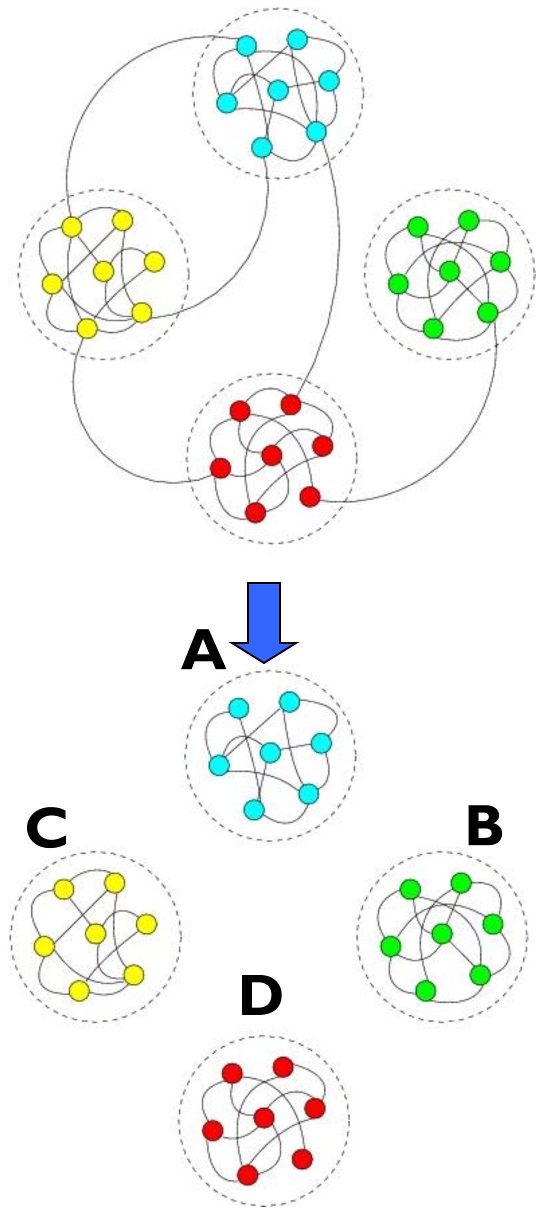
## For directed & weighted networks:

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

**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

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

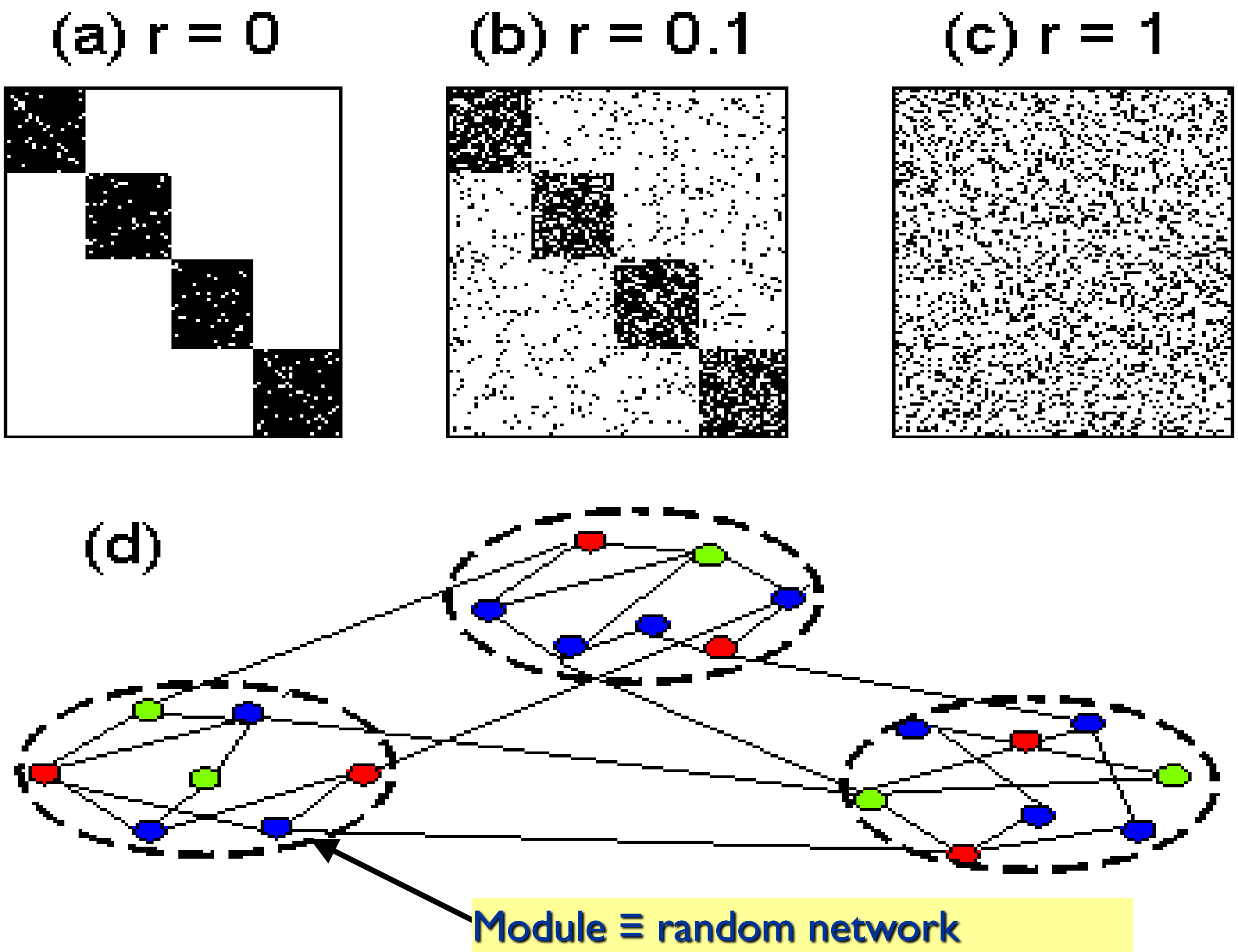
and then assign communities based on the signs of the elements of the eigenvector.

