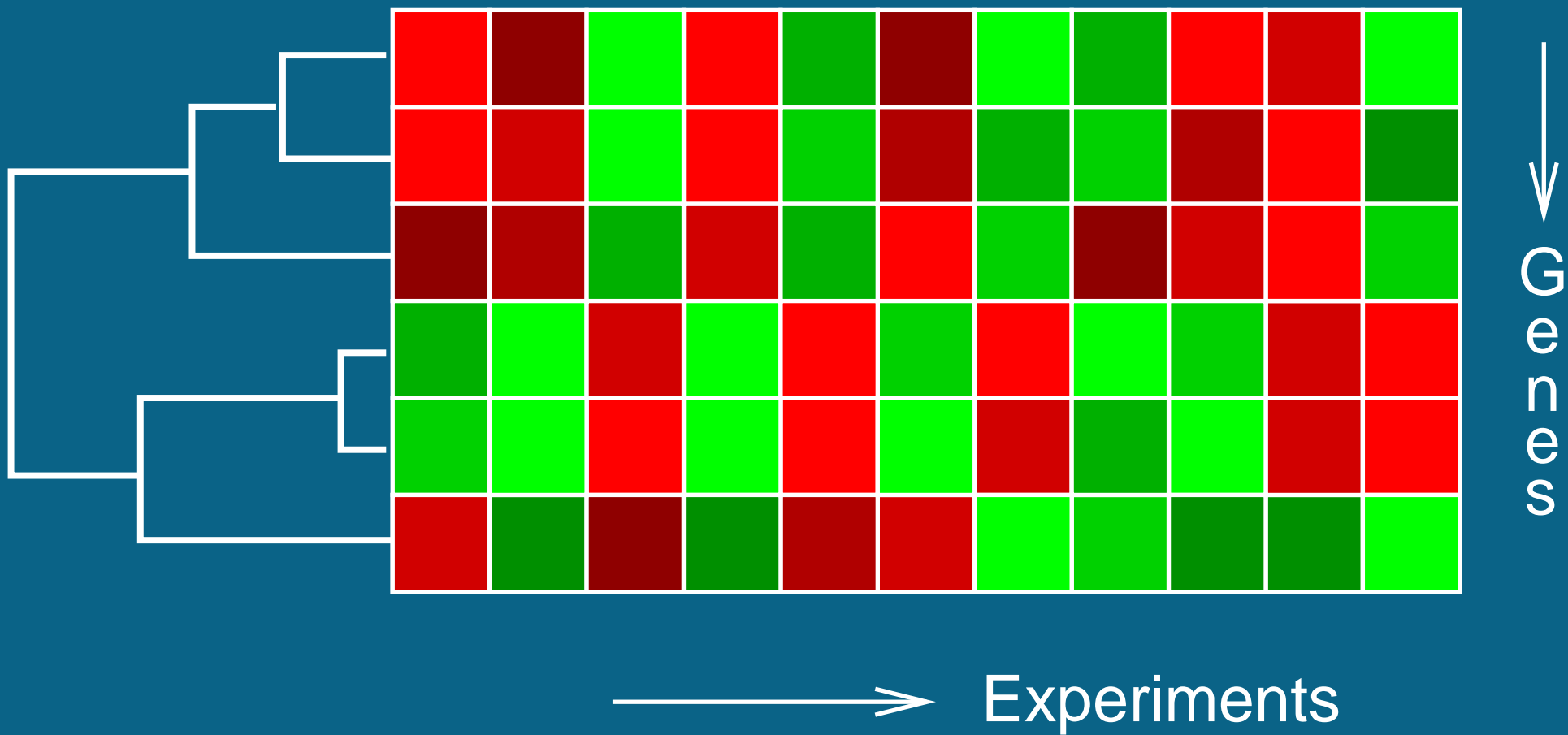

Modelling gene expression data with Bayesian networks

Dirk Husmeier

Biomathematics & Statistics Scotland (BioSS)
JCMB, The King's Buildings, Edinburgh EH9 3JZ
United Kingdom

<http://www.bioss.ac.uk/~dirk>



Advantage of clustering

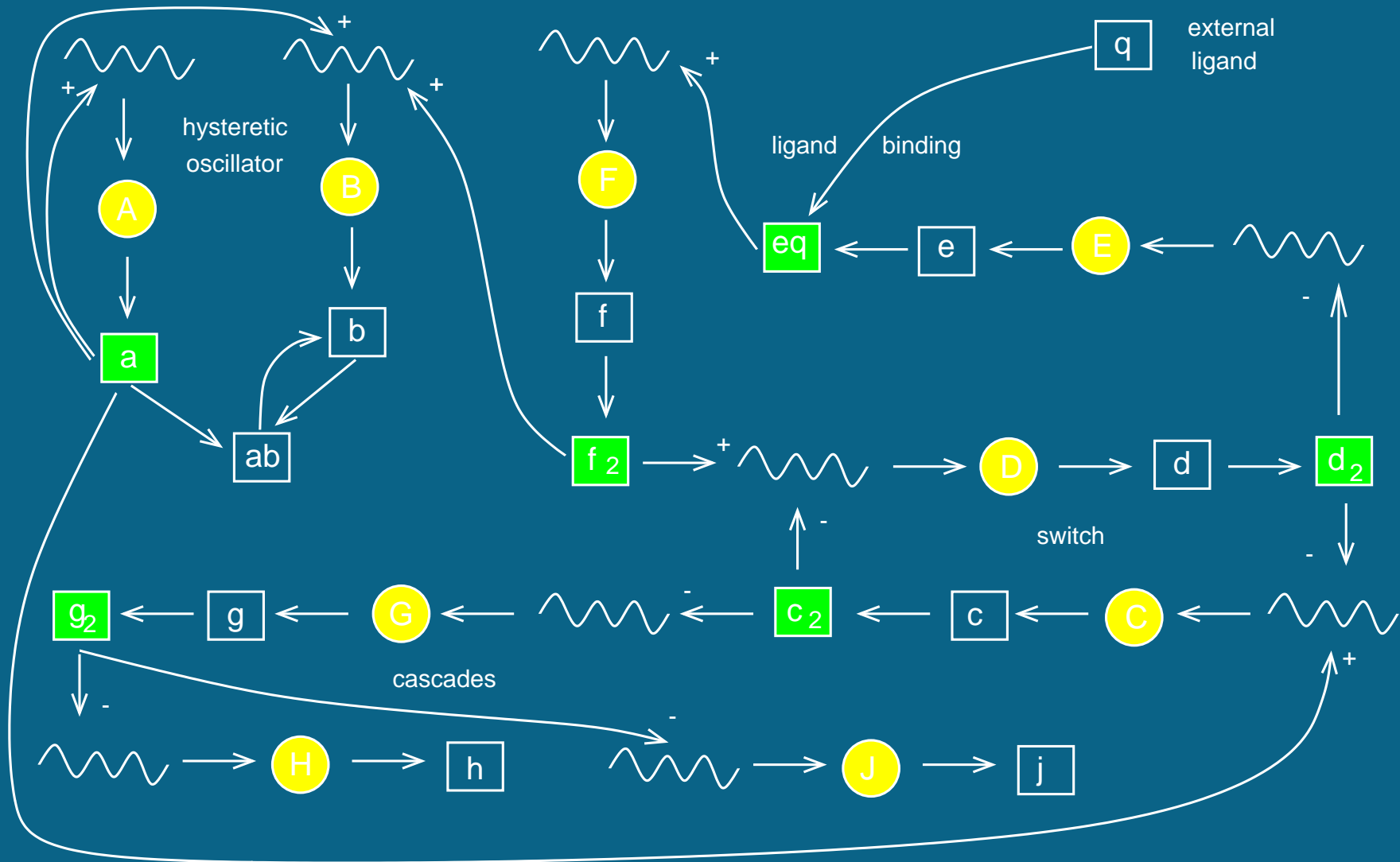
Fast, computationally cheap

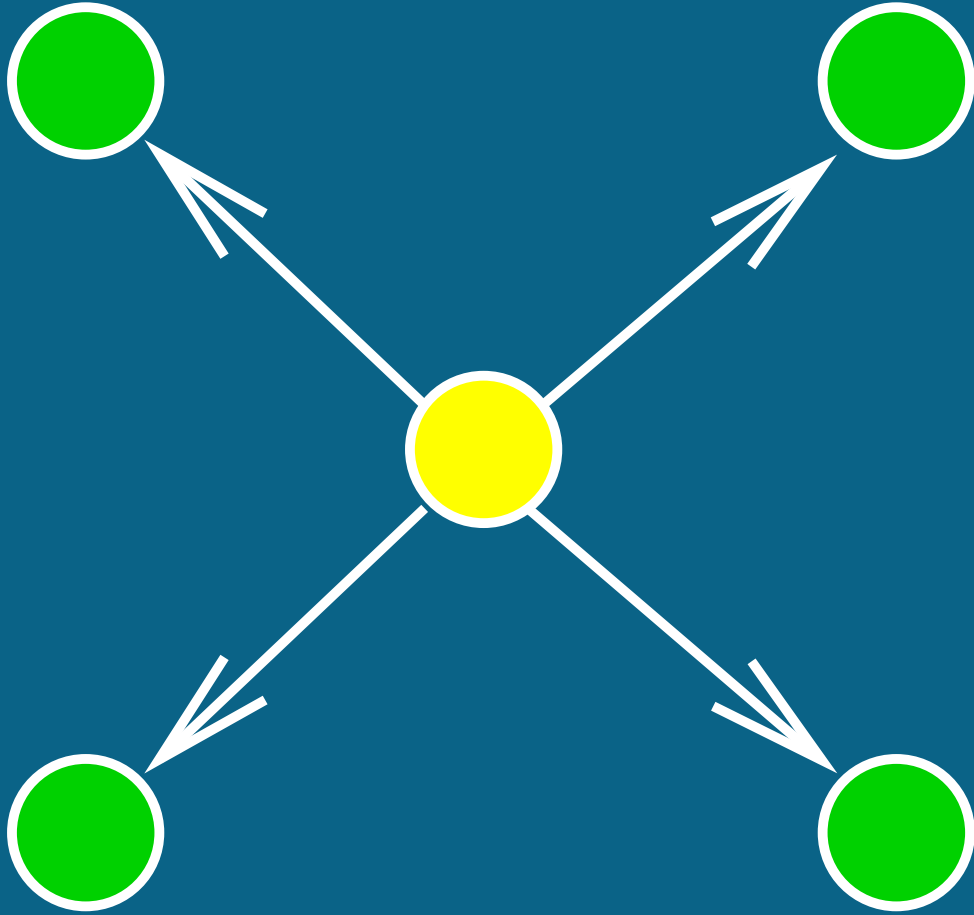
Advantage of clustering

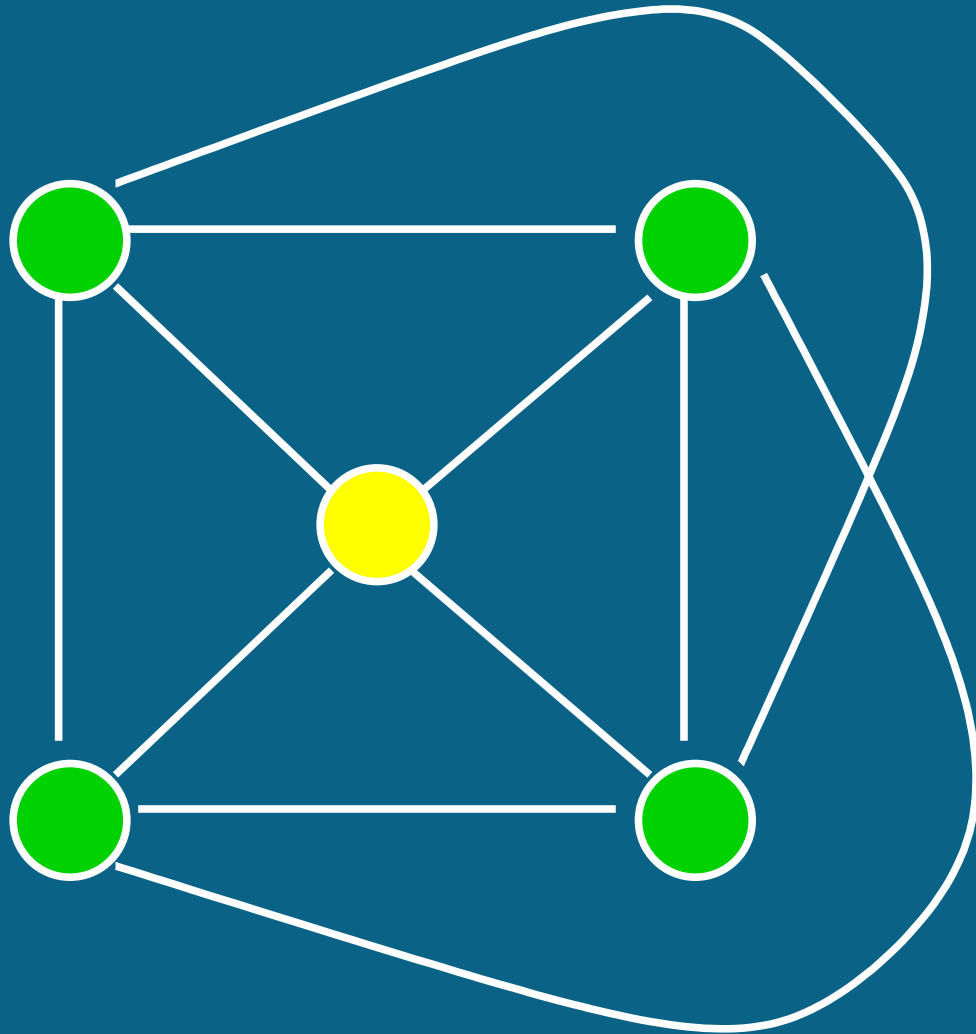
Fast, computationally cheap

Shortcoming of clustering

It is **NOT** reverse engineering







Reverse engineering

Learn the network structure from gene expression data.

Problem: Noise, sparse data

Bayesian networks

Probabilistic framework for
robust inference of interactions
in the presence of noise

Nir Friedman et al. (2000)
Journal of Computational Biology 7: 601-620

Outline of the talk

- Recapitulation: Bayesian networks
- Reverse engineering:
Learning networks from data
- Application to the yeast cell cycle
- Outlook

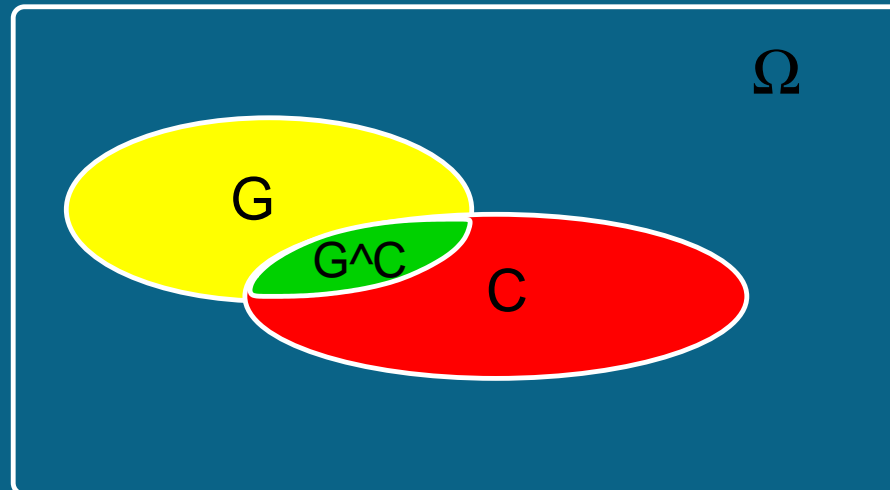
Outline of the talk

- **Recapitulation: Bayesian networks**
- Reverse engineering:
Learning networks from data
- Application to the yeast cell cycle
- Outlook

