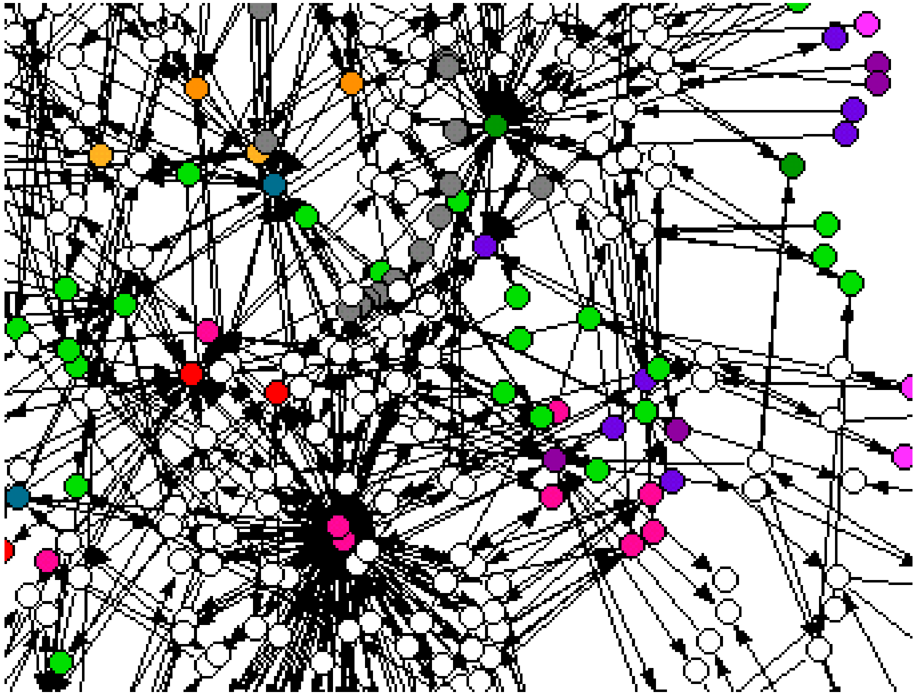
Networks in Molecular Biology

June 2, 2004

Petter Holme

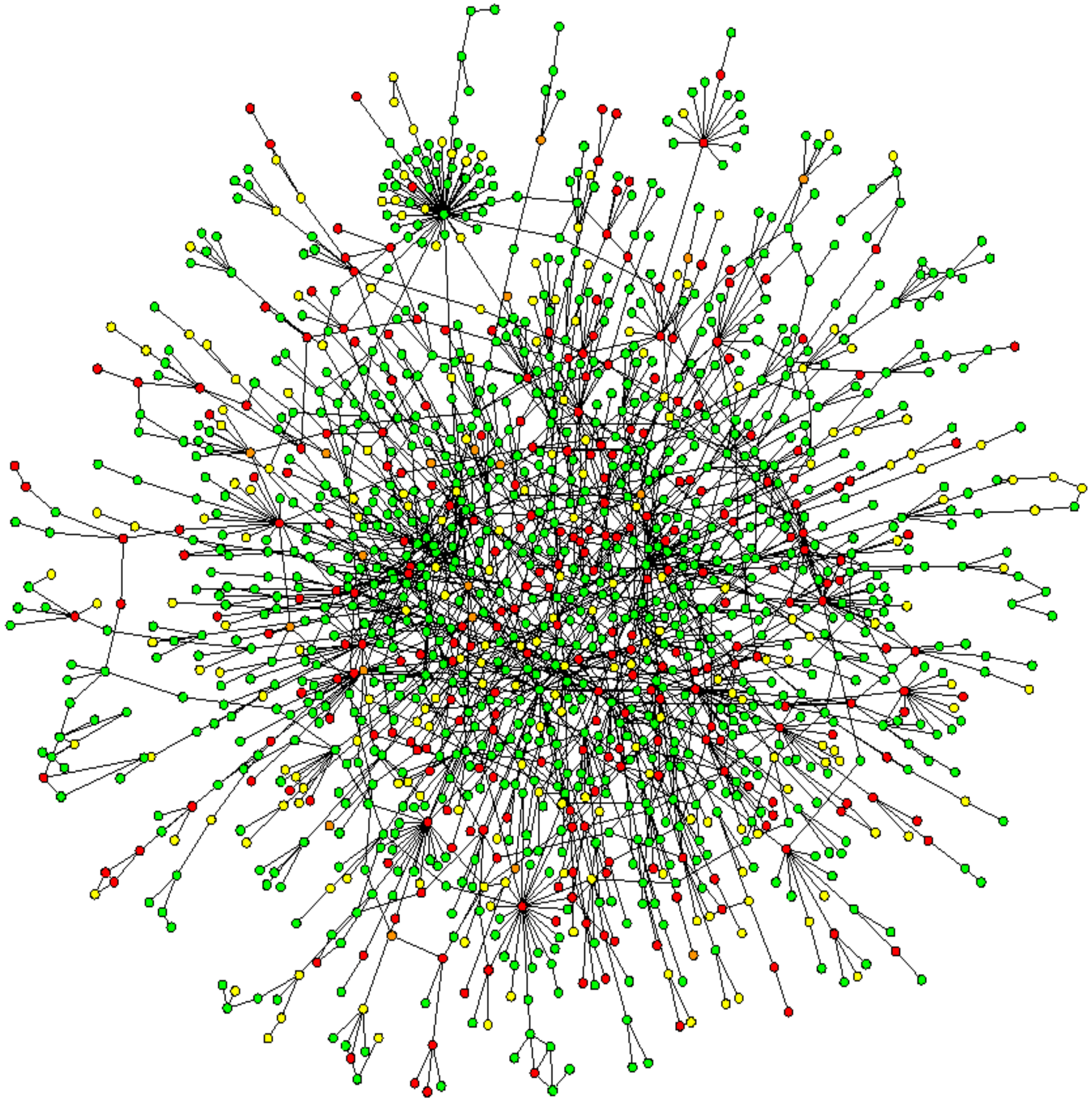
Department of Physics,
Umeå University, Sweden

NORDITA, Copenhagen,
Denmark



INTRODUCTION: NETWORK STRUCTURE

Network structure—the way a network differs from random network (Edrős-Rényi model).



