Chapter 8: The Topology of Biological Networks

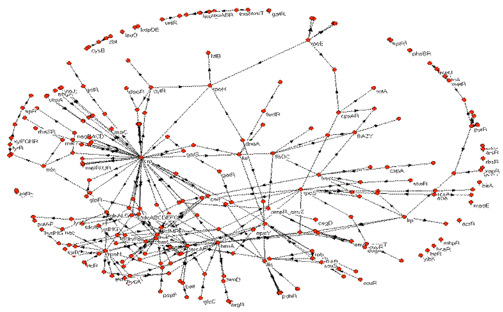
8.2 Network Motifs

Prof. Yechiam Yemini (YY)
Computer Science Department
Columbia University

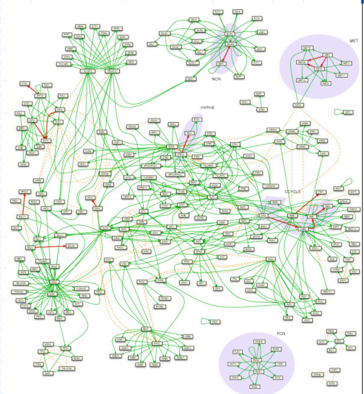
Overview

- This chapter is primarily based on the work of Alon's group
  - http://www.weizmann.ac.il/mcb/UriAlon/
  - The seminal publication:
    S Shen-Orr, R Milo, S Mangan & U Alon,
    "Network motifs in the transcriptional regulation network of Escherichia coli." Nature Genetics, 31:64-68 (2002). Pdf.
  - "An Introduction to Systems Biology/U. Alon; Chapman & Hall; 2007

Are There Underlying Organization Rules?



Regulatory Network of E.Coli



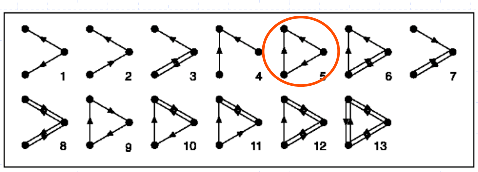
Regulatory Network of Yeast

Thieffry, Collado-Vides, 1998
Shen-Orr, Alon, Nature Genetics 2002

Mazurie et al. Genome Biology 2005 6:R35

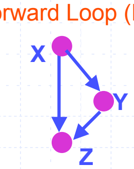
Consider Subgraphs With n Nodes

- n=1 ➔Self-loops and isolated nodes
- n=2 ➔An edge, or a loop of two nodes
- n=3 ➔Potentially 13 types of connected directed graphs



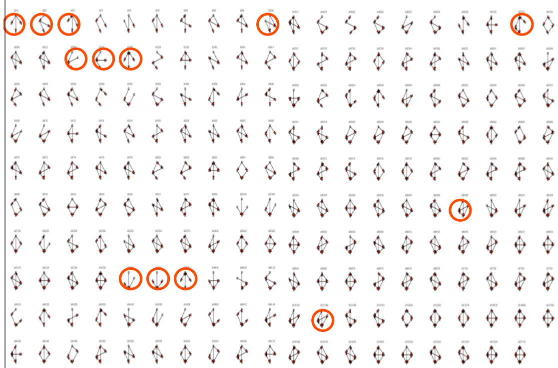
- Surprise: only 1 type shows in E.Coli/Yeast networks:

Feed Forward Loop (FFL)

