Computational Systems Biology

Regulatory networks and integrated network

Hongwu Ma

Computational Systems Biology Group
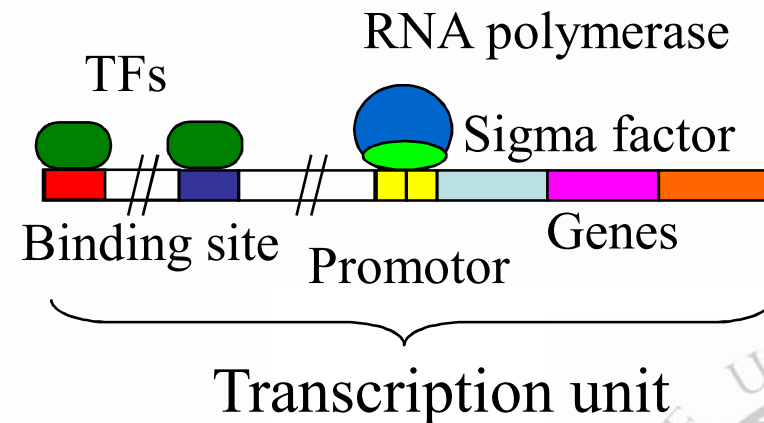
19 February 2010

Outline

- The reconstruction
- Network structure
- Link regulatory network with metabolic network
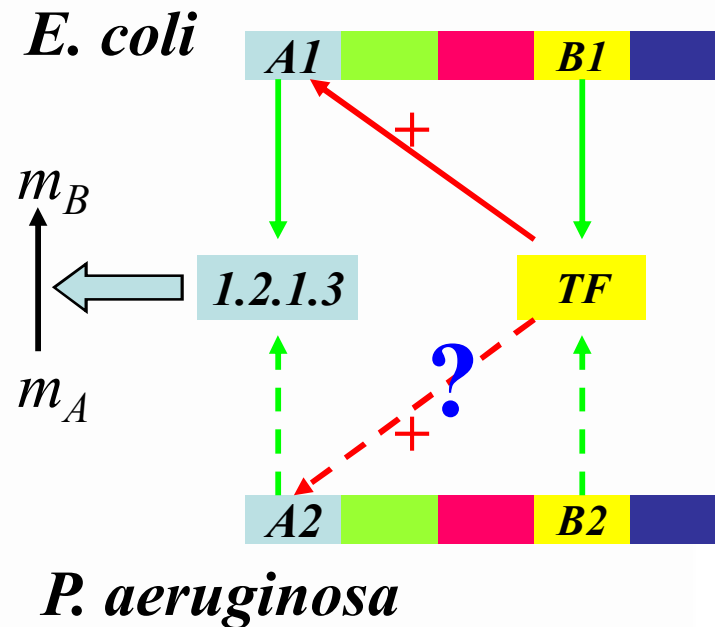- Network Motif analysis

Transcriptional regulation: biological basis

- Transcription unit: operon
- Sigma factor (rpoD, rpoS, rpoN,)
- Transcription factor (TF)
- Promotor
- TF binding site
- More complex in Eukaryotes

Interactions between proteins and genes

Reconstruction of MN and TRN

Are the orthologous genes in different organisms regulated in similar ways?

Transcriptional regulatory networks

- Can not be reconstructed directly from gene annotation information

- data is mainly from database which collect information from literature (for *E. coli* and *B. subtilis*)

- ChIP on chip technology for high throughput reconstruction (mainly yeast network)
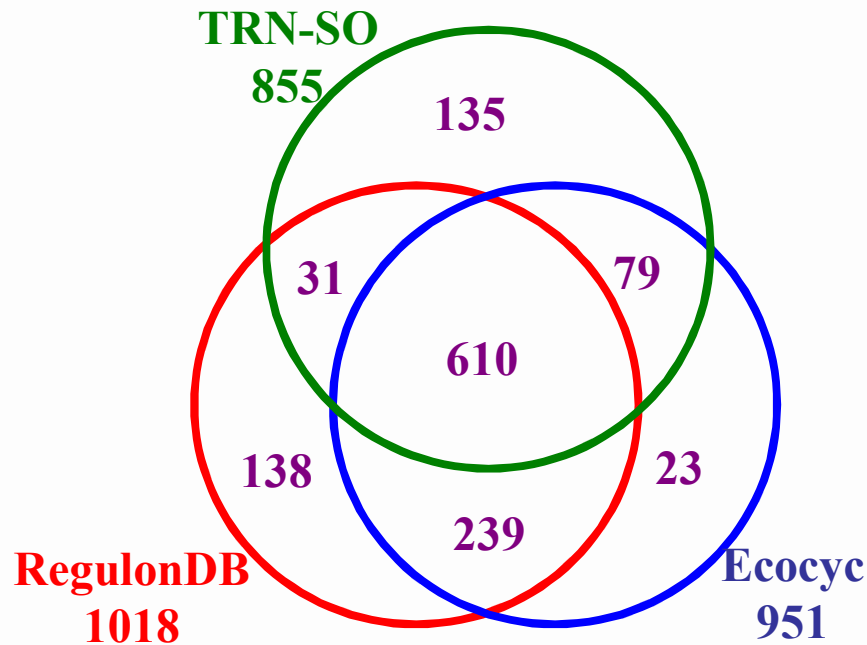
- Reverse engineering from Microarray data

Available databases

- RegulonDB http://regulondb.ccg.unam.mx/

- Ecocyc: ecocyc.org

- DBTBS (*B. subtilis*): dbtbs.hgc.jp

- Prodoric Net (Prokaryote): prodoric.tu-bs.de

- TRANSFAC (Eukaryotes):
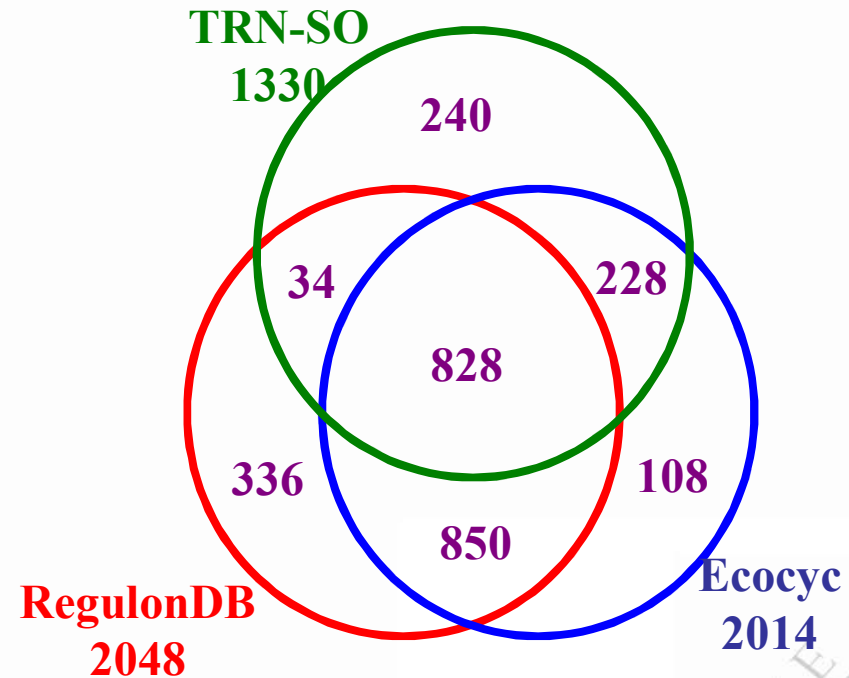 www.biobase.de/pages/products/transfac.html
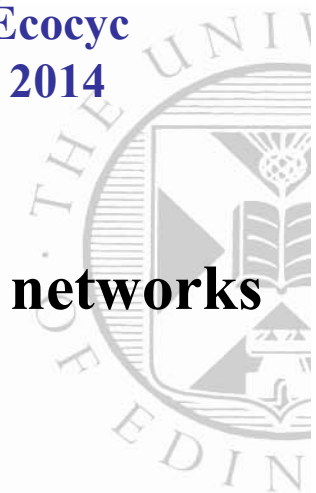
Mainly Use *E. coli* as an example

