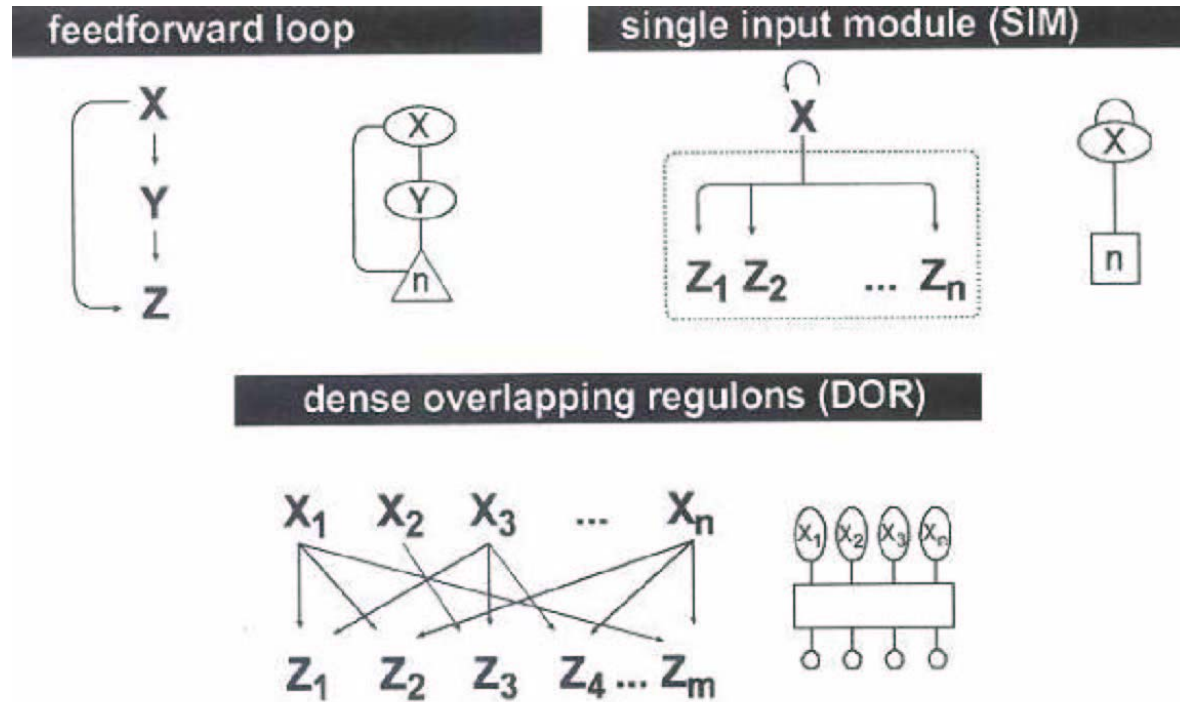# 第8-4章: Network Motifs

- Milo, R., et. al. *Science* **298, 824–827, 2002**.

- Milo, R., et al. *Science* **303, 1538-1542, 2004**.

- Rui Jiang, Zhidong Tu, Ting Chen, and Fengzhu Sun. *Proc. Natl. Acad. Sci. USA,* **103, 9404-9409, 2006.**

Slides for this part are mainly from Fengzhu Sun's talk.

