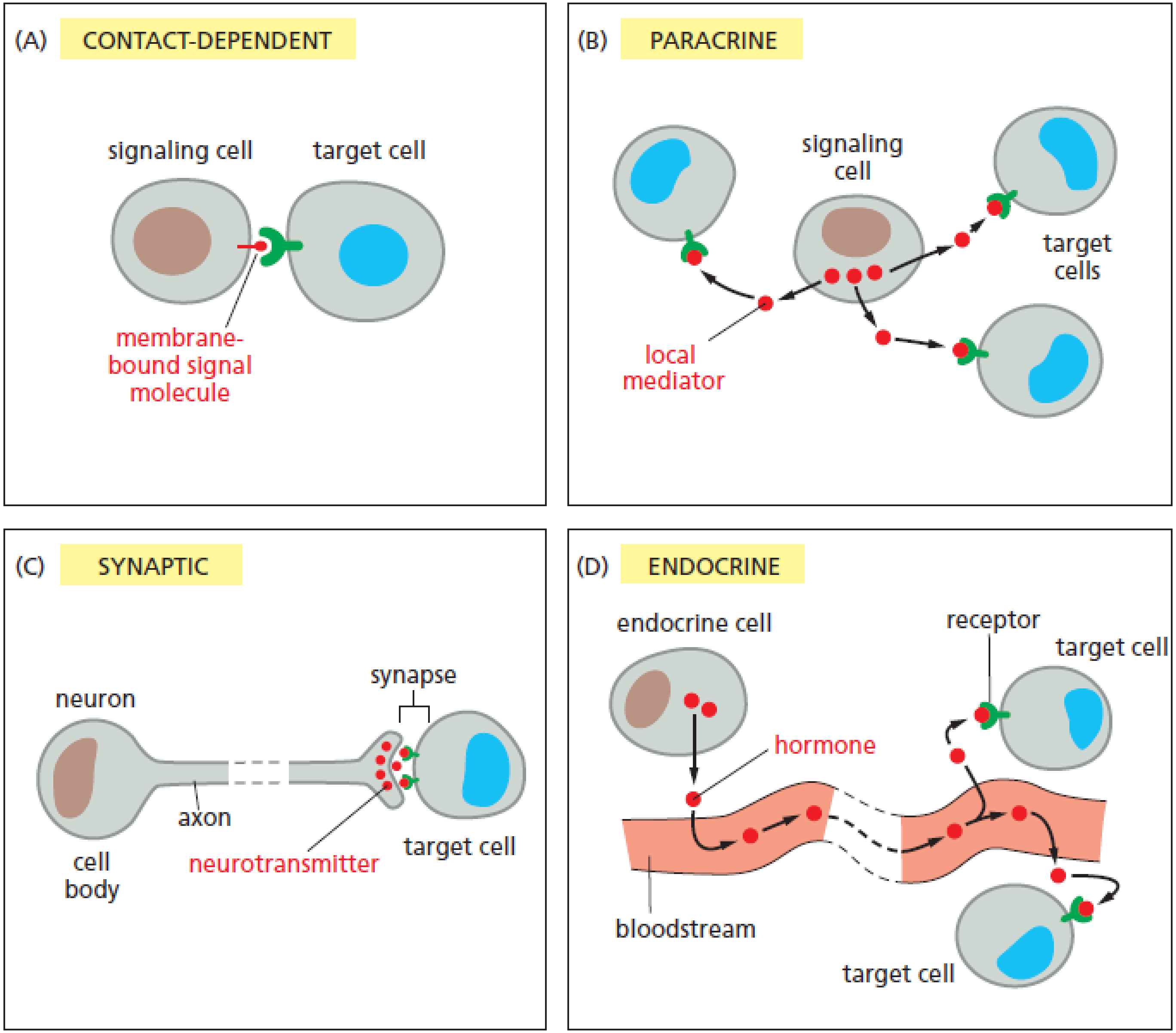
Systems Biology: A Personal View XIII. Modularity and Inter-cellular networks