E. coli regulatory networks

genes in the networks

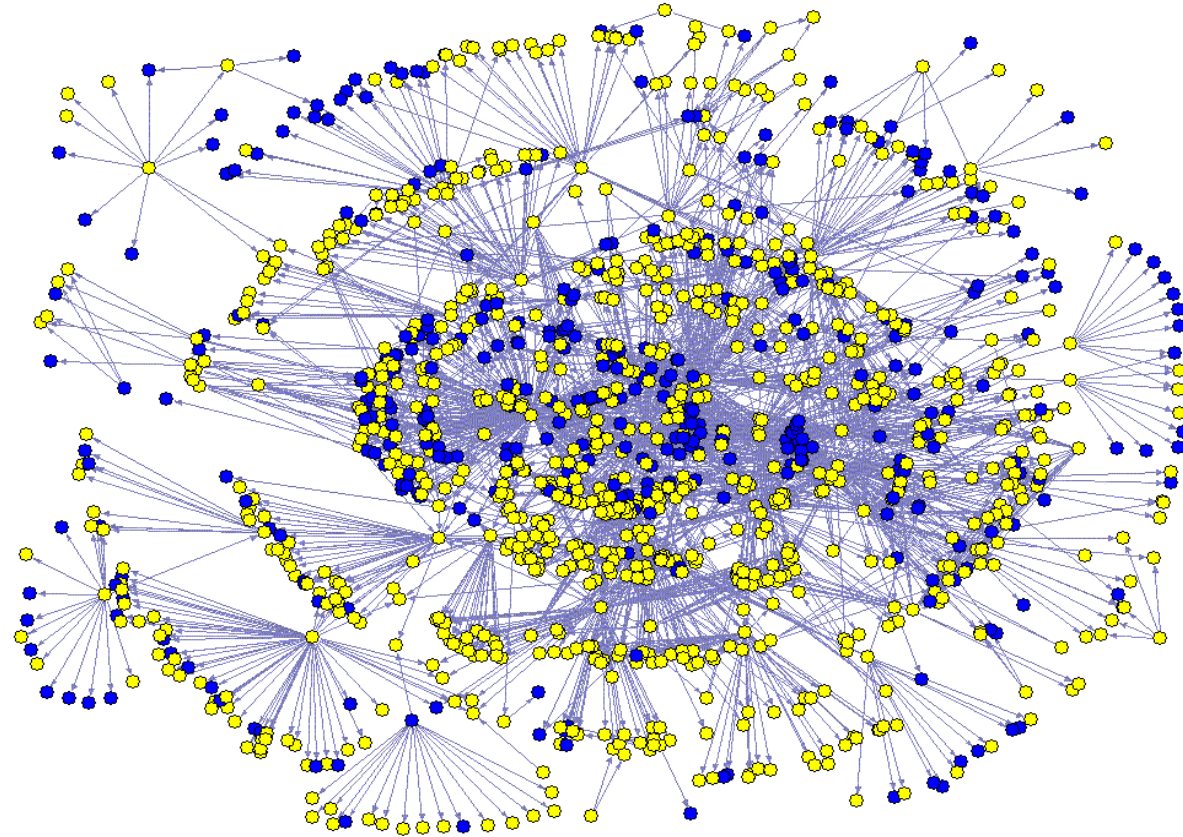# interactions in the networks

E. coli regulatory network

1278 genes
2724 interactions
157 genes for TFs
382 metabolic
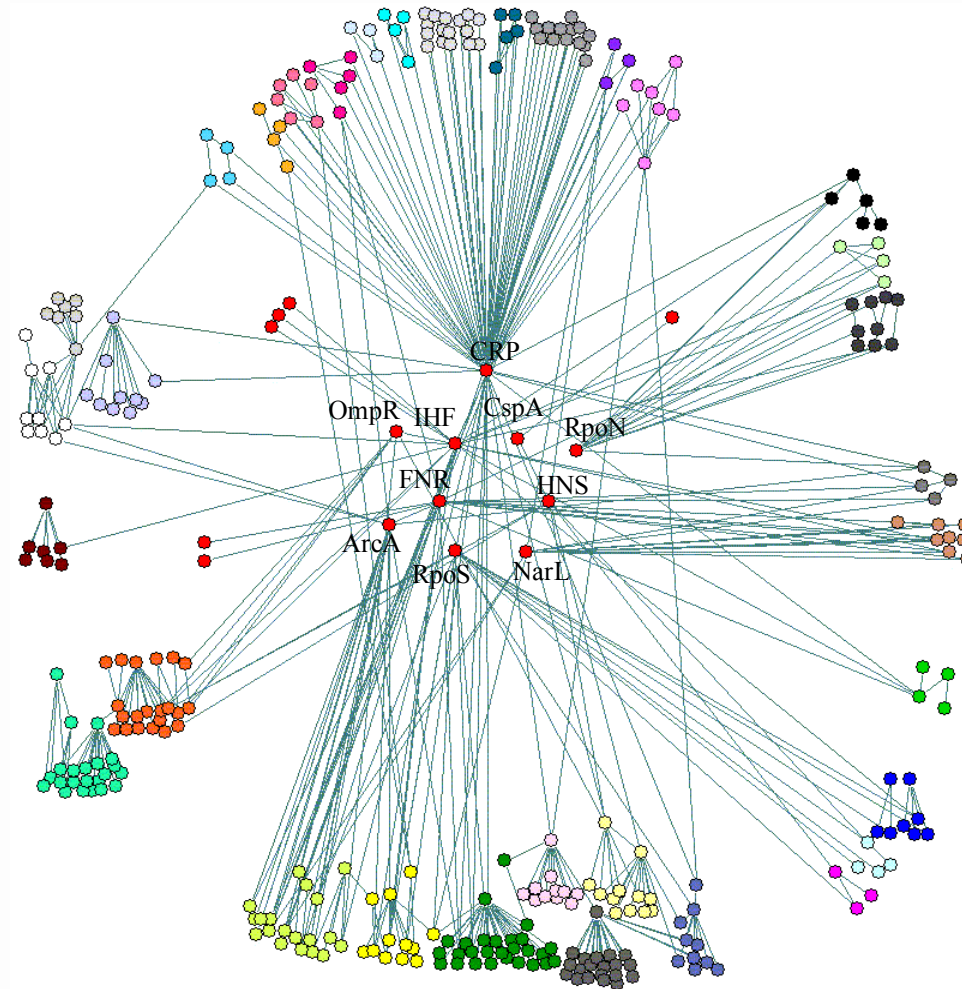**enzyme genes**
(Blue)

Structure analysis

- Degree distribution
  - Out degree distribution follows power law: global regulators regulate many target genes
  - In degree distribution: Maximal In degree is 7. The number of regulators which regulate the same gene is limited by the number of binding sites upstream of the target gene
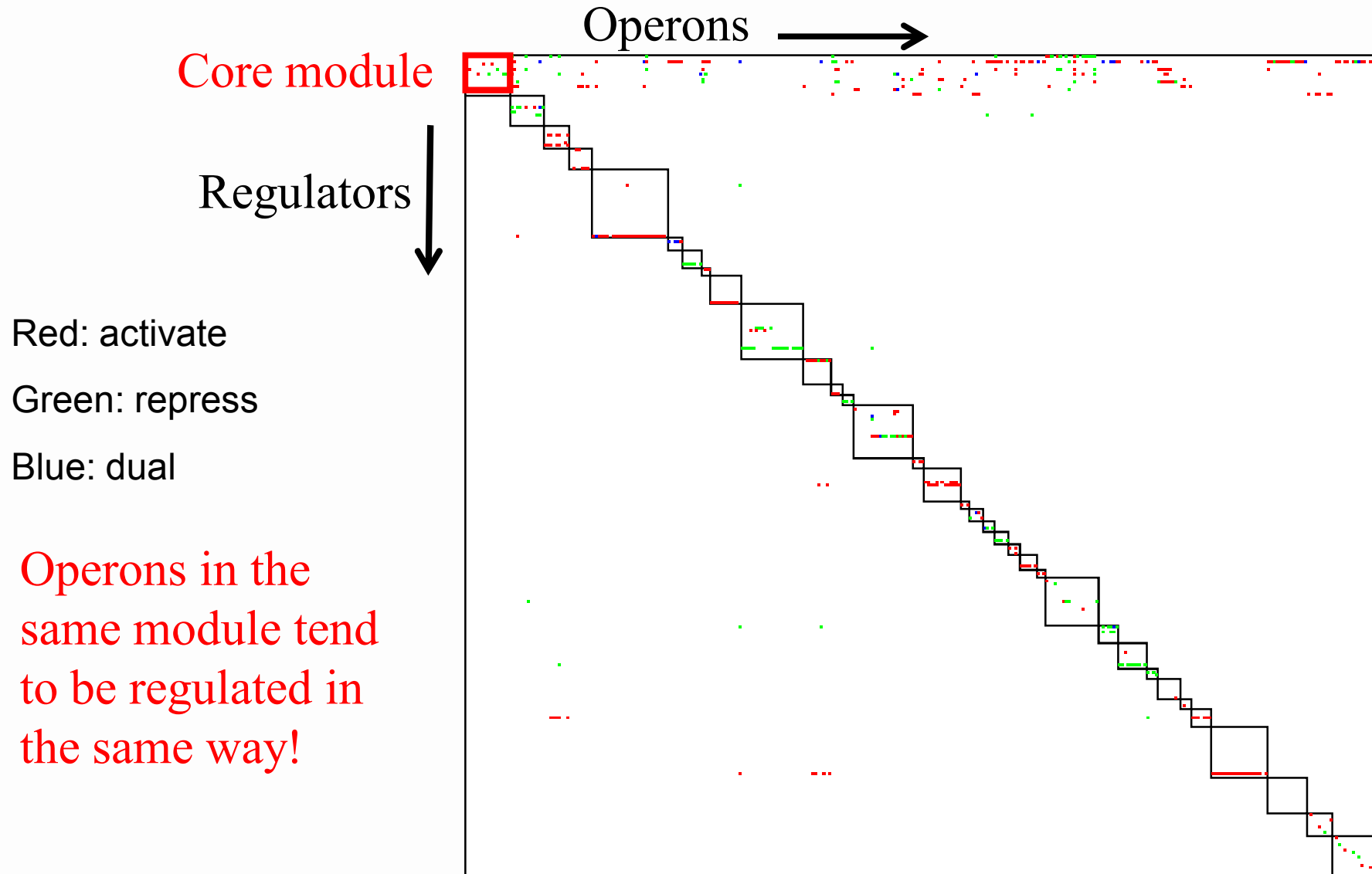- Average path length: less than 2.