Network Motif

- Definition: Patterns of interconnections occurring in complex networks at numbers that are significantly higher than those in randomized networks (Milo, R., et. al. *Science* **298,** 824–827)
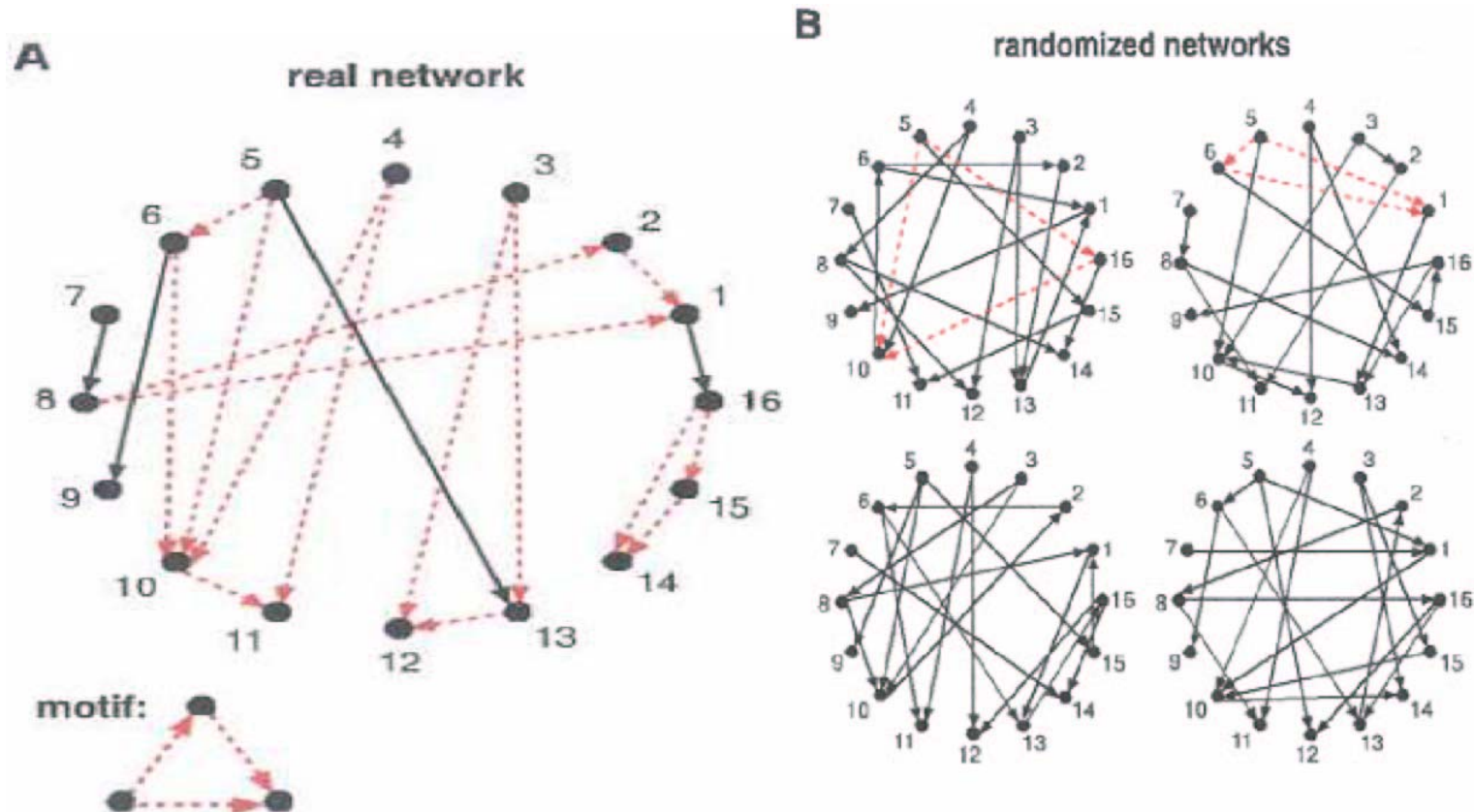
Network Motifs

Existing Methods for Network Motif Identification

- Consider different types of subgraphs separately
- For each type of subgraph
  - Count the number of occurrence of the subgraph in the observed network
  - Shuffle the observed network to obtain a set of randomized networks. Estimate the expected number of occurrence of the same type of subgraph in the randomized networks
  - Calculate a statistic corresponding to the subgraph
- Over-abundant subgraphs are network motifs

Schematic Illustration

Red dashed line indicate edges that participate in the feedforward loop motif, which occur five times in the real network.