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

# A simple model of modular networks

Model parameter *r* :

Ratio of inter- to intra-modular connection density



# Comparison with Watts-Strogatz model

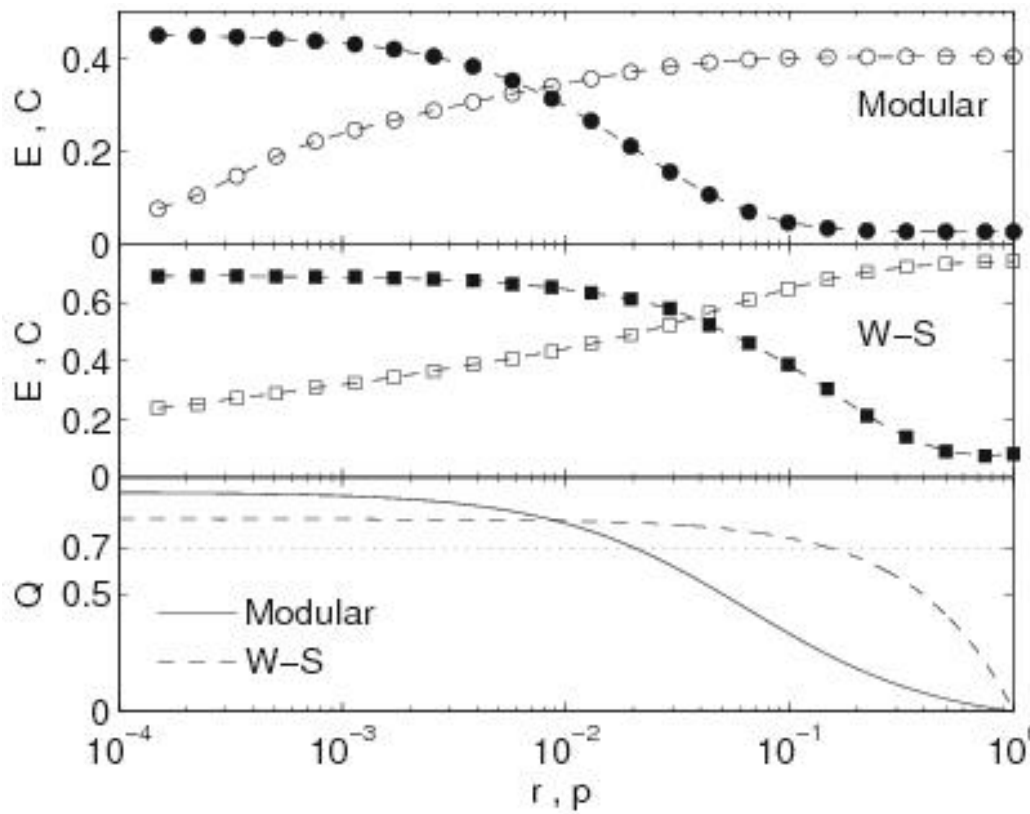
*Structural measures used:*

Communication efficiency $E = [\text{avg path length}, \ell\,]^{-1} = 2/N(N-1) \sum_{i>j} d_{ij}$