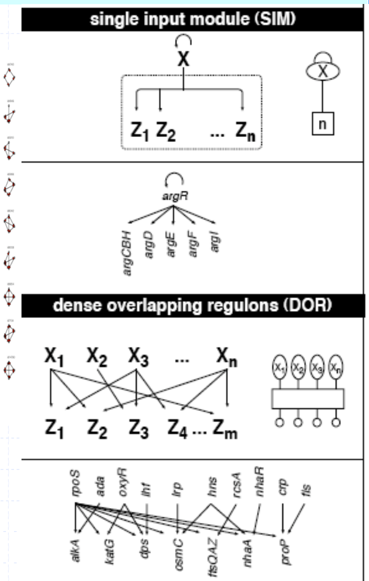


Two More Motifs For n=4

- n=4 ➔ 199 motif candidates

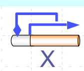


- n=5 ➔ 9364
- n=6 ➔ 1,530,843 motif candidates
  Enumeration is impractical

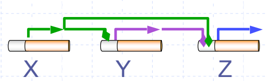


Regulatory Nets Use Motifs

n=1 ➔ Auto-regulation

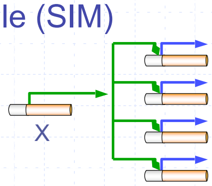


n=3 ➔ Feed-Forward-Loop (FFL)

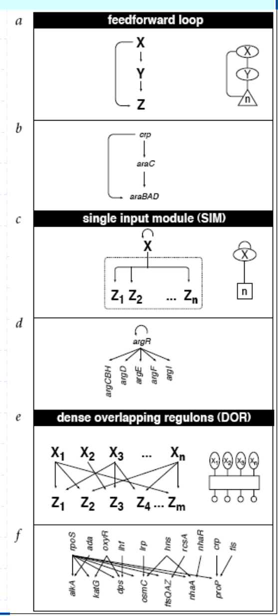


N≥4

➔Single-Input-Module (SIM)



➔Dense Overlapping Regulators (DOR)



Only A Small Number of Motifs Is Used

- n=3 ➔FFL; Coherent type 1 & incoherent type 1 dominate
- n=4 ➔ SIM or DOR

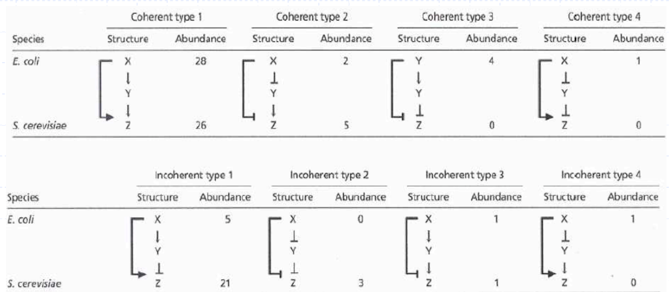


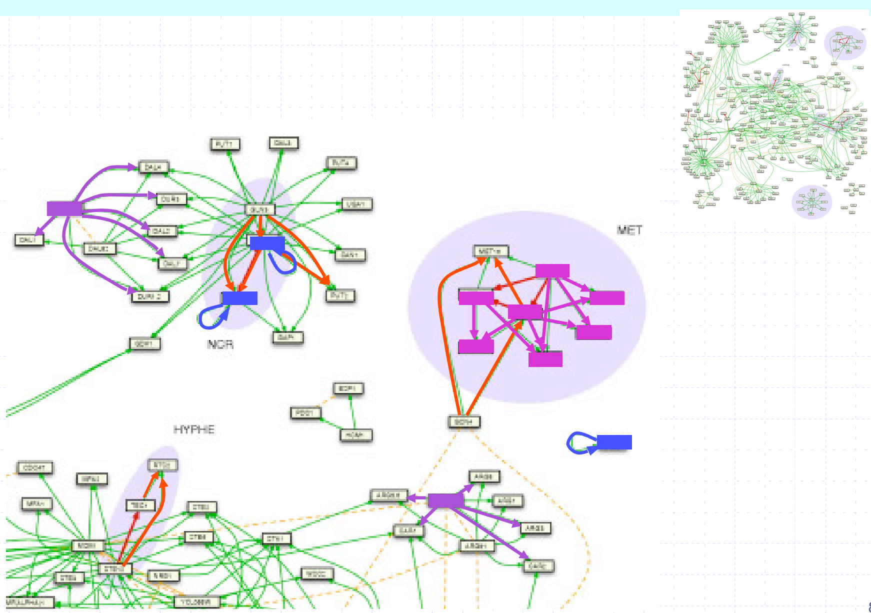
Table 1 • Statistics of occurrence of various structures in the real and randomized networks

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.