Revision: Bayes' Rule

G: A certain gene is over-expressed

C: A patient is suffering from cancer



$$P(G|C) = \frac{P(G, C)}{P(C)}$$

$$P(G, C) = P(G|C)P(C)$$

$$P(G, C) = P(C|G)P(G)$$

$$P(C|G) = \frac{P(G|C)P(C)}{P(G)}$$

A

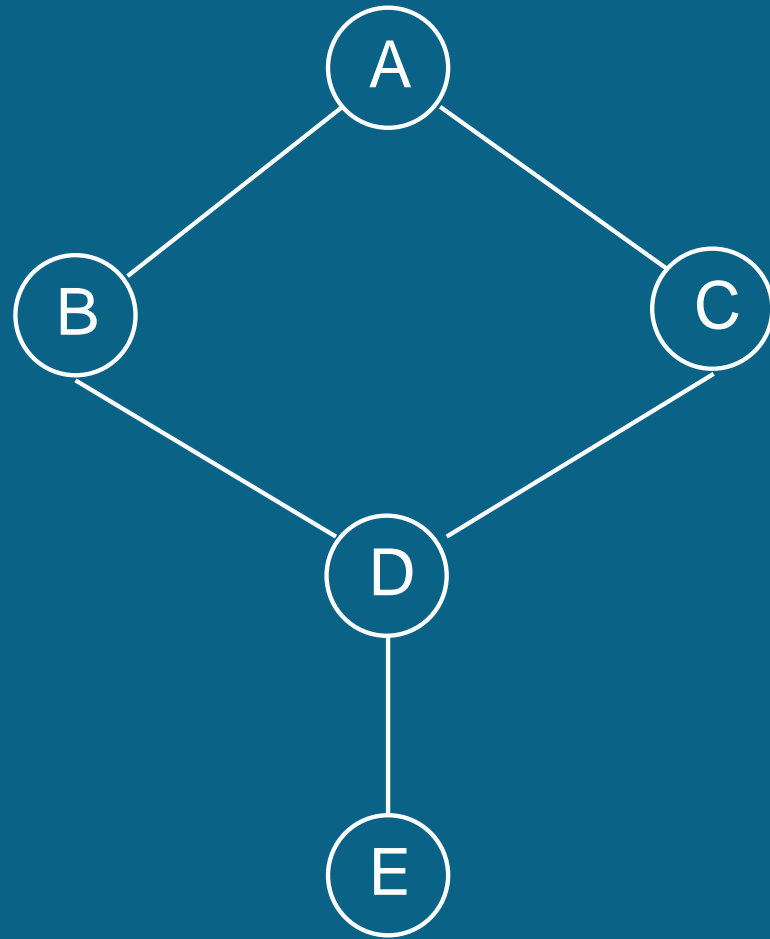
B

C

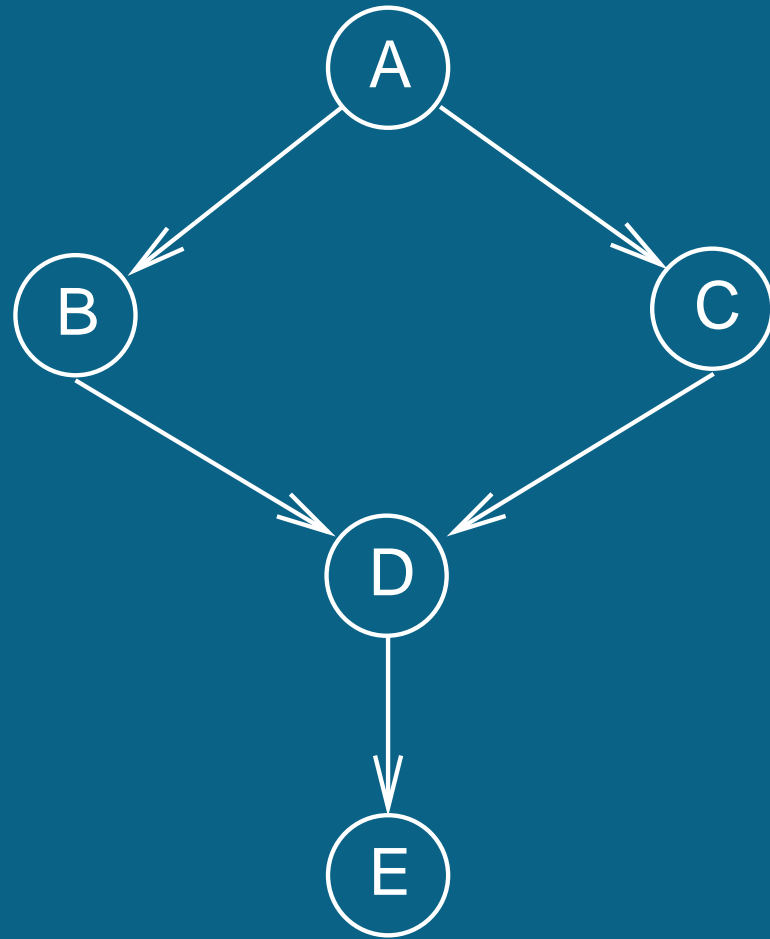
D

E

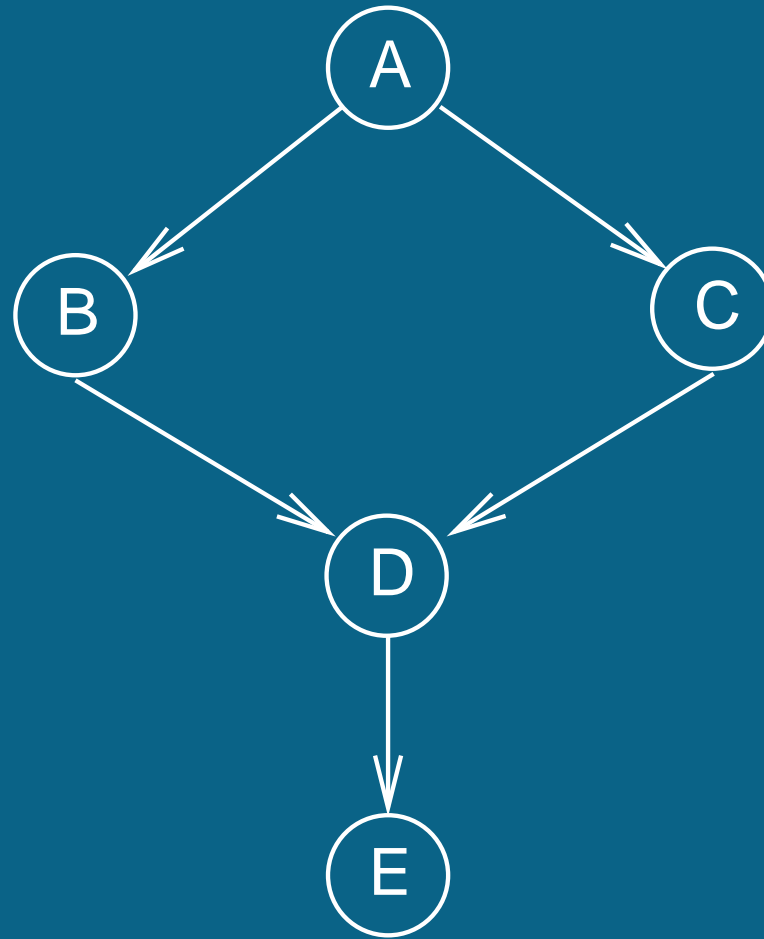
Nodes



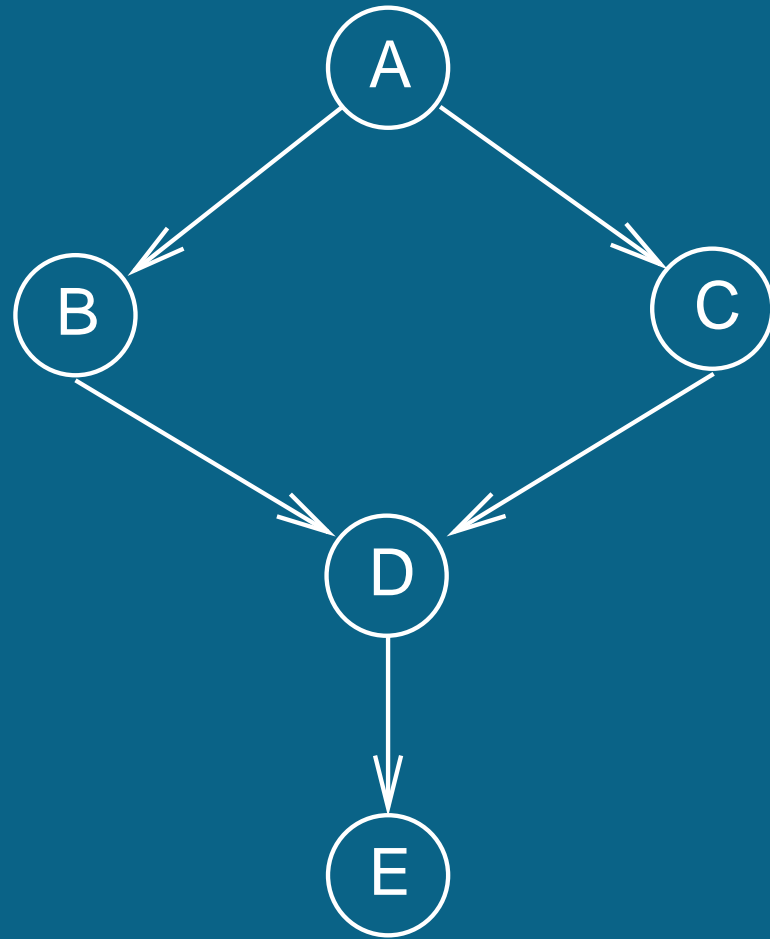
Edges



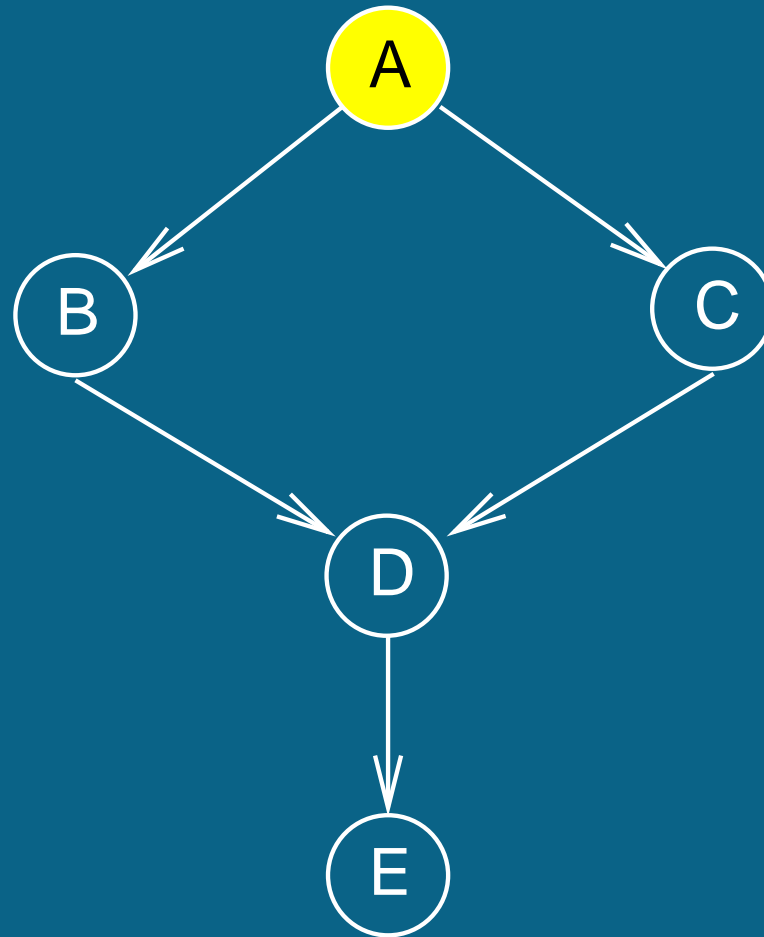
Edges = directed



No directed cycles !