Modular organization: Core-periphery structure

Core module includes the global regulators. Other modules mainly connected through it.

Matrix representation of the connectivity

Operons →

Core module

Regulators ↓

Red: activate

Green: repress

Blue: dual

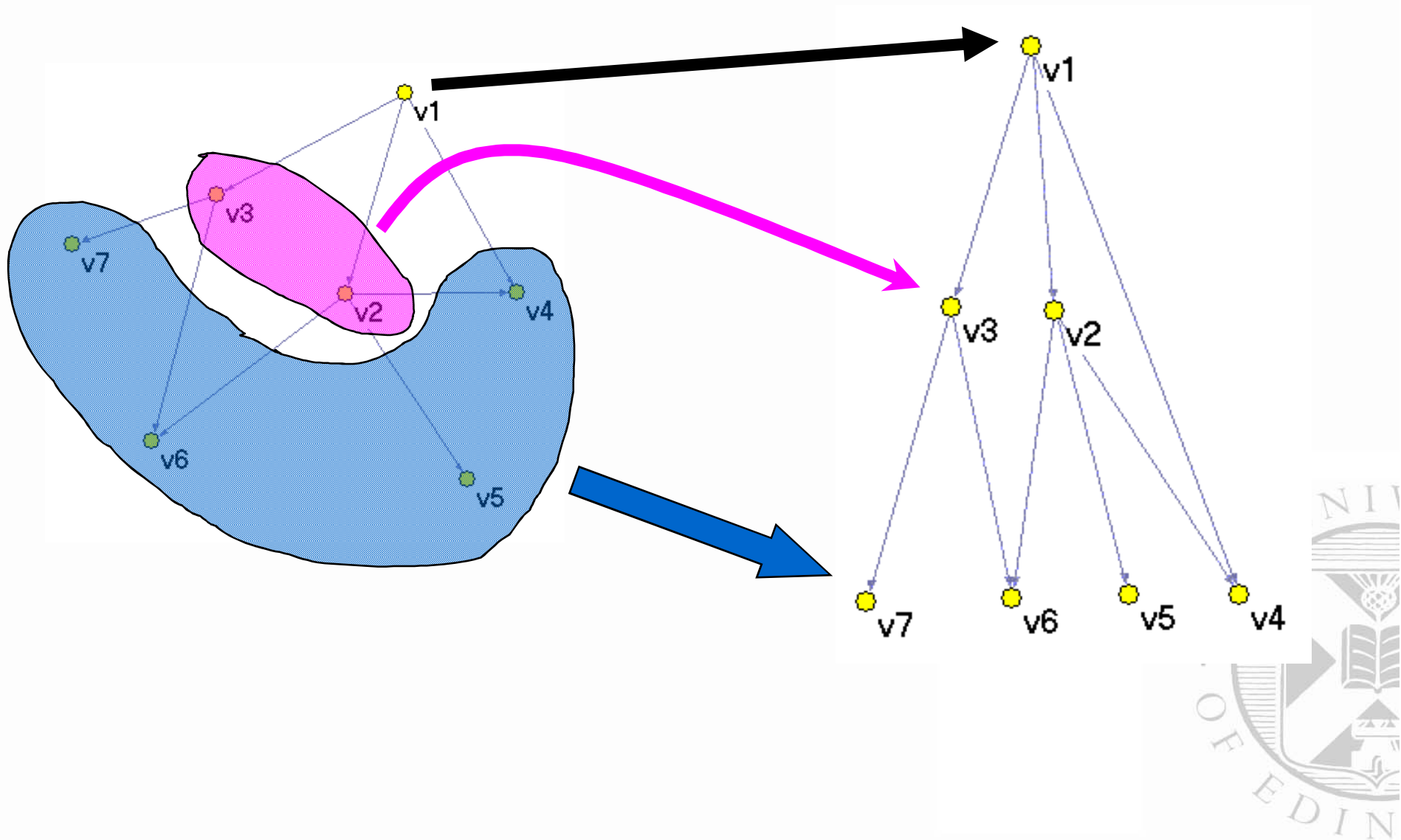Operons in the same module tend to be regulated in the same way!