Clustering coefficient $C$ = fraction of observed to potential triads $= (1/N) \sum_i 2n_i / k_i (k_i - 1)$

WS and Modular networks behave similarly as function of *p* or *r* (Also for between-ness centrality, edge clustering, etc)

In fact, for same N and <k>, we can find *p* and *r* such that the WS and Modular networks have the same "*modularity*" Q




Pan and Sinha, EPL (2009)

# How can you tell them apart ?

Pan and Sinha, EPL (2009)

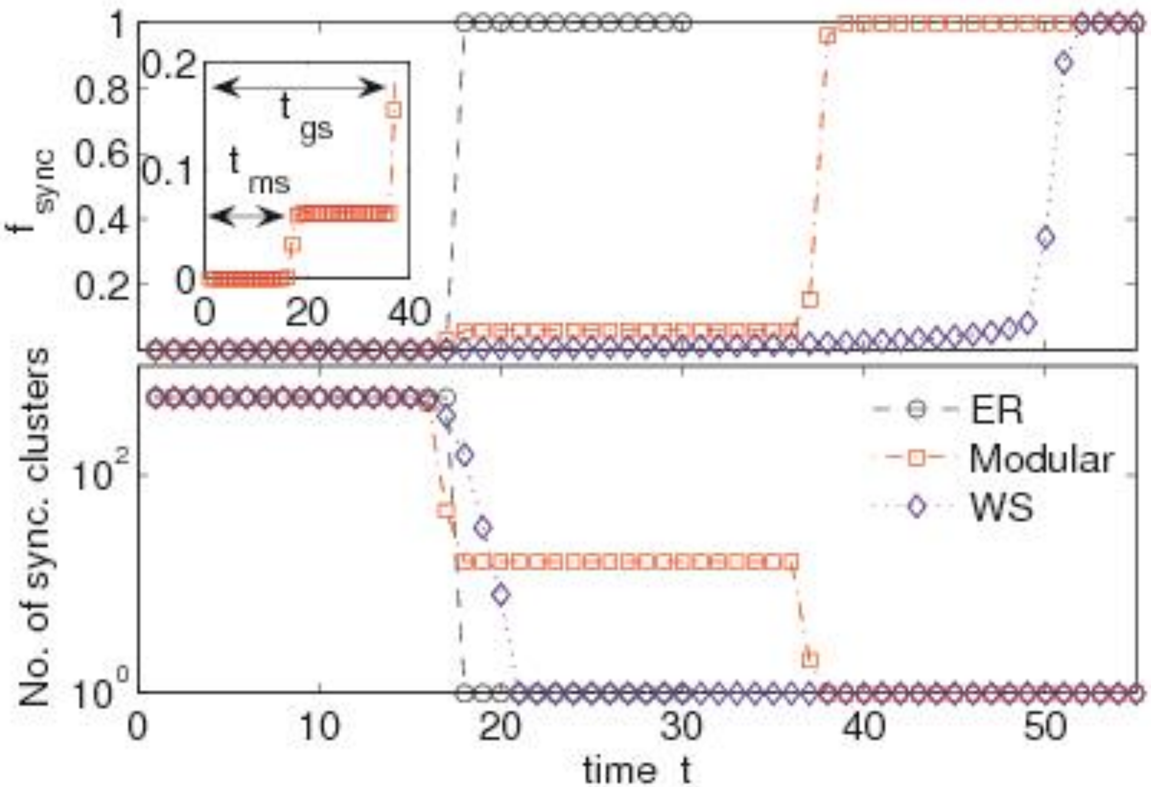
## Dynamics on Modular networks different from that on Watts-Strogatz small-world networks