7

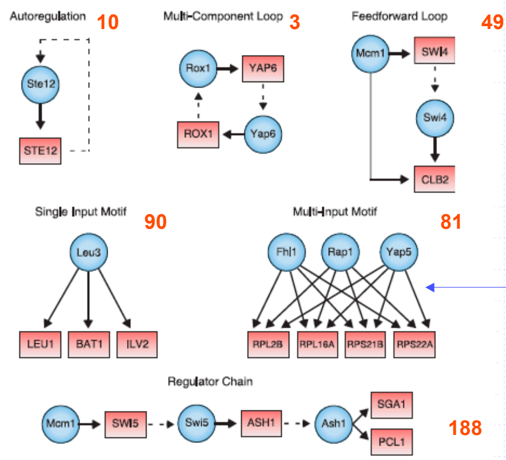
Example: The Yeast Regulatory Network



8

The Yeast Regulatory Network

Young *et. al*: Transcriptional Regulatory Networks in *Saccharomyces cerevisiae; Science* 2002



Numbers represent motif frequencies.

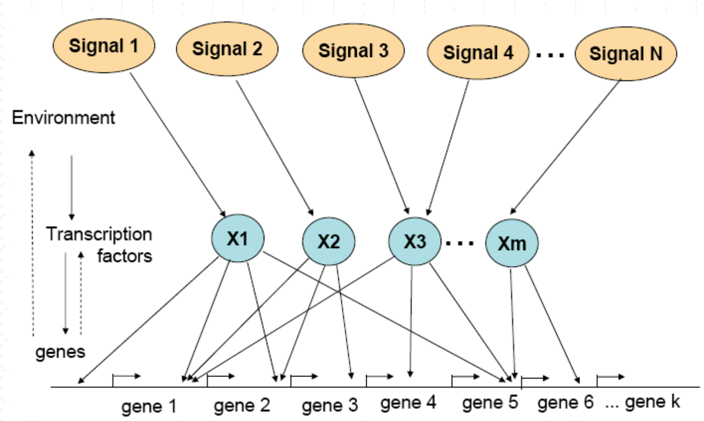
MIM=DOR.

Fig. 3. Examples of network motifs in the yeast regulatory network. Regulators are represented by blue circles; gene promoters are represented by red rectangles. Binding of a regulator to a promoter is indicated by a solid arrow. Genes encoding regulators are linked to their respective regulators by dashed arrows. For example, in the autoregulation motif, the Ste12 protein binds to the *STE12* gene, which is transcribed and translated into Ste12 protein. These network motifs were uncovered by searching binding data with various algorithms. For details on the algorithms used and a full list of motifs found, see (*18*).

9

How Are Motifs Used

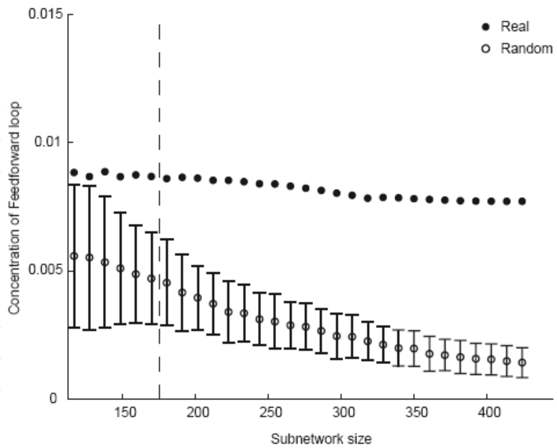
- Example:
  DOR can handle complex processing of related signals