Randomized Network

- It employed a Markov-chain algorithm, based on starting with the real network and repeatedly swapping randomly chosen pairs of connections (X1->Y1, X2 ->Y2 is replaced by X_1->Y_2, X_2->Y_1) until the network is well randomized.
- Switching is prohibited if the either of the connections X_1->Y_2 or X_2->Y_1 already exist.
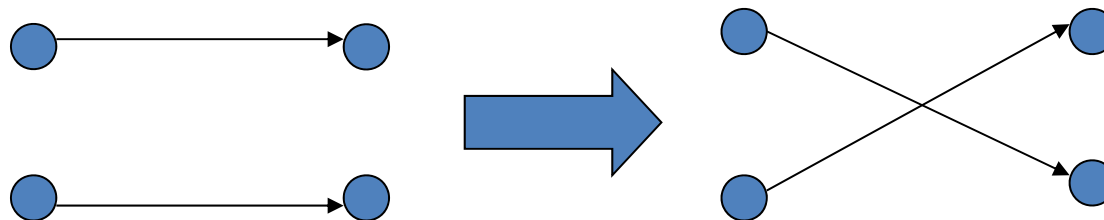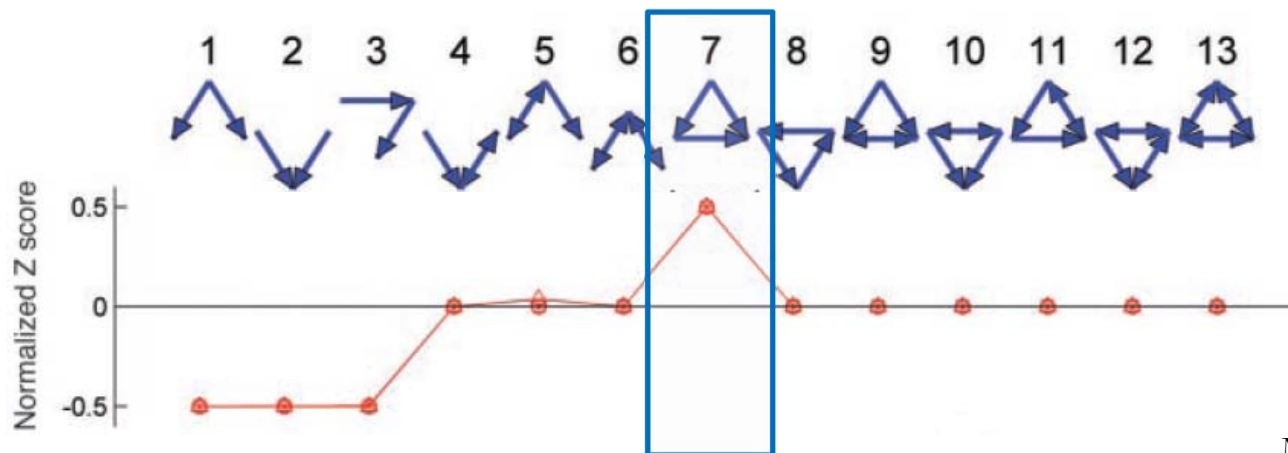
Network Motif Identification in Deterministic Networks

- Hypothesis testing
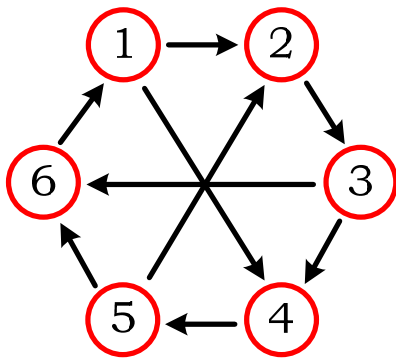
A	1	2	3	4	5	6
1	0	1	0	1	0	0
2	0	0	1	0	0	0
3	0	0	0	1	0	1
4	0	0	0	0	1	0
5	0	1	0	0	0	1
6	1	0	0	0	0	0

时间复杂度

$$mN^n$$

Milo et al. *Science,* **303**, 2004

Network Motif Identification in Stochastic Networks
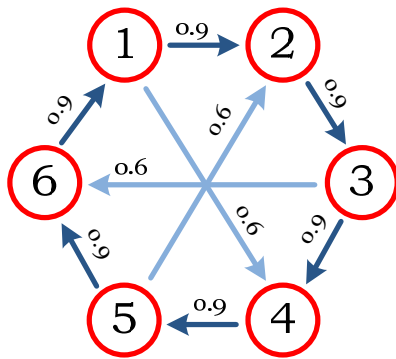
P	1	2	3	4	5	6
1	0	.9	0	.6	0	0
2	0	0	.9	0	0	0
3	0	0	0	.9	0	.6
4	0	0	0	0	.9	0
5	0	.6	0	0	0	.9
6	.9	0	0	0	0	0