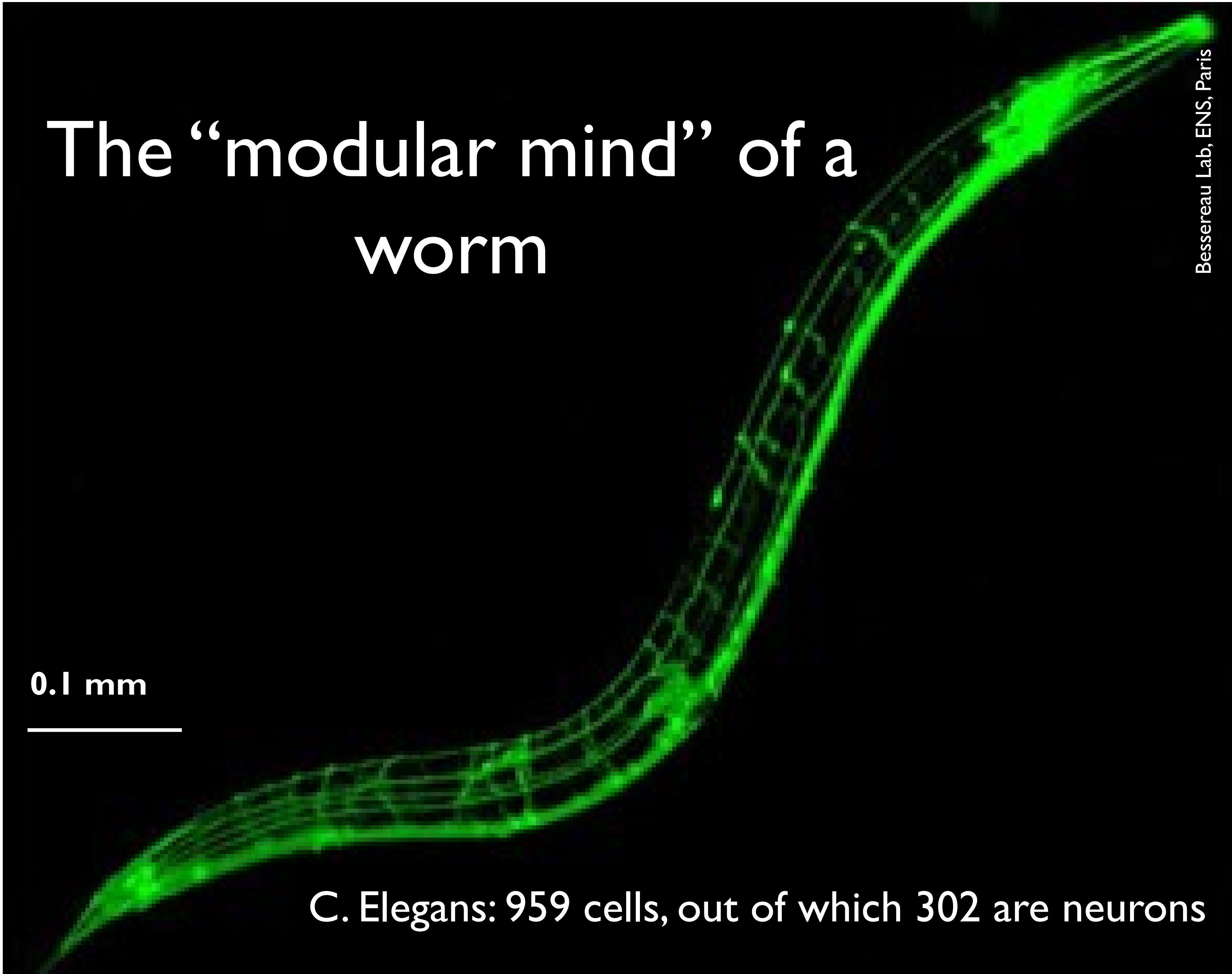
Sitabhra Sinha
IMSc Chennai

Nature of links in inter-cellular networks



Alberts et al. Molecular Biology of the Cell, 6th ed

The “modular mind” of a worm

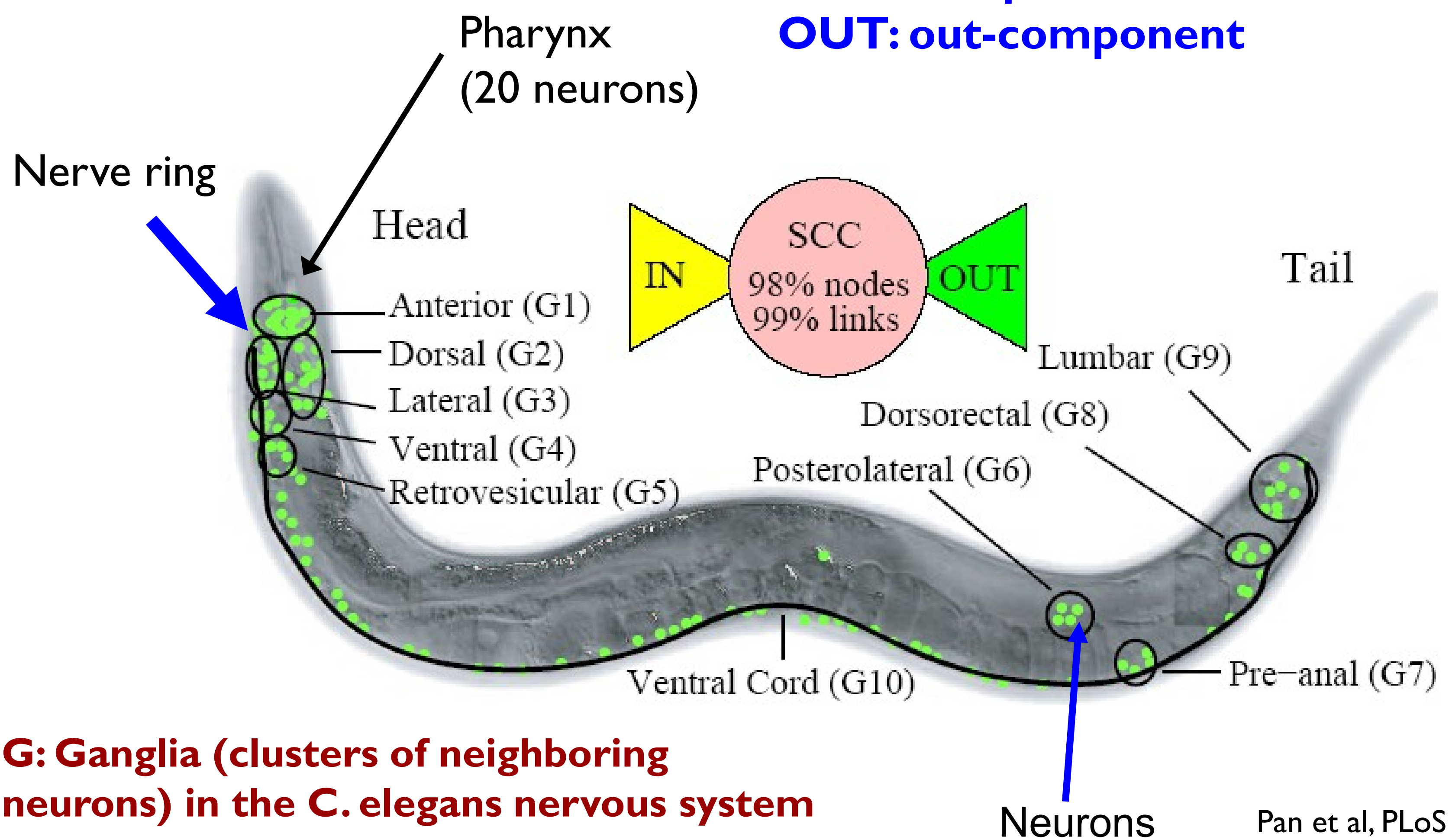


C. Elegans: 959 cells, out of which 302 are neurons

Structure of the nervous system

We concentrate on the 282 non-pharyngeal neurons

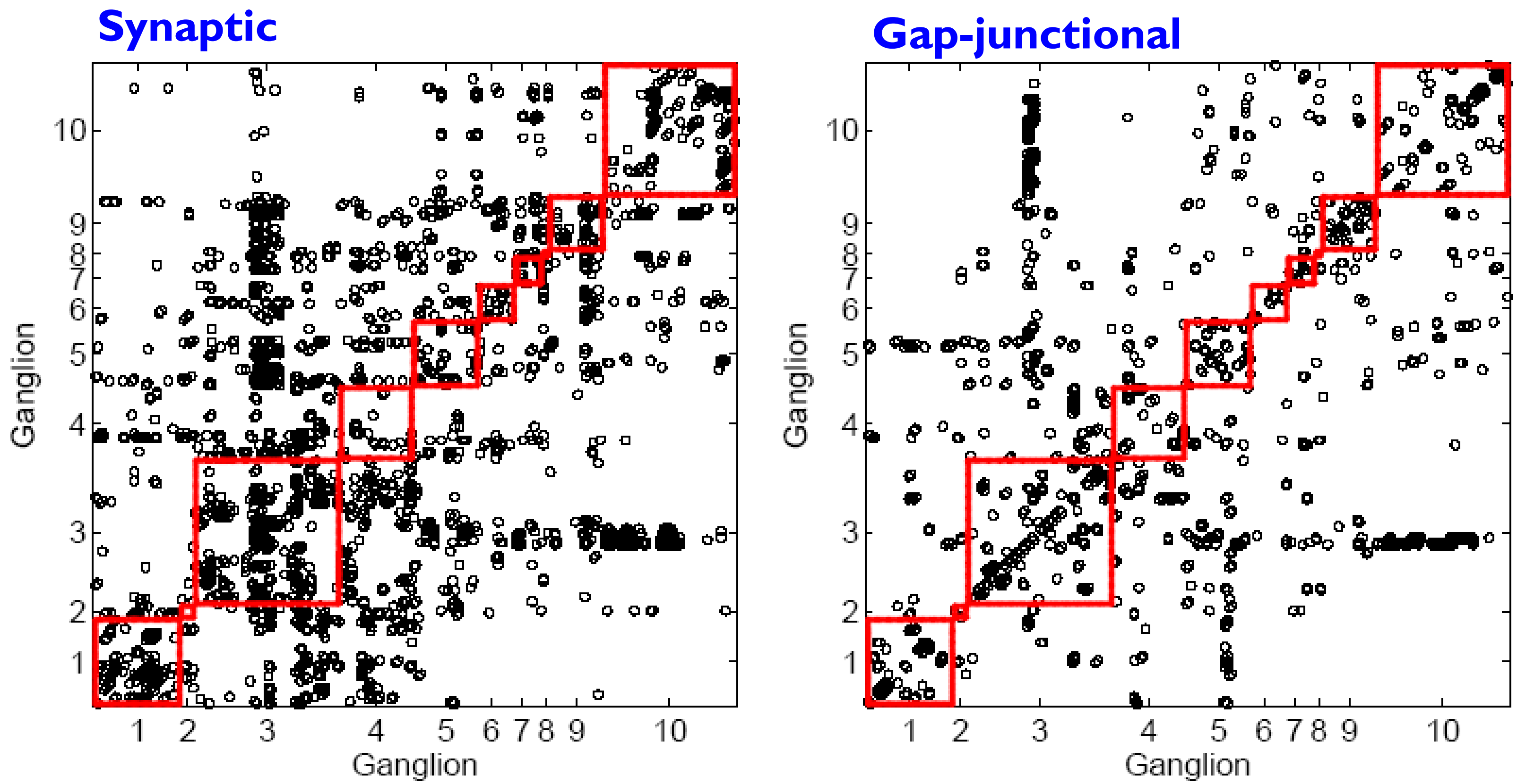
SCC: strongly connected component
IN: in-component
OUT: out-component



G: Ganglia (clusters of neighboring neurons) in the C. elegans nervous system

Pan et al, PLoS ONE (2010)