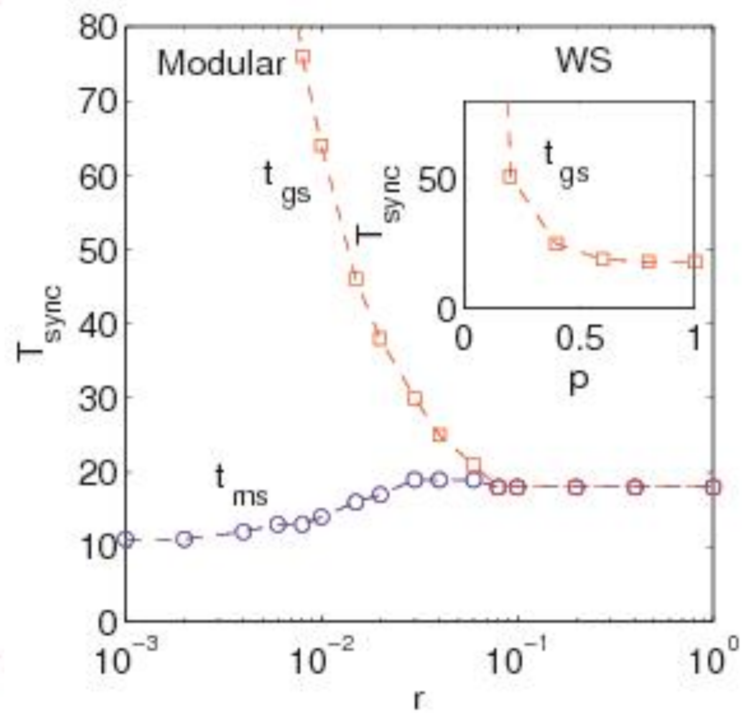
Consider synchronization on modular networks

e.g., phase oscillators: $d\theta_i /dt = w + (1/k_i)\sum K_{ij} \sin(\theta_j - \theta_i)$

*Network topology*

**2 distinct time scales in Modular networks: $t_{modular}$ & $t_{global}$**



# Existence of distinct time-scales in Modular networks

Pan and Sinha, EPL (2009)

Consider linearized dynamics around synchronized state

$d\theta_i /dt = - (\kappa /k_i)\sum_j L_{ij}\, \theta_j$ ,    $(i = 1,\ldots,N)$

L: Laplacian

$\kappa$: coupling strength of oscillators

Focus on the normal modes:

$\phi_i\,(t) = \sum_j B_{ij}\, \theta_j = \phi_i\,(0)\, \exp(-\lambda_i\, t)$,    $(i = 1,\ldots,N)$

B: matrix of eigenvectors
$\lambda_i$ :eigenvalues
} of $L' = D^{-1} L$,
D: diagonal matrix s.t. $D_{ii}=k_i$

$L' \rightarrow L=D^{1/2} L' D^{-1/2}$ is symmetric, normalized Laplacian $\Rightarrow \lambda_i$ real

**Differences in time-scales of modes $\Rightarrow$ gap in spectrum of L**

Mode for smallest $\lambda_i$ : associated with global synchronization

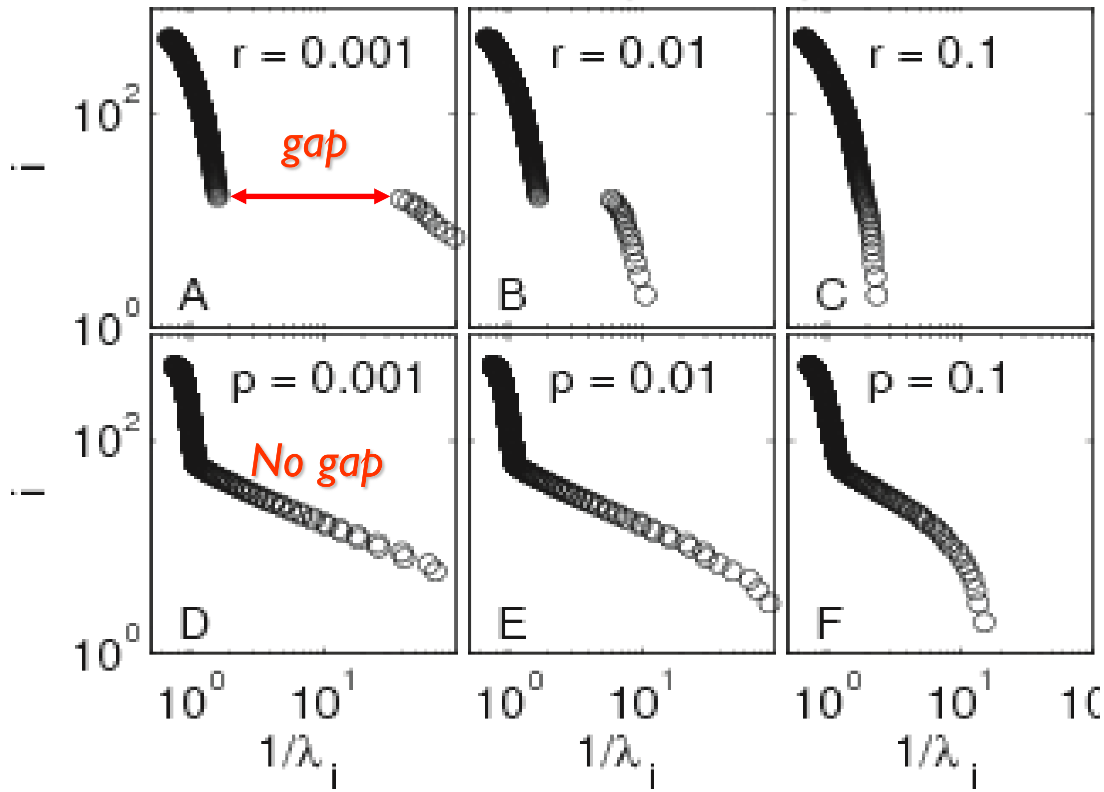
Other modes : synchronization within different groups of oscillators