- Clustering—density of short circuits.
- Degree distribution—often broad (scale-free networks).
- Degree-degree correlations.
- Community structure, "modularity."
- Network motifs.

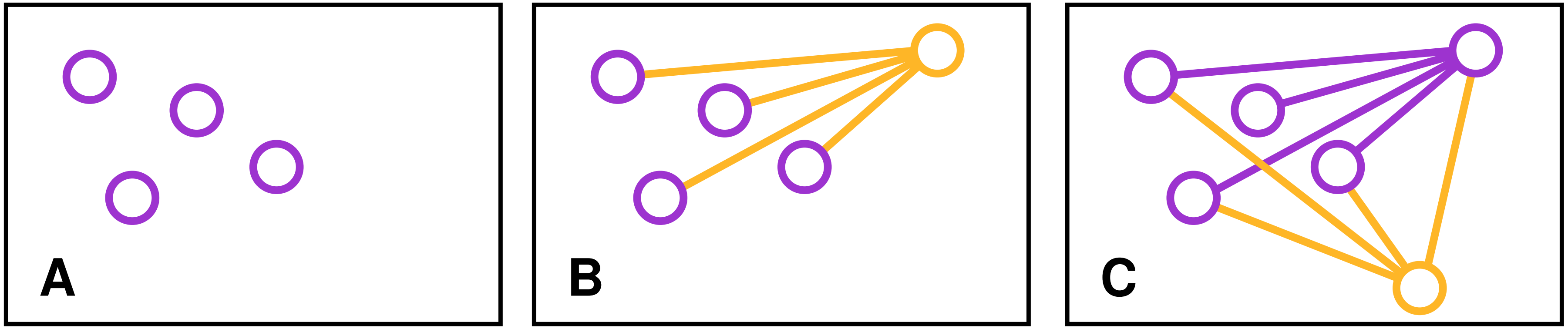
INTRODUCTION: HOW CAN THE STRUCTURE BE UTILIZED?

- How can we protect a network from sinister attacks?
- How can vaccinate a population most efficiently?
- Can we predict the traits of the vertices? (Functional prediction of proteins.)
- Can we (re)build a network to make communication flow more efficiently?
- Can we use biochemical network theory in drug-design?

INTRODUCTION: NETWORK MODELS

- How does structure emerge?
- Can we model this?

One example, the Barabási-Albert model of networks with a power-law degree distribution [Science **286** (1999), pp. 509-512; D. J. S. Price, J. Amer. Soc. Inform. Sci. **27** (1979), pp. 292-306.]



A Start with m isolates. **B-C** Grow the network iteratively: At each time step, add one vertex and m edges and attach the edges to an old vertex v with a probability proportional to the degree of v.

INTRODUCTION: DYNAMIC SYSTEMS ON NETWORKS

Eventually all network studies strives to elucidate: *How does the network structure affect the dynamical systems confined to the network?*

Relevant dynamics:

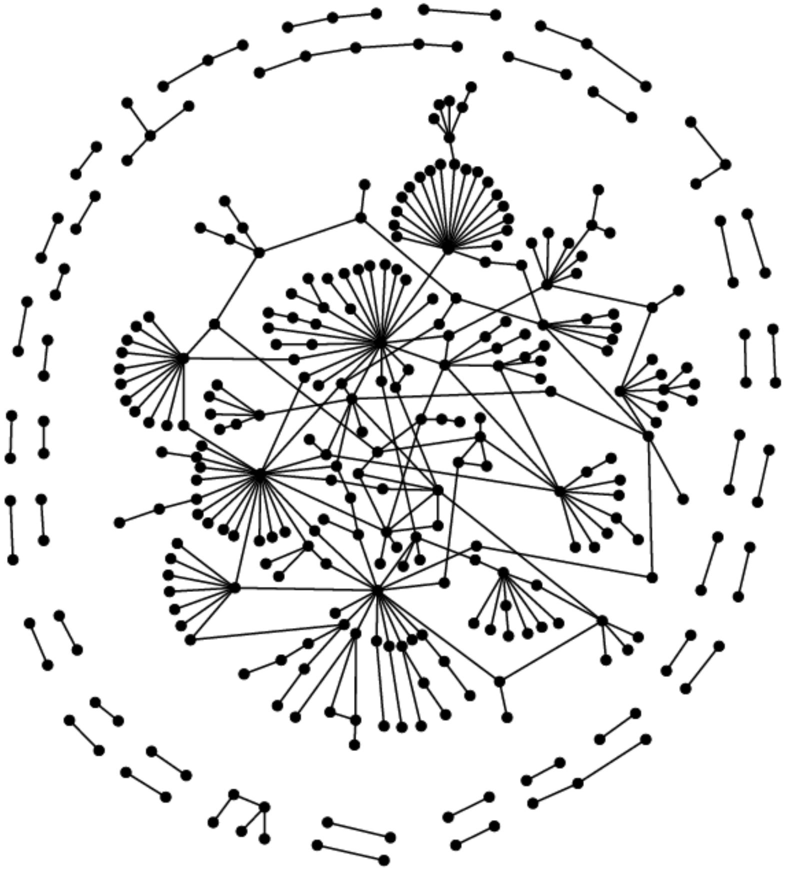
- Social networks: Epidemiological models. Models for social stability & spreading of rumors.
- Computer network: Internet protocol based dynamics.
- . . . and so on.