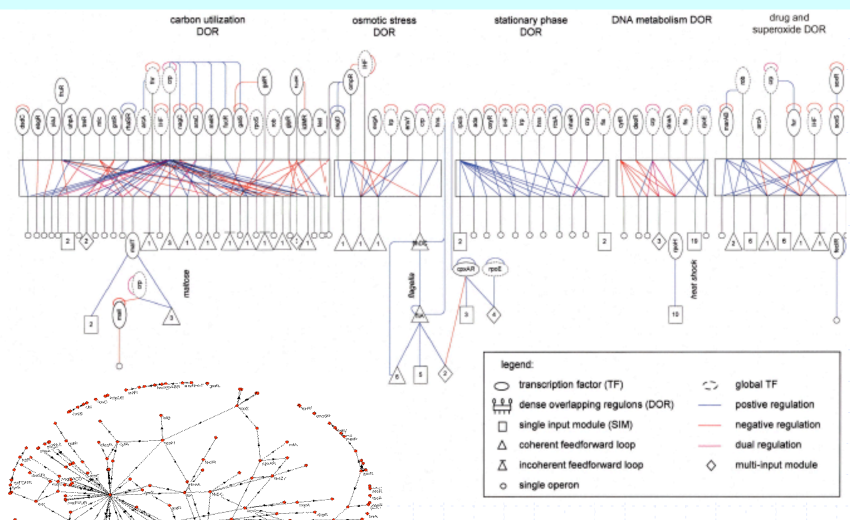
10

Motifs Exhibit Interesting Statistics

- Uniform concentration of FFL
  (Is there a scaling law?)

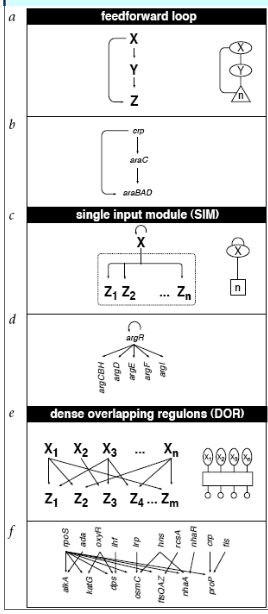


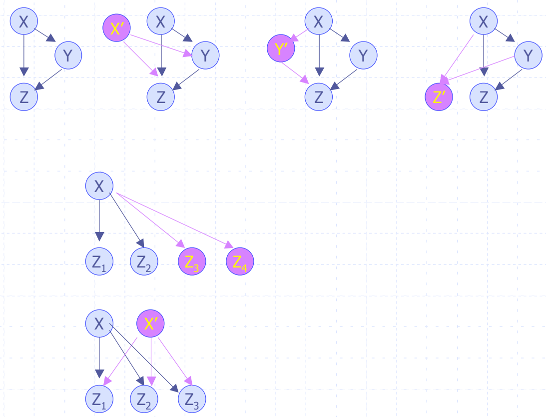
Motif Structure of E.Coli Regulation



http://sabio.villa-bosch.de/motif/

Gene Duplication Conserves Motifs





13

The Challenges

- How do we tell motifs from random sub-graphs?
- What do motifs do? What are they good for?
- How did motif arise? How do they evolve?

14

Discovering Network Motifs

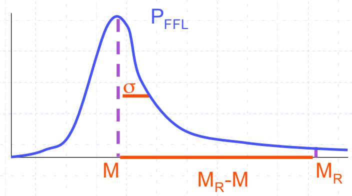
How Do We Tell A Motif?

- Motifs
  - Sequence motif: statistically significant set of homologous sub-sequences
  - Protein motif: statistically significant set of similar folds
- Net Motif=statistically significant set of isomorphic subnets
  - E.g., FFL, SIM, DOR….
  - But how do we decide “Statistically significant”?
  - Recall sequence motifs: compare motif against background statistics
  - Need to compare motif statistics against random graph
  - Which randomness: Erdos-Reneyi (ER)? Scale-free? Small-world? Other?