Pan and Sinha, EPL (2009)

# Eigenvalue spectra of the Laplacian

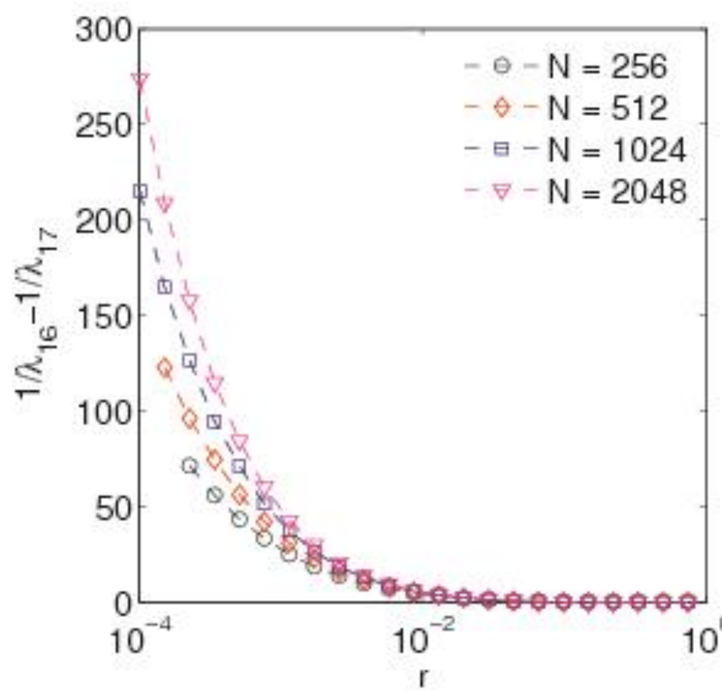
Shows the existence of spectral gap $\Rightarrow$ distinct time scales