Irrelevant dynamics:

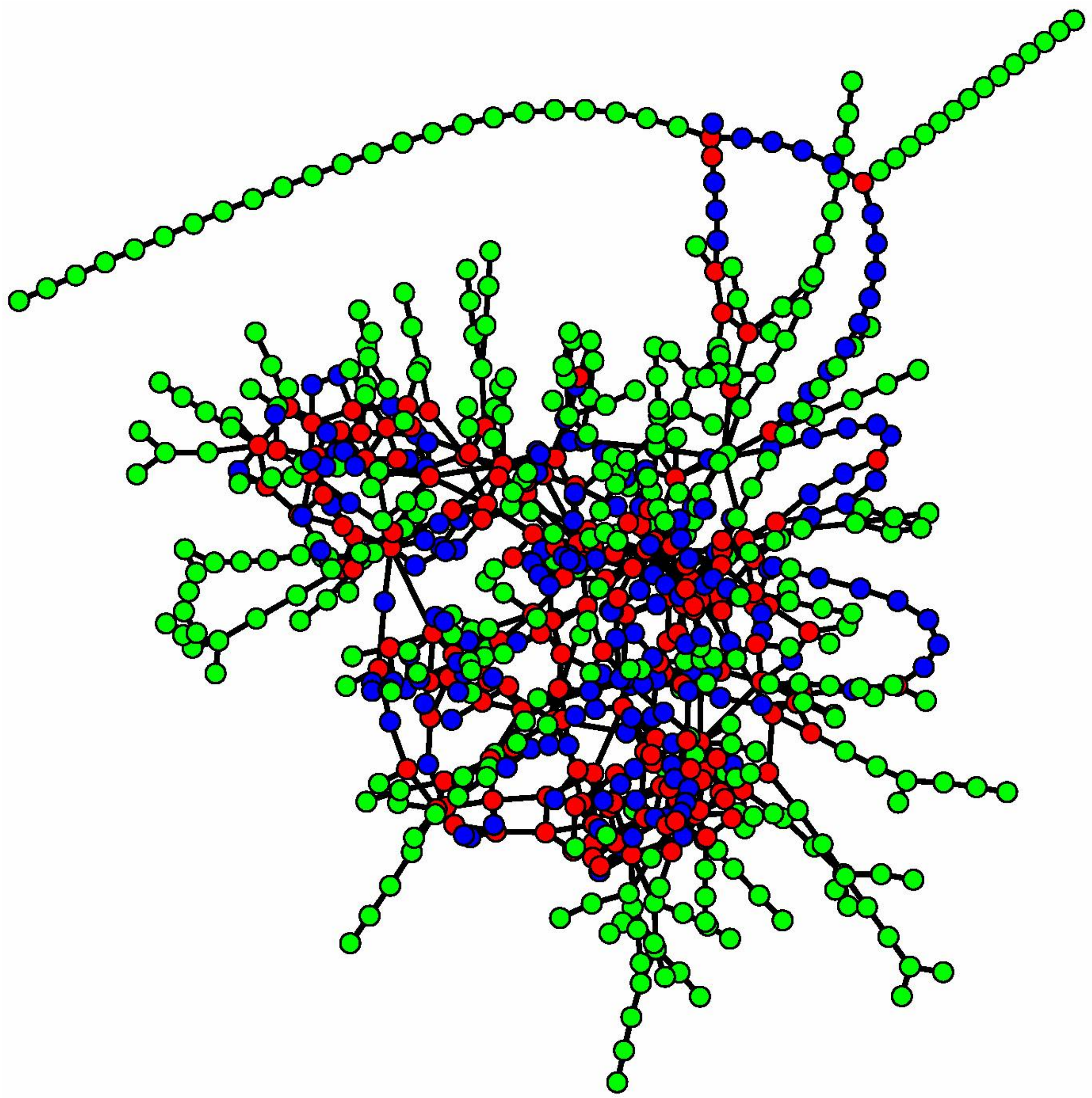
- . . . publishable in Physics Journals.

Protein interaction networks

Vertices: Proteins. **Edges:** Physical interaction between proteins.