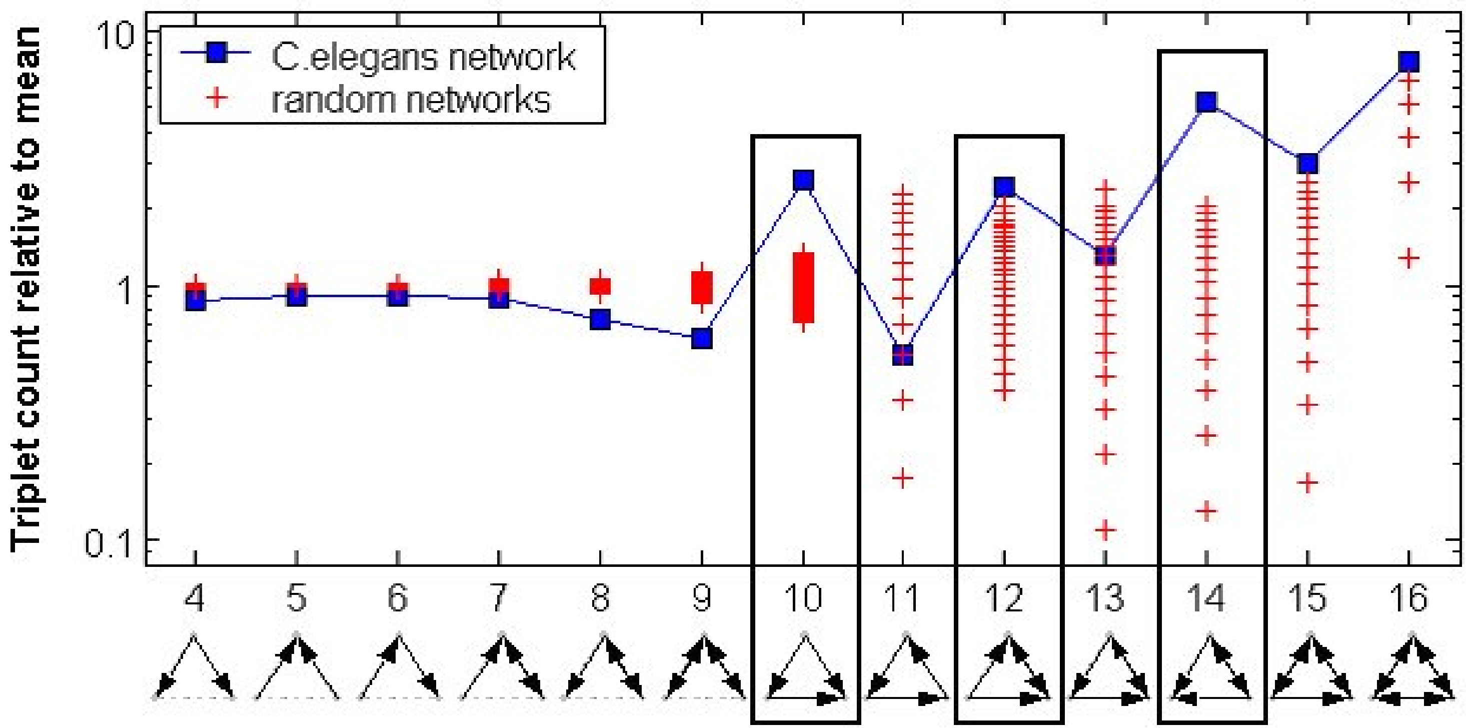
Connectivity of the somatic nervous system



Pan et al, PLoS ONE (2010)

Triplet motifs in *C elegans* neural network

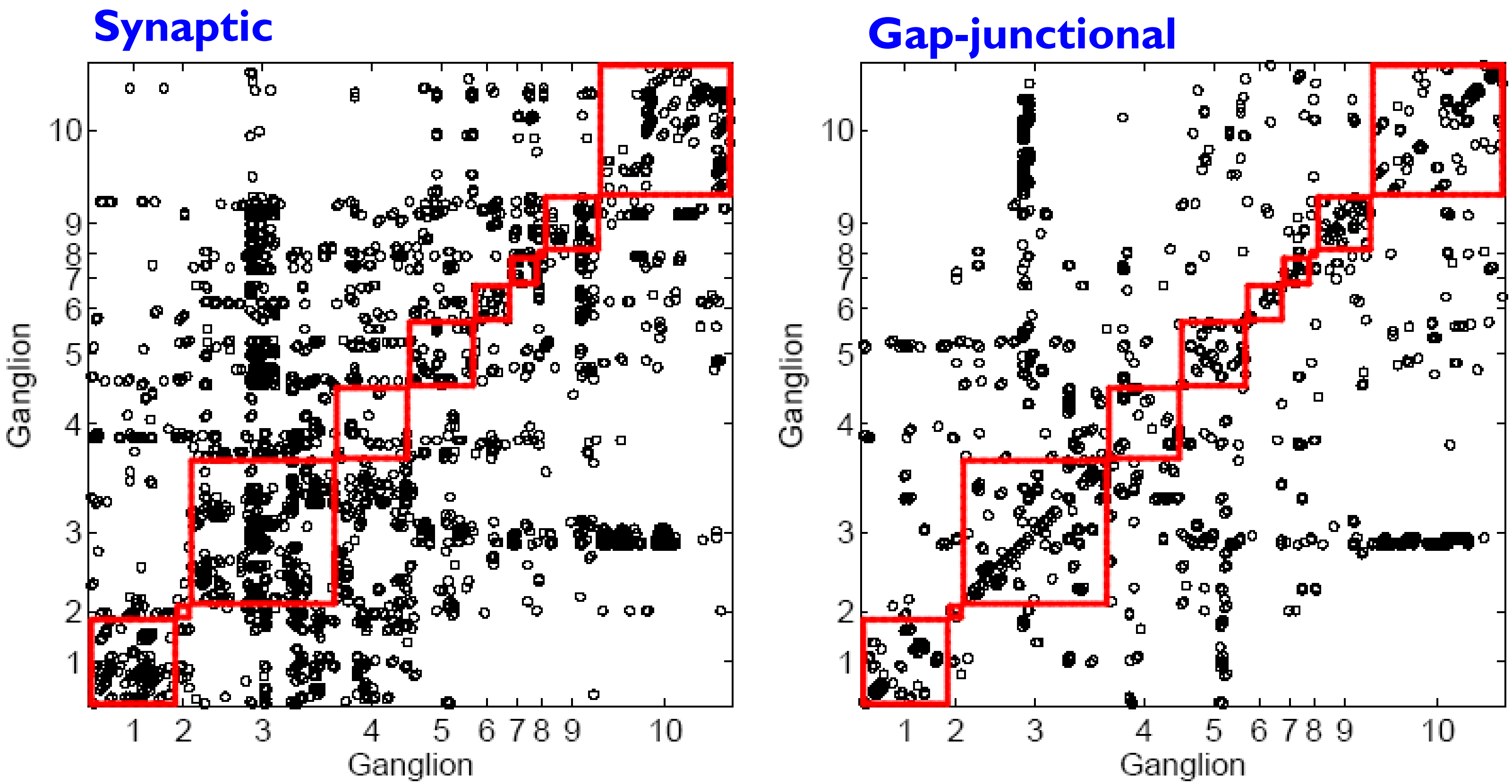
"one of the most consistently over-represented motifs is the feedforward loop"



Reigl *et al. BMC Biology* (2004)

- "The three-layered feedforward neuronal network is not sufficient to account for over-representation of the feedforward loop
- The likelihood of connectivity between nearby neurons may partially account for over-representation of the feedforward loop"

Connectivity of the somatic nervous system



Pan et al, PLoS ONE (2010)

Question:

Is the network modular ? How do you determine the modules if the connections are not localized within corresponding ganglia ?

Measuring modularity: explicit algorithm

First define a modularity matrix B, $$B_{ij} = W_{ij} - \frac{s_i^{\text{in}} s_j^{\text{out}}}{L^W}$$