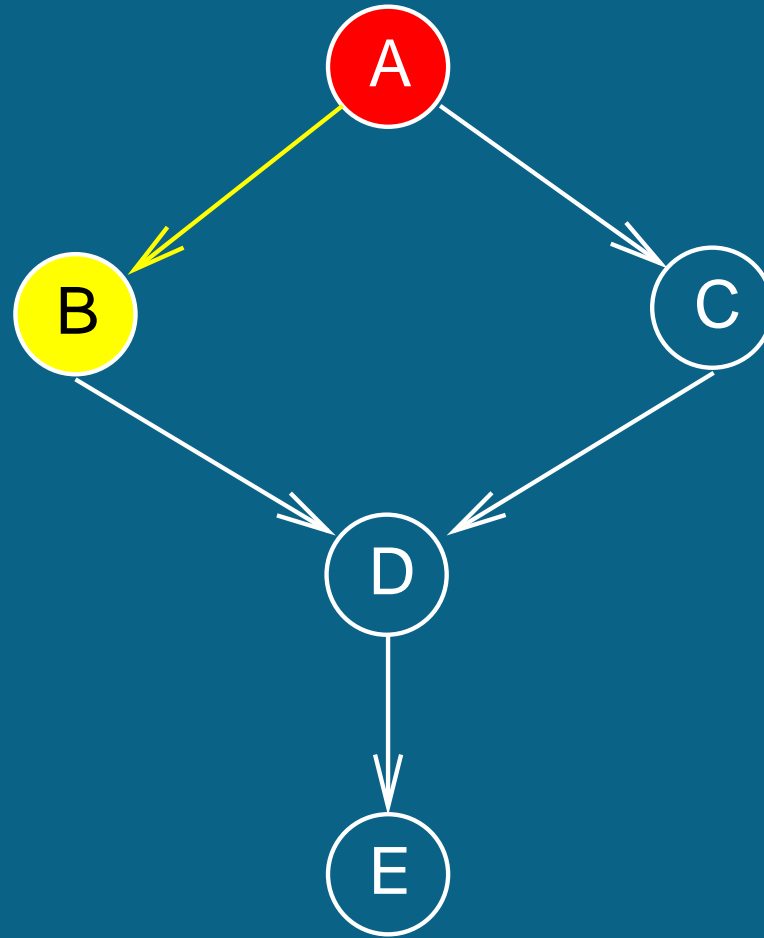


$$P(A, B, C, D, E) = \prod_i P(\text{node}_i | \text{parents}_i)$$

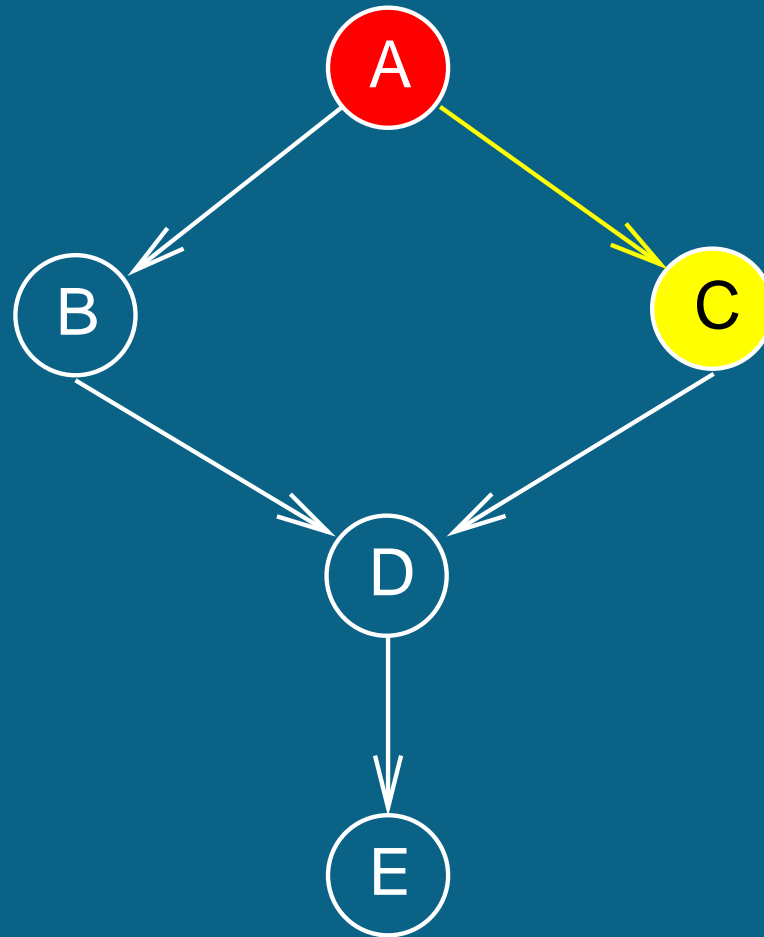


$$P(A, B, C, D, E) =$$

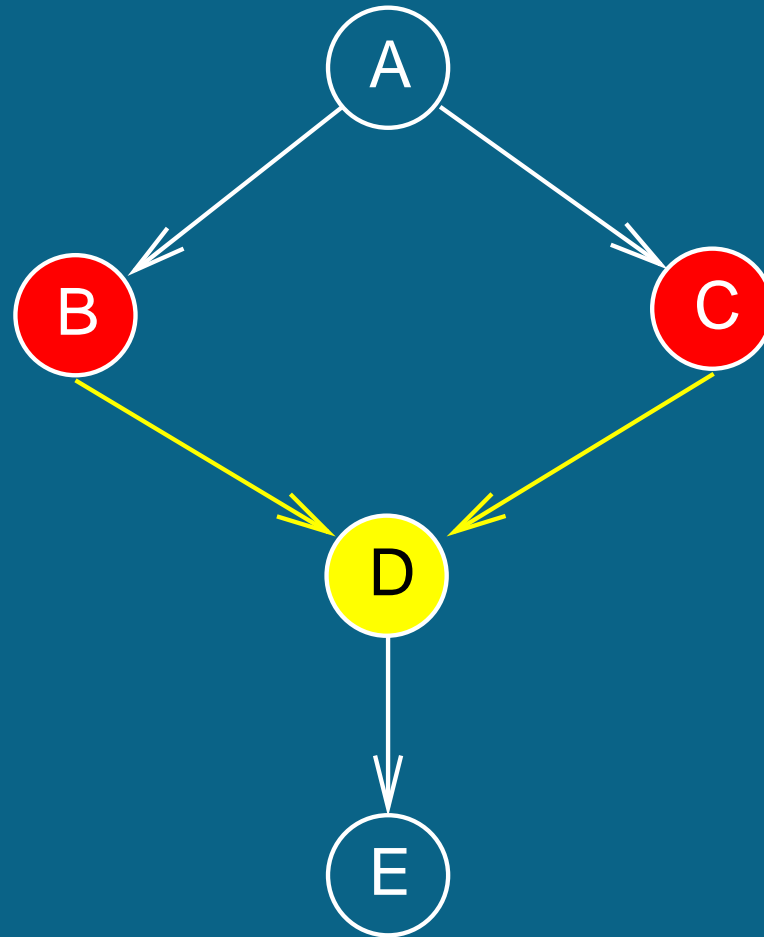
$$P(A)$$



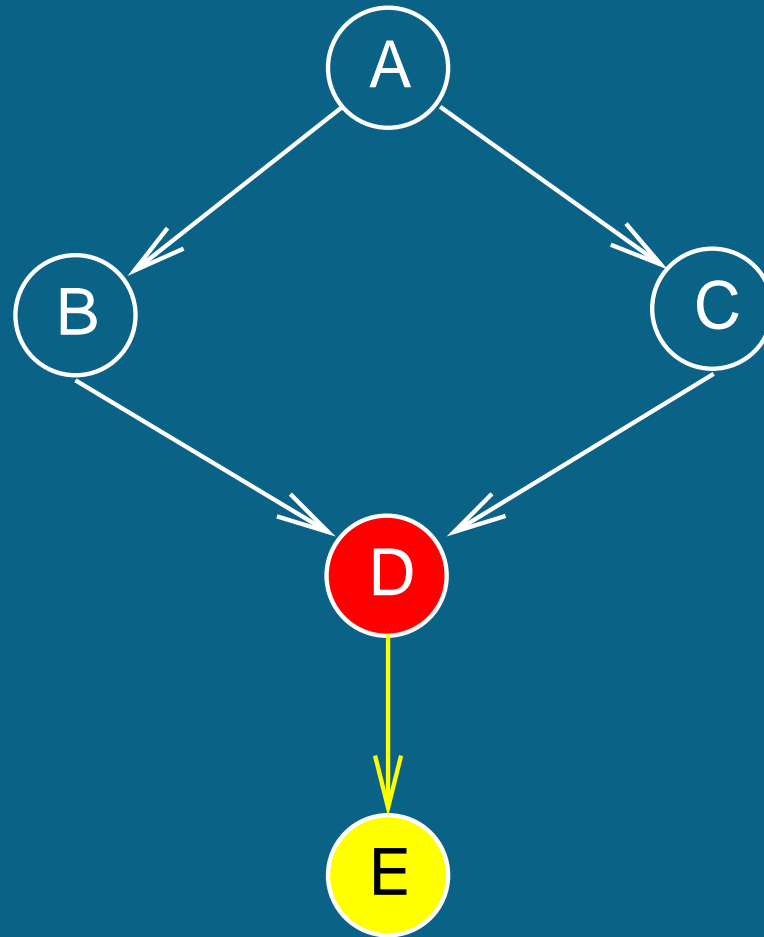
$$P(A, B, C, D, E) = P(A)P(B|A)$$



$$P(A, B, C, D, E) = P(A)P(B|A)P(C|A)$$

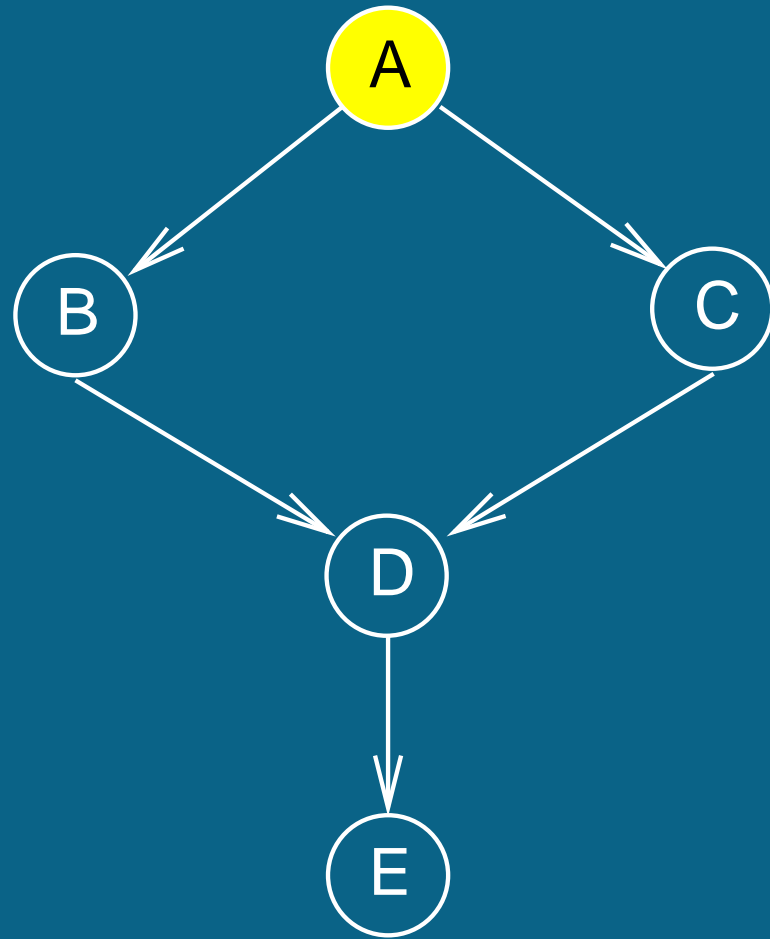


$$P(A, B, C, D, E) = P(A)P(B|A)P(C|A)P(D|B, C)$$

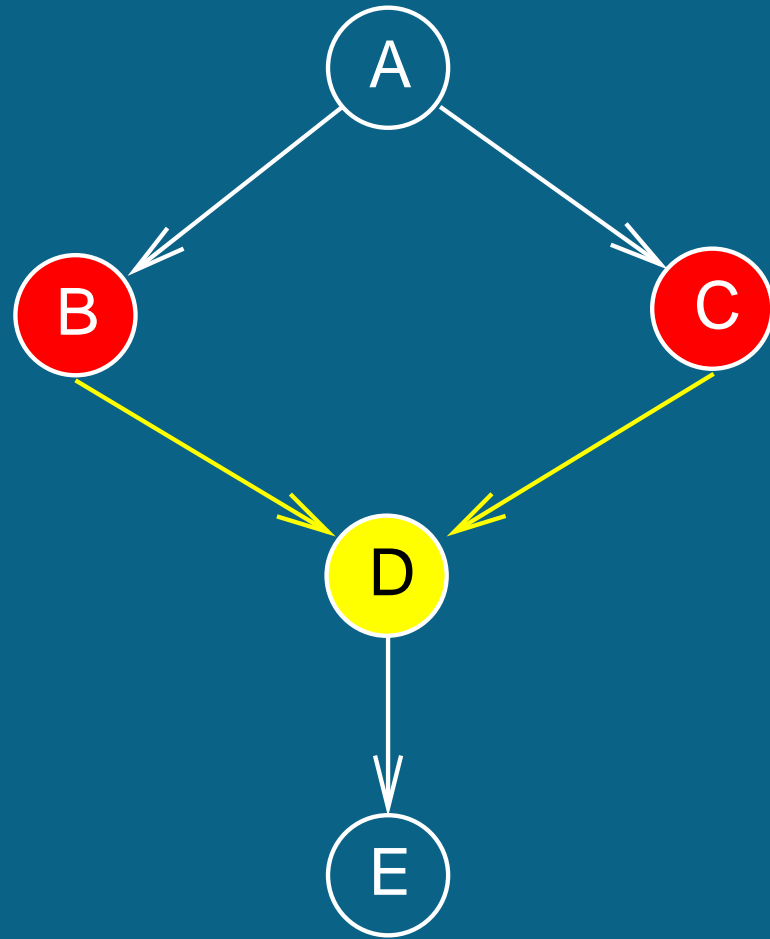


$$P(A, B, C, D, E) = P(A)P(B|A)P(C|A)P(D|B, C)P(E|D)$$

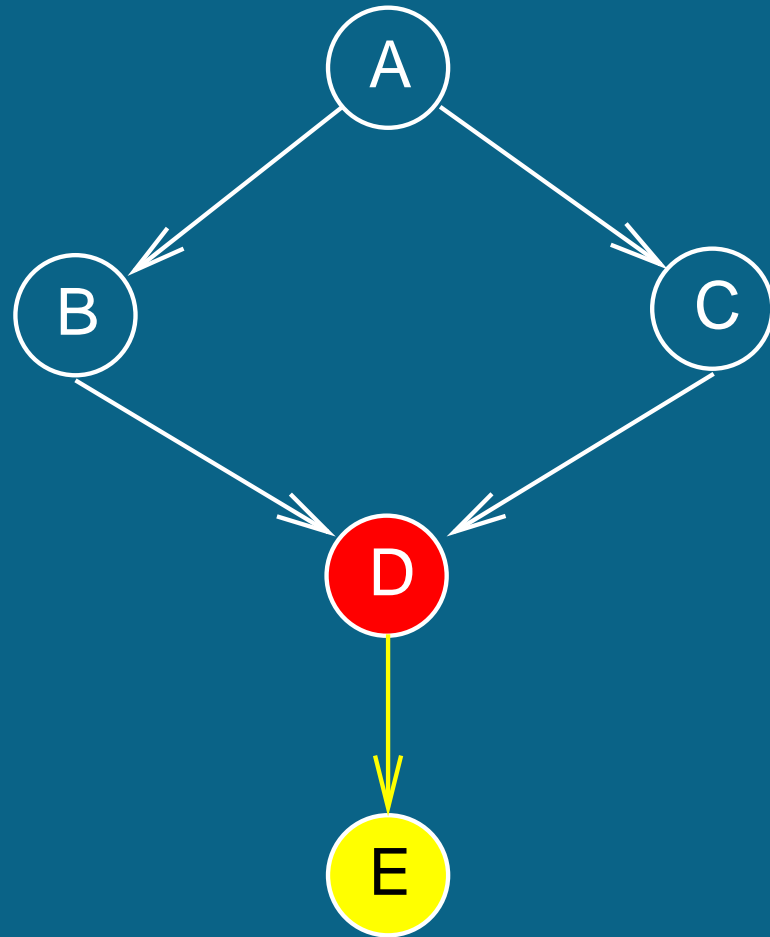
Biological interpretation



Initiation of cell (sub-)cycle

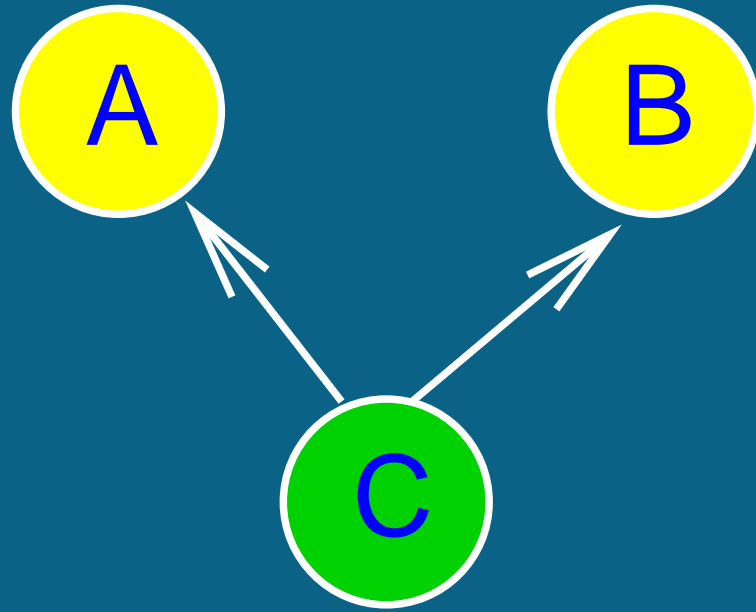


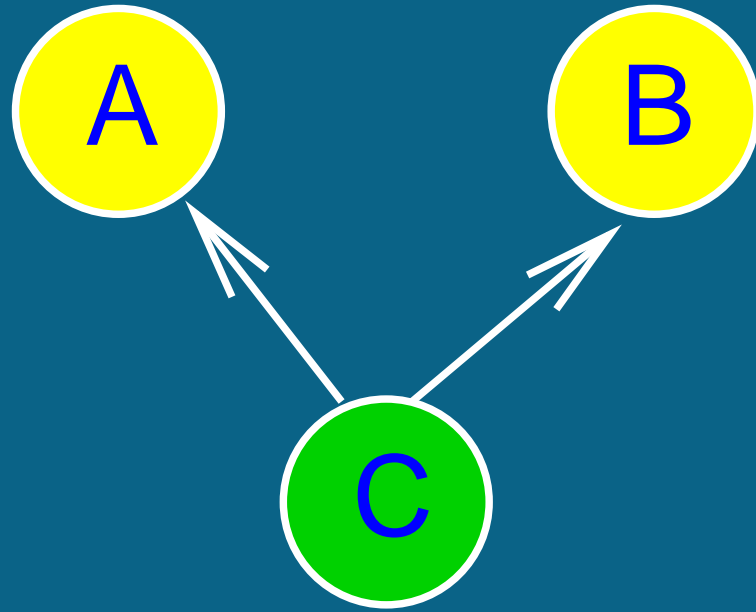
Co-regulation



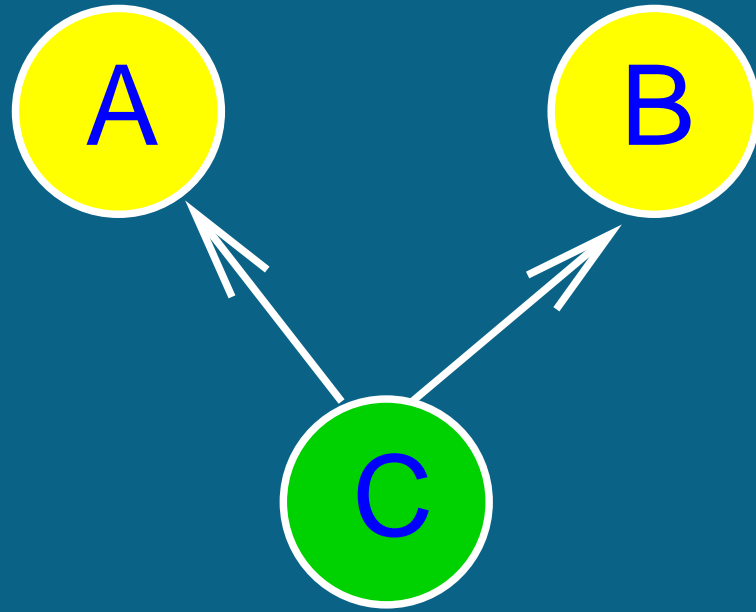
Mediation

Conditional independence relations



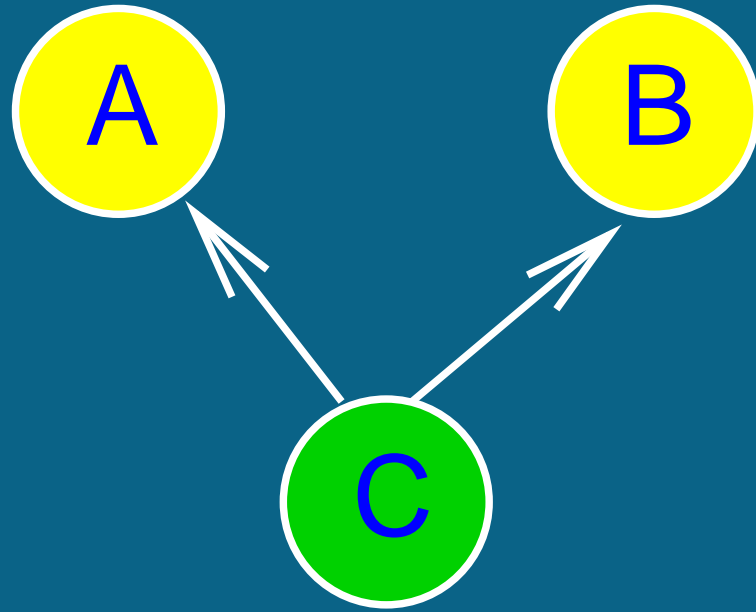


$$P(A, B, C) = P(A|C)P(B|C)P(C)$$



$$P(A, B, C) = P(A|C)P(B|C)P(C)$$

$$P(A, B|C) = \frac{P(A, B, C)}{P(C)} = P(A|C)P(B|C)$$



$$P(A, B, C) = P(A|C)P(B|C)P(C)$$

$$P(A, B|C) = \frac{P(A, B, C)}{P(C)} = P(A|C)P(B|C)$$

But: $P(A, B) \neq P(A)P(B)$

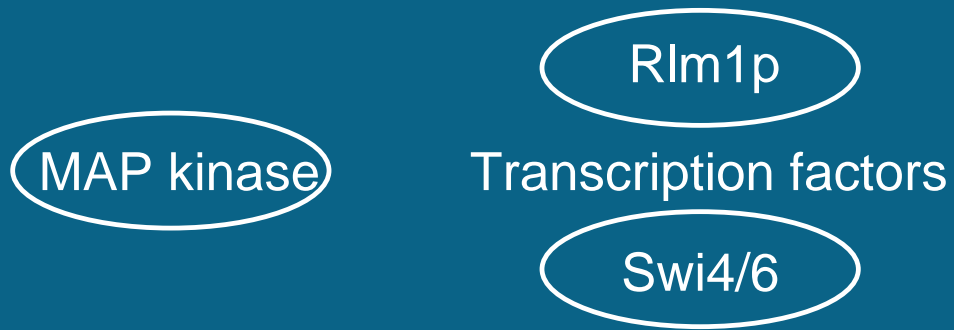
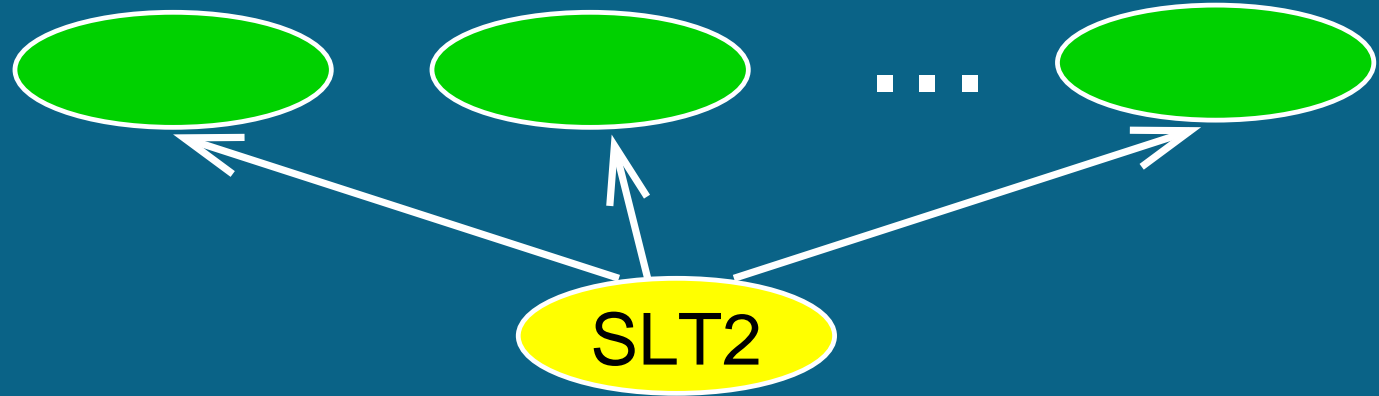
Biological example

Yeast cell cycle

Nir Friedman et al. (2000)

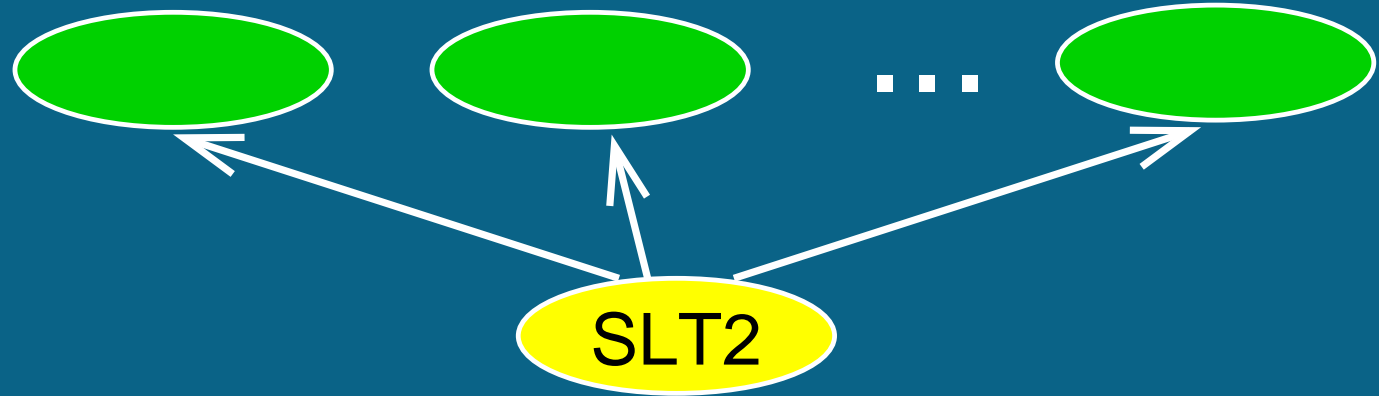
Journal of Computational Biology 7: 601-620

Low osmolarity response genes

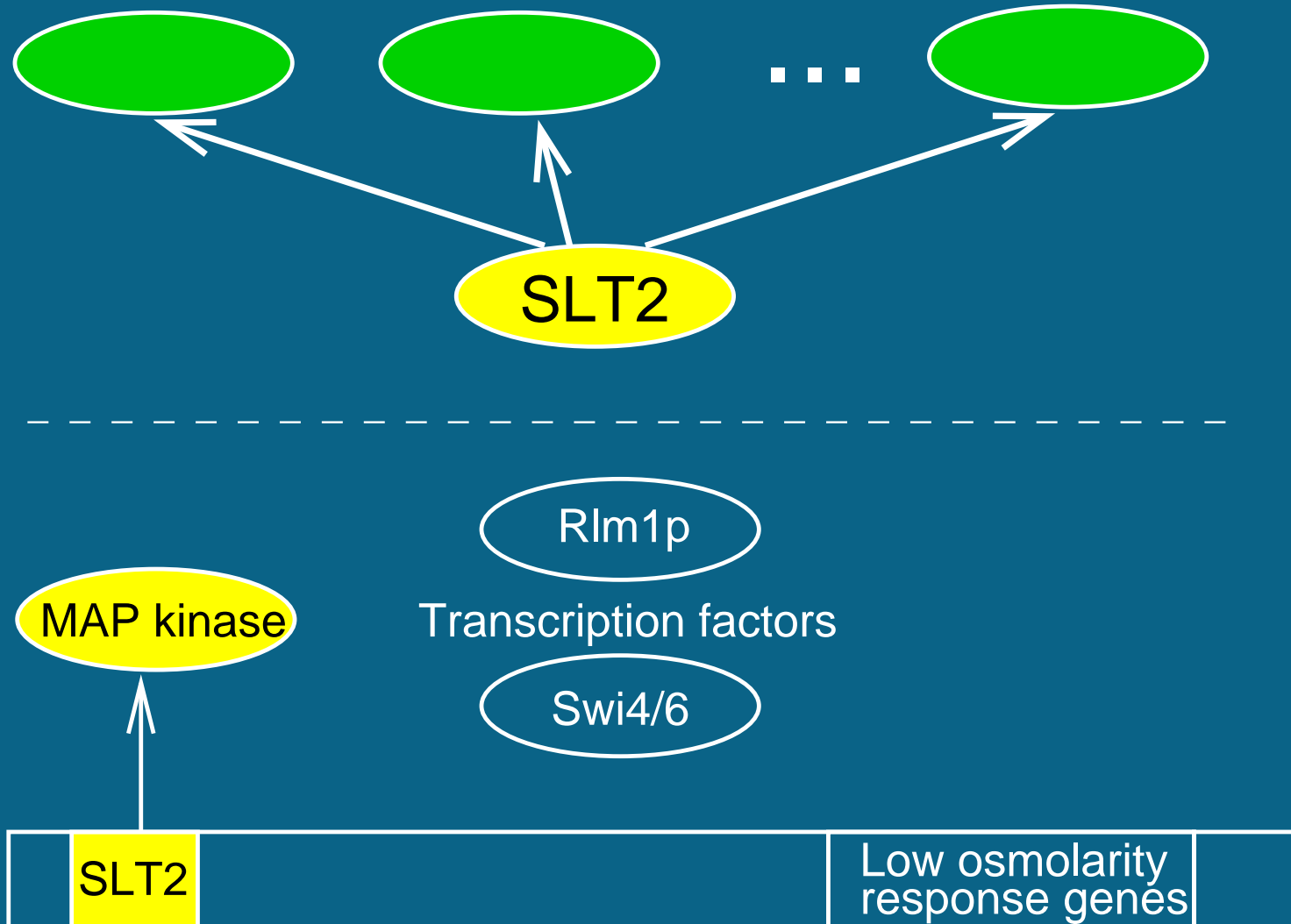


	SLT2		Low osmolarity response genes	
--	------	--	-------------------------------	--

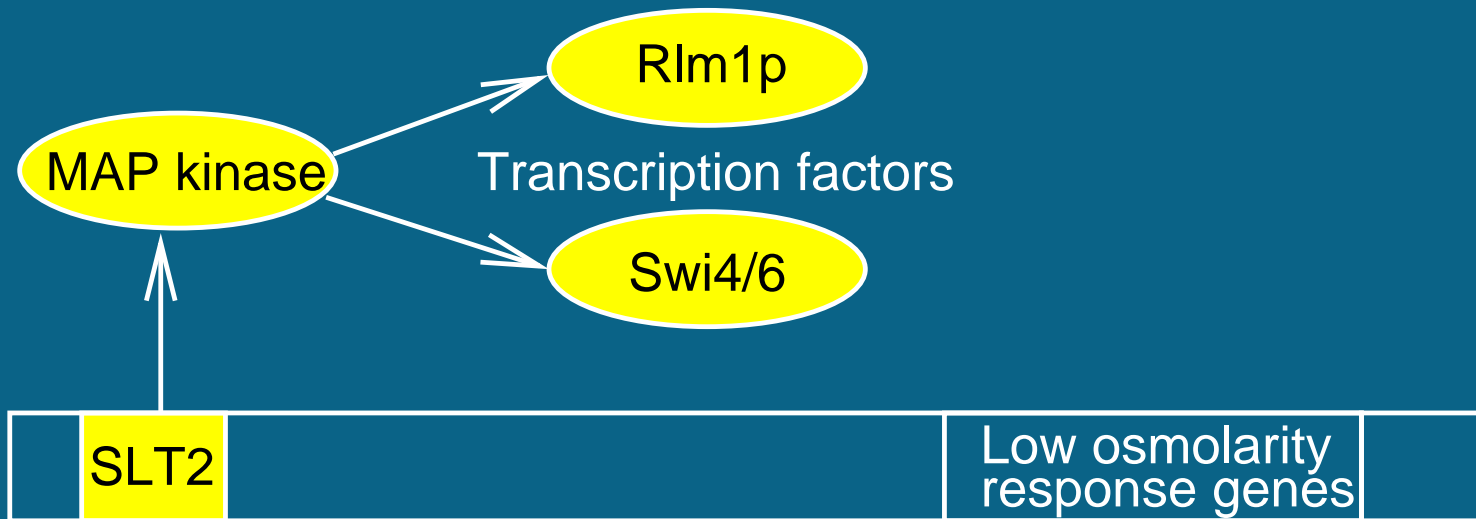
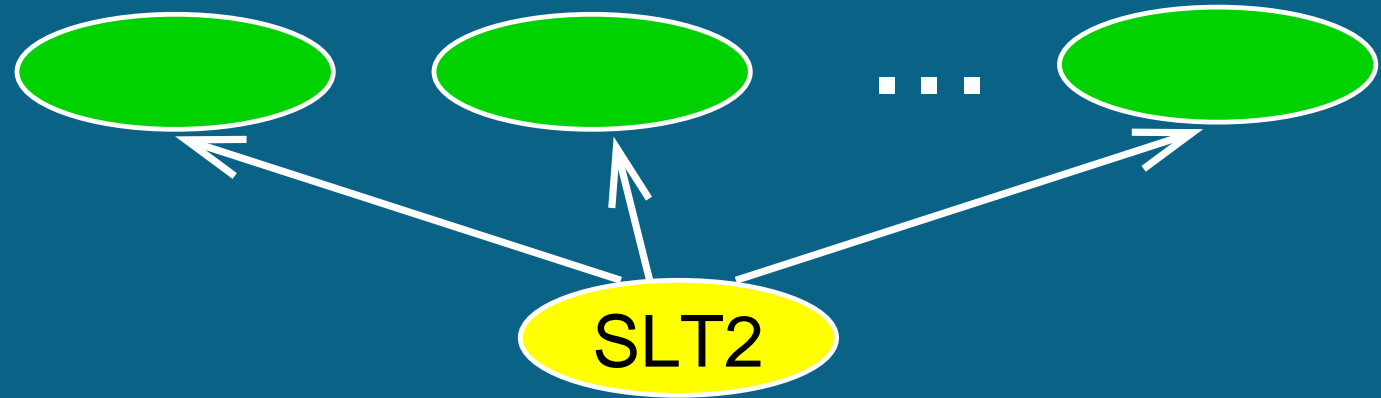
Low osmolarity response genes