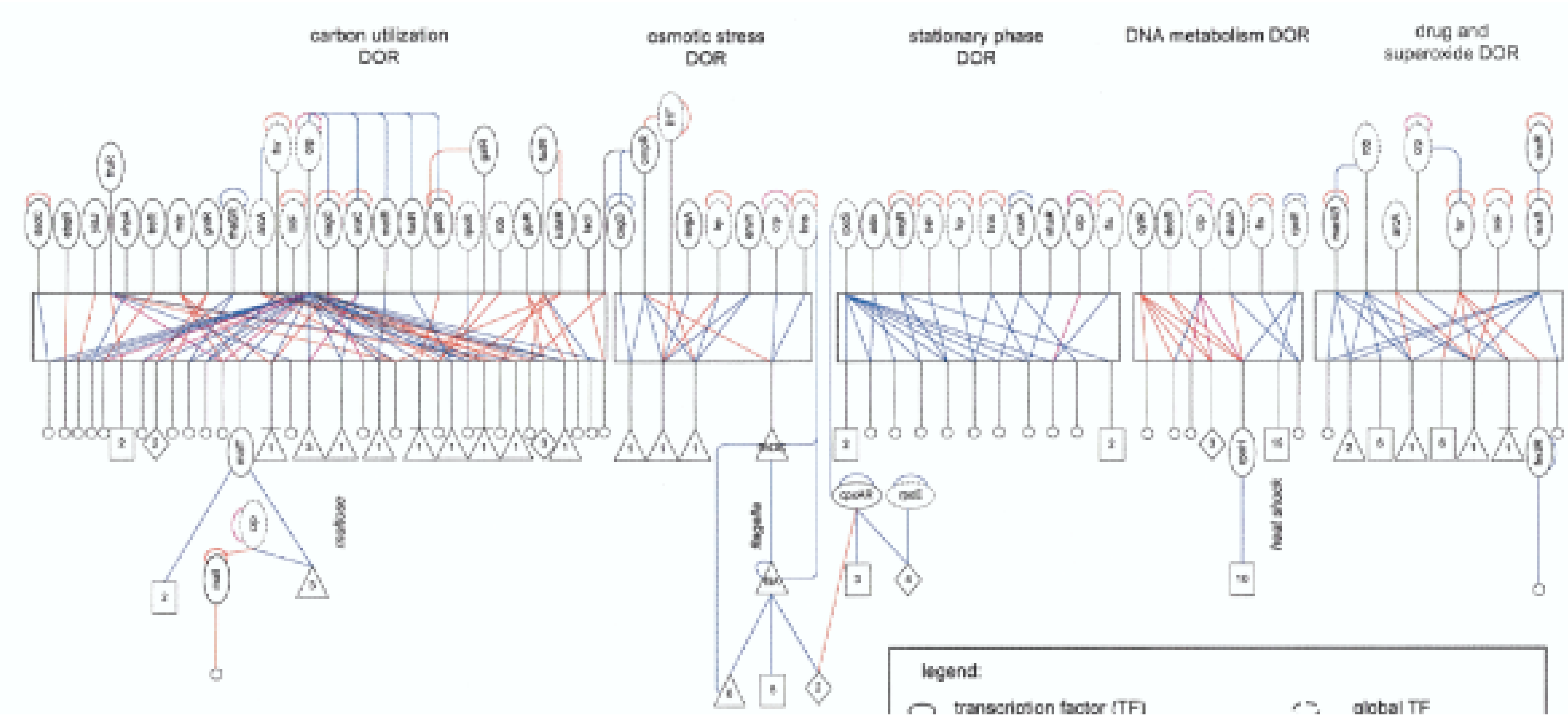
Networks in Molecular Biology



Networks in Molecular Biology *(continued)*

Transcription regulation networks

Vertices: Genes. **Edges:** One gene promotes (or inhibits) the regulation of another.