Connectivity analysis

A giant weakly connected component with 1166 nodes

No strongly connected component exists, indicating no multiple gene regulatory loop exist, impling a somehow acyclic structure.
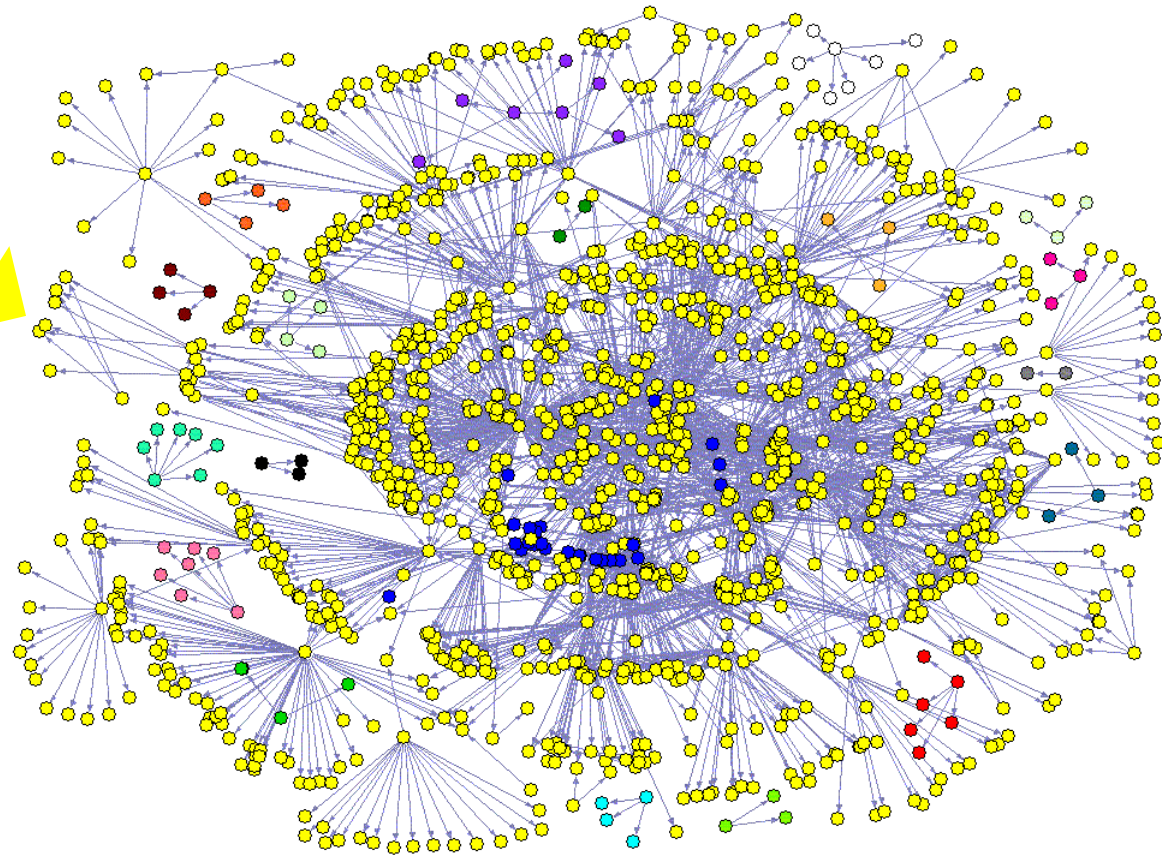
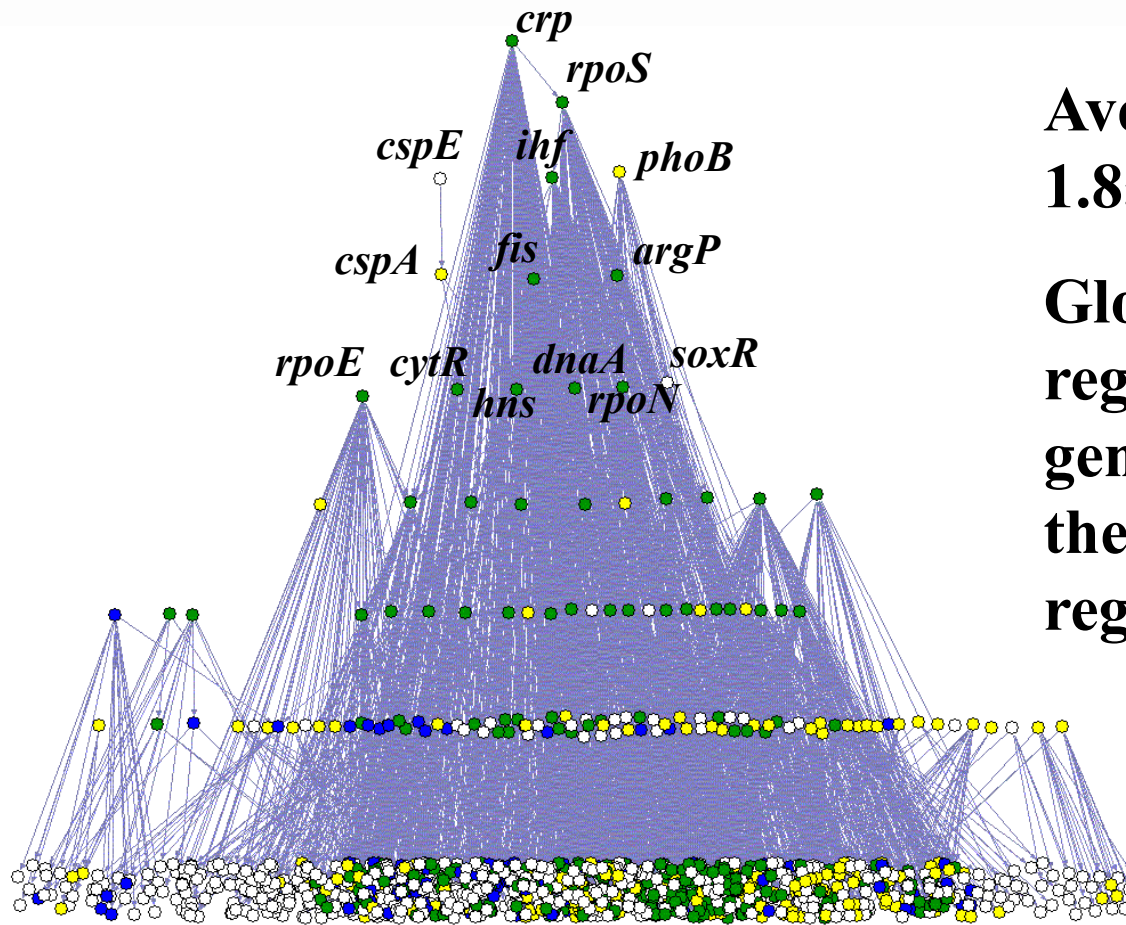# Method to find regulatory hierarchy
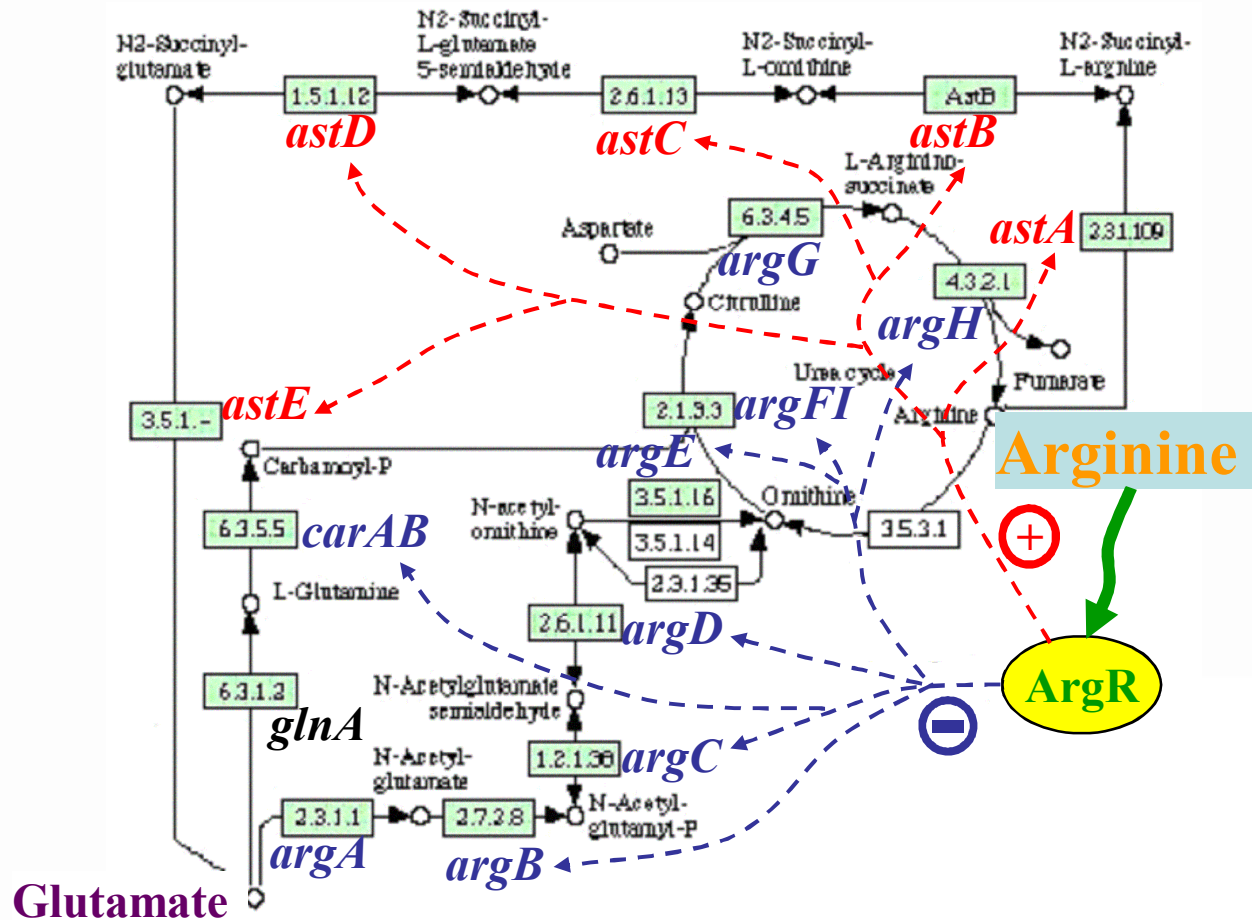
The nine-layer hierarchical structure

Average path length: 1.85

Global regulators regulate the target genes together with the specific regulators

Is it a universal structure? Where is the feedback?
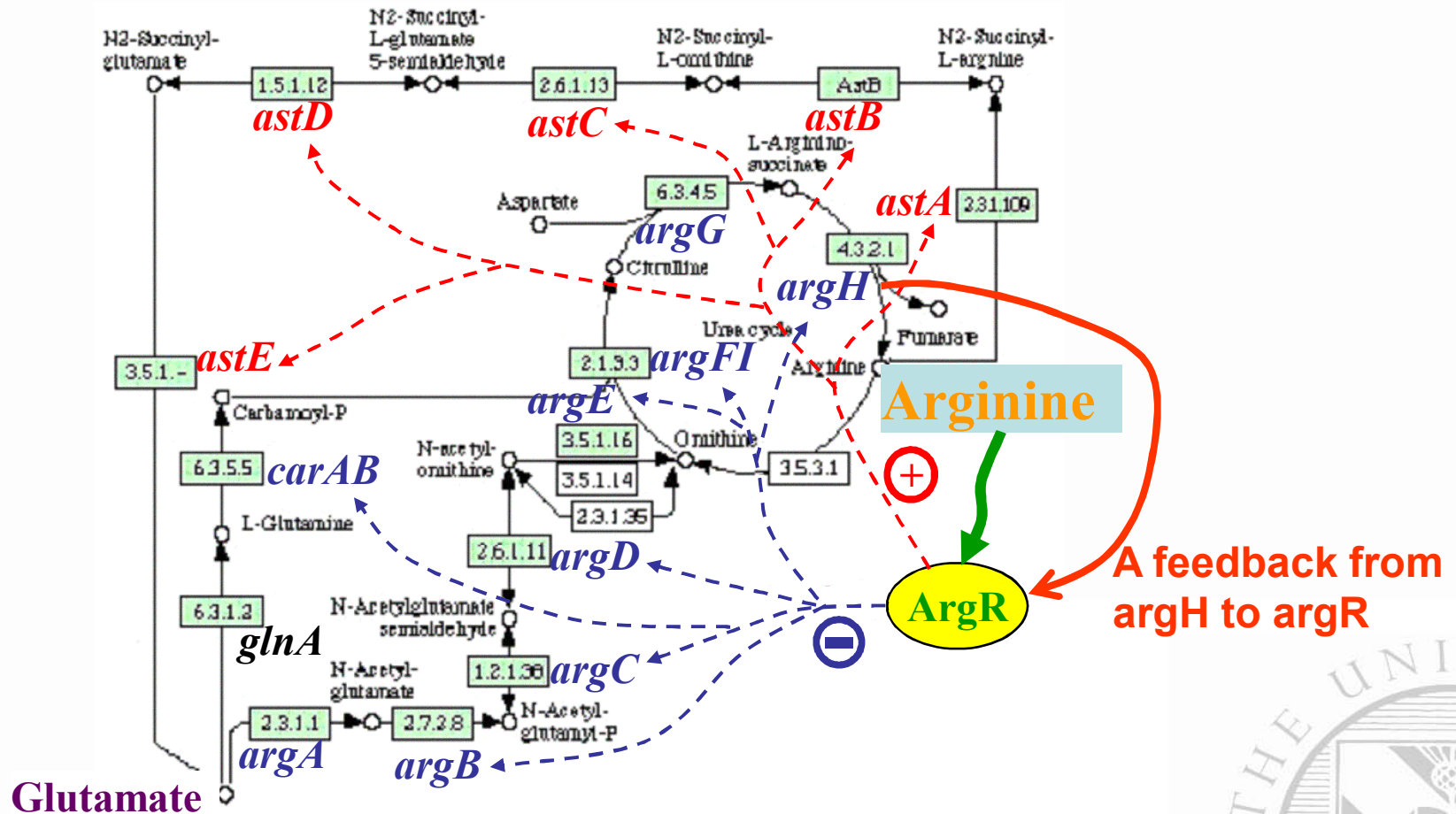
Feedback through metabolites

TF activity is directly affected by metabolites. No complete picture of regulation without metabolite-TF interactions.