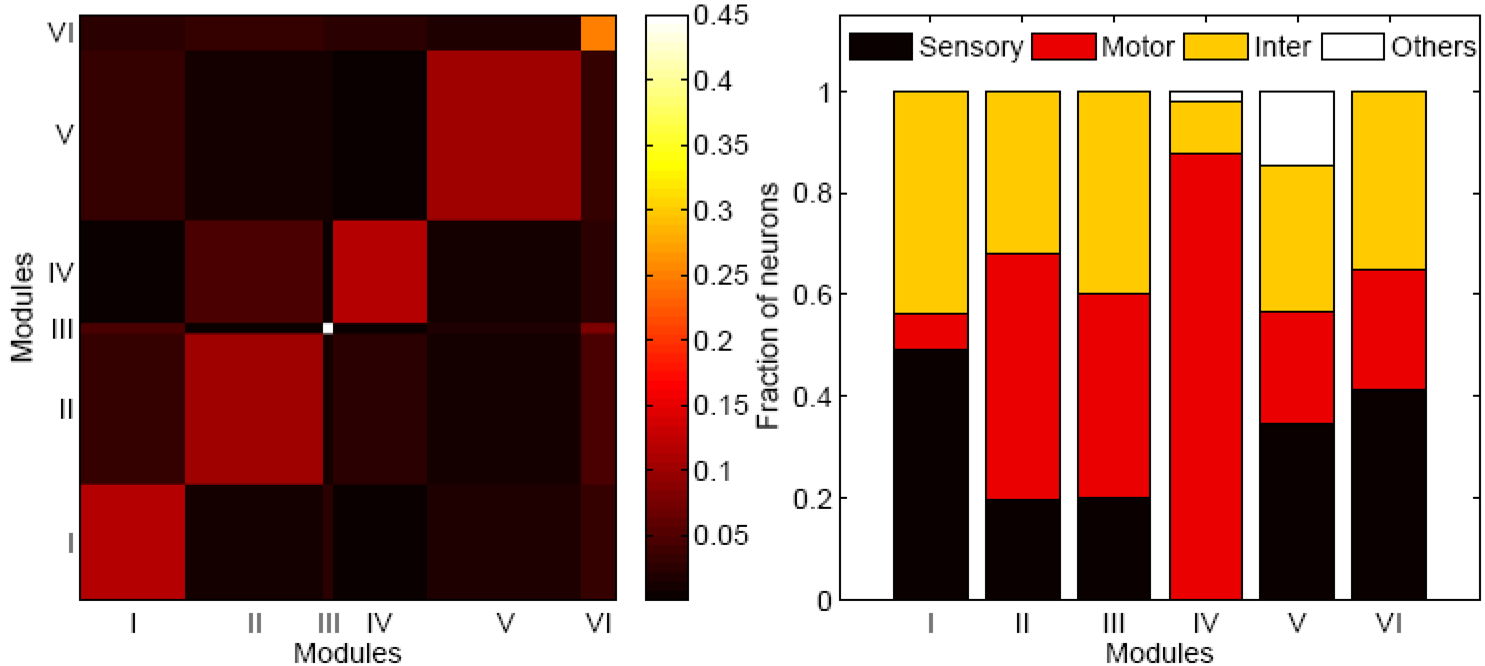
To split the network into modules,

- ❑ the eigenvector corresponding to the largest positive eigenvalue of the symmetric matrix ($B + B^T$) is calculated
- ❑ the communities are assigned based on the sign of the elements of the eigenvector.
- ❑ This divides the network into two parts, which is refined further by exchanging the module membership of each node in turn if it results in an increase in the modularity.
- ❑ The process is then repeated by splitting each of the two divisions into further subdivisions.
- ❑ This recursive bisection of the network is carried out until no further increase of Q is possible.

The Modular Structure of the Network

Optimal decomposition of the somatic nervous system into 6 modules

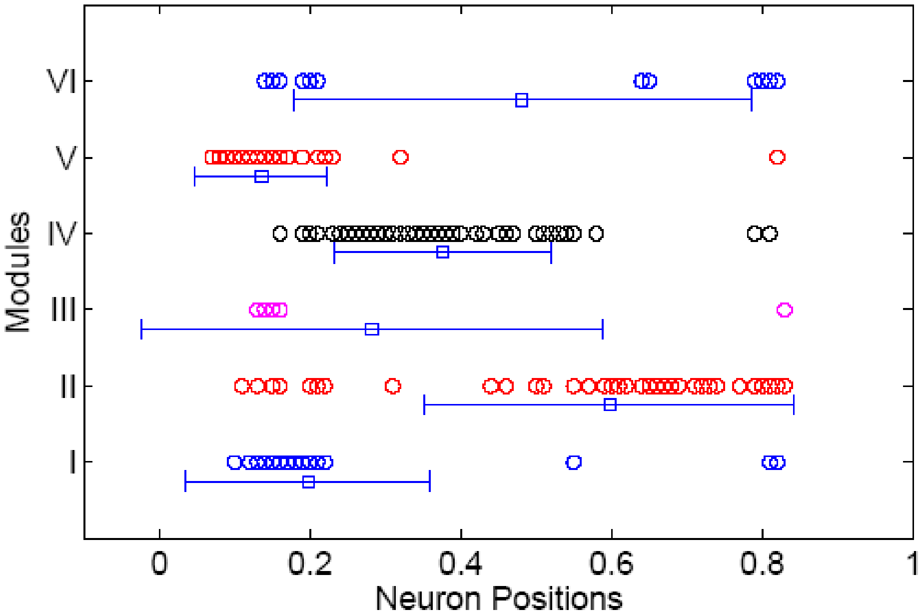
Pan et al, PLoS ONE (2010)



- Dense interconnectivity within neurons in a module, relative to connections between neurons in different modules
- The modules are not simply composed of one type of neurons (e.g., a purely sensory neuron or motor neuron or interneuron module does not exist)

Modules and Spatial Localization

Pan et al, PLoS ONE (2010)

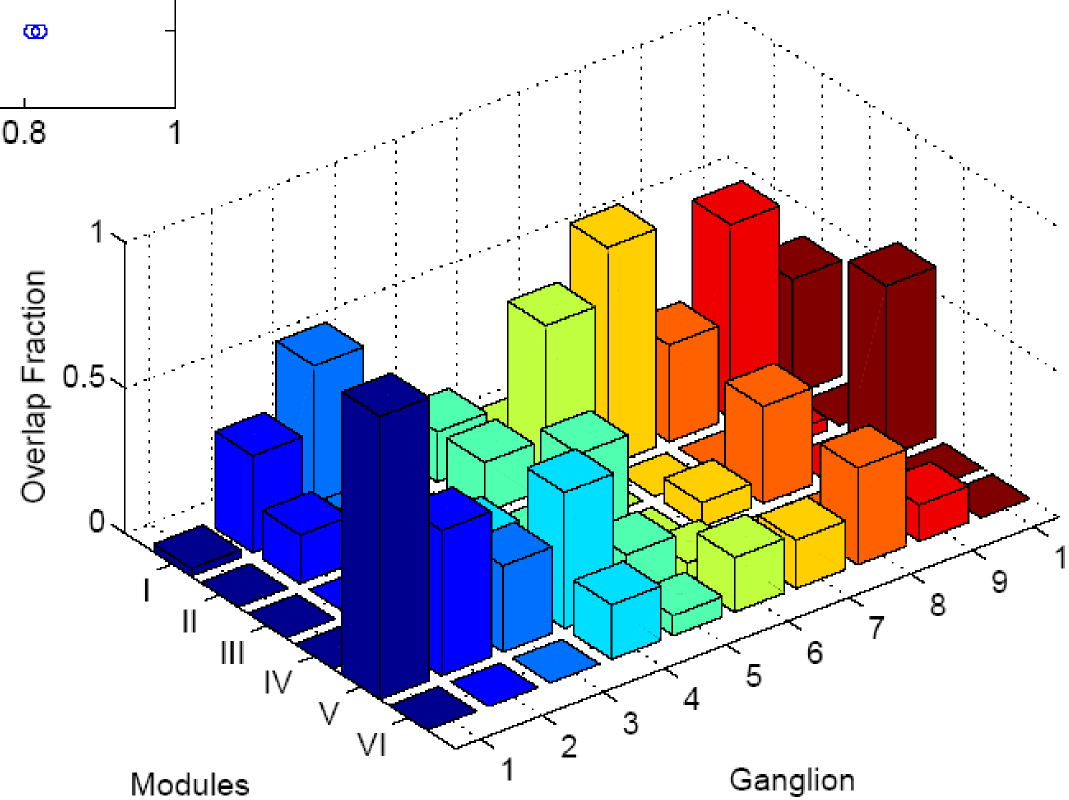


Q. Do constraints related to physical adjacency of neurons (e.g., minimization of wiring length) completely explain the modular organization ?