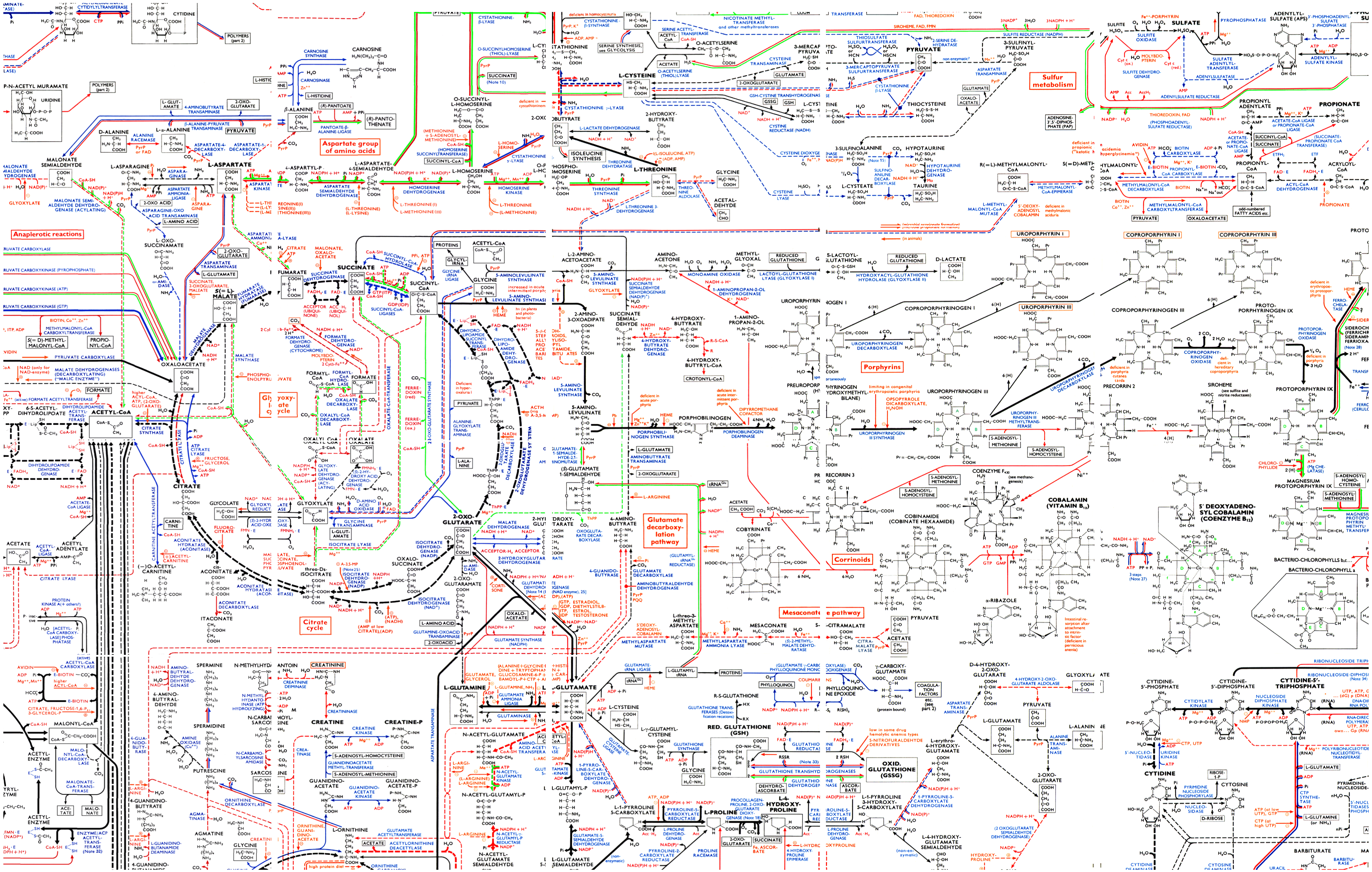
Networks in Molecular Biology

Biochemical networks

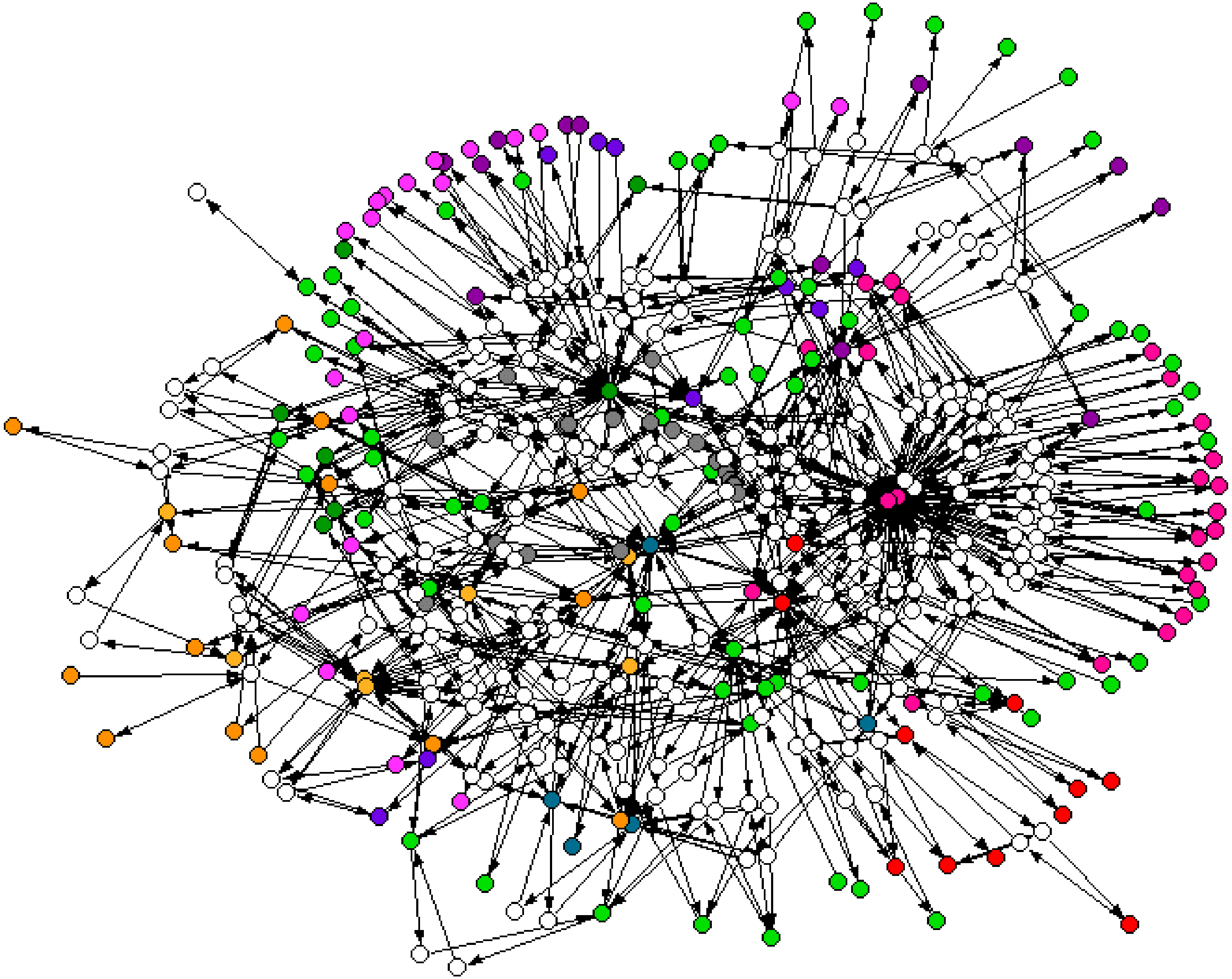
Vertices: Chemical substrates. **Edges:** One substrate can be converted to another.

…pictures in a moment …

SUBNETWORK HIERARCHIES OF BIOCHEMICAL PATHWAYS



Subnetwork hierarchies of biochemical pathways



metabolic pathways of *Borrelia burgdorferi* (a bacterium)

Subnetwork hierarchies of biochemical pathways

P. Holme, M. Huss and H. Jeong, *Subnetwork hierarchies of biochemical pathways*, Bioinformatics **19** (2003), pp. 533-538.

Motivation:

Complexity

Even *E. coli* has a metabolism involving over 850 substances and 1500 reactions ⇒

- The coarsest level of description—the graph representation—is needed, at least as a complement.
- One would like to decompose the graph into functional subunits. Both for conceptual and analytical purposes.

Our work

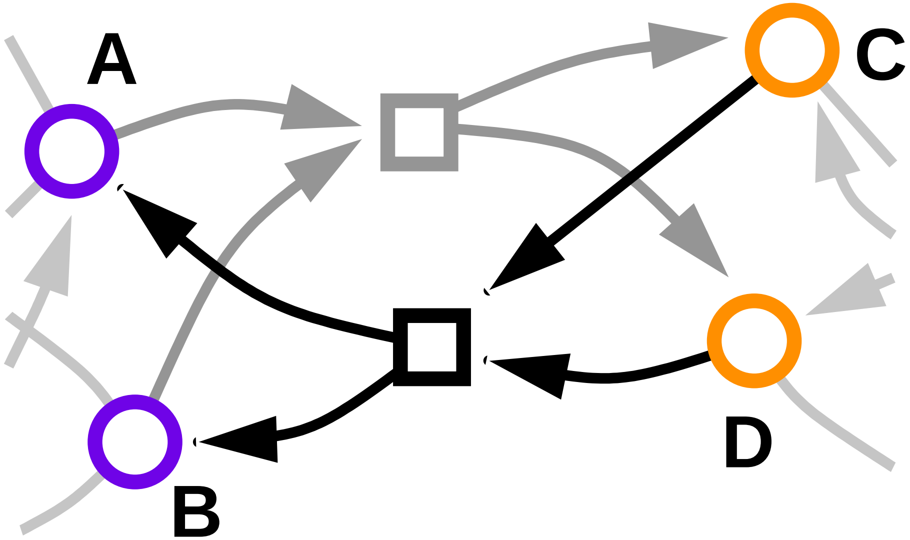
- Earlier algorithms have been based on local algorithms that may miss some large-scale features.
- Not much known about how the large scale subnetwork ordering looks like. Can the network easily be decomposed into autonomous subnetworks? How independent are the modules? Is it useful to talk about modules at all?

The basic assumption

If we find a subnetwork that is well-connected within, and sparsely connected to the outside, then it is likely to be a relatively autonomously functioning subnetwork.