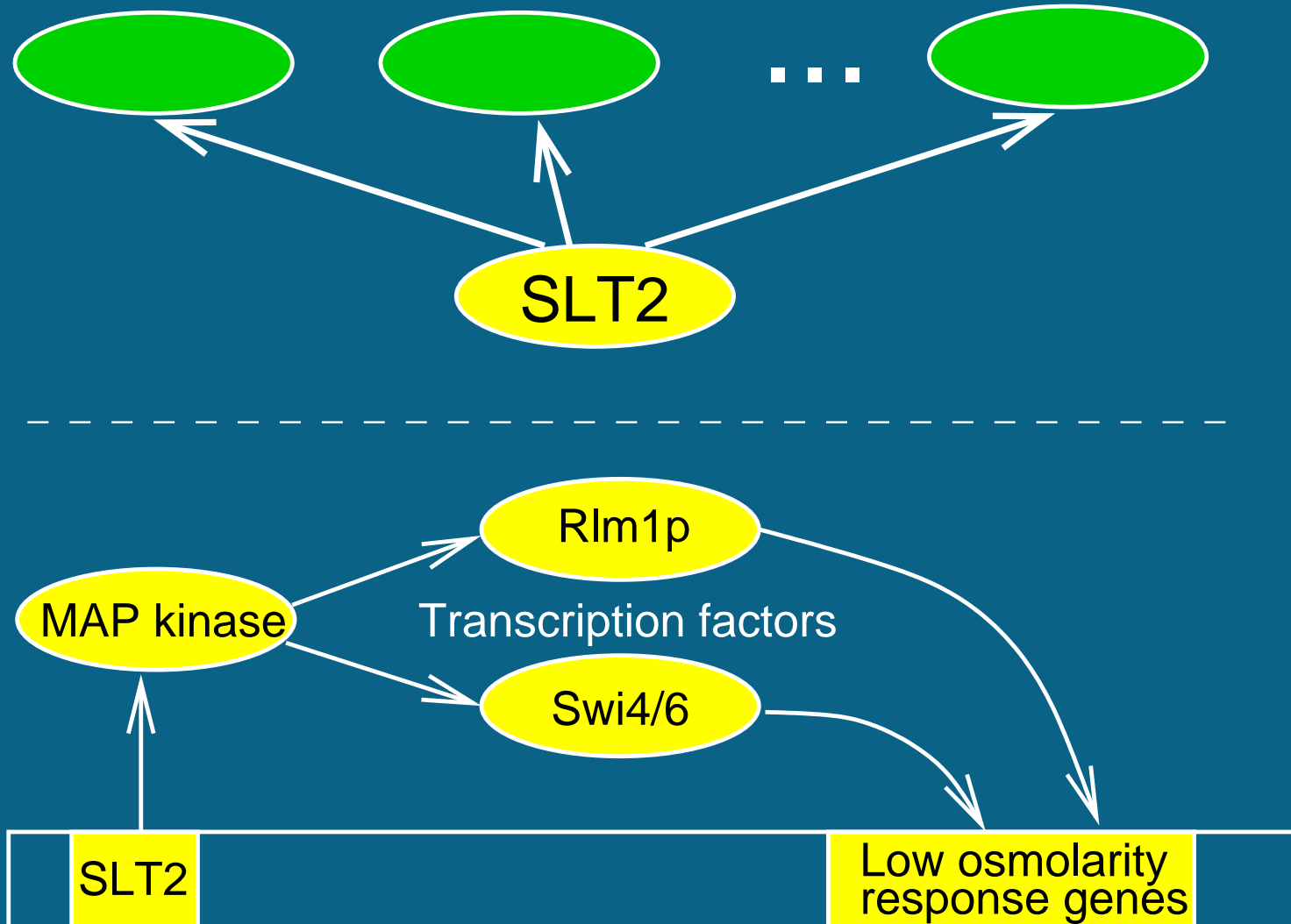
Low osmolarity response genes



Low osmolarity response genes



Low osmolarity response genes

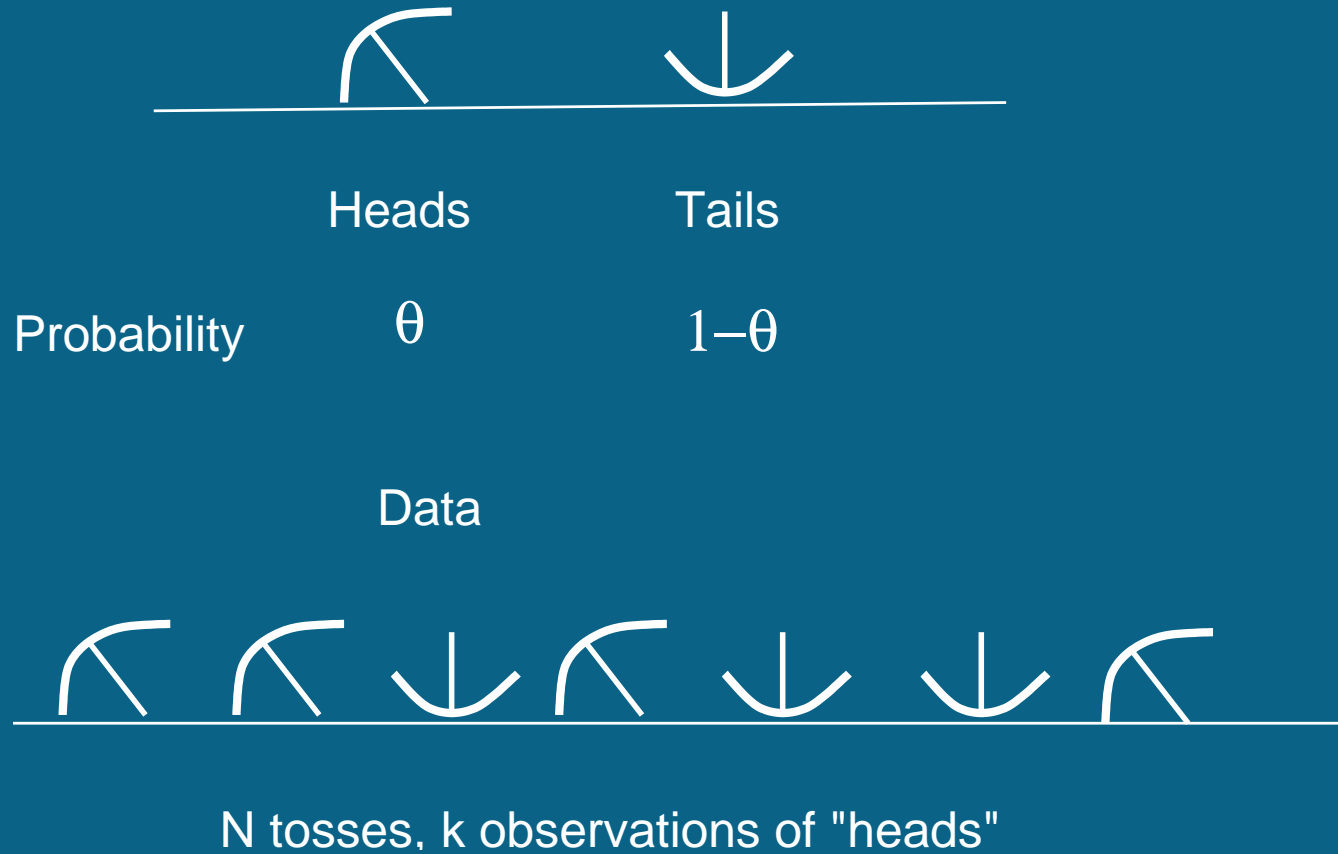


Outline of the talk

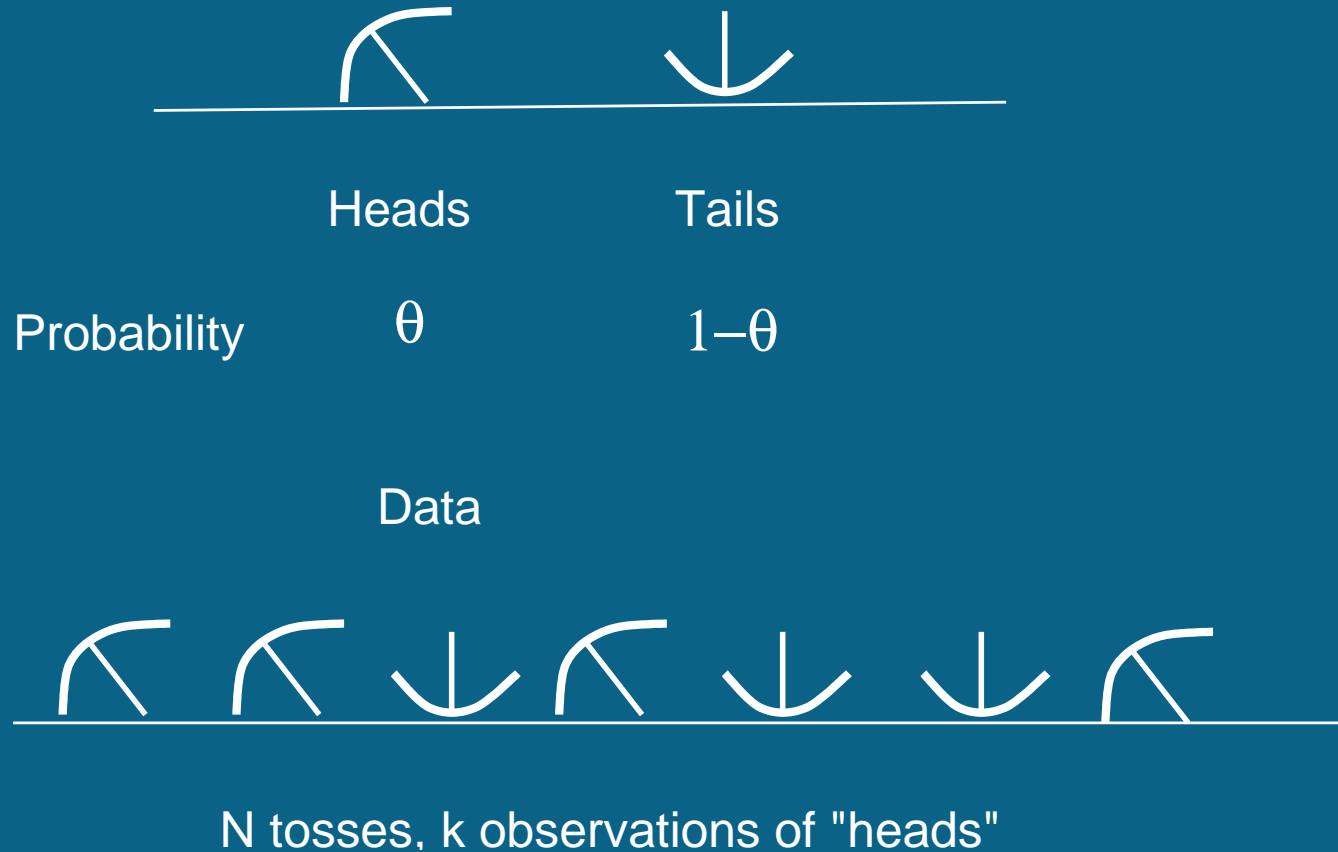
- Recapitulation: Bayesian networks
- **Reverse engineering:**
Learning networks from data
- Application to the yeast cell cycle
- Outlook

Learning from data

Learning from data



Learning from data



$$P(\theta|D) \propto P(D|\theta)P(\theta)$$

$$P(\theta|D) \propto P(D|\theta)P(\theta)$$

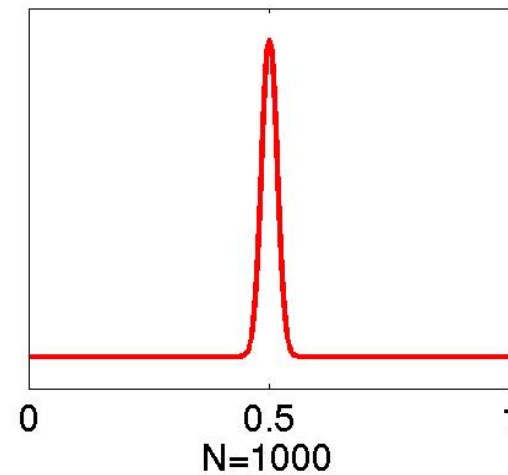
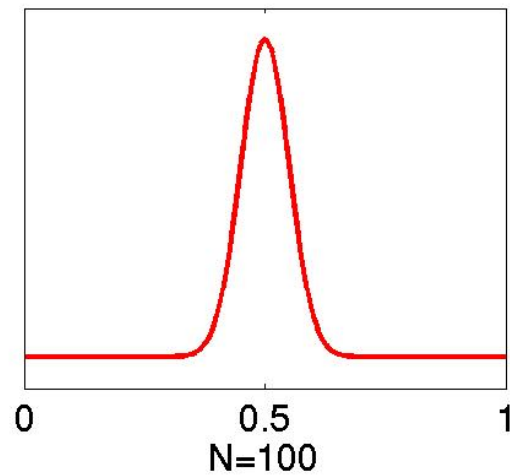
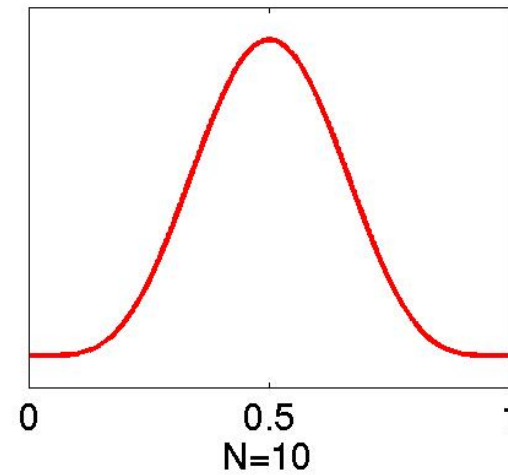
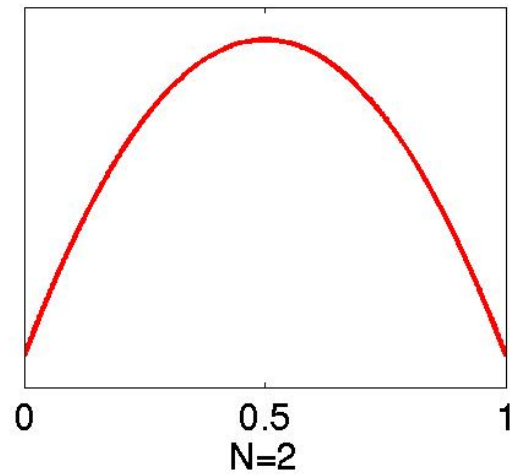
$$P(D|\theta) \propto \theta^k (1 - \theta)^{N-k}$$

$$P(\theta) = B(\alpha, \beta) \theta^{\alpha-1} (1 - \theta)^{\beta-1},$$

$$B^{-1}(\alpha, \beta) = \int_0^1 \theta^{\alpha-1} (1 - \theta)^{\beta-1} d\theta$$

$$P(\theta|D) = B(k + \alpha, N - k + \beta) \theta^{k+\alpha-1} (1 - \theta)^{N-k+\beta-1}$$

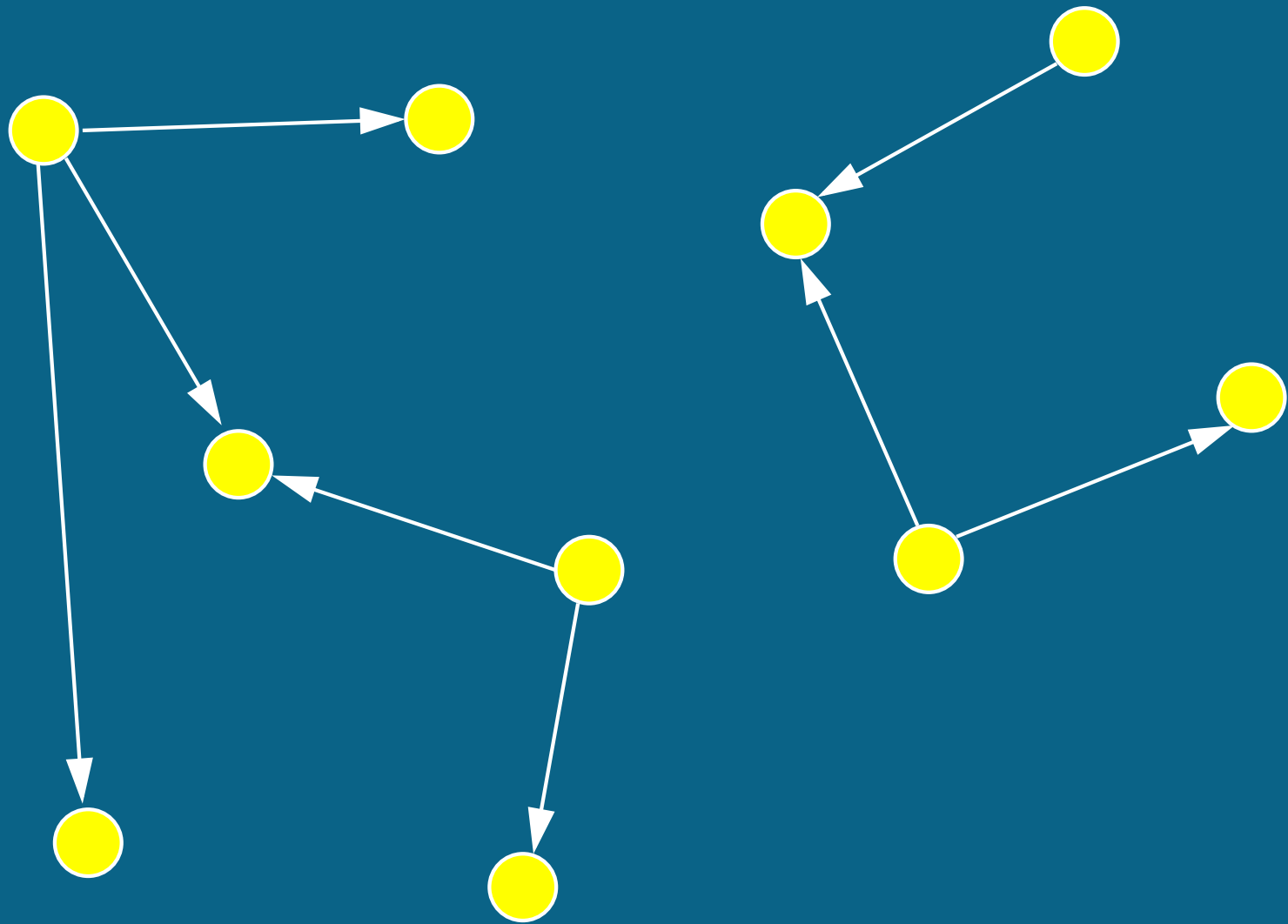
Example: $P(\theta|D)$ for equal numbers of heads and tails