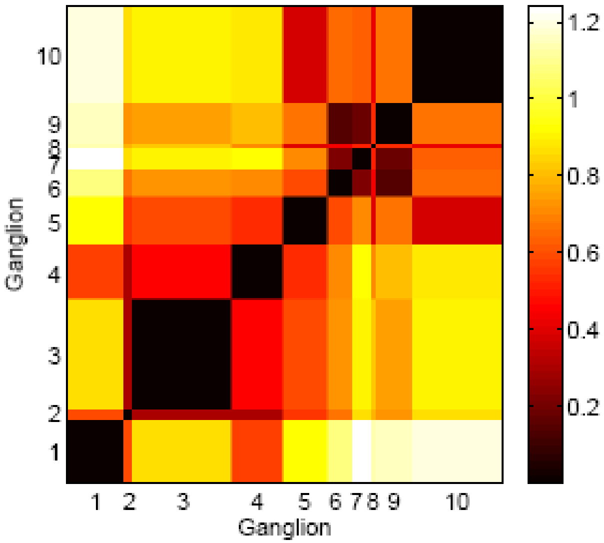
Ans. No

Q. How far does the existence of ganglia explain the modules ?

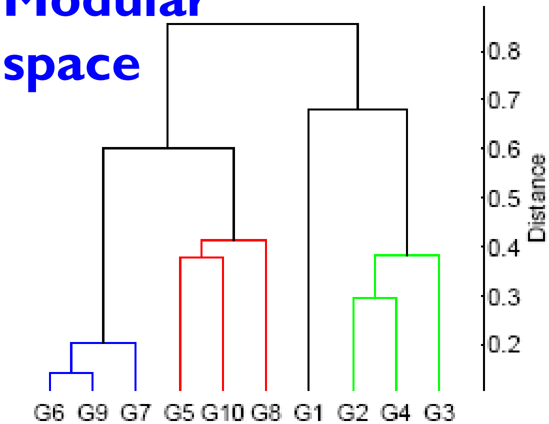
Ans. The overlap between modules and ganglia indicates that most ganglia are composed of neurons belonging to many different modules



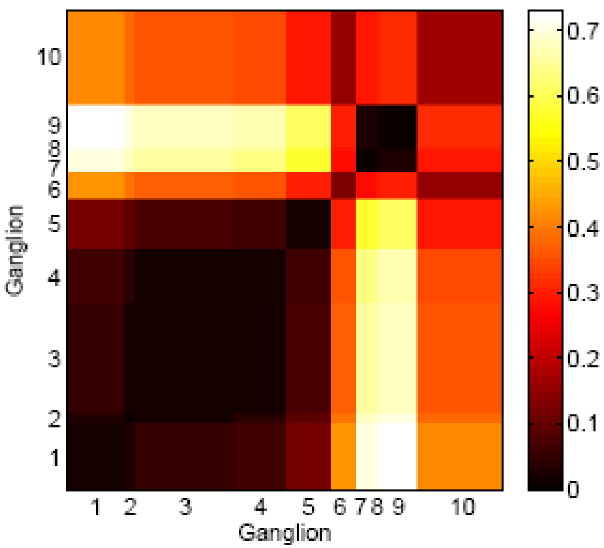
Pan et al, PLoS ONE (2010)



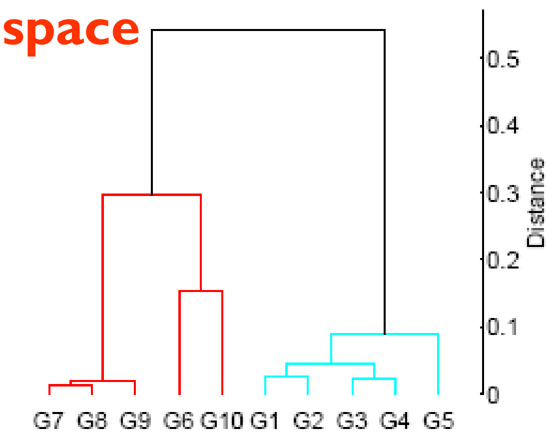
Modular space



We can define a *modular decomposition profile* for each ganglia : the distribution of the neurons in each ganglion into the 6 modules



Physical space



Two ganglia are *close* to each other, if they have similar profiles

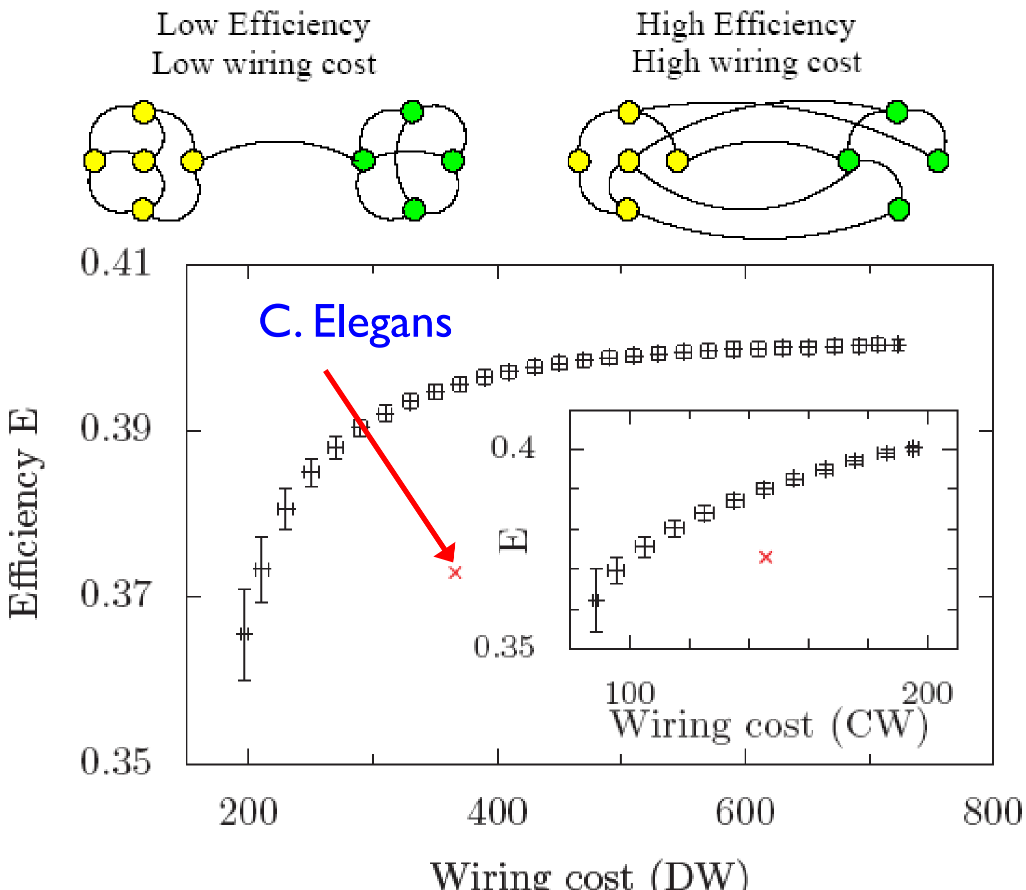
Inter-Ganglion distance in physical space and in the "modular" space show interesting differences !

Optimizing for wiring cost and communication efficiency

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

Wiring cost $DW = \sum_{i>j} d_{ij}$ for all connected neurons

("dedicated wire" model)



Trade-off between increasing communication efficiency and decreasing wiring cost

The network is sub-optimal !
$\Rightarrow$ presence of other constraints (possibly related to function) governing network organization

Pan et al, PLoS ONE (2010)