Θ_1	1	2	3
1	θ_{11}	θ_{12}	θ_{13}
2	θ_{21}	θ_{22}	θ_{23}
3	θ_{31}	θ_{32}	θ_{33}

时间复杂度

$$lmN^n$$
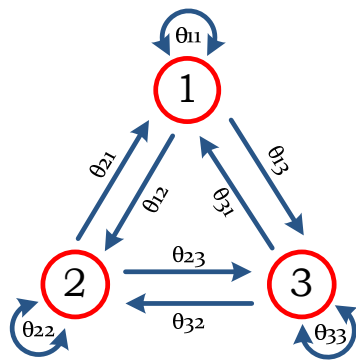
- Mixture model
- EM algorithm
- Likelihood ratio test

Jiang et al. *Proc Natl Acad Sci USA,* **103**, 2006

Subgraph

- Isomorphic structures of a subgraph
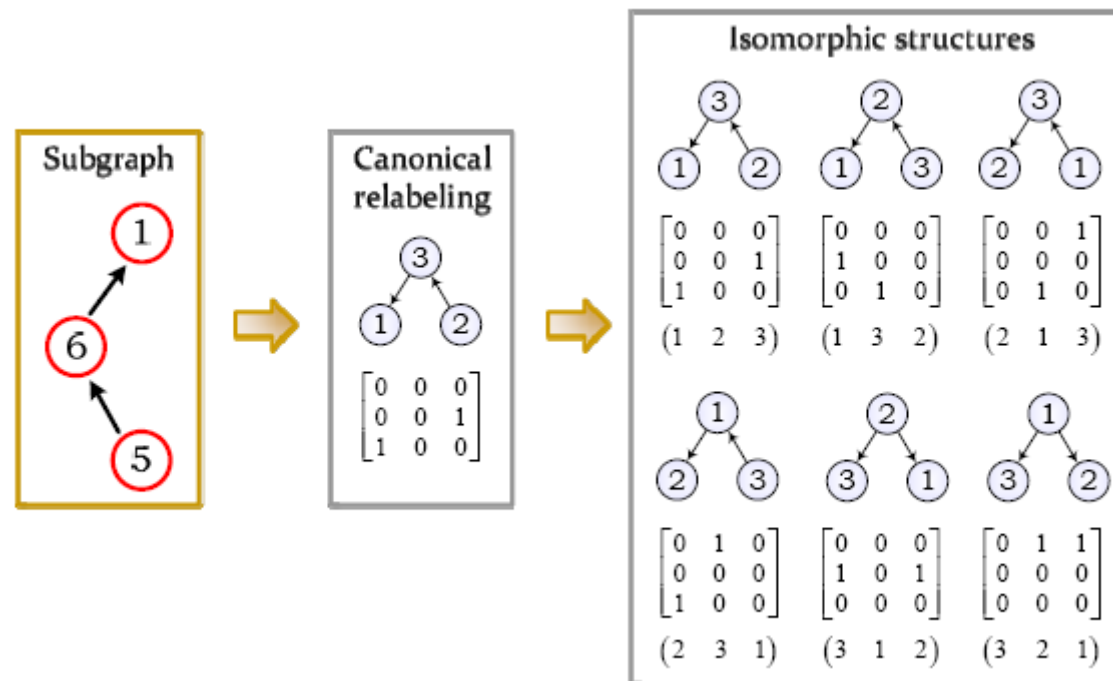
Probabilistic Network Motifs
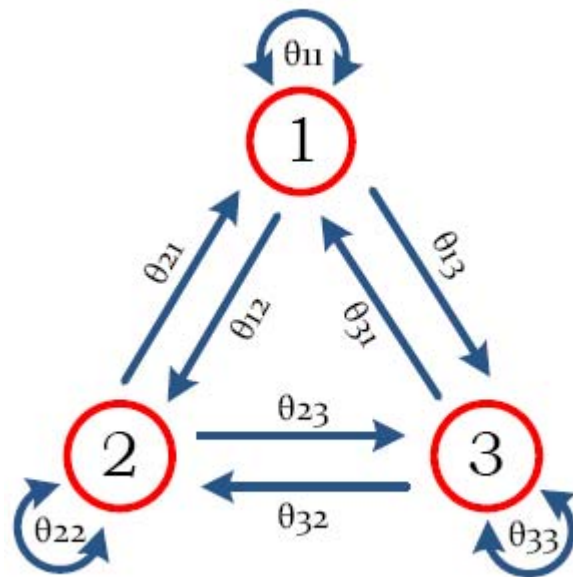
Visual presentation

$$\Theta_1 = (\theta_{ij})_{n \times n}, \ 0 \le \theta_{ij} \le 1$$

Matrix presentation

Probabilistic Network Motifs

- How to match an isomorphic structure of a subgraph to a probabilistic network motif?