Metabolite-TF interactions

- EcoTFs database: 50 TFs and their cofactors

- Four types of interactions

  - bind to active an activator: 20

  - bind to deactive a repressor: 14

  - Bind to deactive an activator: 5

  - Bind to active a repressor: 11

- RegulonDB: 78 interactions for 66 TFs, including 18 phosphorylation reactions. Many are phospyorylated by two component systems
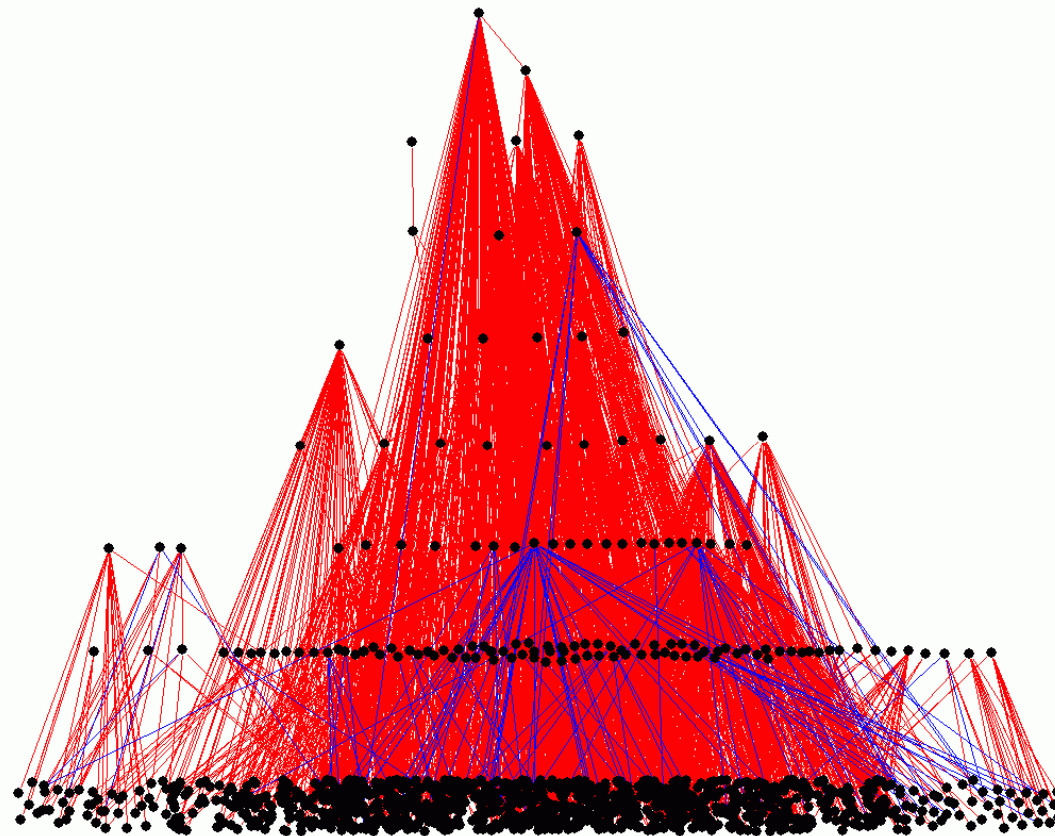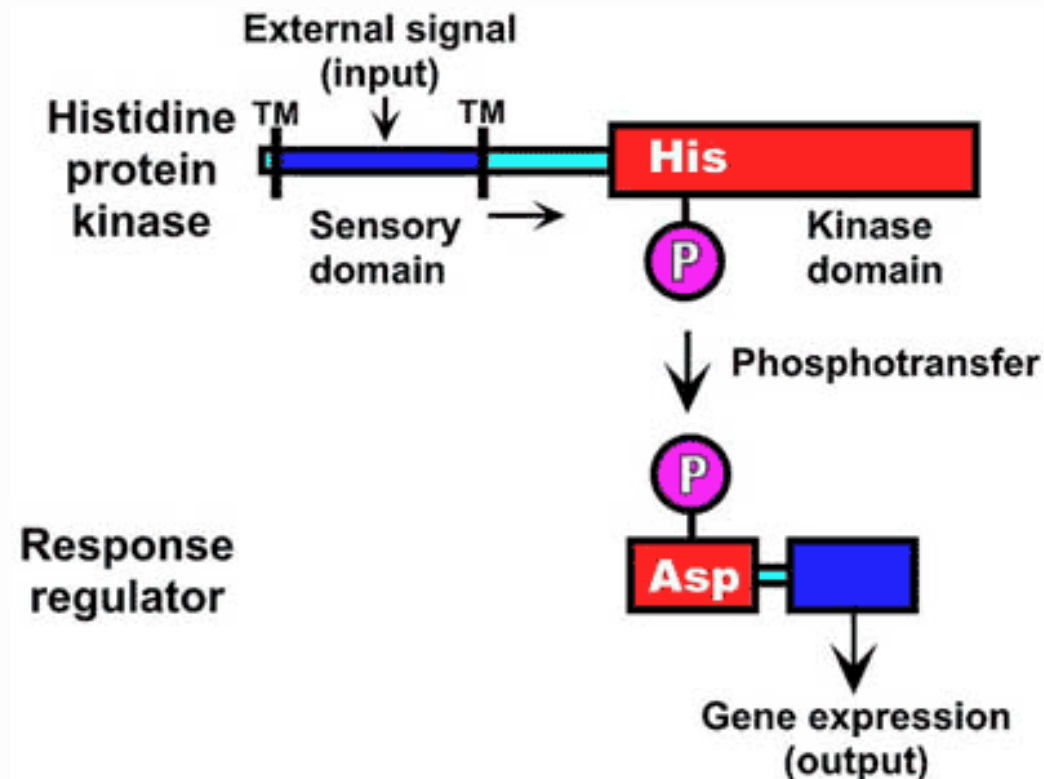
Graph representation of the Feedback

Feed-backs through metabolite-protein interactions
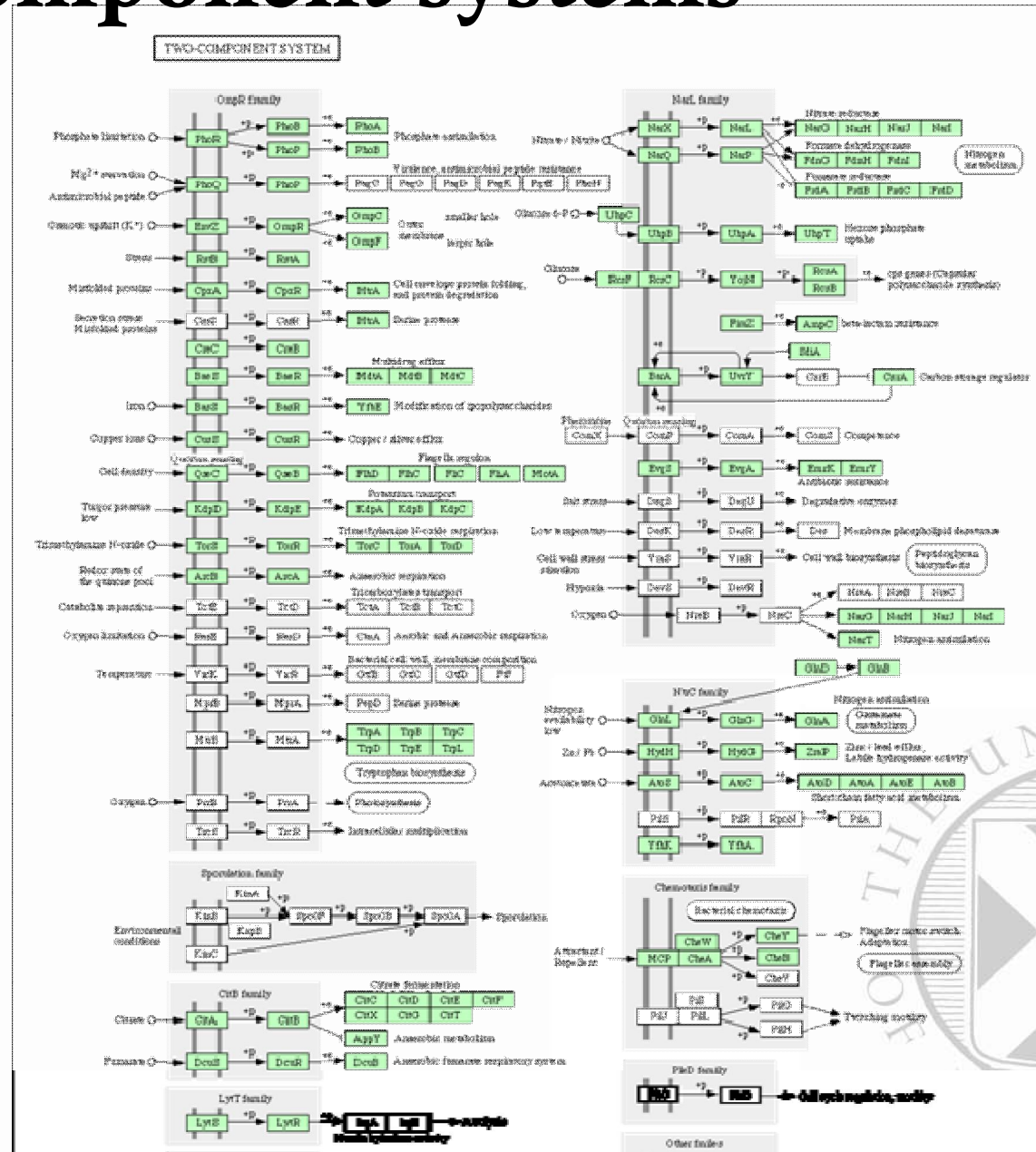
Other regulation systems