16

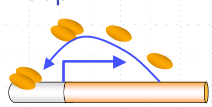
Finding Motifs in ER Random Graphs

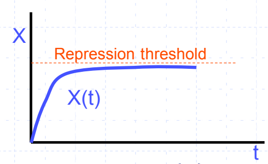
- Compare the observed network against a respective ER network
  - Let R=<N,E> be the observed network; N=#nodes, E=#edges
  - A comparison ER network is the random graph G(N,p) where $p=E/N^2$
- Given a motif, let P(m)= probability of m motif occurrences in G(N,p)
  - P(m) defines the statistics for finding the motif in an ER random network
  - Let M be the expected value of P and let σ be its standard deviation.
- Statistical significance can be evaluated by standard Z-score or p-value
  - $Z=(M_R-M)/\sigma$
  - M_R is the # of occurrences of the motif in the observed network R
- Computational challenges
  - Given a motif, how to compute M,σ and MR?
  - Given a network, how do we discover motifs?



Auto-regulation is A Motif

- Auto-regulation= self-loop
  - Negative feedback

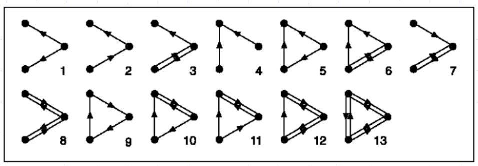




- P(m)= probability of m self loops in G(N,p): $P(m)=B(m,p)=\binom{N}{m} p^m(1-p)^{N-m}$
  - Expected # of self loops $=pN=(E/N^2)N=E/N$
  - Standard deviation $\sigma=\sqrt{E/N}$
- For E.coli N=424, E=519
  - A random graph would have E/N~1.2 self loop and σ~1.1
  - But E.coli has 40 self-loops
- The Z-score: Z=(40-1.2)/1.1~35
  - Conclusion: Self-loop is a motif

Discovering Small Motifs

- Represent the network as an adjacency matrix A
  - A(i,j)= { 1 if gene j activates gene i; -1 if gene j represses gene I; 0 otherwise }
- Scan A for all nxn sub-matrices
  - Count motif frequencies
- E.g., for n=3 there are 13 possible motifs
  - Motifs = non-isomorphic directed graphs on 3 nodes
  - Exhaustive search is useful only for small motifs



19

Computing p-Value

- Two challenges:
  - How to generate "good" random networks
  - How to compute motif frequencies for each motif
- How to generate comparison random networks?
  - Key idea: use the real network R to provide background statistics
  - Randomly switch edges of R
  - Preserve the # of subnets of size 3,4,…n-1 (when searching motifs of size n)
  - (Generalizing ER; ER considers only n=2)
  - Variants: use Metropolis (Gibbs) sampling to switch edges (Switch edges with temperature-dependent probability exp(-E/T))

20

Probabilistic Algorithm For Motif Finding

- Challenge: how to reduce complexity
- Key-idea: sample the network to detect motif frequency

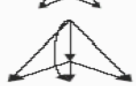
Subgraph Sampling Algorithm:

1. Initialize: start an n-subgraph by selecting a random edge
2. Iterate:
   select a random edge connecting subgraph to a new node
   add new node, until subgraph has n nodes.
3. Repeat 1-2 to collect a set of n-subgraphs
4. Compute weighted concentration of distinct n-subgraphs

Kashtan *et al.:* "Efficient sampling algorithm for estimating subgraph concentrations and detecting network motifs"; *Bioinformatics* 2004.