X	1	2	3
1	x_{11}	x_{12}	x_{13}
2	x_{21}	x_{23}	x_{23}
3	x_{31}	x_{32}	x_{33}

Θ_1	1	2	3
1	θ_{11}	θ_{12}	θ_{13}
2	θ_{21}	θ_{22}	θ_{23}
3	θ_{31}	θ_{32}	θ_{33}

$$\Pr(\mathcal{P}|\Theta_1) = \prod_{i=1}^{n} \prod_{j=1}^{n} \theta_{ij}^{x_{ij}}(1 - \theta_{ij})^{1-x_{ij}}$$

Isomorphic Structure

- A permutation of the canonical labels corresponding to a certain isomorphic structure.
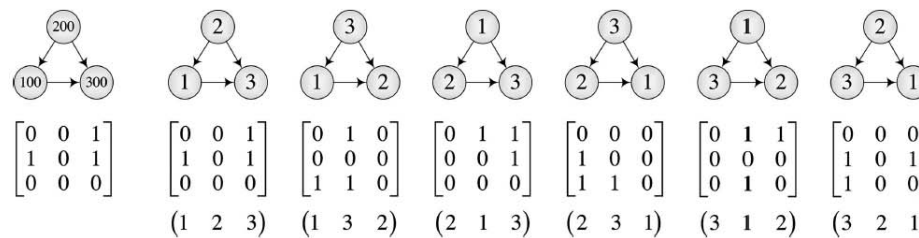

Fig. 3. A subgraph and its isomorphic structures. (*Left*) A 3-node subgraph labeled by (100, 200, 300) in a certain graph. The adjacency matrix is obtained by relabeling nodes as (1, 2, 3), respectively. (*Right*) Isomorphic structures of the subgraph. The isomorphic structures (top) have the same connectivity as the subgraph but different adjacency matrices (middle), which is generated by permuting the node labels (indices of the matrices), as shown at the bottom.

Probabilistic Network Motifs

- How to match a subgraph to a network motif?

$$\Pr(\mathcal{S}|\Theta_1) = \sum_{p=1}^{P_w} \prod_{i=1}^{n} \prod_{j=1}^{n} \theta_{ij}^{x_{ij}^p} (1-\theta_{ij})^{1-x_{ij}^p}$$
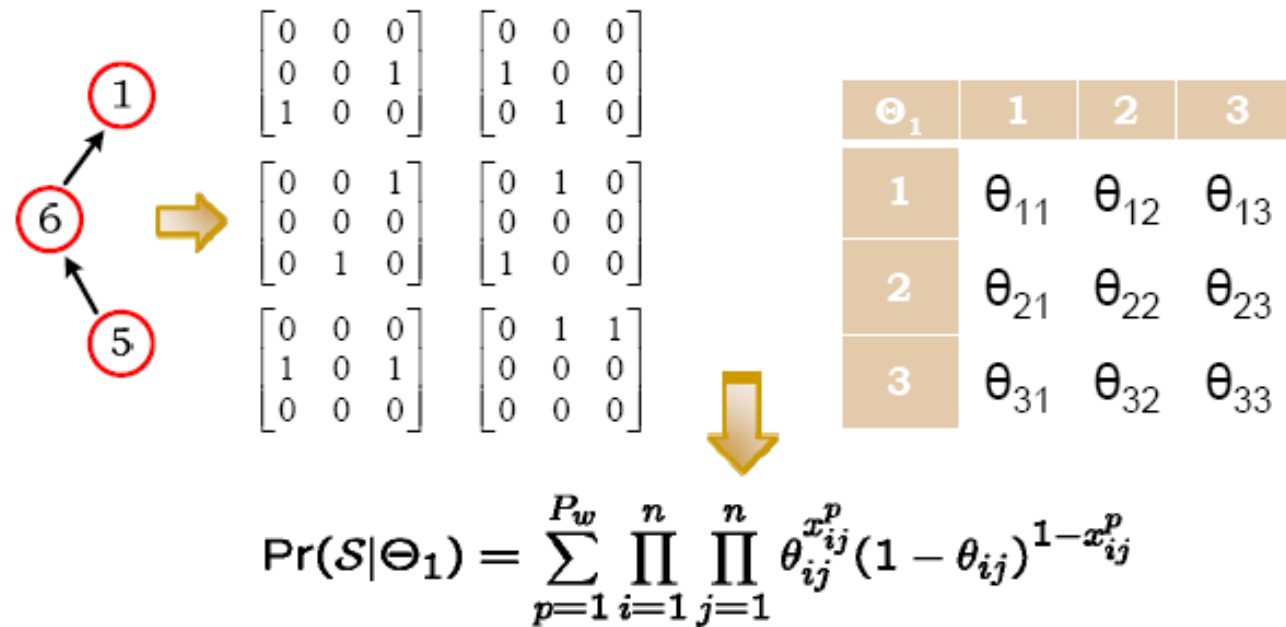
Where Pw is the number of isomorphic structures.