Two component system: Two proteins work together, the Kinase senses the environmental signal and phosphorylated the regulator. (the signal pathway in bacteria)
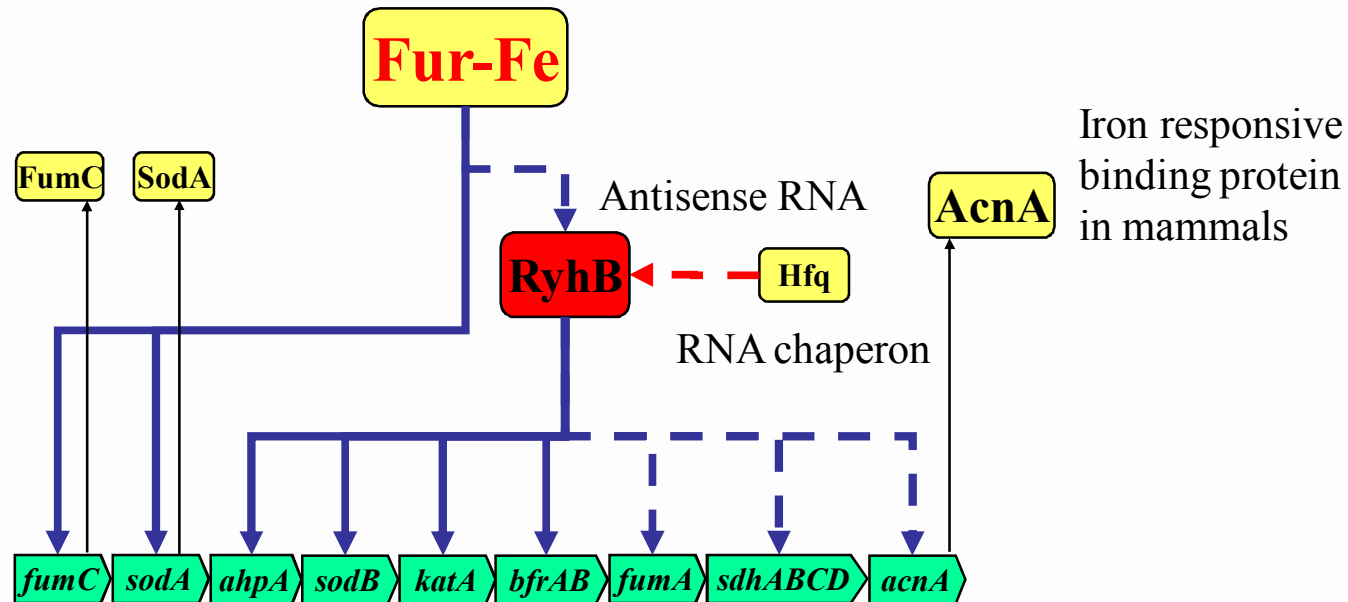
Two component systems

26 from KEGG, including information about the environmental signal

Altogether 32 TCSs obtained from databases and literature with known function.
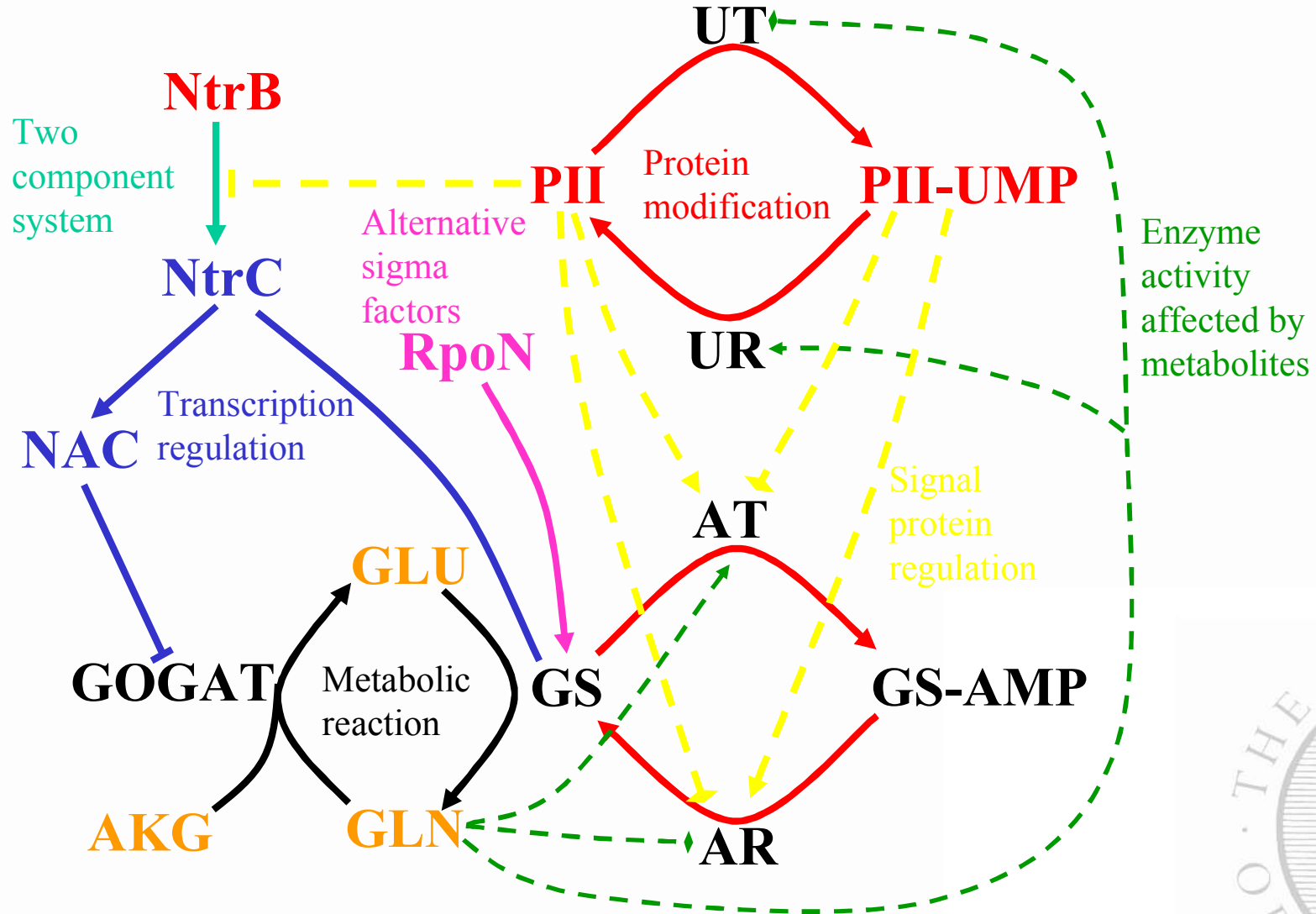
Small RNA mediated Fur regulation