Subnetwork hierarchies of biochemical pathways

Biochemical networks as directed bipartite networks

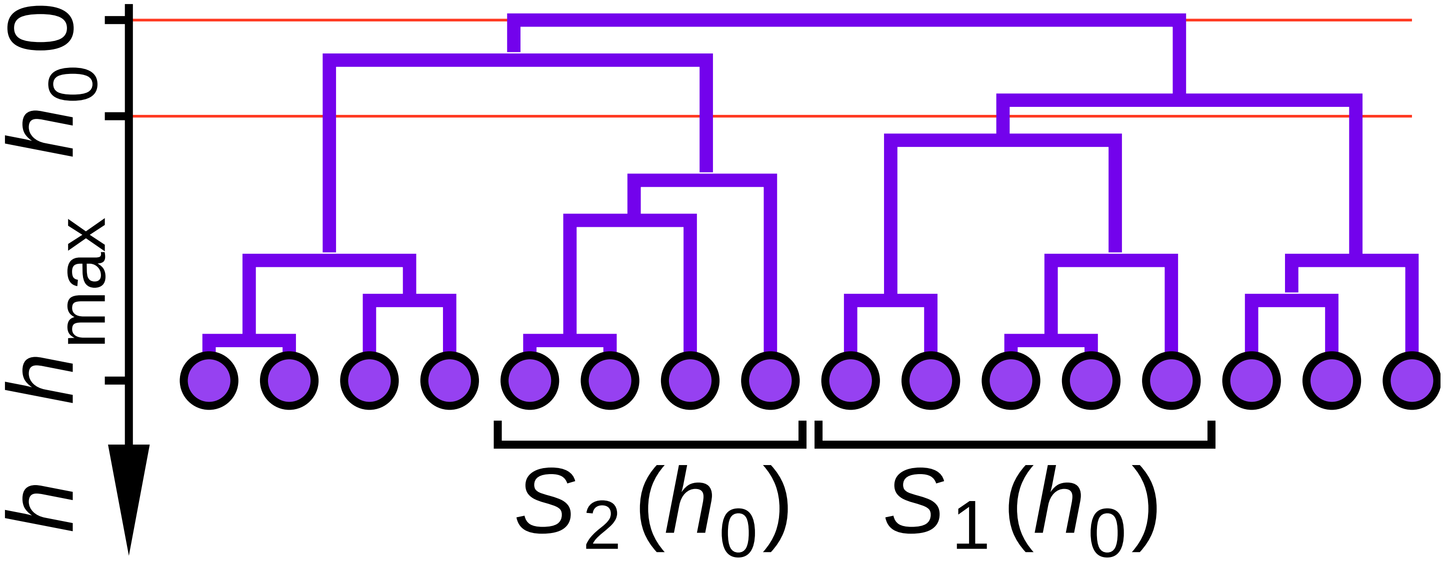


The reaction A + B ↔ C + D in a directed bipartite representation:

- Two types of vertices, representing substrates and chemical reactions.
- Arcs (arrows) between different types of vertices

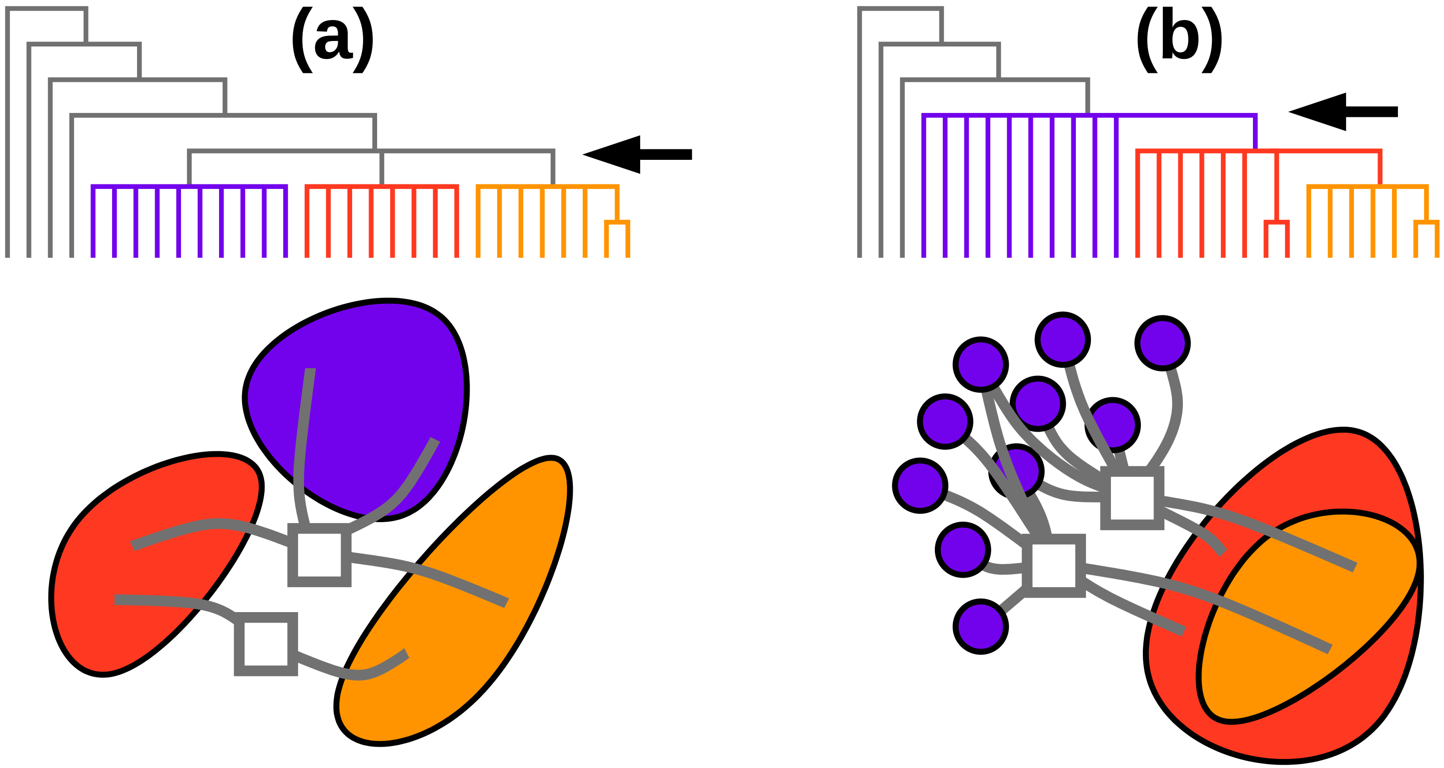
We denote the set of chemical substances by S and the set of reaction vertices by R.