A Mixture Model for Network Motif Identification

- A stochastic network = Foreground probabilistic network motif (λ) +Background random ensemble ($1-\lambda$)

- Network motif identification procedure
  - Sample from the stochastic network to obtain a set of subgraphs — Observation
  - For each subgraph, determine whether it comes from the foreground or the background — Missing
  - For each subgraph, determine a proper isomorphic structure to derive the foreground motif pattern or to match the background random ensemble — Missing

Mathematical Formulae

- Observation

- Missing data 1

- Missing data 2

- Parameters

$$\mathbf{X} = \{\mathbf{X}_p^w\}$$

$$\mathbf{Z} = \{Z_h^w\}$$

$$\mathbf{Y} = \{Y_{hp}^w\}$$

$$\Theta = \{\lambda, \Theta_0, \Theta_1\}$$

- Pseudo-likelihood function

$$
\begin{aligned}
L(\Theta|\mathbf{X}, \mathbf{Y}, \mathbf{Z}) &= \Pr(\mathbf{X}, \mathbf{Y}, \mathbf{Z}|\Theta) \\
&= \prod_{w=1}^{W} \prod_{h=0}^{1} [\lambda_h \prod_{p=1}^{P_w} \Pr(\mathbf{X}_p^w|\Theta_h)^{Y_{hp}^w}]^{Z_h^w}
\end{aligned}
$$

- Maximize L to estimate parameters

An EM Algorithm

- The parameters associated with the background ($\Theta 0$) can be estimated offline in advance

- E-step: Calculate the expectation of the log pseudo-likelihood, conditional on the observation **X and the current estimation** of the parameters λ and $\Theta 1$

- M-step: Maximize the expectation of the log pseudo-likelihood to update the estimation of the parameters λ and $\Theta 1$

- Iterate over the E-step and M-step

Statistical Significance of the Identified Network Motifs

- A hypothesis testing problem

  H_0: $\lambda = 0$ versus H_1: $\lambda > 0$

- A pseudo-likelihood ratio statistic

  LR=-2log(L_0/L_1)

- A simulation method
  - Calculate LR* for the given stochastic network
  - Simulate a number of randomized networks, apply the EM algorithm, calculate LRs
  - Count how many LR from the randomized networks are greater than LR* from the given stochastic network