Iron storage proteins and proteins require iron as cofactors are expressed at high iron concentration through an RyhB mediated positive regulatory mechanism in *E. coli*.
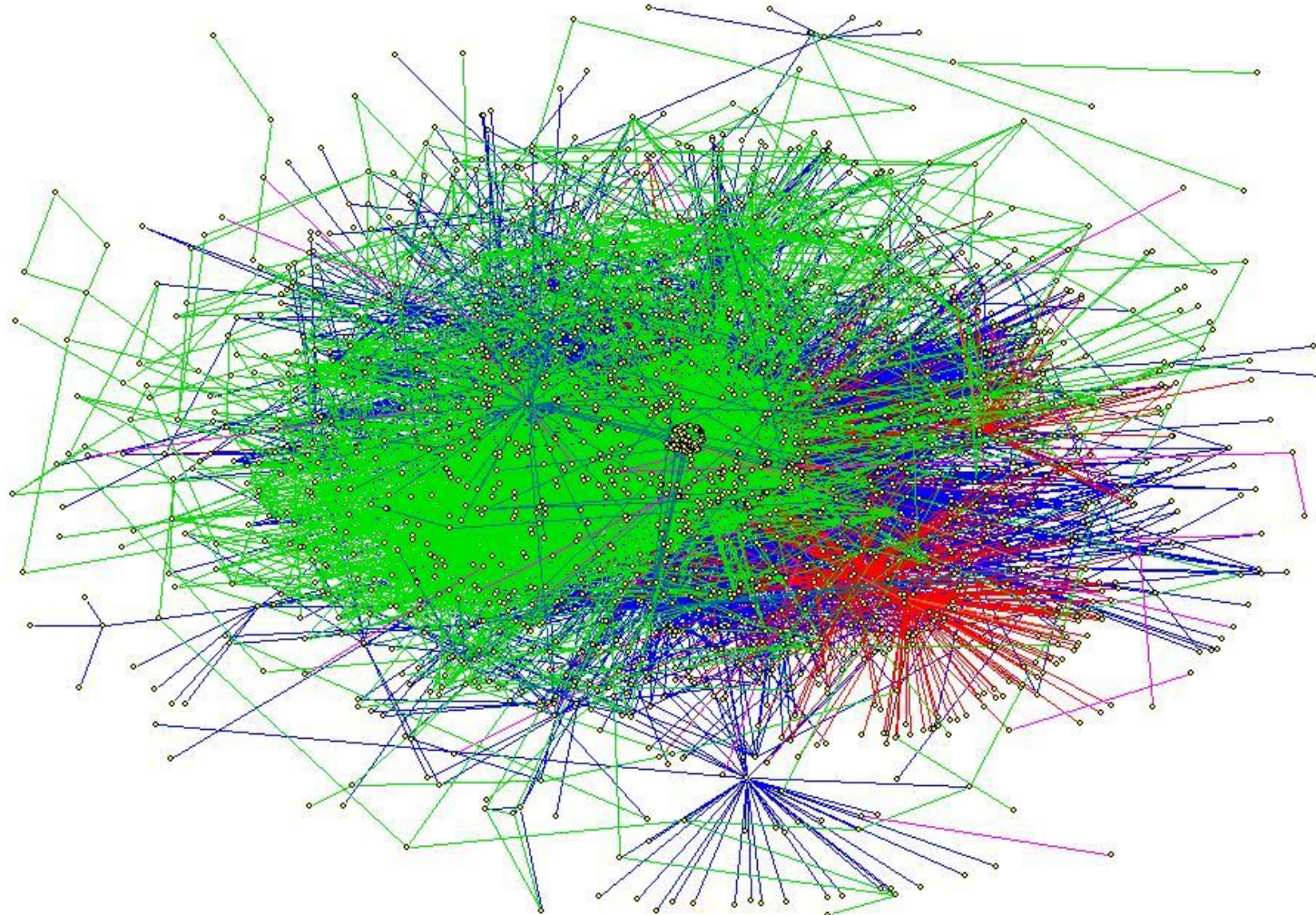
Bacteria version of RNA interference?

Regulation of ammonium assimilation

Regulation in Bacteria may be more complex than expected

E. Coli integrated network

2000 genes and 7000 interactions

Network motif analysis

Network motif: Subgraphs that appear in the network at frequencies much higher than those found in randomized networks (Shen-Oorr et al., Nature Genetics, 31:64).

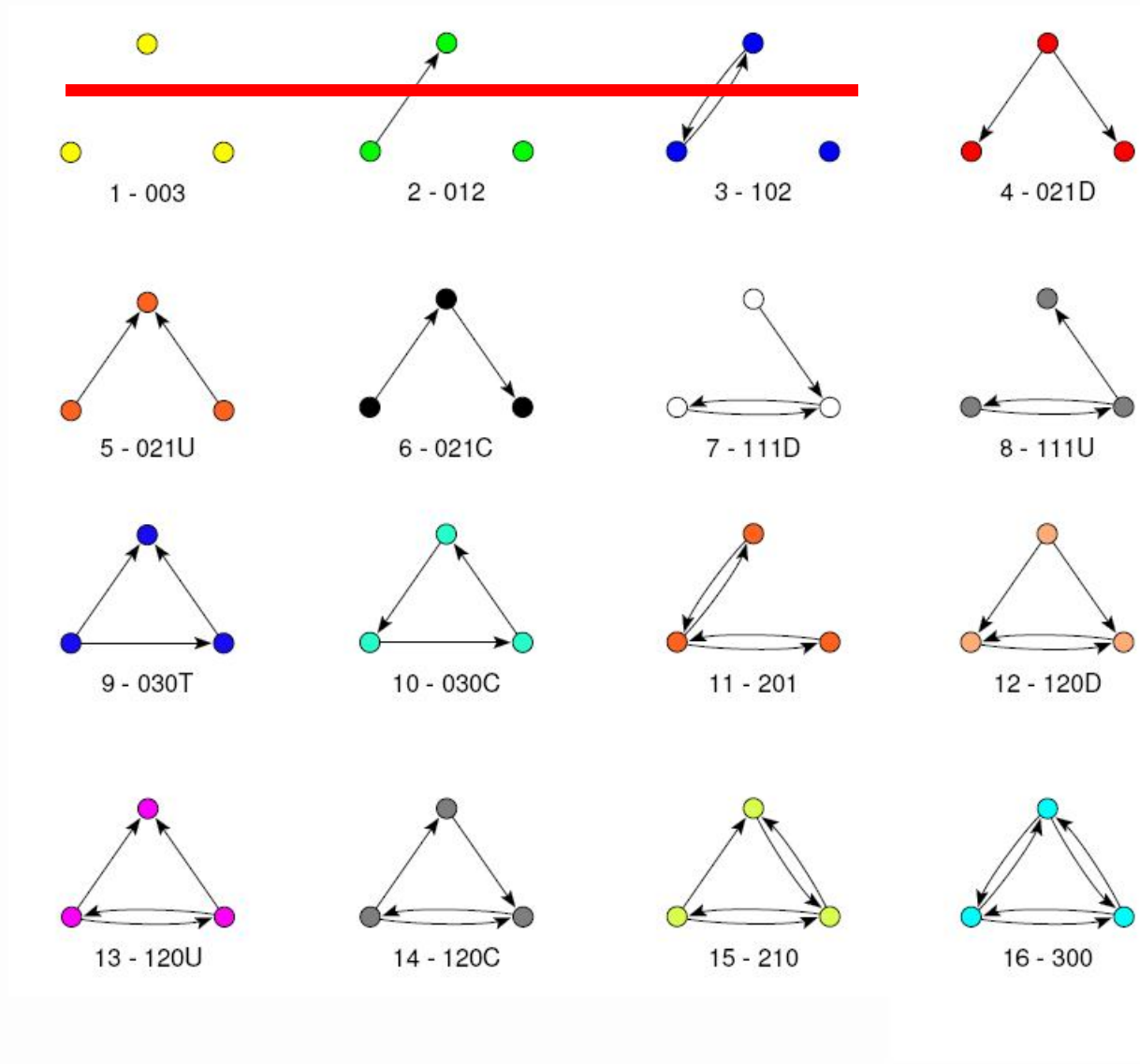 Regarded as elementary building blocks of network

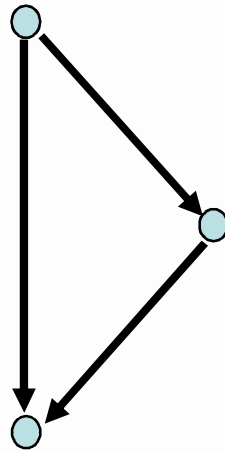Milo et al. *Network motifs: simple building blocks of complex networks.*
Science. 2002, 298(5594):824-7.
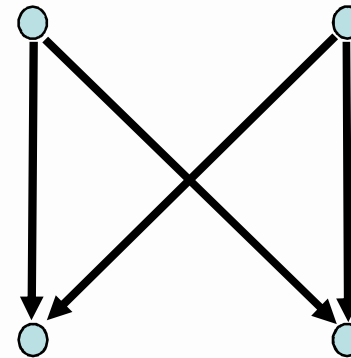
Three node subgraphs in the network