*Modular network Laplacian spectra*




*WS network Laplacian spectra*

Spectral gap in modular networks diverges with decreasing r

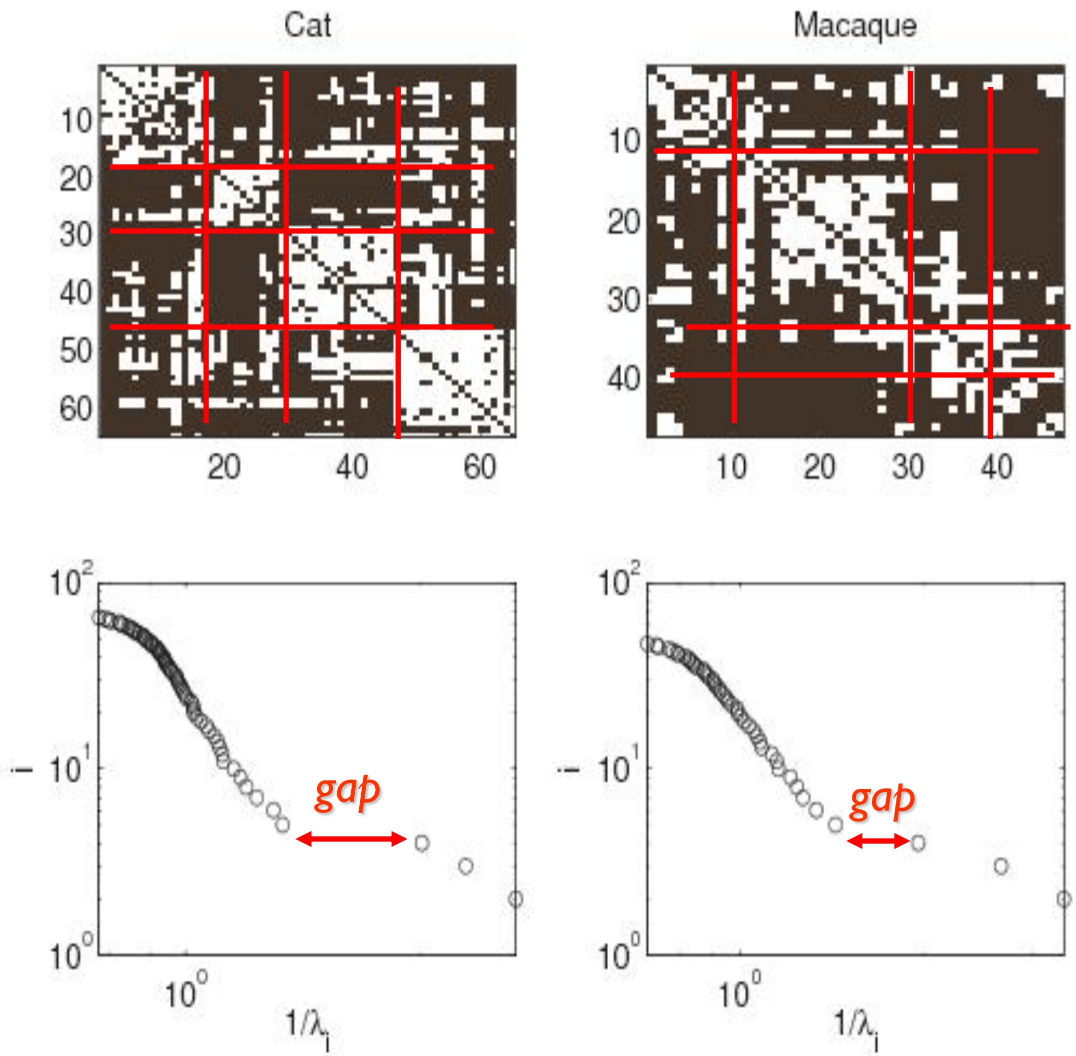



Existence of distinct time-scales in Modular networks

No such distinction in Watts-Strogatz small-world networks

# How about "real" SW networks ?

Pan and Sinha, EPL (2009)



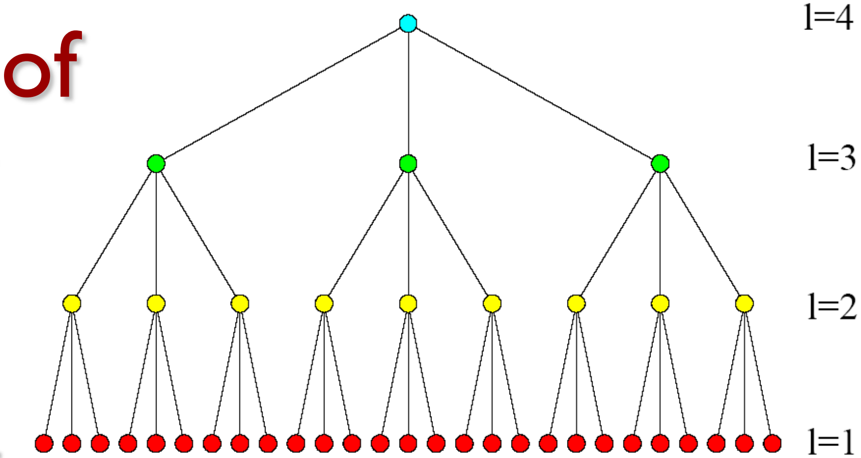

The networks of cortical connections in mammalian brain have been shown to have *small-world* structural properties

Our analysis reveals their dynamical properties to be consistent with modular "small-world" networks

Fast synchronization of neuronal activity within a module :
The mechanism for efficient neural information processing ?