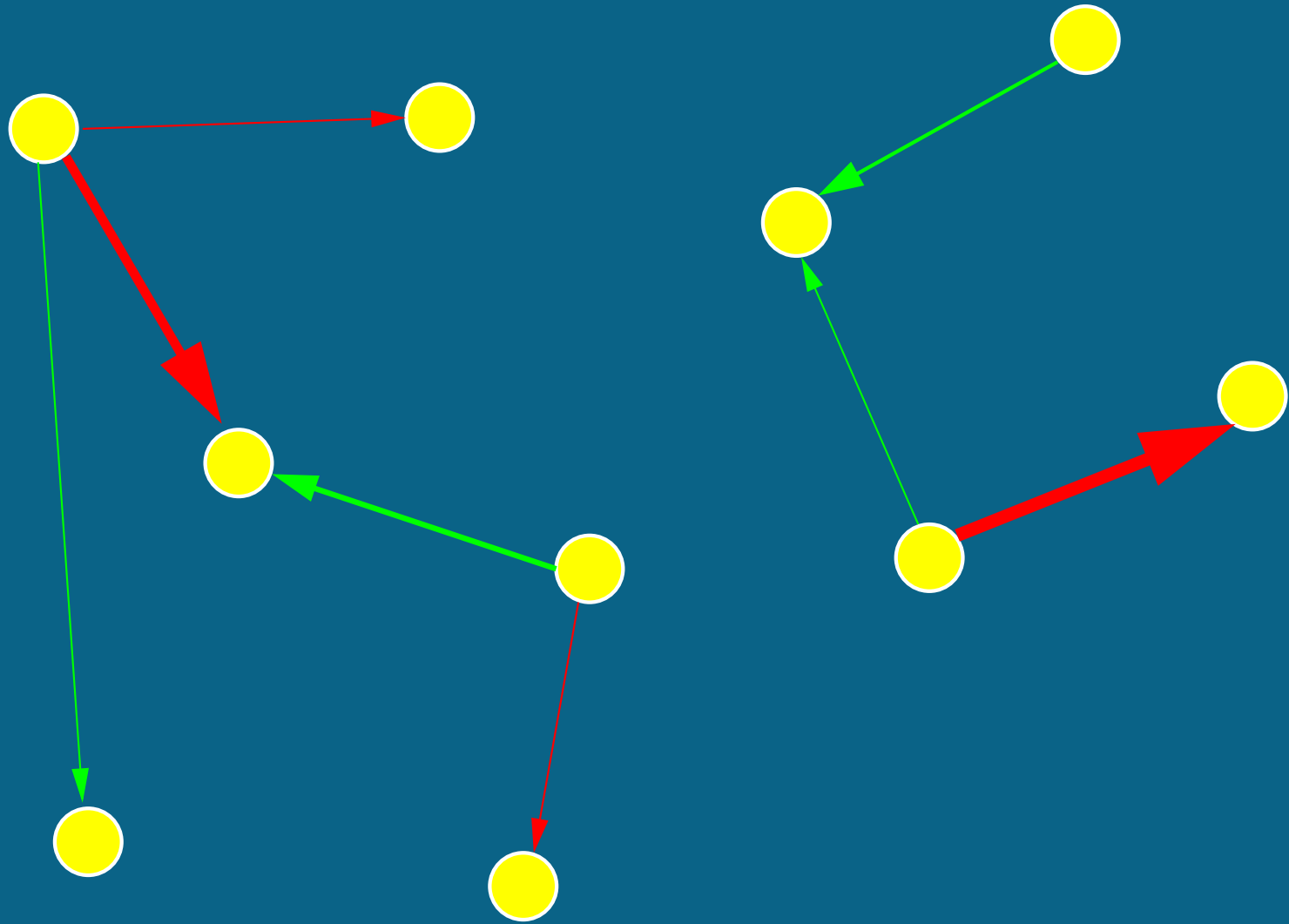


Genetic networks:

Probability distribution depends on the
network structure (M)
and its
parameters (θ).







Classical learning paradigm

Classical learning paradigm

Find the best network structure M :

$$M^* = \operatorname{argmax}\{P(M|D)\}$$

Classical learning paradigm

Find the **best network structure** M :

$$M^* = \operatorname{argmax}\{P(M|D)\}$$

Find the **best parameters** θ^*

$$\theta^* = \operatorname{argmax}\{P(\theta|D, M^*)\}$$

Find the best model M , that is, the best network

$$P(M|D) \propto P(D|M)P(M)$$

Find the best model M , that is, the best network

$$P(M|D) \propto P(D|M)P(M)$$

$$P(D|M) = \int P(D|\theta, M)P(\theta|M)d\theta$$

When is the integral **analytically tractable**?

Find the best model M , that is, the **best network**

$$P(M|D) \propto P(D|M)P(M)$$

$$P(D|M) = \int P(D|\theta, M)P(\theta|M)d\theta$$

When is the integral **analytically tractable**?

- Complete observation: **No missing values.**
- $P(D|\theta, M)$ and $P(\theta|M)$ must satisfy certain regularity conditions.
- Examples: **Multimodal** with a Dirichlet prior, **linear Gaussian** with a normal-gamma prior.

Naive approach

- Compute $P(M|D)$ for all possible network structures M .
- Select network structure M^* that maximizes $P(M|D)$

Naive approach

- Compute $P(M|D)$ for all possible network structures M .
- Select network structure M^* that maximizes $P(M|D)$

Problem 1:

Number of different network structures increases super-exponentially with the number of nodes.

N of nodes	2	4	6	8	10
N of structures	3	543	3.7×10^6	7.8×10^{11}	4.2×10^{18}

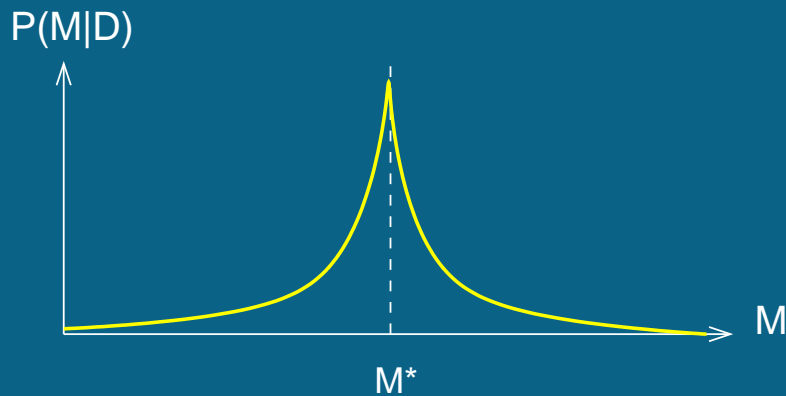
→ Optimization problem intractable for large N of nodes

Naive approach

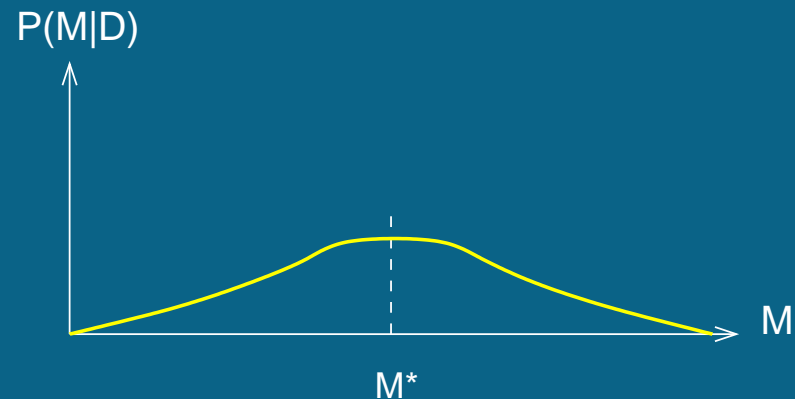
- Compute $P(M|D)$ for all possible network structures M .
- Select network structure M^* that maximizes $P(M|D)$

Problem 2:

Data are sparse \rightarrow Intrinsic uncertainty of inference

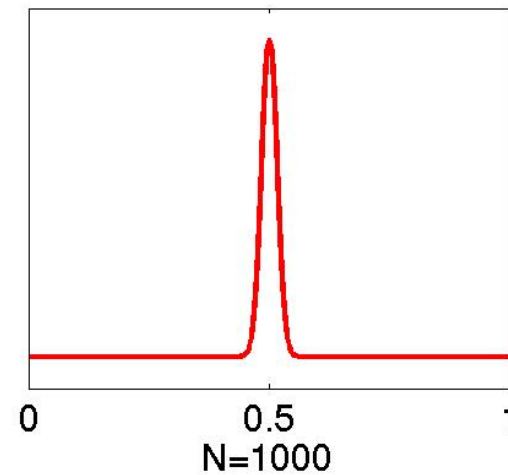
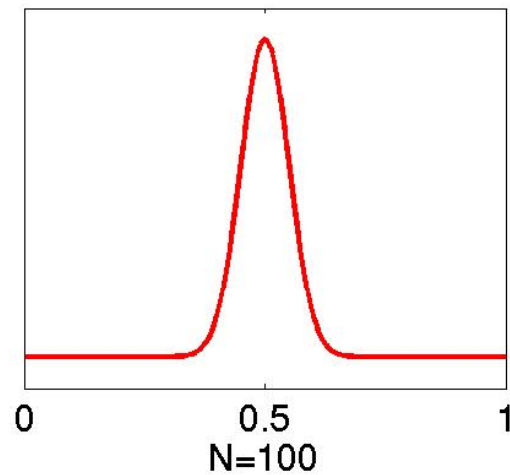
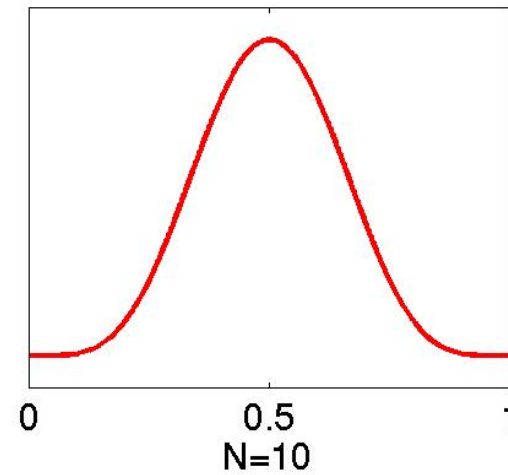
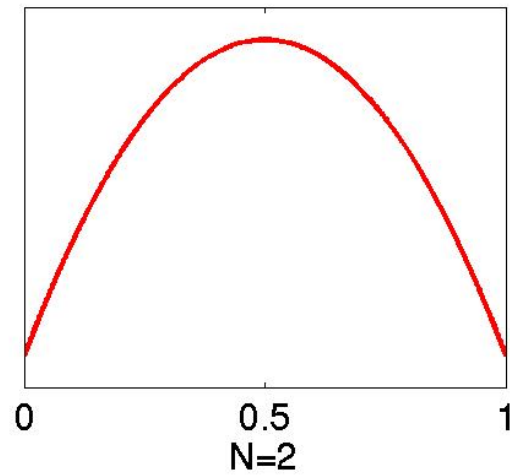


Large data set D :
Best network structure M^* well defined



Small data set D :
Intrinsic uncertainty about M^*

Example: $P(\theta|D)$ for equal numbers of heads and tails



Objective: Sample from the posterior distribution

$$P(M_k|D) = \frac{P(D|M_k)P(M_k)}{\sum_i P(D|M_i)P(M_i)}$$

Direct approach intractable due to $\sum_i P(D|M_i)P(M_i)$

Objective: Sample from the posterior distribution

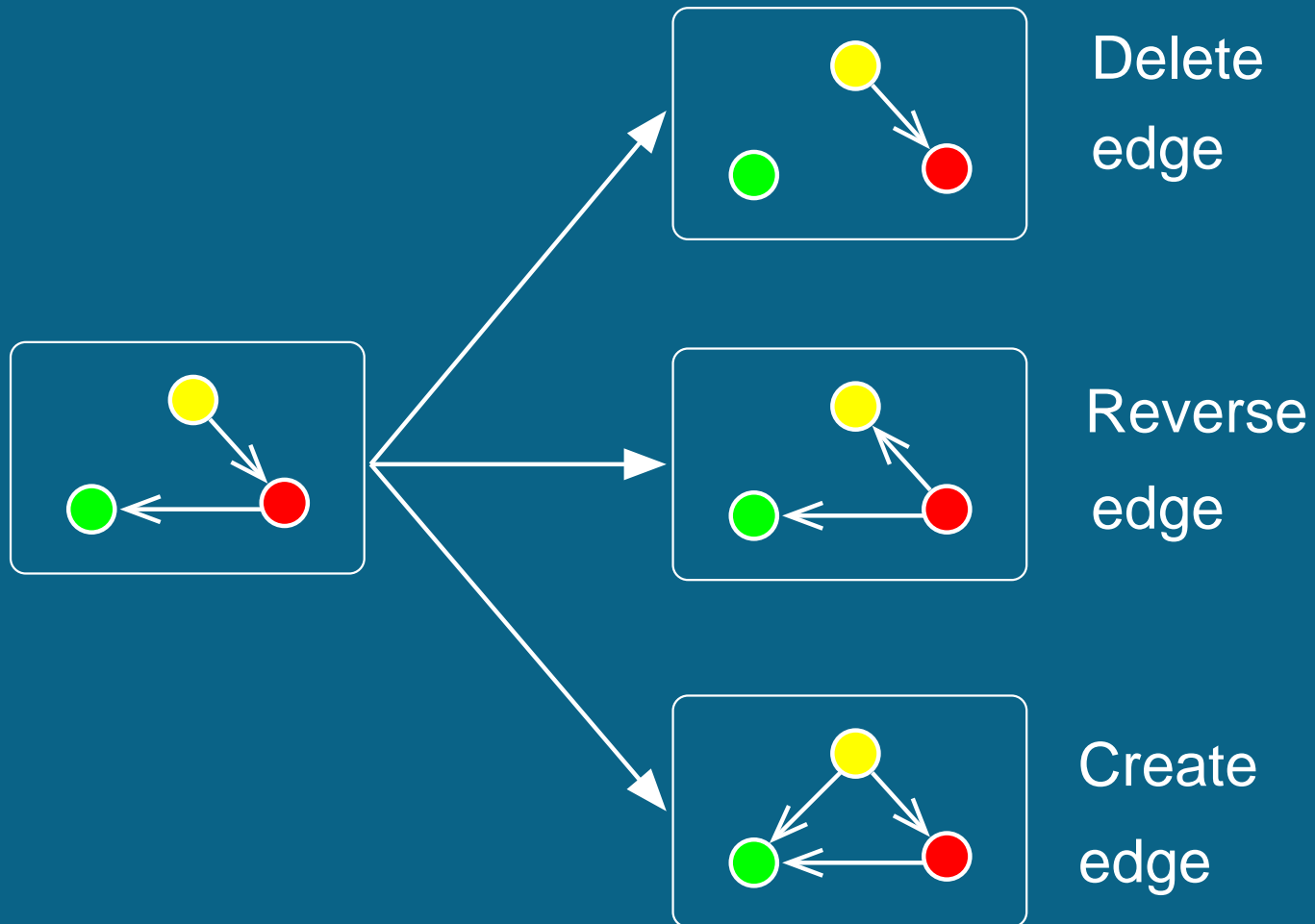
$$P(M_k|D) = \frac{P(D|M_k)P(M_k)}{\sum_i P(D|M_i)P(M_i)}$$

Direct approach intractable due to $\sum_i P(D|M_i)P(M_i)$

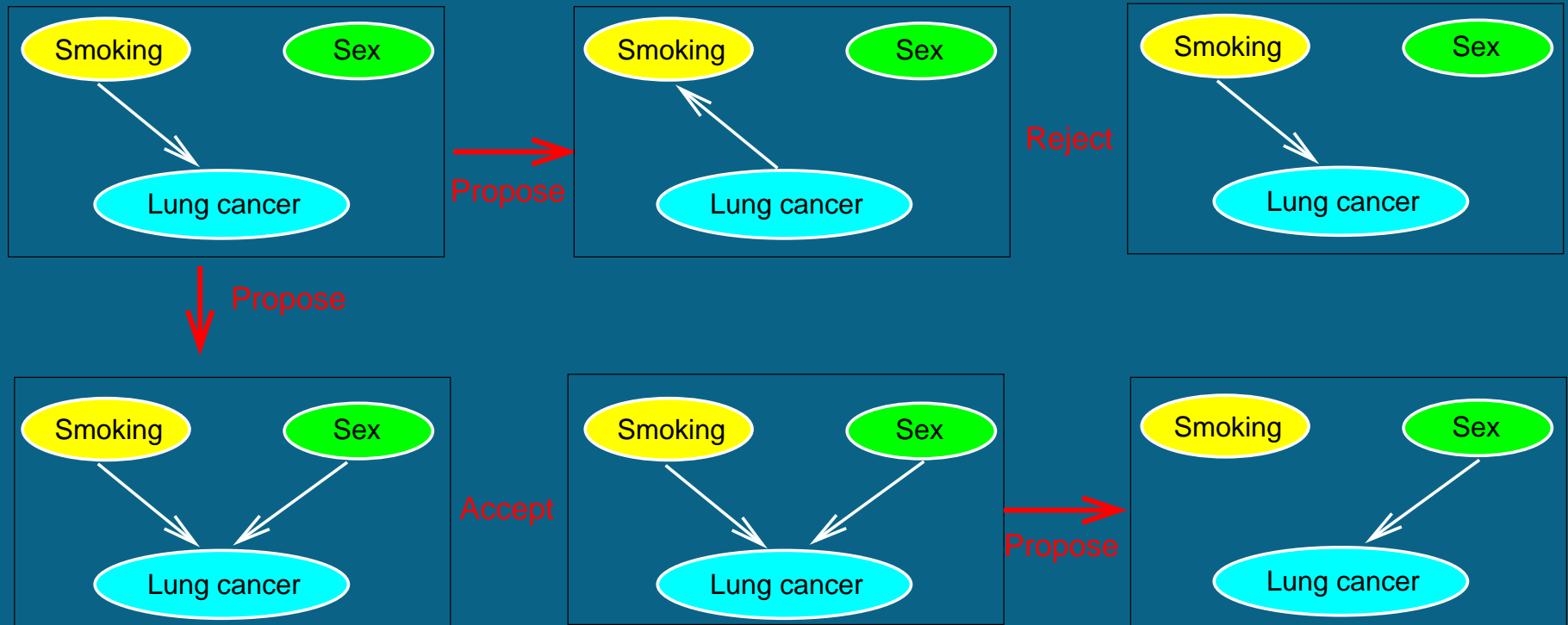
Markov chain Monte Carlo (MCMC):

- **Proposal move:** Given network M_{old} , propose a new network M_{new} with probability $Q(M_{new}|M_{old})$.
- **Acceptance/Rejection:** Accept this new network with probability $\min \left\{ 1, \frac{P(D|M_{new})P(M_{new})}{P(D|M_{old})P(M_{old})} \times \frac{Q(M_{old}|M_{new})}{Q(M_{new}|M_{old})} \right\}$

MCMC moves



Markov chain Monte Carlo (MCMC)



Accept move with probability: $\min \left\{ 1, \frac{P(M_{new}|D)}{P(M_{old}|D)} \times \frac{Q(M_{old}|M_{new})}{Q(M_{new}|M_{old})} \right\}$

Convergence of MCMC simulation

Burn-in: T MCMC steps

Sampling: T MCMC steps

How large do we have to choose T ?