Network motifs in the regulatory network

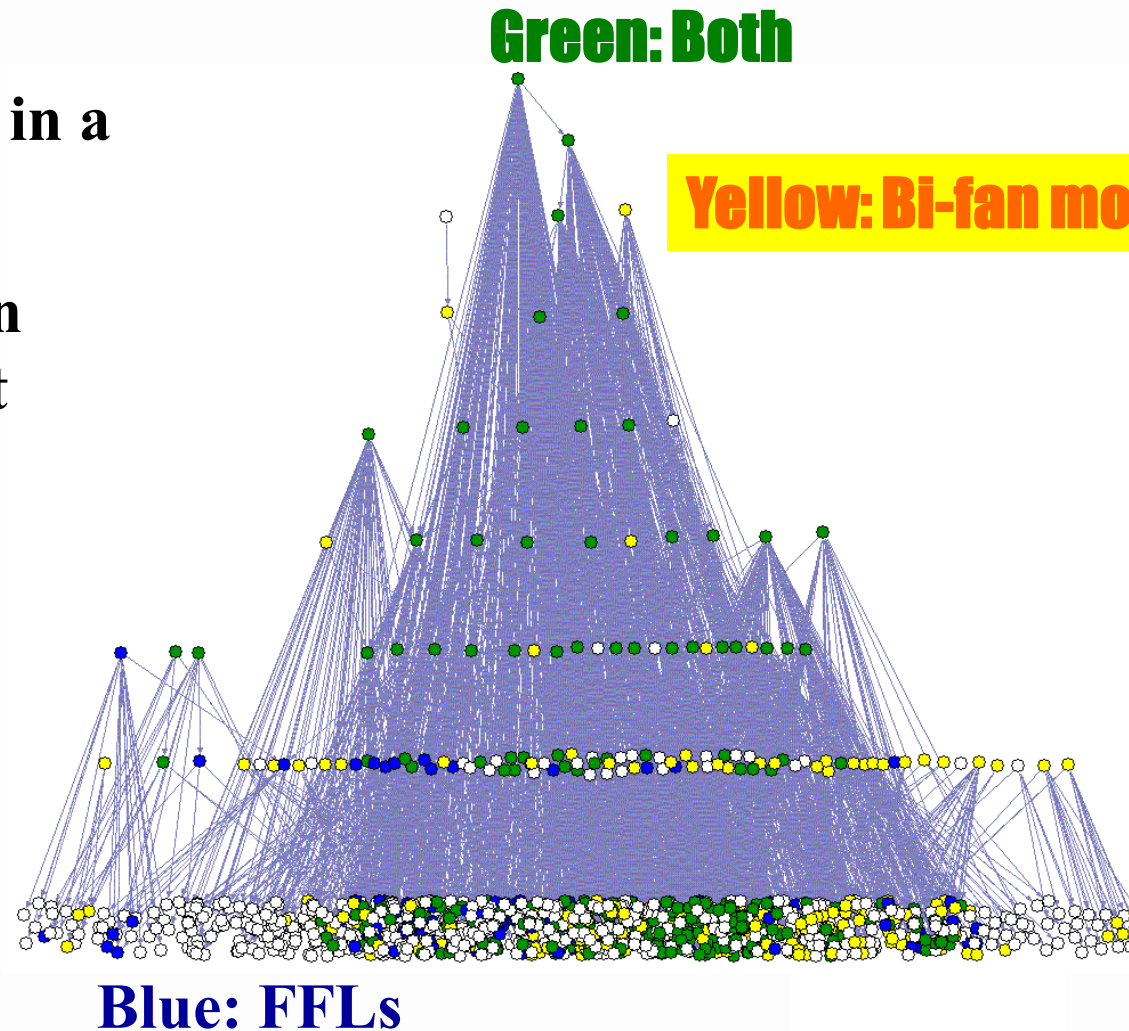**Feed forward loops**

712 in the network

Bi-fan motif

11935

Motifs in the network

Green: Both

All but 11 FFLs in a giant cluster

All but 10 Bi-fan motifs in a giant cluster.
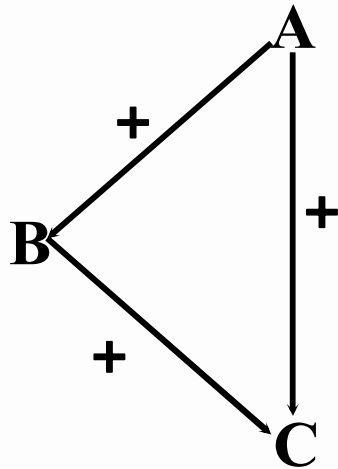
Yellow: Bi-fan motifs

Blue: FFLs

The multi-layer structure determined the FFL motifs!

Two kinds of FFLs

Mangan and Alon, (2003) PNAS, 100:11980

Coherent FFLs

Incoherent FFL

Direct effect of the upper regulator *A* on the target gene *C* is consistent with its indirect effect through *B*.
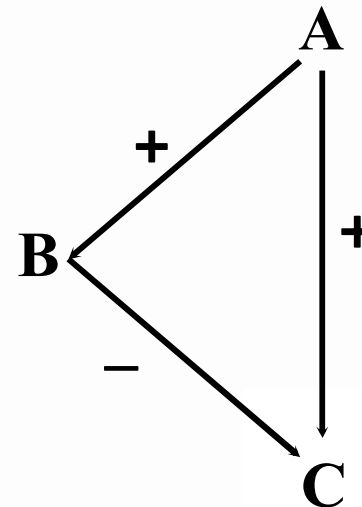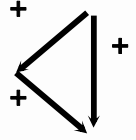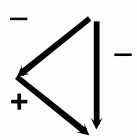
Direct effect is inconsistent with the indirect effect.
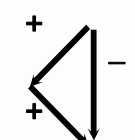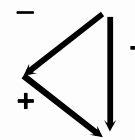
Motif distribution
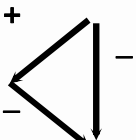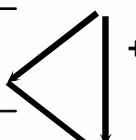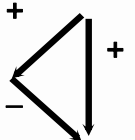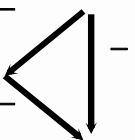
Coherent FFLs: 330   **Incoherent FFLs: 152**

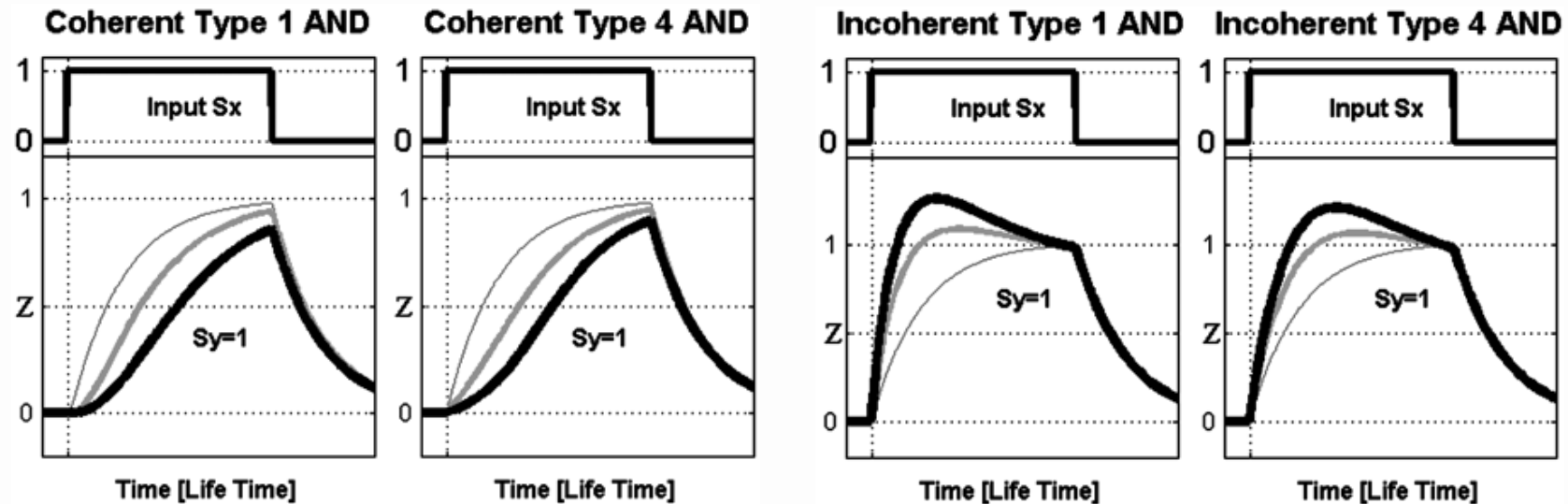| type								
number	**265**	28	7	30	**119**	9	8	16
examples	*flhC-fliA-fliGH*	*cpxR-csgD-csgA*	*fis-hns-cysG*	*fnr-narL-dcuB*	*crp-nagC-manXYZ*	*arcA-betI-betAB*	*ihf-flhD-nrfA*	*fnr-narL-moeAB*

Dominant motifs!

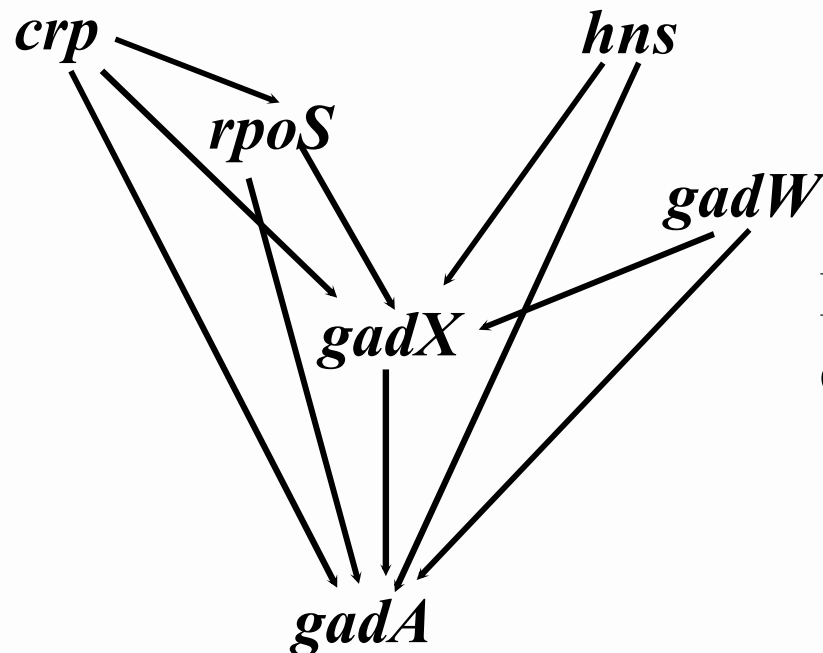# Function of network motifs

Mangan and Alon, (2003) PNAS, 100:11980

Delay the responses of the target gene so that it only response to a persistent signal. Robust to noise!

speed up the responses of the target gene

How when it is like this?

By both coherent FFLs and in coherent FFLs

Using kinetic model to study the behavior of the target gene