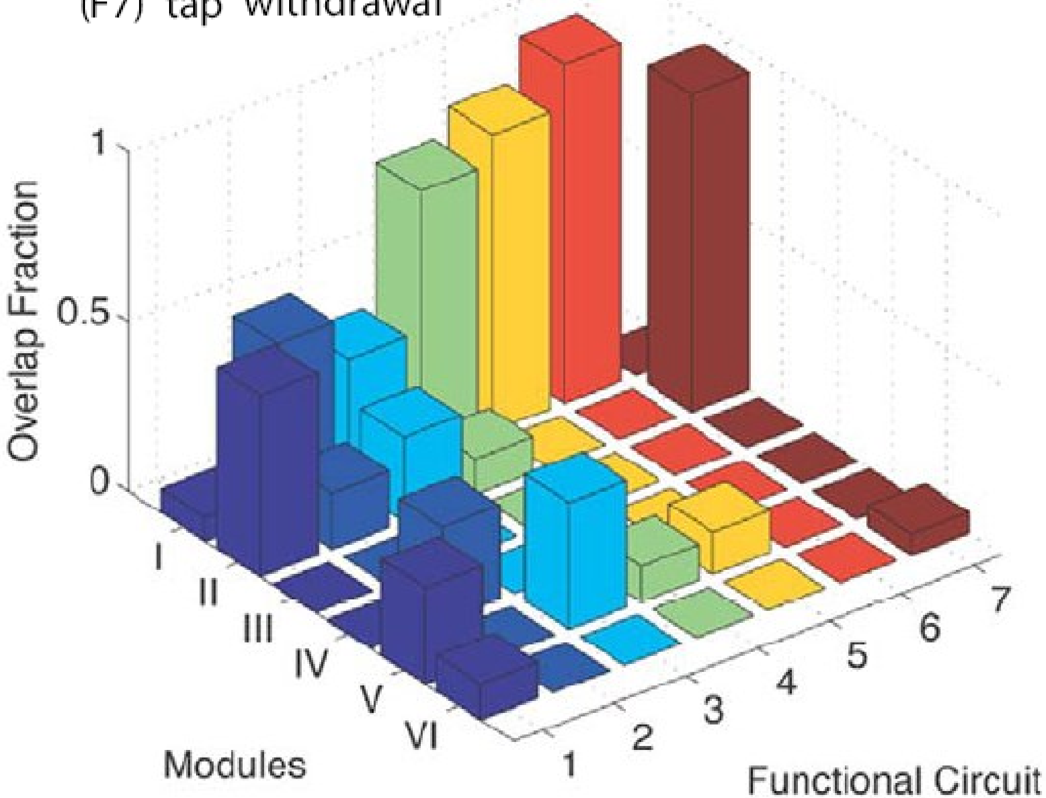
Modules and Functional Circuits

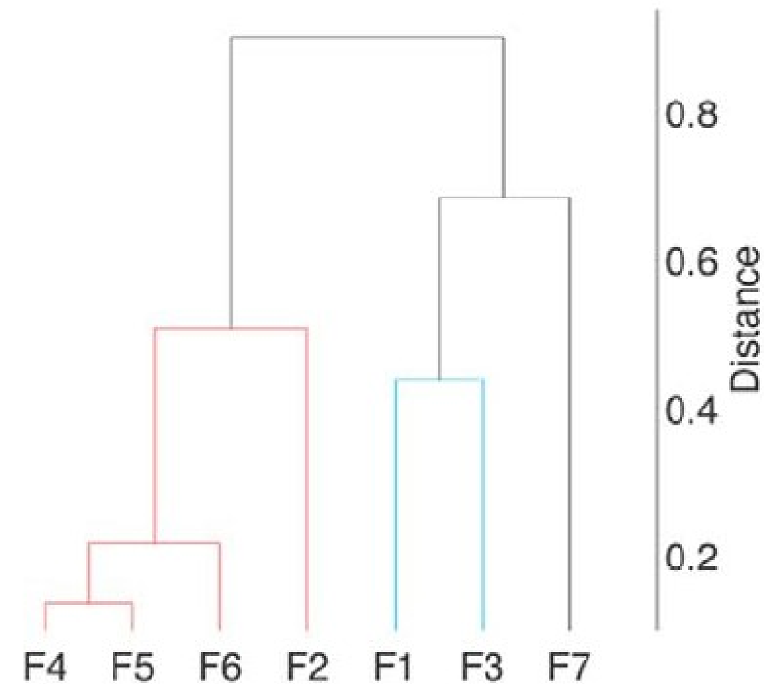
Pan et al, PLoS ONE (2010)

(F1) mechanosensation
(F2) egg laying
(F3) thermotaxis
(F4) chemosensation
(F5) feeding
(F6) exploration
(F7) tap withdrawal

Overlap between module & functional circuit measured by fraction of neurons common

Closeness among functional ckts in 6-D "modular" space
⇒F2 close to (F4,F5,F6) *Supported by exptl observation*: presence of food detected through chemosensory neurons modulates the egg-laying rate in C. elegans





Classification in terms of modular role

Guimera and Amaral, Nature (2005)

- Nodes can be classified in terms of functional roles according to their pattern of intra- and inter-module connections.

- Intra-modular connectivity defined in terms of within-module degree z-score:

$$z_i = \frac{k_i - \bar{k}_{si}}{\sigma_{si}}$$

k_i: number of links of node i to other nodes in its module s_i,
$\bar{k}_{si}$: average of k over all the nodes in s_i
σ_{si} : the standard deviation of k in s_i.

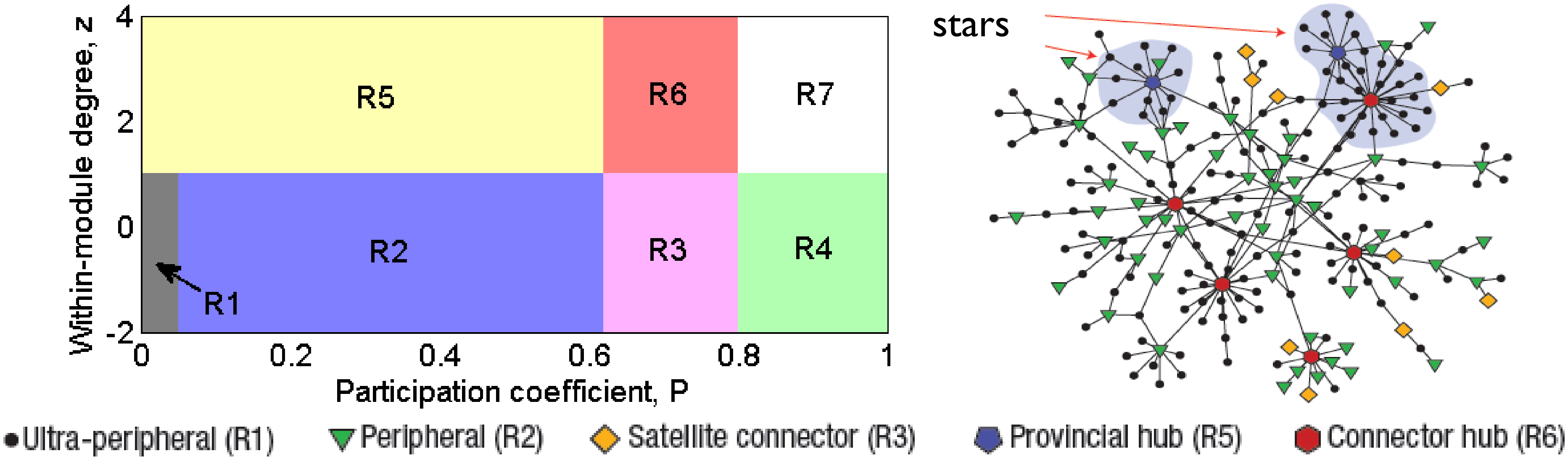
- Inter-modular connectivity defined in-terms of the participation coefficient P_i of node i:

k_{is}: number of links of node i to nodes in module s
k_i : total number of links of node i.

$$P_i = 1 - \sum_{s=1}^{N_M} \left(\frac{k_{is}}{k_i} \right)^2$$

P $\rightarrow$ 1 for a node if links are uniformly distributed among all modules
P $\rightarrow$ 0 if all its links are within its own module.

What do different regions in P-z space mean ?