Surprise: Discovering Motifs With A Few Samples

Comparison with exhaustive search

Shape	Full enumeration Appearances (Z-score)	Concentration ($\times 10^{-3}$)	Sampling method Concentration ($\times 10^{-3}$) (Z-score)
	42 ($z = 10$)	8.07	8.69 ($z = 10$)
	209 ($z = 9$)	2.49	2.69 ($z = 8$)
	51 ($z = 15$)	0.61	0.65 ($z = 15$)
	54 ($z = 120$)	0.038	0.035 ($z = 30$)
	271 ($z = 16$)	0.189	0.196 ($z = 11$)
	20 ($z = 18$)	0.014	0.013 ($z = 8$)
	18 ($z = 12$)	0.013	0.014 ($z = 8$)

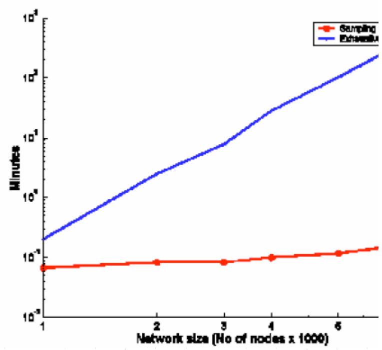
1K (~5K total three-node subgraphs)

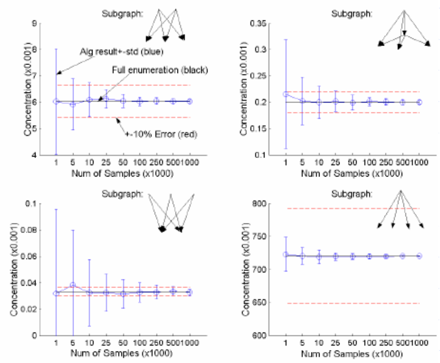
10K (~85K total four-node subgraphs)

50 K (~1.4M total five-node subgraphs)

High-Speed Motif Finder

- Runtime is almost independent of net size
- Rapid convergence to real concentration
- Apply to discover larger motifs



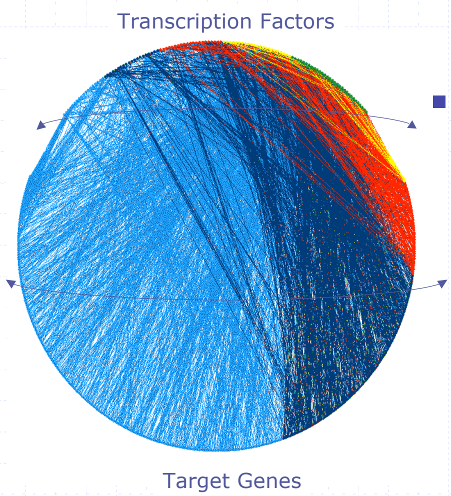


23

Yeast Regulatory Network Motifs & Functions

Luscombe NM, Babu MM, Yu H, Snyder M, Teichmann SA & Gerstein M (2004)
Genomic analysis of regulatory network dynamics
reveals large topological changes.
Nature 431: 308-312.