A Novel Stochastic Bi-fan Motif
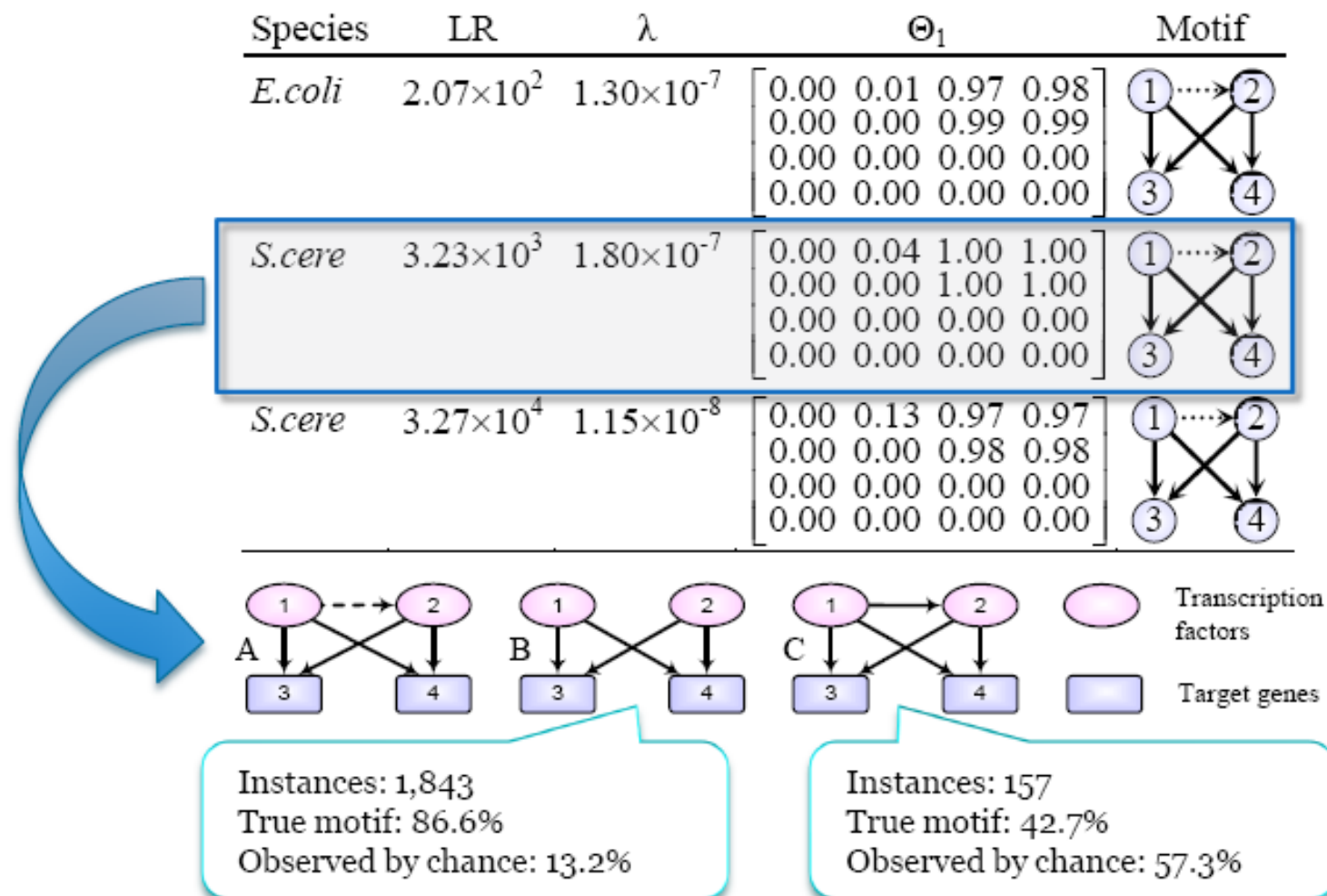
The high-osmolarityglycerol (HOG) Stress Response Pathway
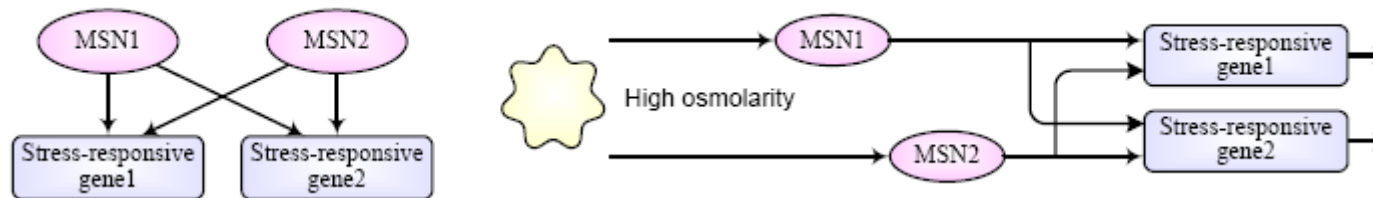
The "backup" combinatorial transcriptional regulation

- MSN2 (with MSN4) can activate many high-osmolarity stress responsive (HOSR) genes
- MSN1 can also activate the HOSR genes when MSN2 is not present
- MSN1 acts as a "backup" of MSN2 in the activation of HOSR genes

– Martinez-Pastor, M. T., et al. (1996) *The EMBO Journal* **15**, 2227–2235.
– Martijn, R., et al. (1999) *Molecular and Cellular Biology* **19**, 5474–5485.