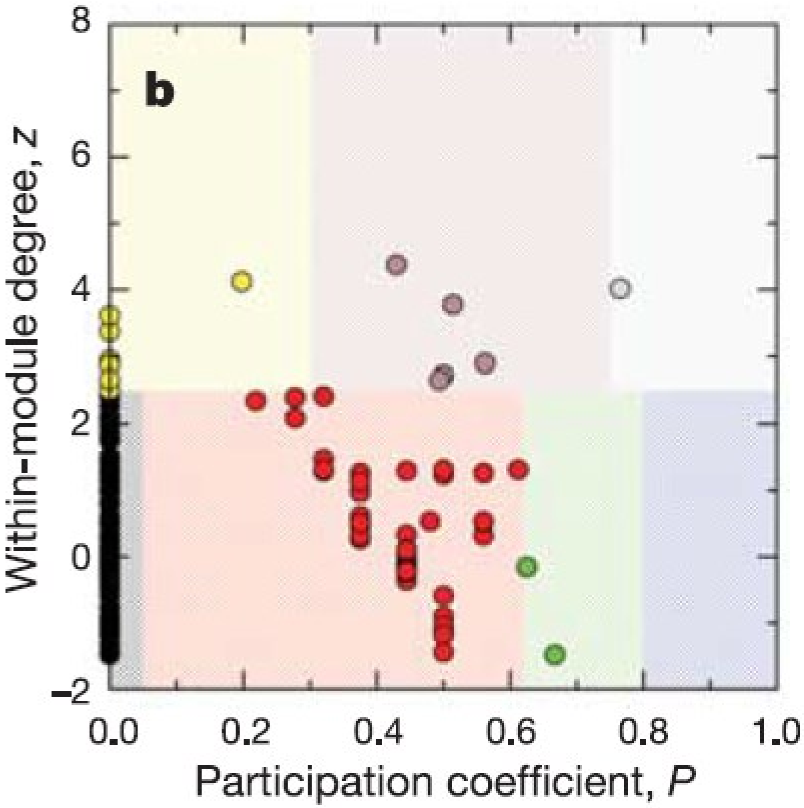


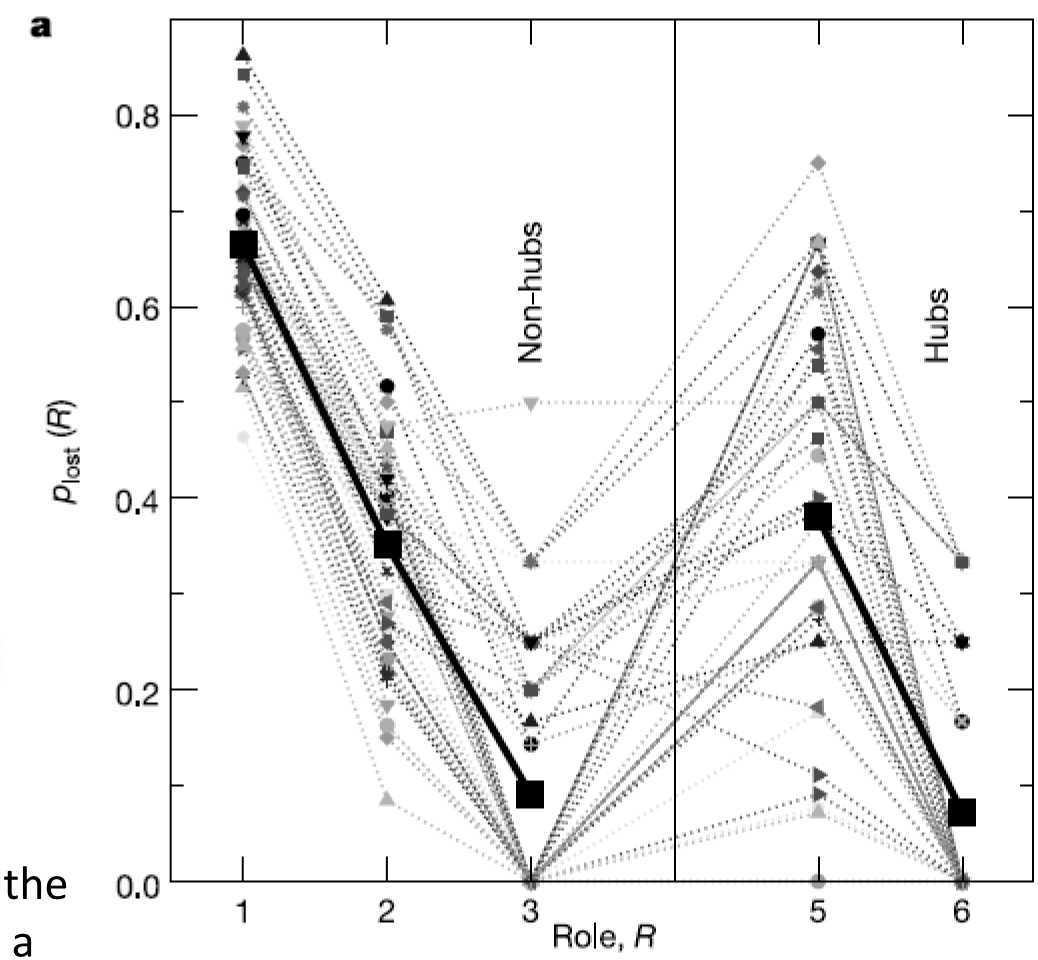
Seven different universal roles, each defined by a different region in the P-z parameter space. (Guimera and Amaral, 2005)

- ❑ The within-module degree z defines **hubs** (nodes with z ≥ 1) and **non-hubs** (z< 1).
- ❑ **Non-hub** nodes are divided into four different roles:
- **(R1) ultra-peripheral nodes**: *all* their links within their own module (P ≤ 0.05)
- **(R2) peripheral nodes**: *most* links within their module (0.05 <P ≤ 0.62)
- **(R3) non-hub connector nodes**: *many* links to other modules (0.62 < P ≤ 0.80)
- **(R4) non-hub kinless nodes**: links *homogeneously distributed* among all modules (P>0.80)
- ❑ **Hub** nodes are divided into three different roles:
- **(R5) provincial hubs**: *most* links within their own module (P ≤ 0.62)
- **(R6) connector hubs:** *many* links to most of the other modules (0.62< P ≤ 0.8)
- **(R7) global hubs**: links *homogeneously distributed* among all modules (P> 0.8)

Modular roles in *E coli* metabolic network

Guimera and Amaral, Nature (2005)





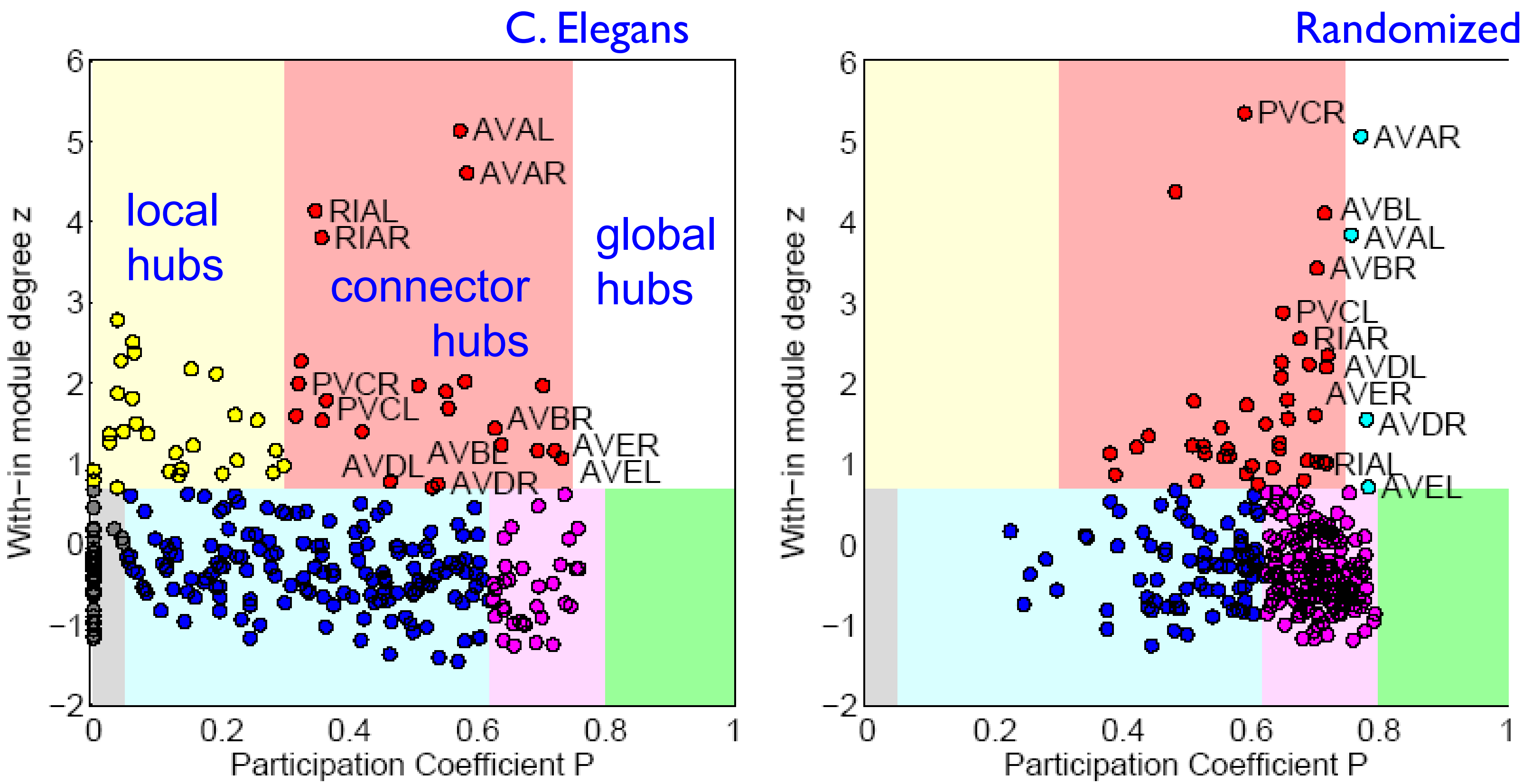
Modular role and conservation:

For a pair of species, A and B, loss rate is the probability p_{lost} (R)= $p(R_A = 0|R_B = R)$ that a metabolite is not present in one species (A) given that it plays role R in other species (B).