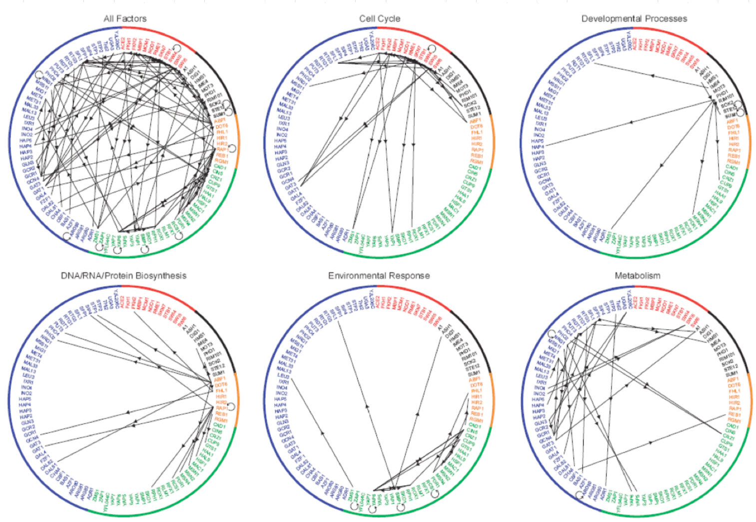
Comprehensive Dataset Available



- Very complex network
  - 3420 genes, 142 TFs
  - 7074 regulatory interactions

Yeast Regulatory Network Motifs

Lee et al, Science 2002

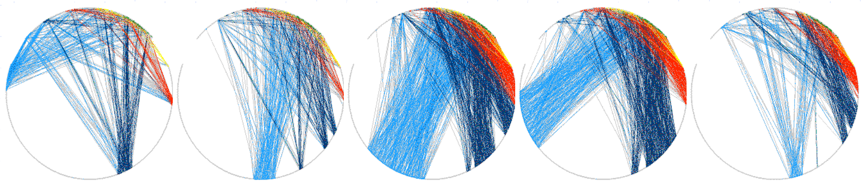


Cell Cycle | Developmental | Biosynthesis DNA/RNA/Prot | Environment | Metabolism

Activity Subnets

Luscombe NM, Babu MM, Yu H, Snyder M, Teichmann SA & Gerstein M (2004) Genomic analysis of regulatory network dynamics reveals large topological changes. Nature 431: 308-312.

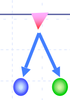
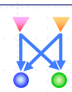
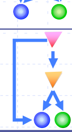
Cell cycle **Sporulation** **Diauxic shift** **DNA damage** **Stress**



Multi-stage activities

Binary state

Motifs Statistics Depend On The Task

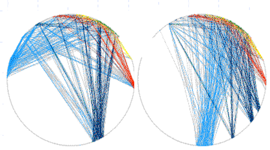
Motifs		Cell cycle	Sporulation	Diauxic shift	DNA damage	Stress response
SIM		32.0%	38.9%	57.4%	55.7%	59.1%
MIM		23.7%	16.6%	23.6%	27.3%	20.2%
FFL		44.3%	44.5%	19.0%	17.0%	20.7%

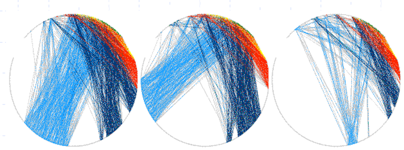
Observations

network motif	description	example
SIM *MIM*	Simultaneous regulation of multiple genes such as those involved in the same pathway or macromolecular complex. They appear suited for controlling large-scale turnover of genes observed in exogenous conditions.	**DNA damage.** Rpn2 regulates three proteosomal subunits Rpt2, Rpt4, and Rpt6.
FFL	Regulatory buffer that respond only to persistent input signals from the primary TF, and allows for rapid shutdown when signal ceases. It appears suited for endogenous conditions as cells will only enter a new phase once the regulatory signal from the previous one has stabilised. The signal will also terminate quickly once the cell has entered a new phase.	**Sporulation.** Rim1 acts as the primary and Ime1 as the secondary TF to regulate Ime2 in the early phase. Ime2 is a kinase that stimulates about 20 further TFs in the middle and late phases; it ensures a quick shutdown of the regulatory cascade through phosphorylation of Ime1.

29

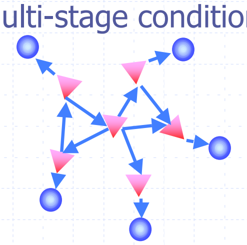
Architectural Rationale

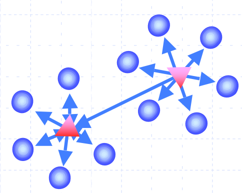




multi-stage conditions

binary conditions





- fewer target genes
- longer path lengths
- more inter-regulation between TFs

- more target genes
- shorter path lengths
- less inter-regulation between TFs

30

Conclusions

- Motifs are fundamental units of regulation
- Gene duplication conserves motifs
- Motifs have respective functional roles (considered in the next section)

31