Convergence of MCMC simulation

Burn-in: T MCMC steps

Sampling: T MCMC steps

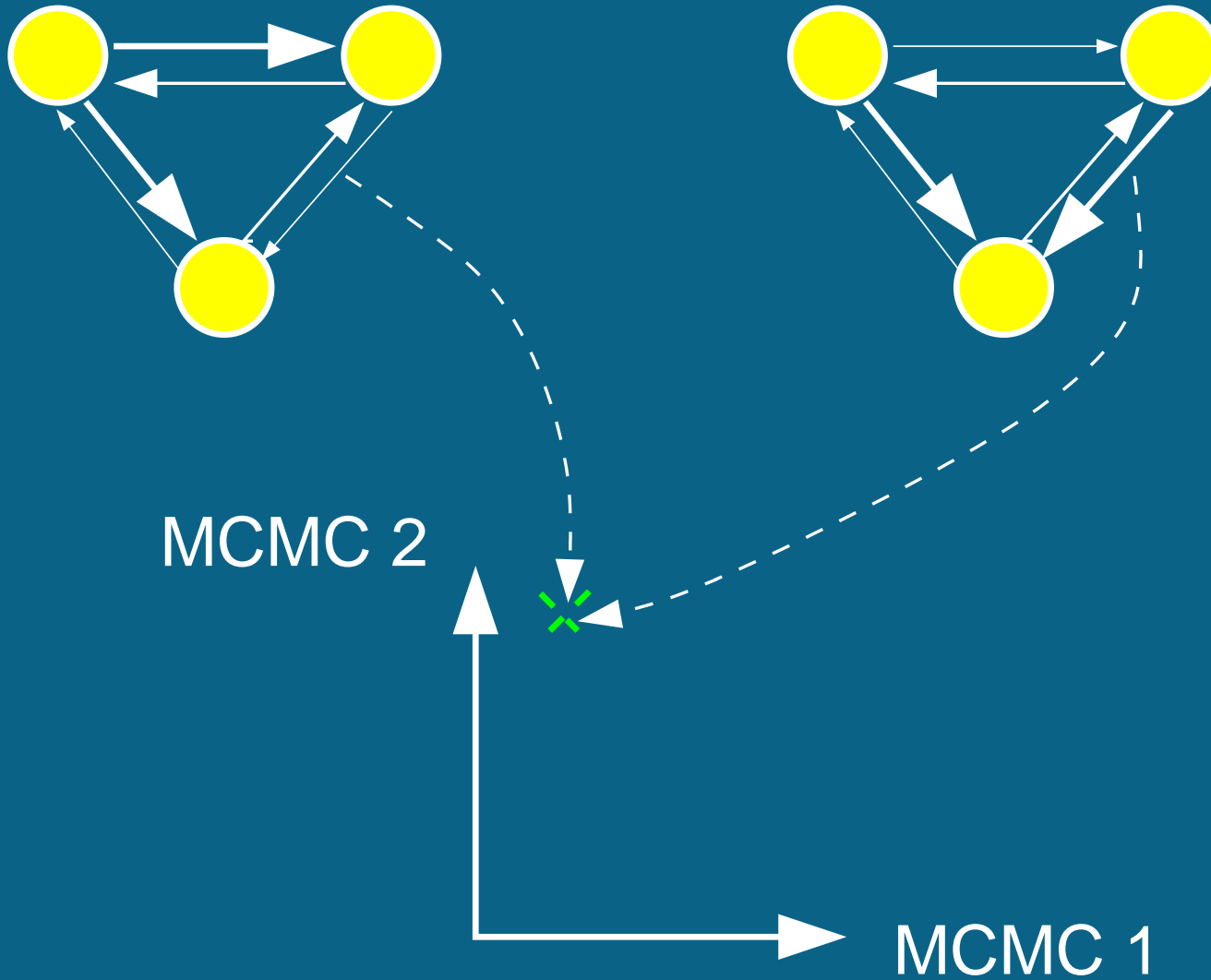
How large do we have to choose T ?

Repeat MCMC simulations with different random number generator seeds

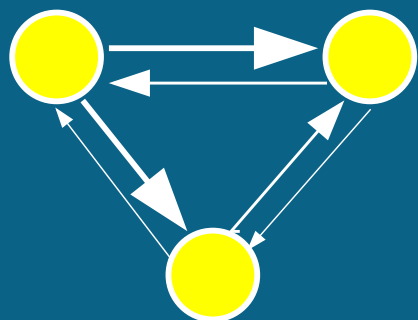
Scatter plots of posterior probabilities of edges

MCMC simulation 1

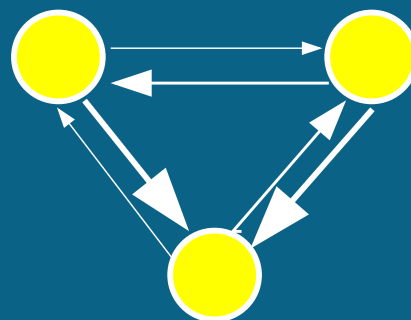
MCMC simulation 2



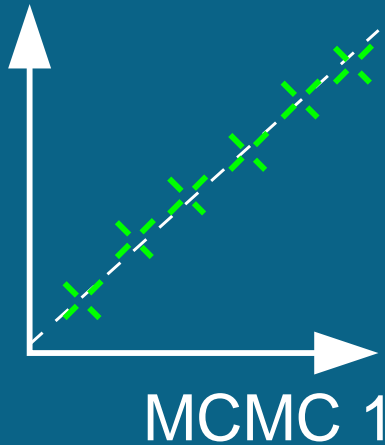
MCMC simulation 1



MCMC simulation 2

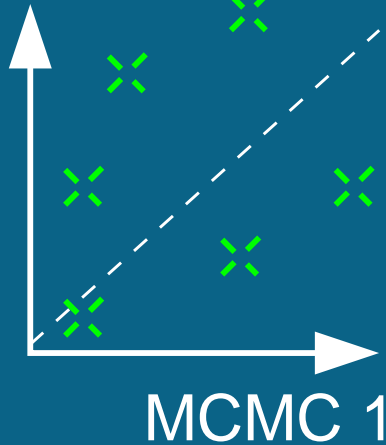


MCMC 2



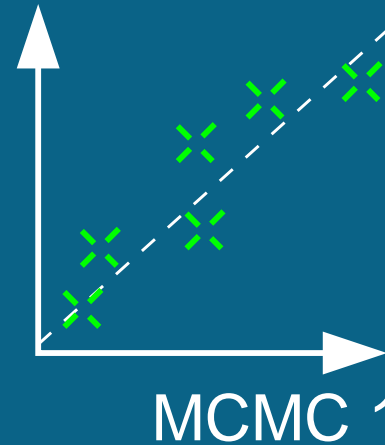
T infinite

MCMC 2



T too short

MCMC 2



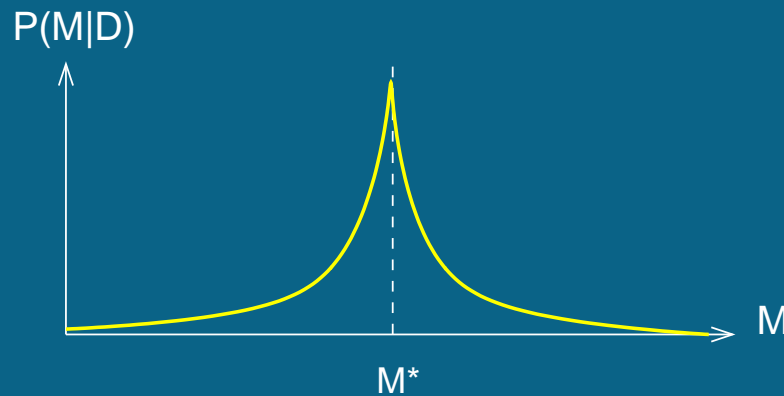
T long enough

Problem: Statistical significance of the networks

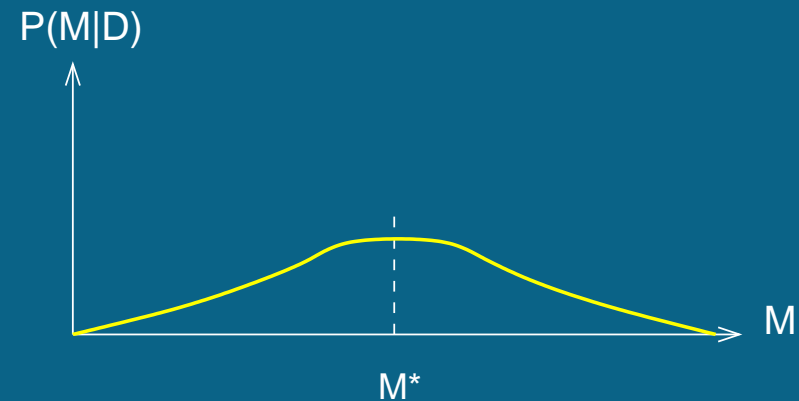
- **Complex models:** Transcript levels of thousands of genes.
- **Sparse data:** Typically a few dozen samples.

Problem: Statistical significance of the networks

- **Complex models:** Transcript levels of thousands of genes.
- **Sparse data:** Typically a few dozen samples.



Large data set D:
Best network structure M^* well defined



Small data set D:
Intrinsic uncertainty about M^*

- Posterior probability $P(M|D)$ diffuse: **Global network** inference is **meaningless**.

Solution: Focus on **features** and **subnetworks**

Feature: Indicator variable for a property of interest,
e.g.: Are X and Y close neighbours in the network?

$$f(M) = \begin{cases} 1 & \text{if } M \text{ satisfies the feature} \\ 0 & \text{otherwise} \end{cases}$$

Solution: Focus on **features** and **subnetworks**

Feature: Indicator variable for a property of interest,
e.g.: Are X and Y close neighbours in the network?

$$f(M) = \begin{cases} 1 & \text{if } M \text{ satisfies the feature} \\ 0 & \text{otherwise} \end{cases}$$

Posterior probability of features: $P(f|D) = \sum_M f(M)P(M|D)$

Solution: Focus on features and subnetworks

Feature: Indicator variable for a property of interest,
e.g.: Are X and Y close neighbours in the network?

$$f(M) = \begin{cases} 1 & \text{if } M \text{ satisfies the feature} \\ 0 & \text{otherwise} \end{cases}$$

Posterior probability of features: $P(f|D) = \sum_M f(M)P(M|D)$

Approximate this sum with MCMC: $P(f|D) = \frac{1}{T} \sum_{i=1}^T f(M_i)$

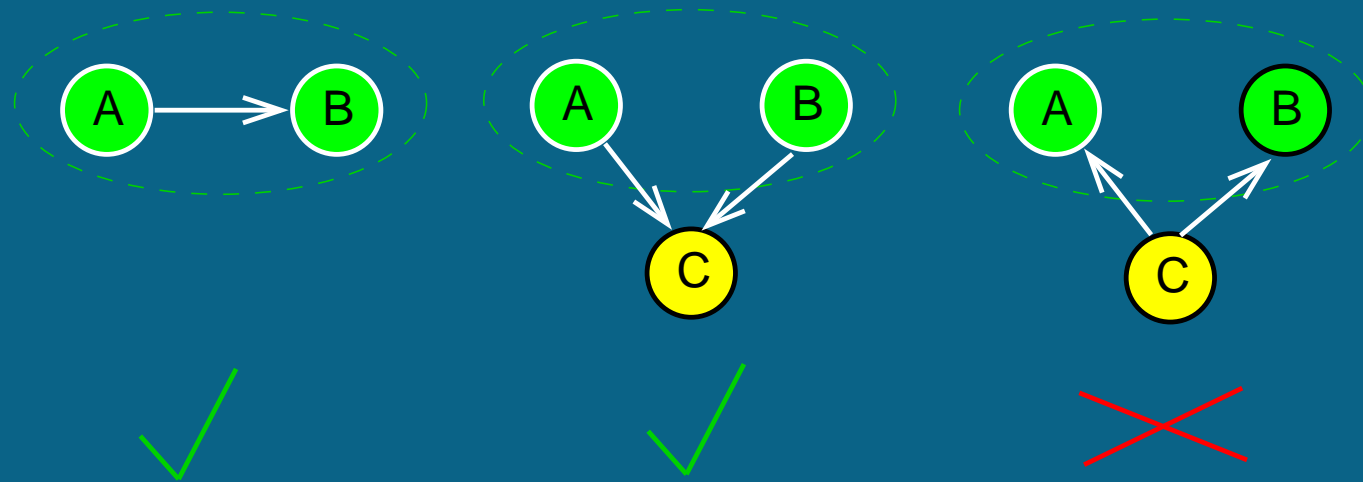
where $\{M_i\}$ is a sample from the posterior obtained with MCMC.

N. Friedman et al. (2000) and D. Pe'er et al (2001)

- Markov neighbours
- Separators
- Order relations

Markov neighbours

- Variables that are not separated by any other measured variable in the domain.



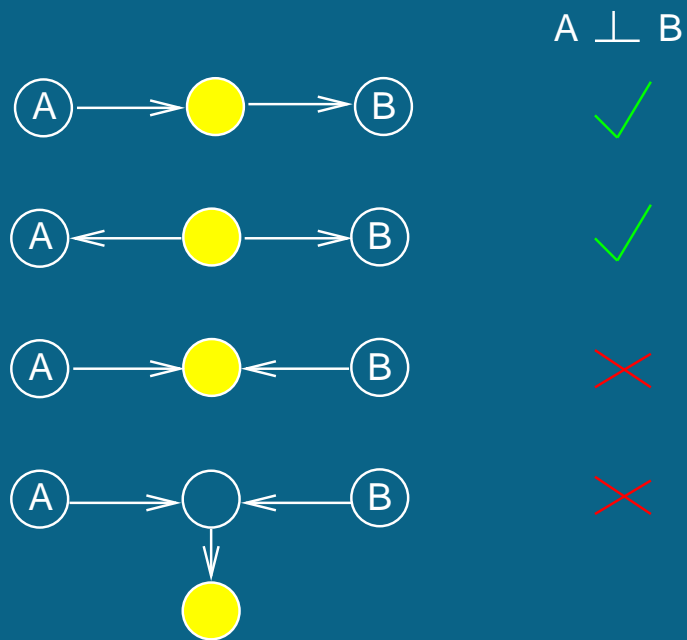
- Indication that two genes are related in some **joint biological interaction or process**.
- Parent-child**: One gene regulating another.
- Spouse relations**: Two genes co-regulating another.

Separators

- Given that A and B are indirectly dependent, **which factors *mediate* this dependence?**

Separators

- Given that A and B are indirectly dependent, **which factors mediate this dependence?**
- Search for a **set of variables Z** such that $A \perp B | Z$.
- Z explains all the dependencies between A and B.



Order relations

- Is A an **ancestor** of B in all the networks of a given equivalence class?
- Does the **network** contain a **directed path** from A to B ?

Order relations

- Is A an **ancestor** of B in all the networks of a given equivalence class?
- Does the **network** contain a **directed path** from A to B ?
Indication that A might be a **causal ancestor** of B .

Outline of the talk

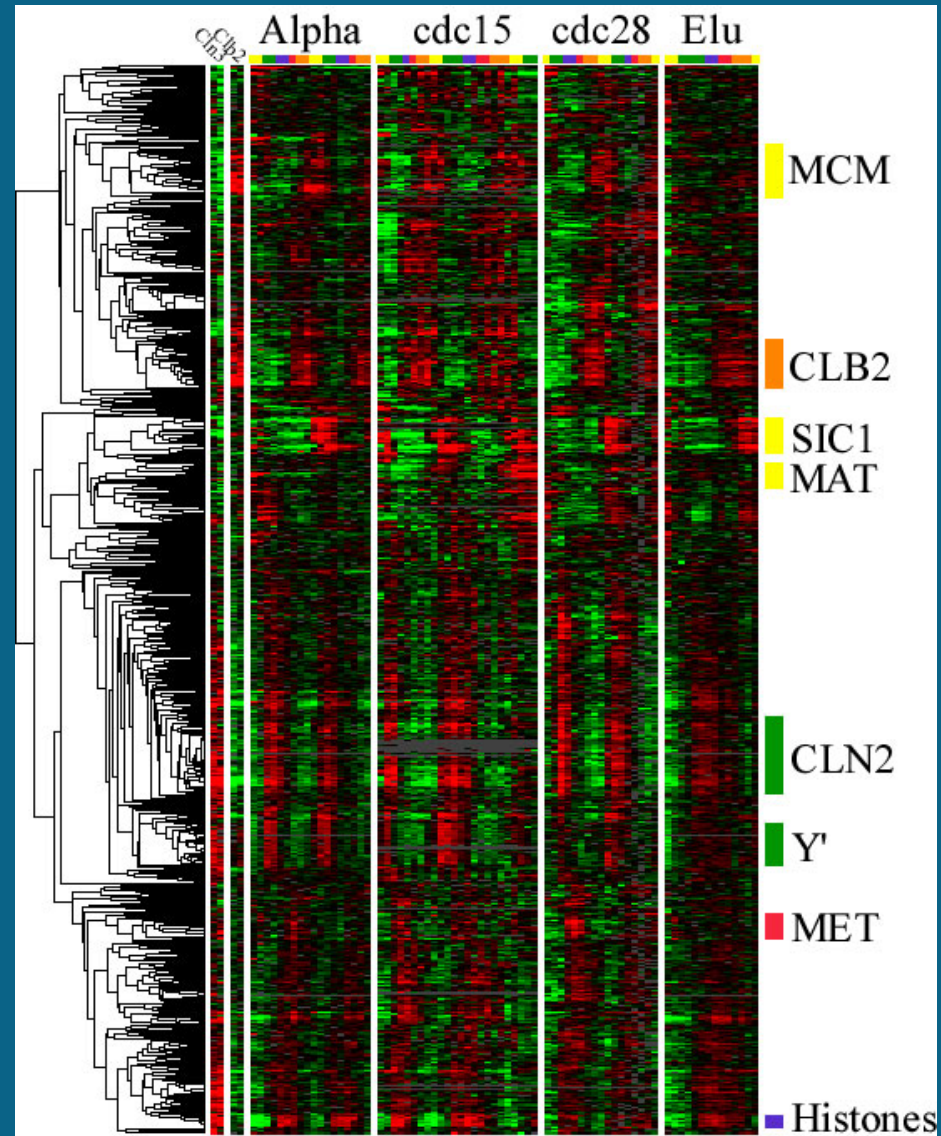
- Recapitulation: Bayesian networks
- Reverse engineering:
Learning networks from data
- **Application to the yeast cell cycle**
- Outlook

Experimental Results

- Friedman, Linial, Nachman, Pe'er (2000)
Journal of Computational Biology 7 (3/4): 601-620
<http://www.cs.huji.ac.il/labs/compbio/expression/#papers>
- Pe'er, Regev, Elidan, Friedman (2001)
Bioinformatics S1: 215-224
<http://www.cs.huji.ac.il/labs/compbio/ismb01/>
- Spellman, Sherlock, Zhang, Iyer, Anders, Eisen, Brown, Botstein, Futcher (1998)
Molecular Biology of the Cell 9 (12) :3273-97
<http://cellcycle-www.stanford.edu/>

- Yeast cell cycle (*S. cerevisiae*).
- Six time series under different experimental conditions, altogether 76 gene expression measurements.
- 800 genes.

- **Yeast** cell cycle (*S. cerevisiae*).
- Six time series under different experimental conditions, altogether **76 gene expression measurements**.
- **800 genes**.
- No biological **prior knowledge**.
- Do not take into account the **temporal aspect** of the measurements. Introduce an additional root node representing the cell cycle phase.
- **Discretization**: Underexpressed (-1), normal (0), overexpressed (1).

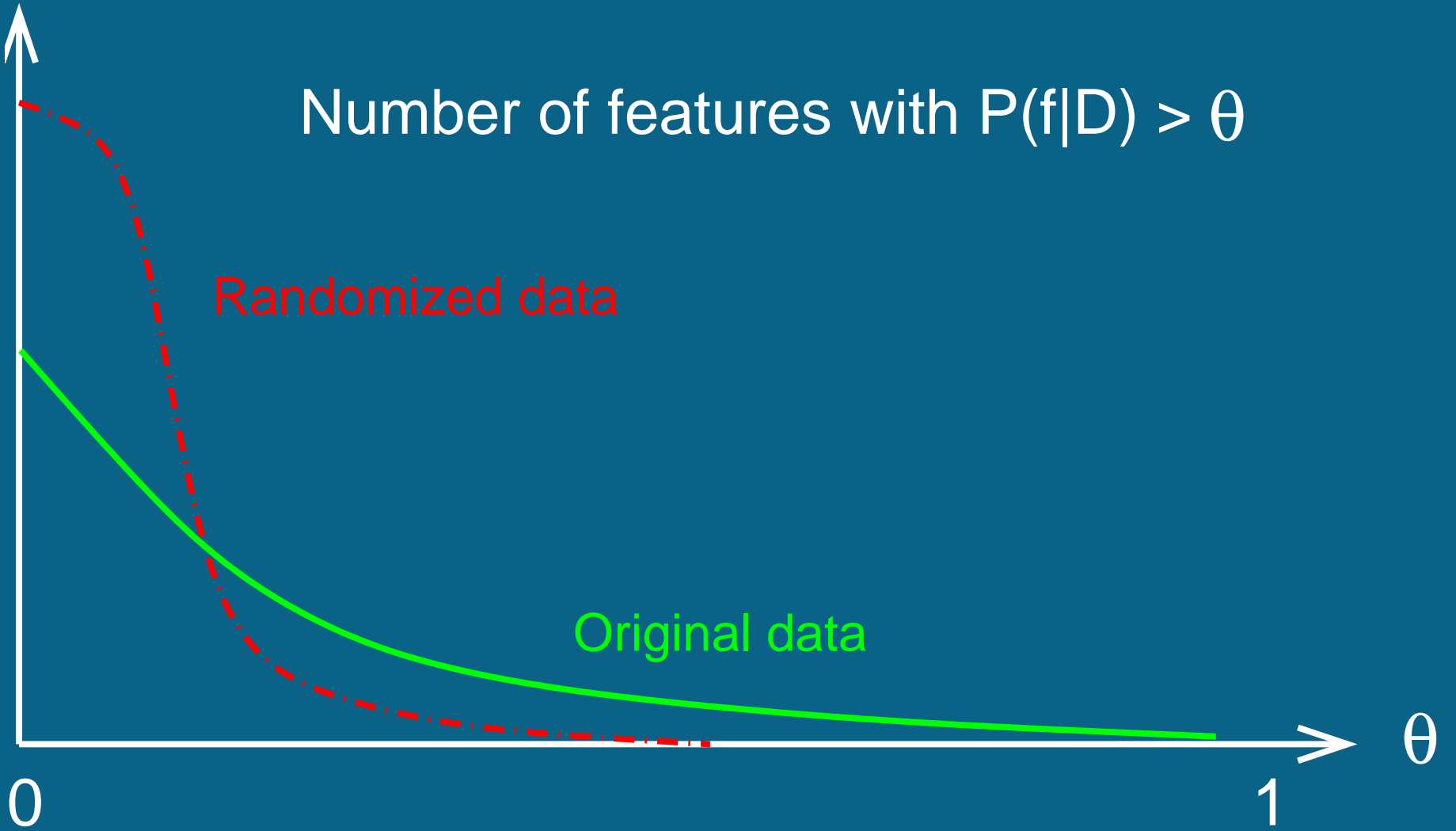


From Spellman et al., <http://cellcycle-www.stanford.edu/>

Number of features with $P(f|D) > \theta$

Randomized data

Original data



Order relations

Confidence in X being an ancestor of Y :

$$P(X \rightarrow Y | D)$$

Dominance score of X : $\sum_Y P(X \rightarrow Y | D)$

Genes with high dominance scores are **indicative** of potential **causal** sources of the cell cycle process.