# How about other kinds of mesoscopic structures ?

## E.g., Hierarchy

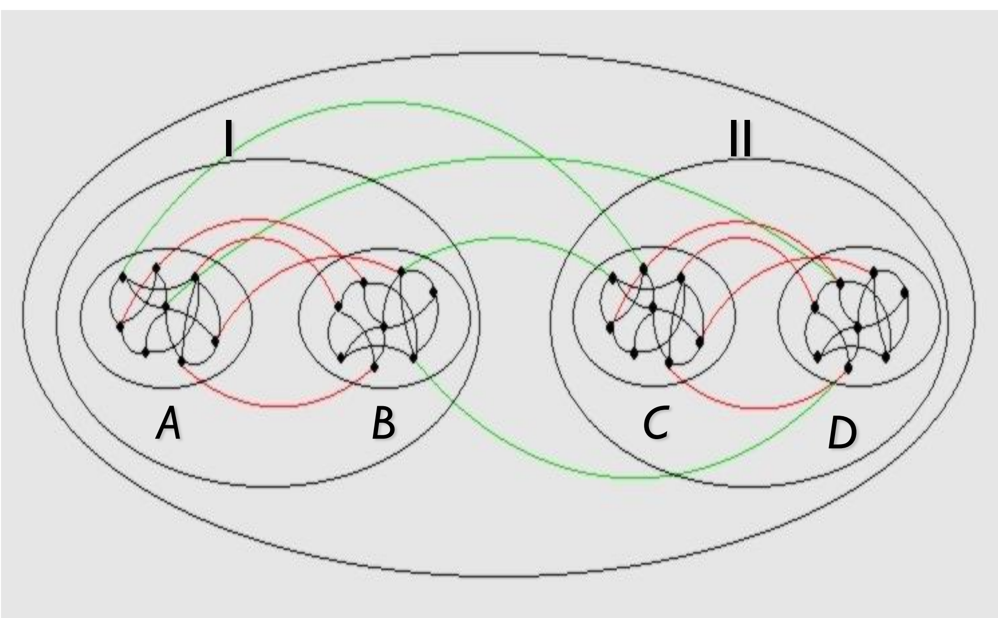



## Hierarchical Modular networks

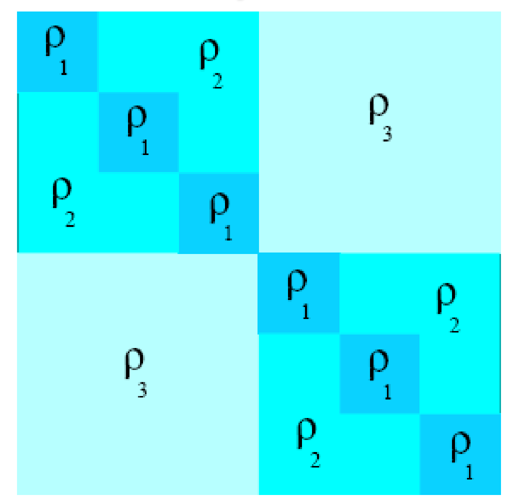
Modules may occur at different levels of hierarchy

*Level 1*: Modules *A, B, C, D*

*Level 2*: Meta-Modules I, II




- $r$ = 1 : randomly coupled network.
- $r$ = 0: isolated sub-networks (modules)
- 0 < $r$ < 1 : hierarchically structured network.