Structurally relevant modular roles have low values of p_{lost}(R)

How mesoscopic network structure can alert us to critical functional role of neurons

Pan et al, PLoS ONE (2010)



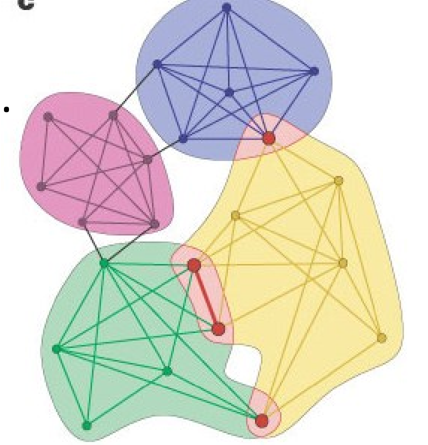
Importance of connector hubs: possibly integrating local activity to produce coherent response, 21 out of the 23 already implicated in critical functions

Prediction: AVKL and SMBVL are likely important for some as yet undetermined function

Palla et al, Nature (2005)

Clique Percolation
Detecting overlapping communities

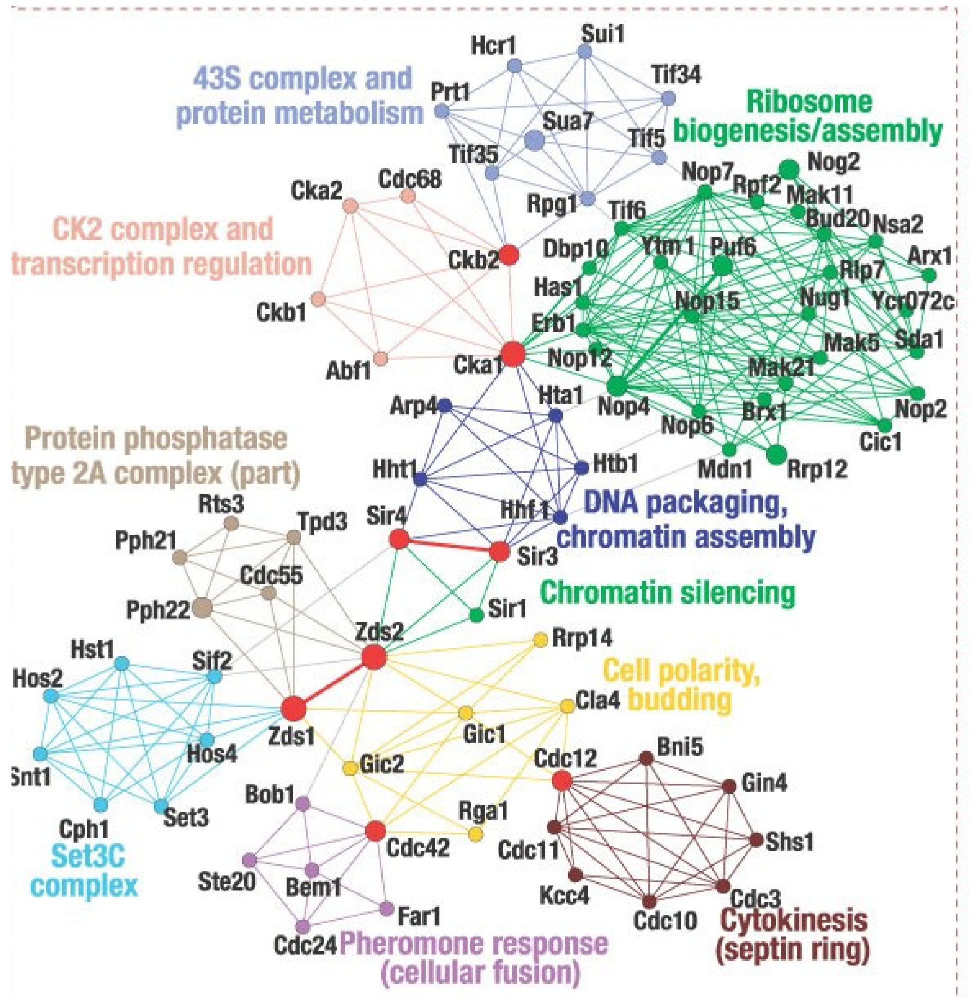
overlapping *k*-clique communities at *k* = 4.



Existing community-finding techniques assume that modules are non-overlapping and non-nested – however, in many networks a node may belong to multiple communities

E.g., a large fraction of proteins belong to several protein complexes simultaneously

***k*-clique community**: a union of all *k*-cliques (complete subgraphs of size *k*) that can be reached from each other through a series of adjacent *k*-cliques. Two *k*-cliques are adjacent if they share *k* - 1 nodes.



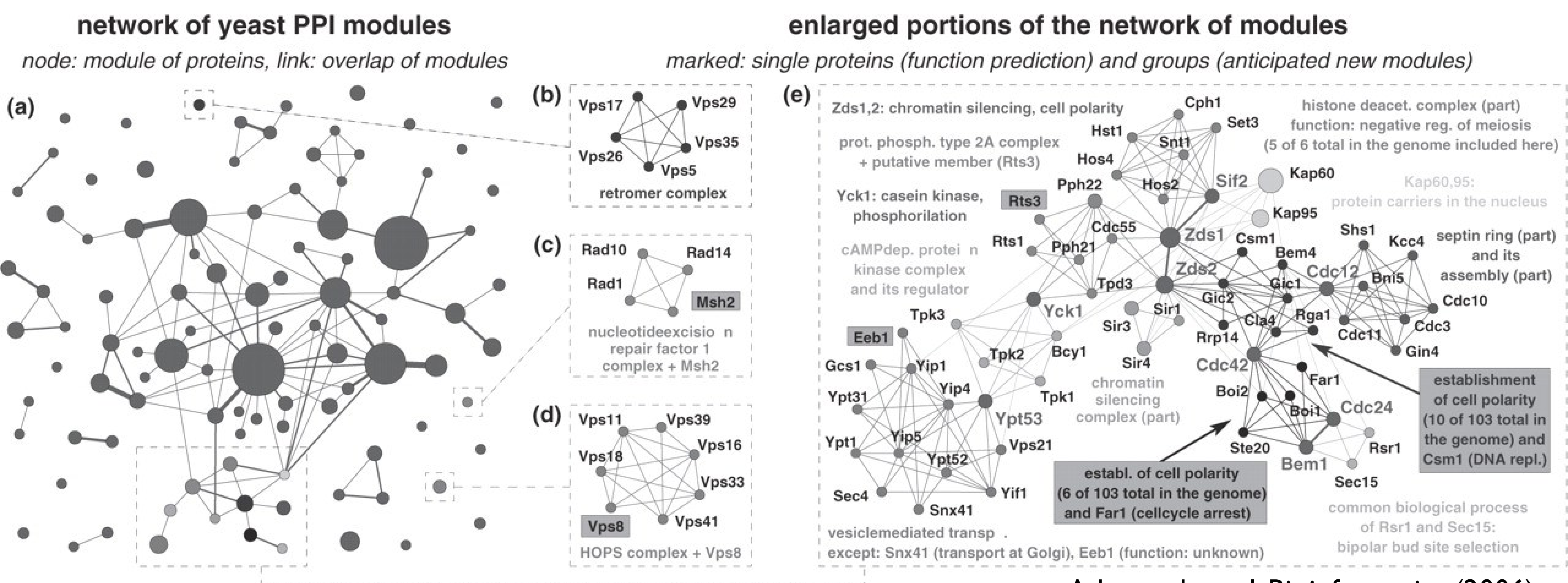
Communities in the protein-protein interaction network of *S. cerevisiae* (DIP database) for k=4. Node size and link widths are proportional to total number of communities they belong to.

Cfinder: Finding overlapping modules in networks

Computational implementation of Clique Percolation Method for identifying to locate the *k*-clique percolation clusters of a network that are interpreted as modules.

CFinder used to

- ❑ predict function of single nodes (e.g., protein) in biological networks based on their membership in modules (“guilt by association”)
- ❑ to identify new modules i.e., groups of densely interconnected nodes, possibly involved in a specific function (“a gossiping group must be upto something”)
- ❑ locating the cliques of large sparse graphs



Adamcsek et al, Bioinformatics (2006)