Subnetwork hierarchies



- The substrates are at the base of the tree.
- If a horizontal line is drawn across the tree, the vertices below are connected at that particular level of the hierarchy.
- Clusters that are isolated high in the hierarchy (close to the bottom of the tree) are more entangled in other pathways.

Subnetwork hierarchies of biochemical pathways *(continued)*

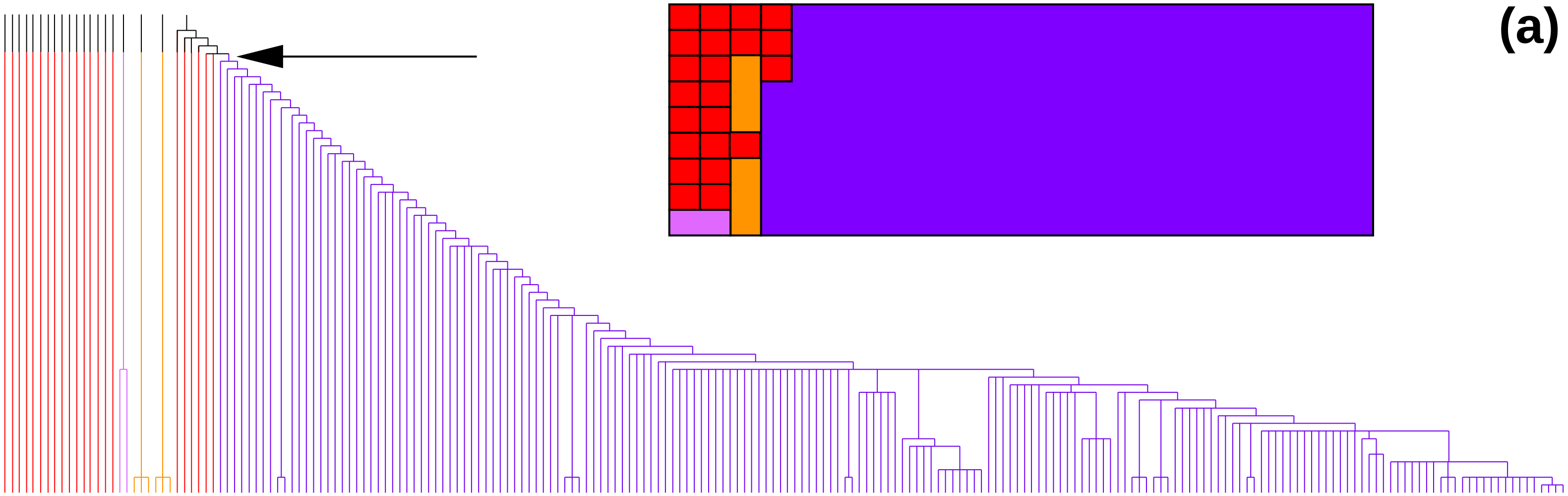


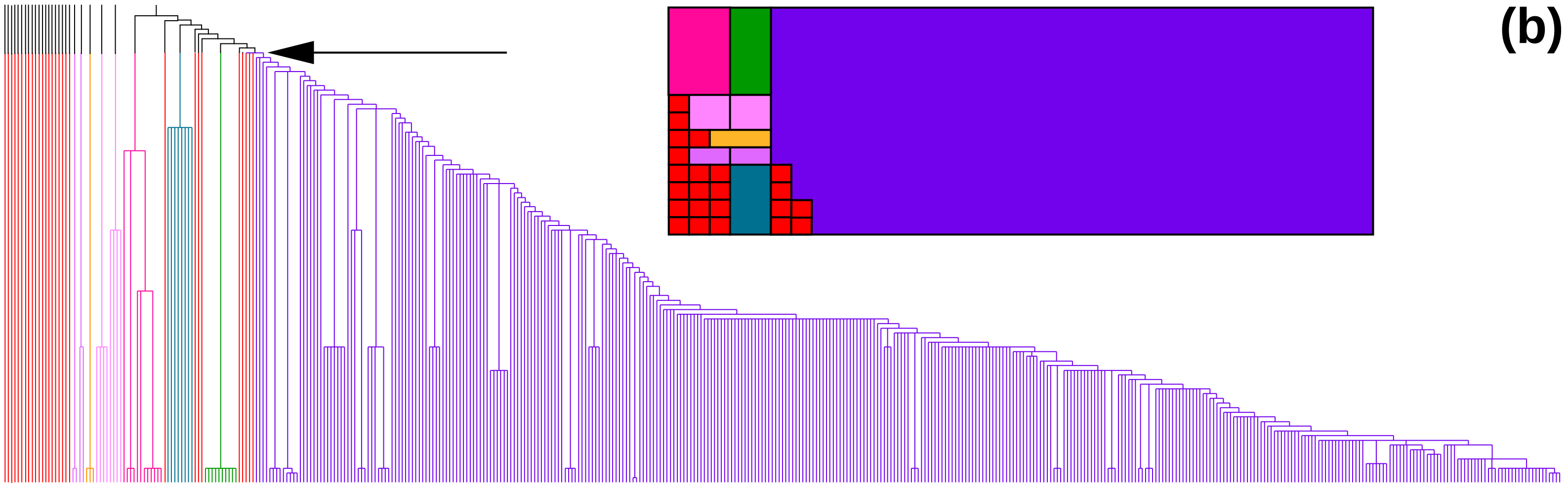
- **(a)** Clusters that get isolated at the same level are more highly wired within, than to its surrounding (and therefore a candidate to a functional module).
- **(b)** Vertices that becomes isolated at the same level forms an outer shell of the cluster in question.