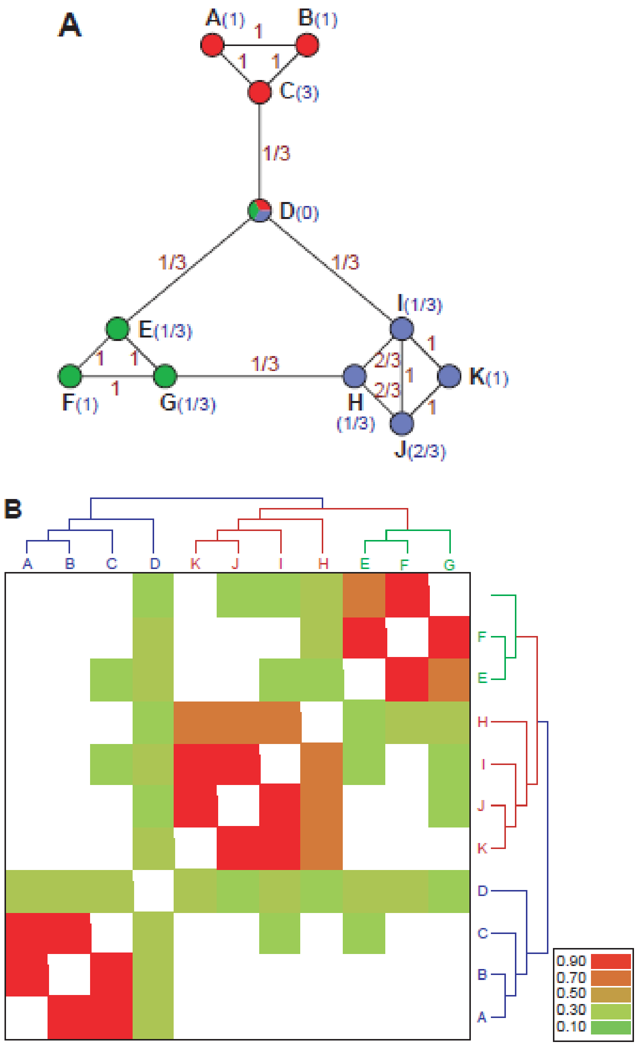


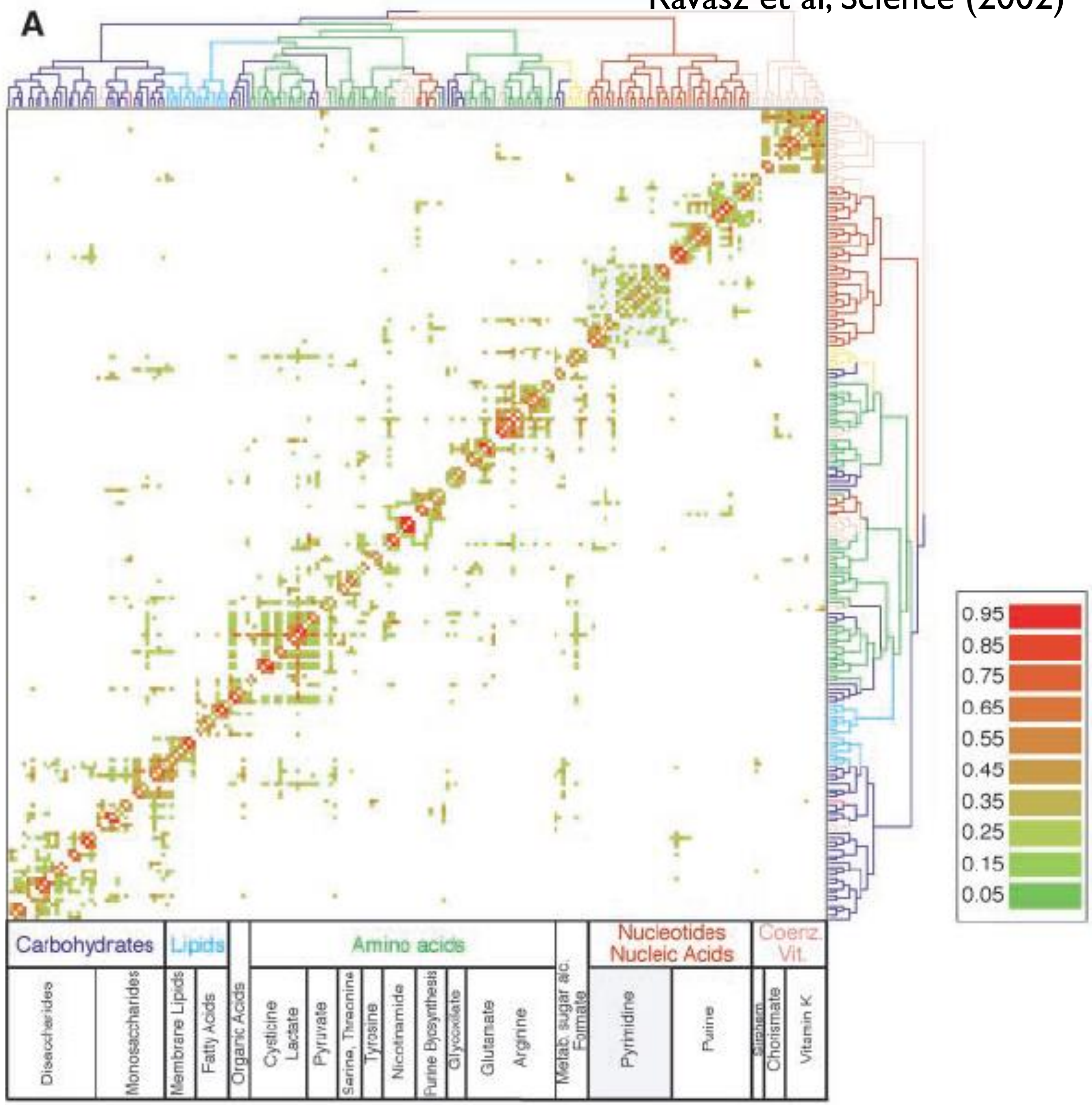
# Hierarchical modularity in metabolic network

Topological overlap:

Ravasz et al, Science (2002)

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



# Hierarchical Modular Networks exhibit several distinct time-scales – equal to the number of hierarchical levels (Sinha & Poria, 2011)

Synchronization of phase oscillators in hierarchical modular network show as many distinct time-scales as number of hierarchical levels …
Reflected in the eigenvalue spectra