Order relations

Confidence in X being an ancestor of Y :

$$P(X \rightarrow Y | D)$$

Dominance score of X : $\sum_Y P(X \rightarrow Y | D)$

Genes with high dominance scores are **indicative** of potential **causal** sources of the cell cycle process.

Finding: Only a few genes dominate the order.

Dominant genes in the ordering relations

CLN1	Role in cell cycle start
CLN2	Role in cell cycle start
CDC5	Cell cycle control , required for exit from mitosis
RAD53	Cell cycle control : checkpoint function
RFA2	Involved in nucleotide excision repair
PLO30	Required for DNA replication and repair
MSH6	Required for mismatch repair in mitosis and meiosis

DNA repair is associated with **transcription initiation**: DNA areas which are more active in transcription are also repaired more frequently.

Markov relations

$P(X \leftrightarrow Y|D)$: Indication that genes are functionally related.

Markov relations

$P(X \leftrightarrow Y|D)$: Indication that genes are **functionally related**.

- Most Markov pairs: **Intracluster pairings** with high correlation in their expression.
- **But:** Genes where $P(X \leftrightarrow Y|D)$ is high and correlation is low.

Markov relations

$P(X \leftrightarrow Y|D)$: Indication that genes are **functionally related**.

- Most Markov pairs: **Intracluster pairings** with high correlation in their expression.
- **But:** Genes where $P(X \leftrightarrow Y|D)$ is high and correlation is low.

FAR1	Role in a mating type switch
ASH1	Role in a mating type switch
LAC1	GPI transport protein
YNL300W	Modified by GPI
SAG1	Induces the mating process
MF-ALPHA-1	Participates in the mating process

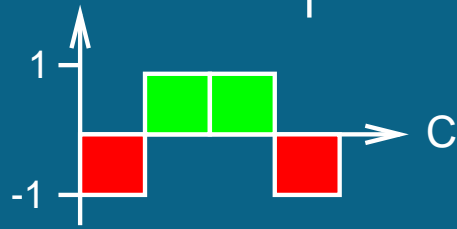
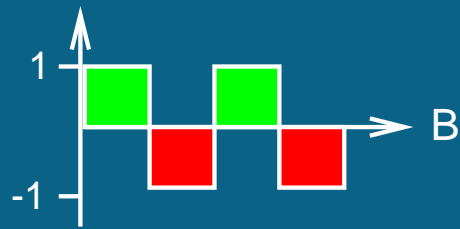
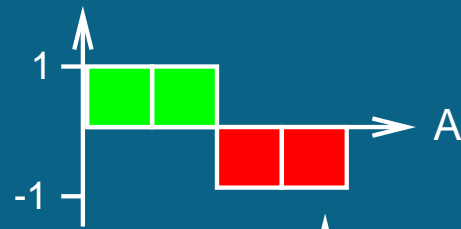
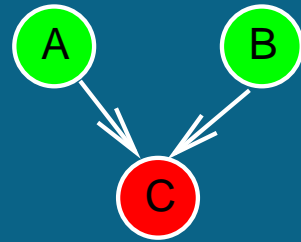
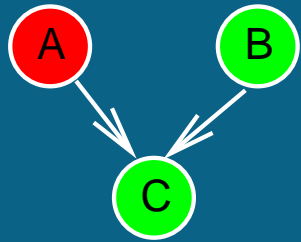
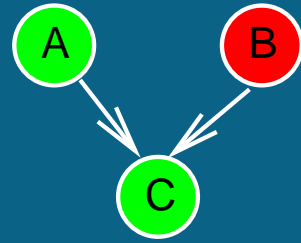
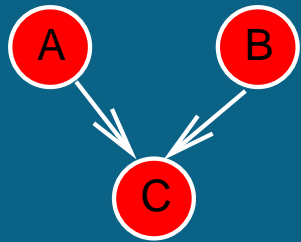
Markov relations

$P(X \leftrightarrow Y|D)$: Indication that genes are **functionally related**.

- Most Markov pairs: **Intracluster pairings** with high correlation in their expression.
- **But:** Genes where $P(X \leftrightarrow Y|D)$ is high and correlation is low.

FAR1	Role in a mating type switch
ASH1	Role in a mating type switch
LAC1	GPI transport protein
YNL300W	Modified by GPI
SAG1	Induces the mating process
MF-ALPHA-1	Participates in the mating process

Advantage of Bayesian networks: **context-specific** and **non-linear**.



Separator relations

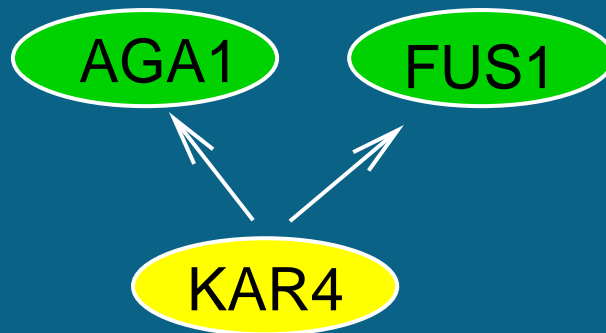
Explain away dependencies.

- **Separators**: Known transcriptional regulators
- **Separated genes**: Share functional roles

Separator relations

Explain away dependencies.

- **Separators:** Known transcriptional regulators
- **Separated genes:** Share functional roles



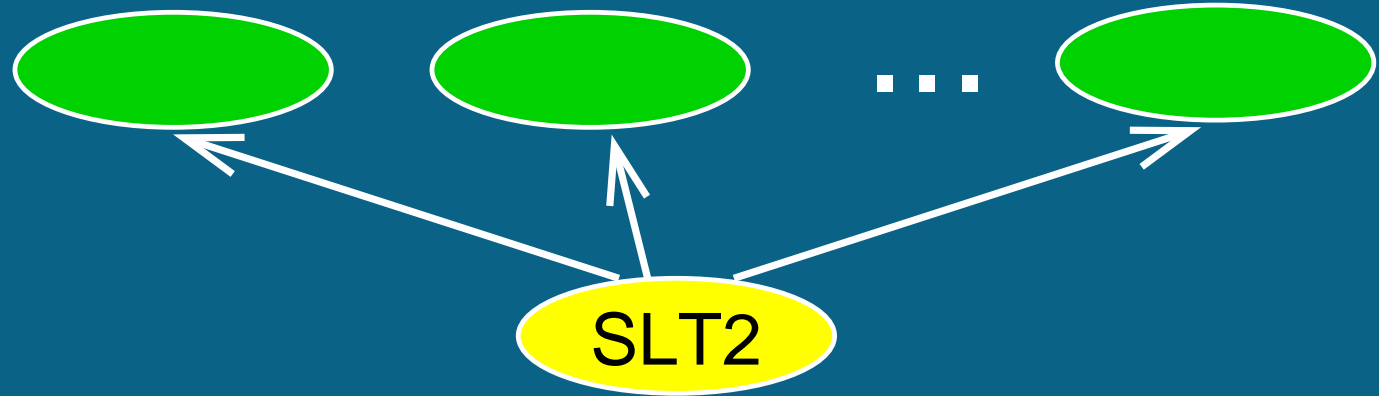
KAR4

AGA1, FUS1

Transcriptional regulator of nuclear fusion genes.

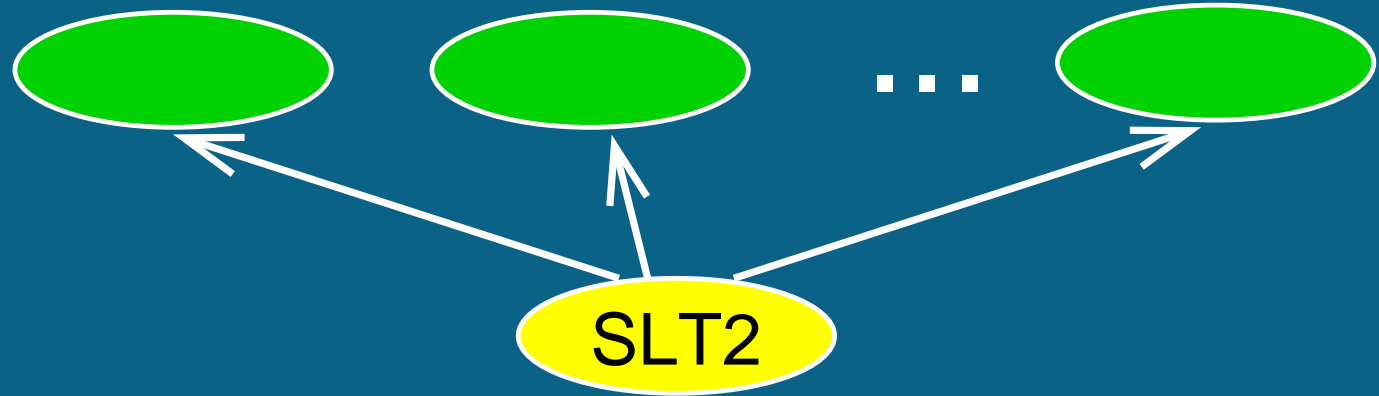
Cell fusion genes

Low osmolarity response genes

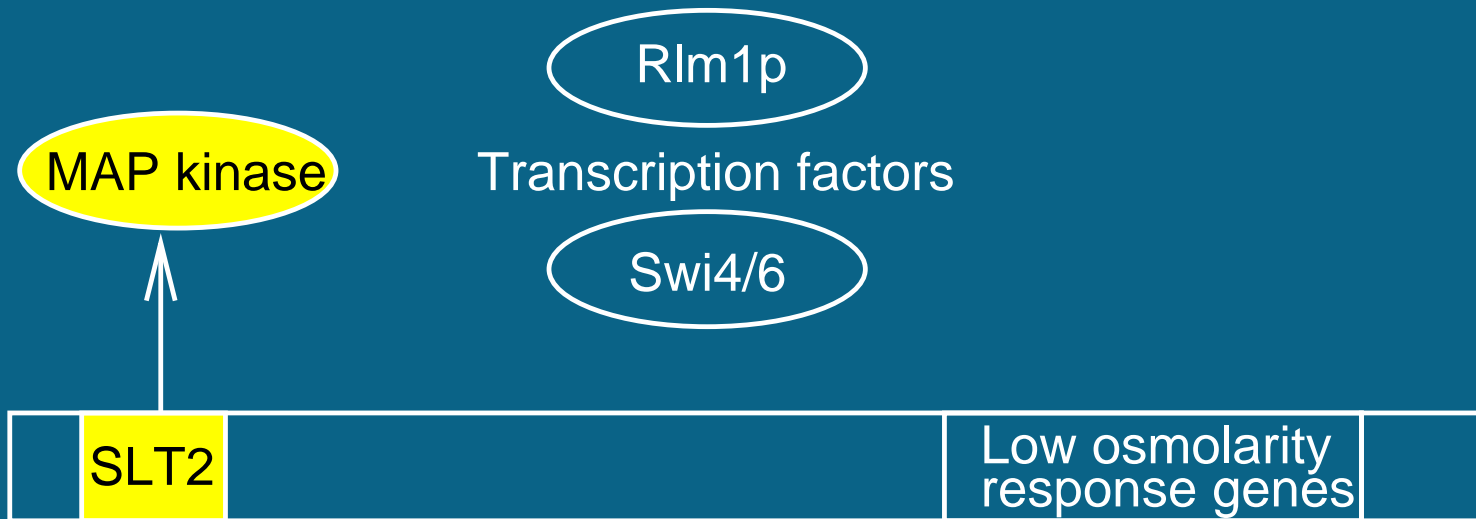
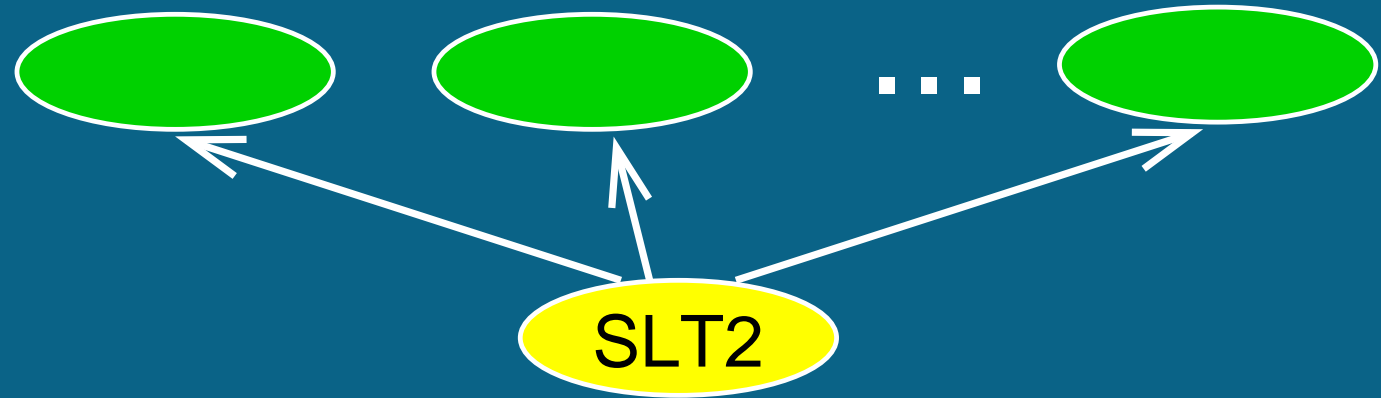


	SLT2		Low osmolarity response genes	
--	------	--	-------------------------------	--

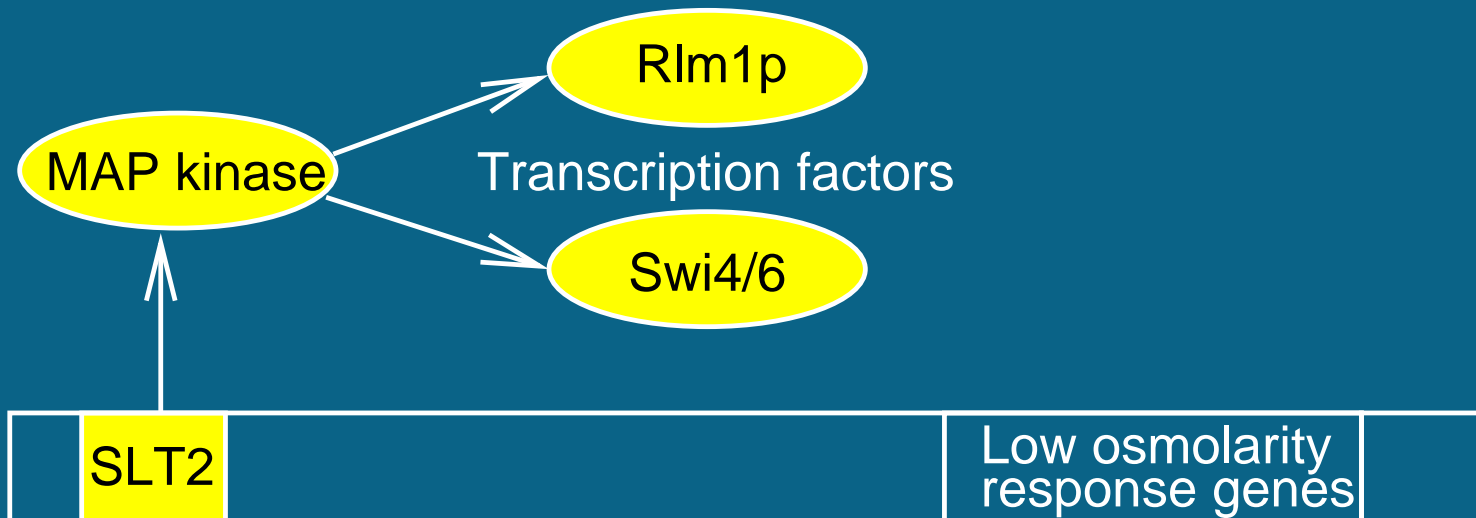
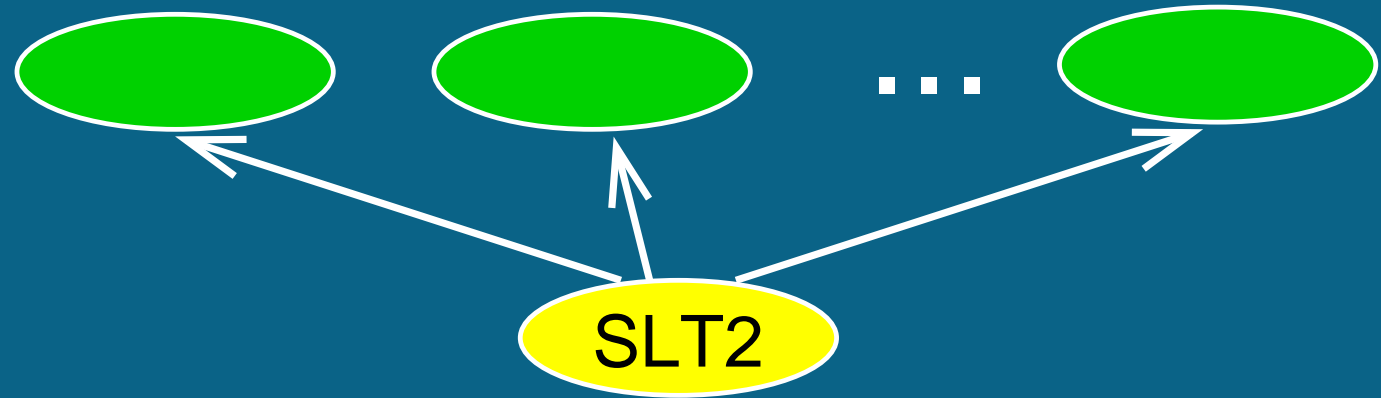
Low osmolarity response genes