Subnetwork hierarchies of biochemical pathways

(continued)

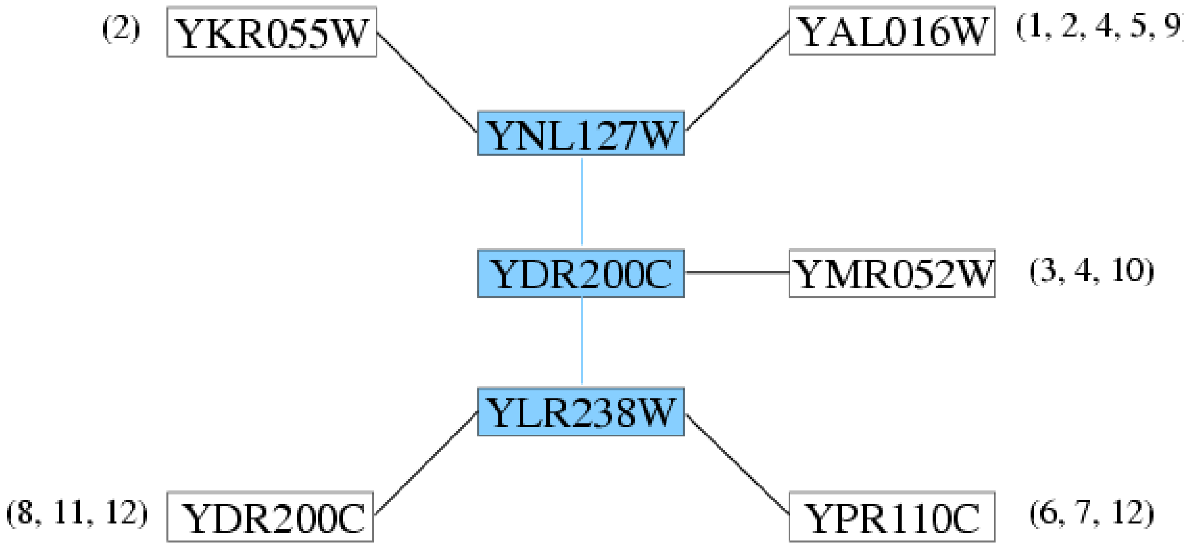
Examples: *Treponema pallidum*





NETWORK BASED FUNCTIONAL PREDICTION

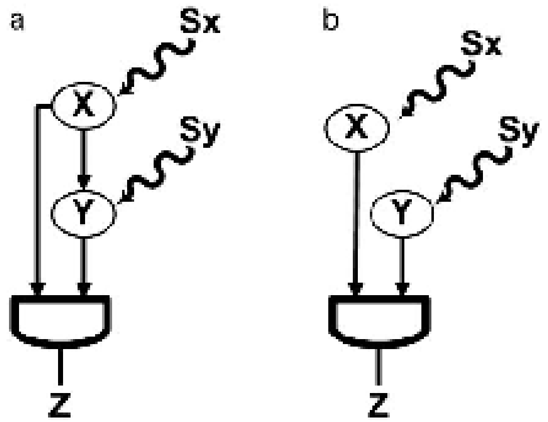
A. Vazquez *et al.*, *Global protein function prediction in protein-protein interaction networks*, Nature Biotech. **21** (2003), 697-700.



- Before: Majority rule.
- A. Vazquez *et al.*'s idea: Minimization of Hamiltonian-like cost function.

NETWORK MOTIFS IN TRANSCRIPTION REGULATION

S. Mangan and U. Alon, PNAS **100** (2003), pp. 11980-11985; S Shen-Orr *et al.*, Nature Genetics **31** (2002), pp. 64-68.



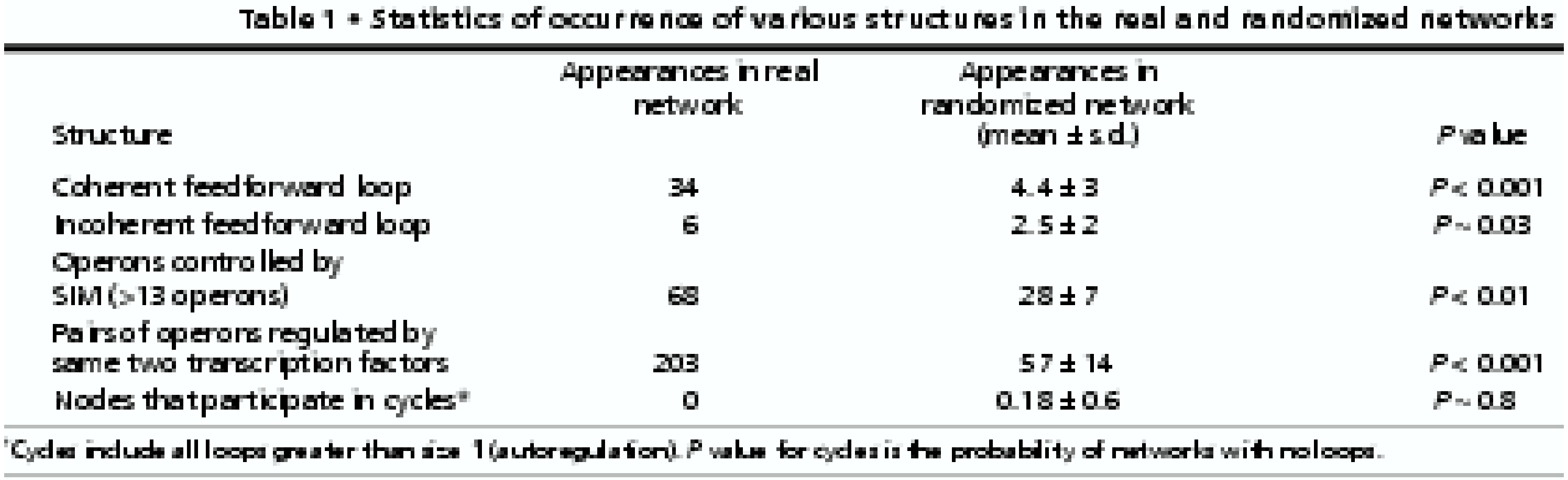


Table 1 • Statistics of occurrence of various structures in the real and randomized networks

Structure	Appearances in real network	Appearances in randomized network (mean ± s.d.)	P value
Coherent feedforward loop	34	4.4 ± 3	$P < 0.001$
Incoherent feedforward loop	6	2.5 ± 2	$P \sim 0.03$
Operons controlled by SIM (>13 operons)	68	28 ± 7	$P < 0.01$
Pairs of operons regulated by same two transcription factors	203	57 ± 14	$P < 0.001$
Nodes that participate in cycles*	0	0.18 ± 0.6	$P \sim 0.8$

*Cycles include all loops greater than size 1 (autoregulation). P value for cycles is the probability of networks with no loops.