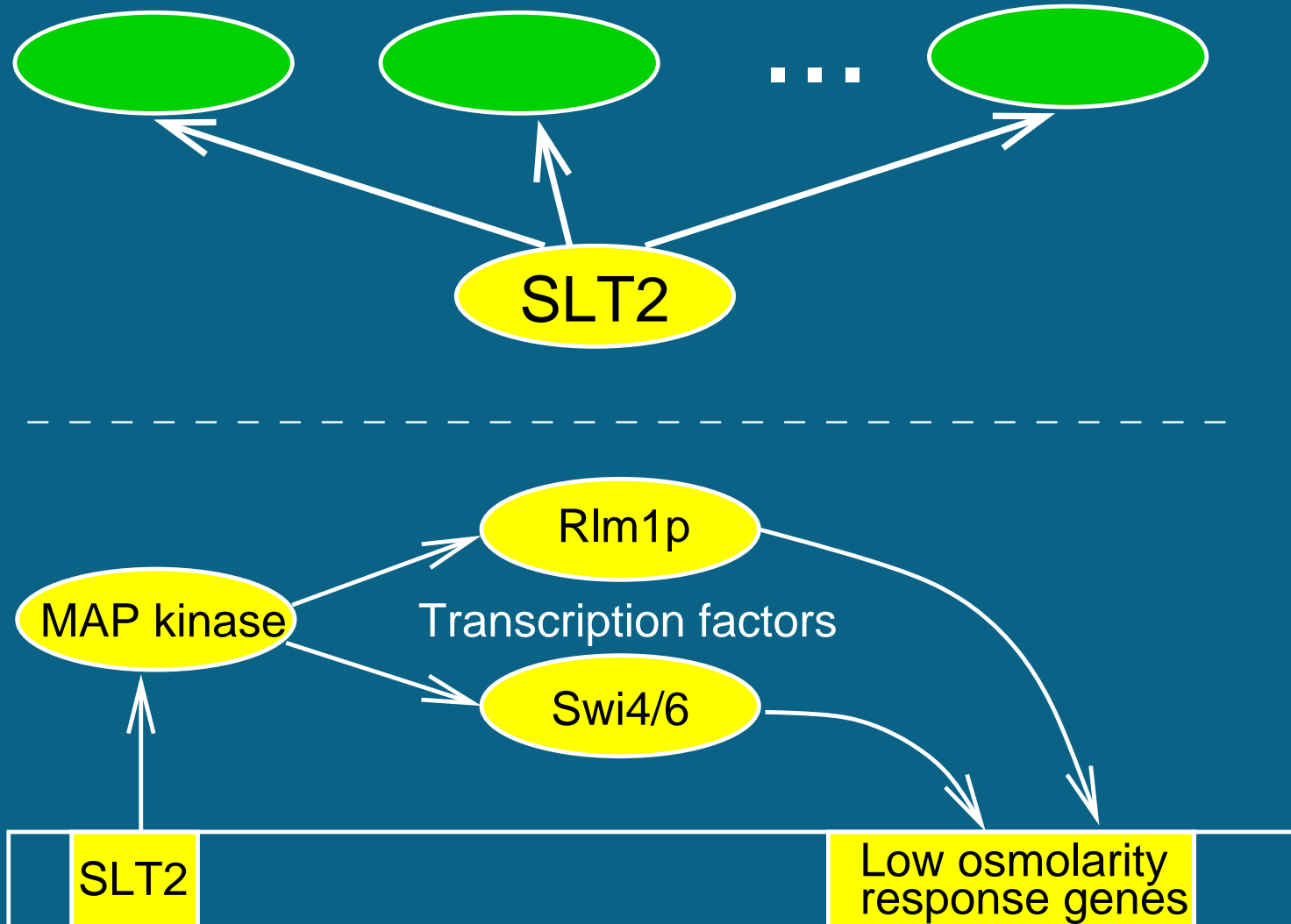
Low osmolarity response genes



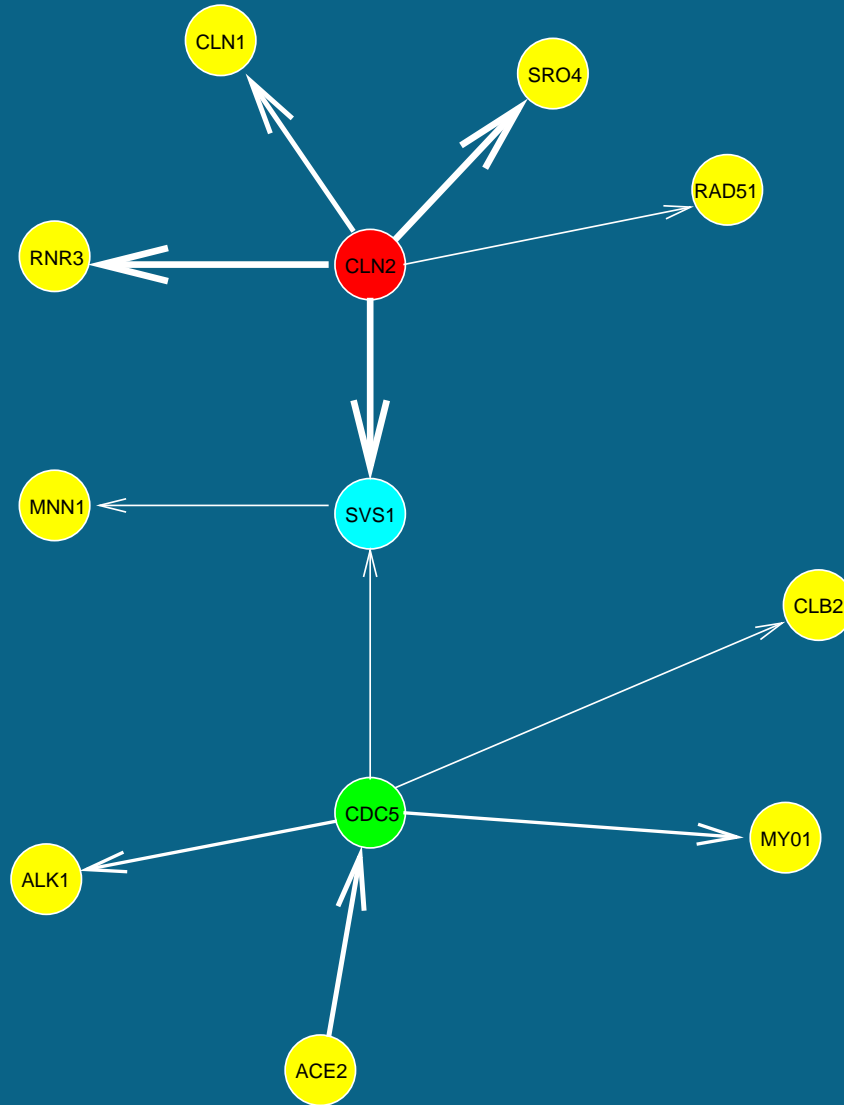
Low osmolarity response genes



Low osmolarity response genes



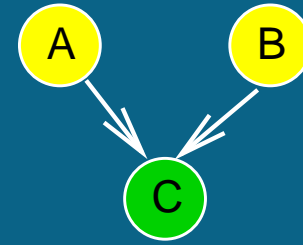
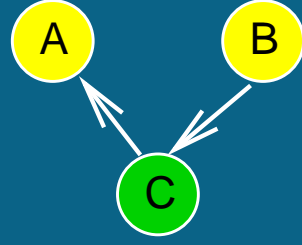
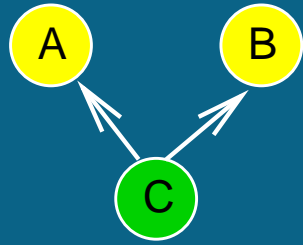
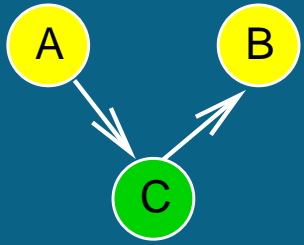
Subnetworks

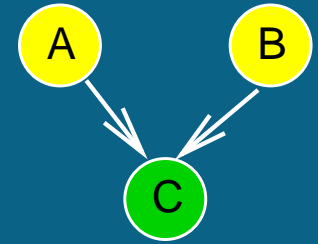
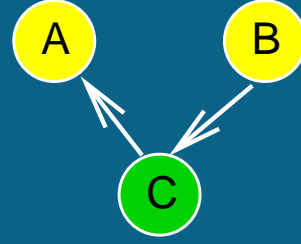
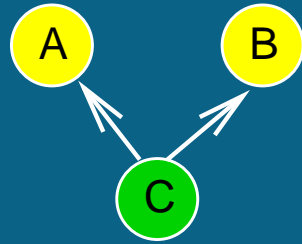
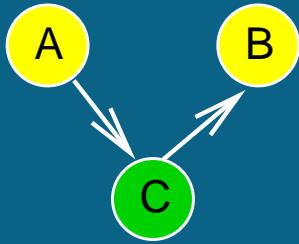


Outline of the talk

- Recapitulation: Bayesian networks
- Reverse engineering:
Learning networks from data
- Application to the yeast cell cycle
- **Outlook**

Causal (directed) interactions





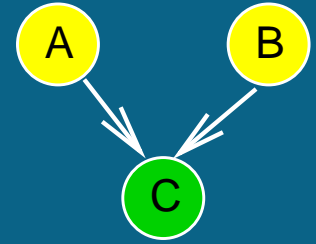
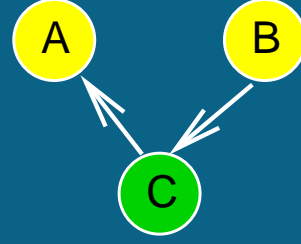
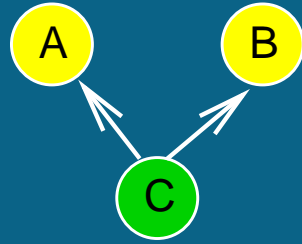
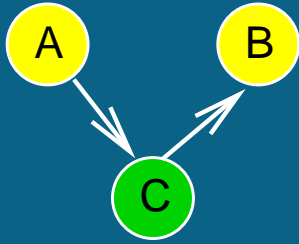
$P(A,B,C) =$

$P(B|C) P(C|A) P(A)$

$P(A|C) P(B|C) P(C)$

$P(A|C) P(C|B) P(B)$

$P(C|A,B) P(A) P(B)$



$P(A,B,C) =$

$$P(B|C) P(C|A) P(A)$$

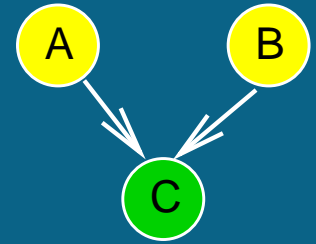
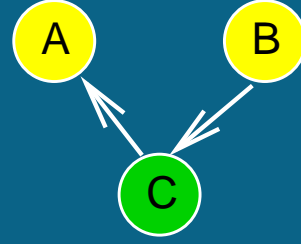
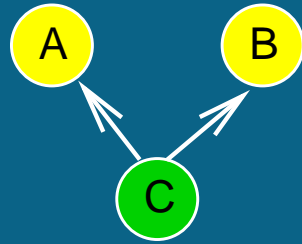
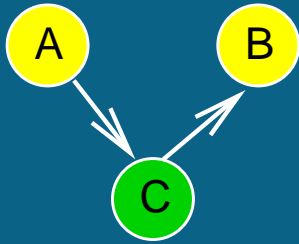
$\underbrace{\hspace{10em}}$
 $P(A|C) P(C)$

$$P(A|C) P(B|C) P(C)$$

$$P(A|C) P(C|B) P(B)$$

$\underbrace{\hspace{10em}}$
 $P(B|C) P(C)$

$$P(C|A,B) P(A) P(B)$$



$P(A,B,C) =$

$$P(B|C) P(C|A) P(A)$$

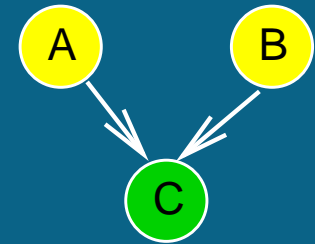
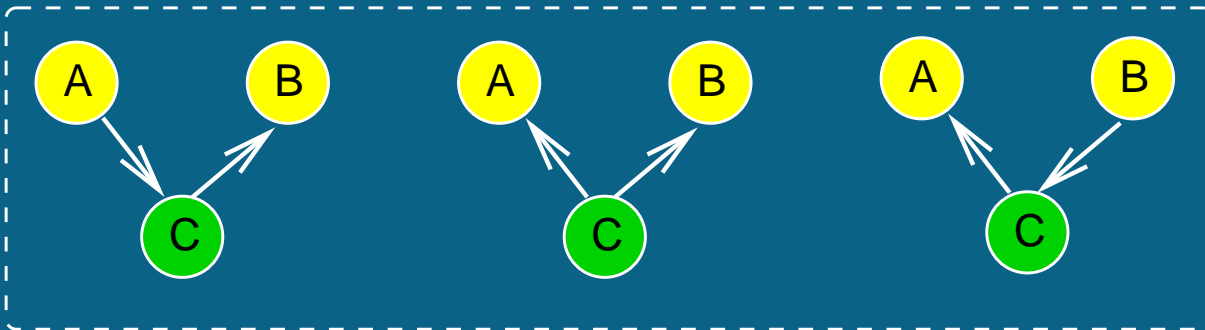
$$P(A|C) P(B|C) P(C)$$

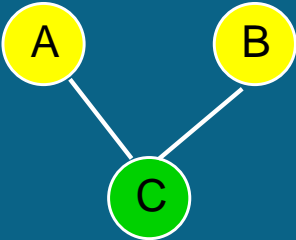
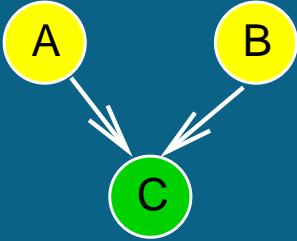
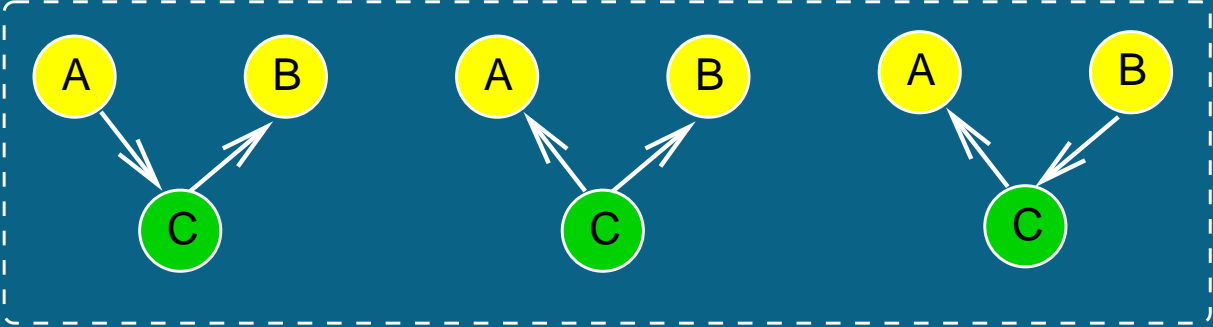
$$P(A|C) P(C|B) P(B)$$

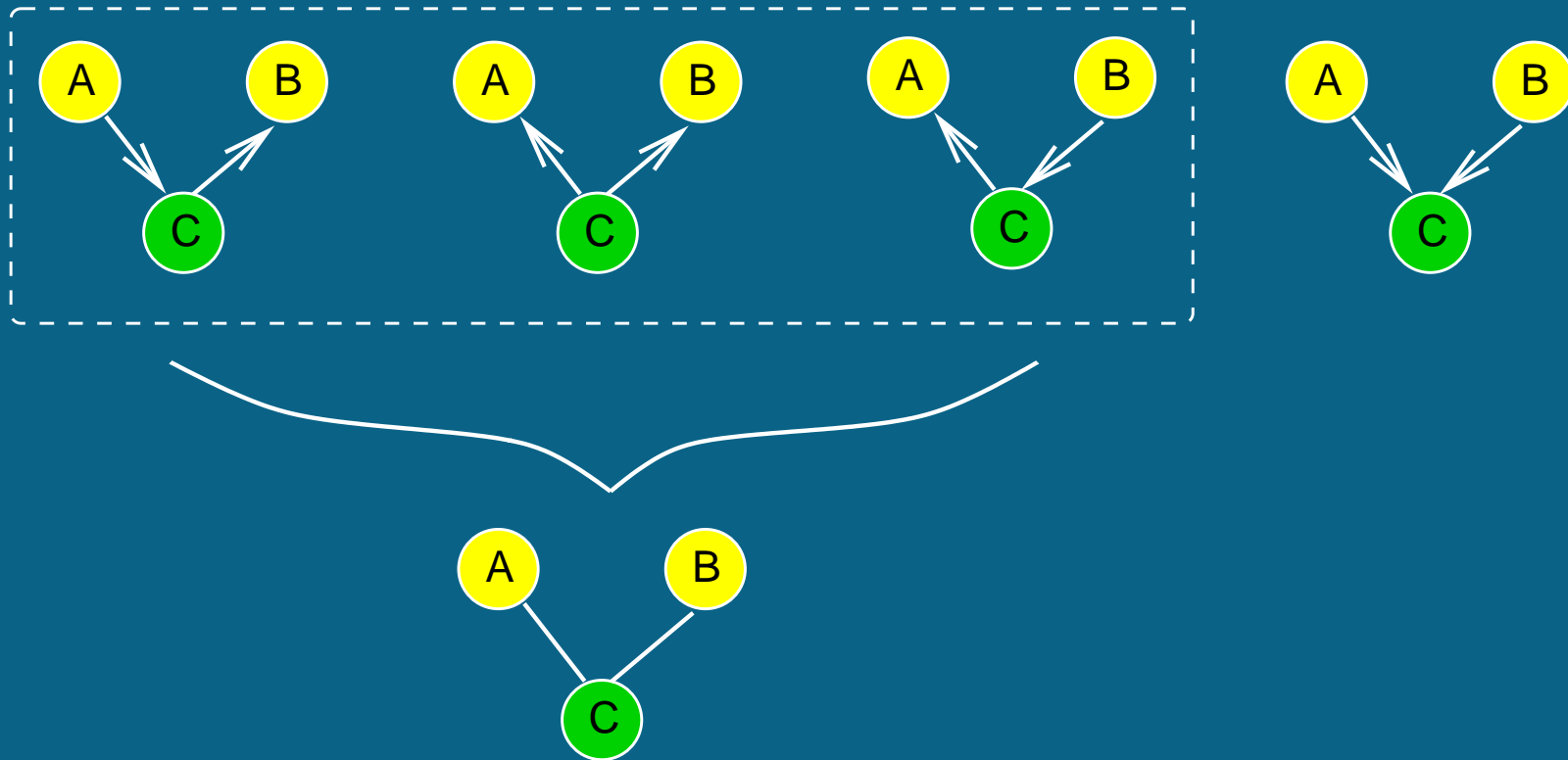
$$P(C|A,B) P(A) P(B)$$

$$\underbrace{\hspace{10em}}_{P(A|C) P(C)}$$

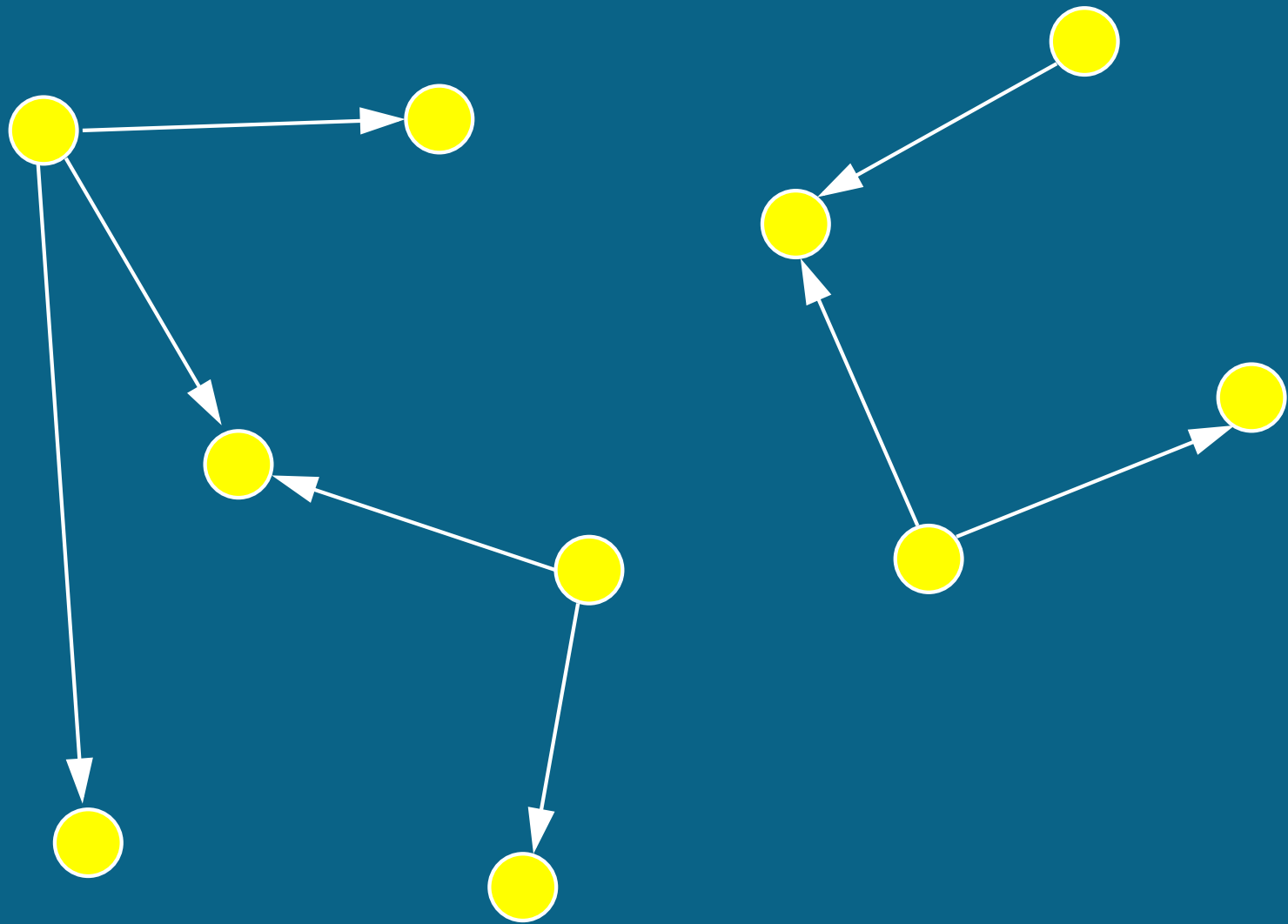
$$\underbrace{\hspace{10em}}_{P(B|C) P(C)}$$

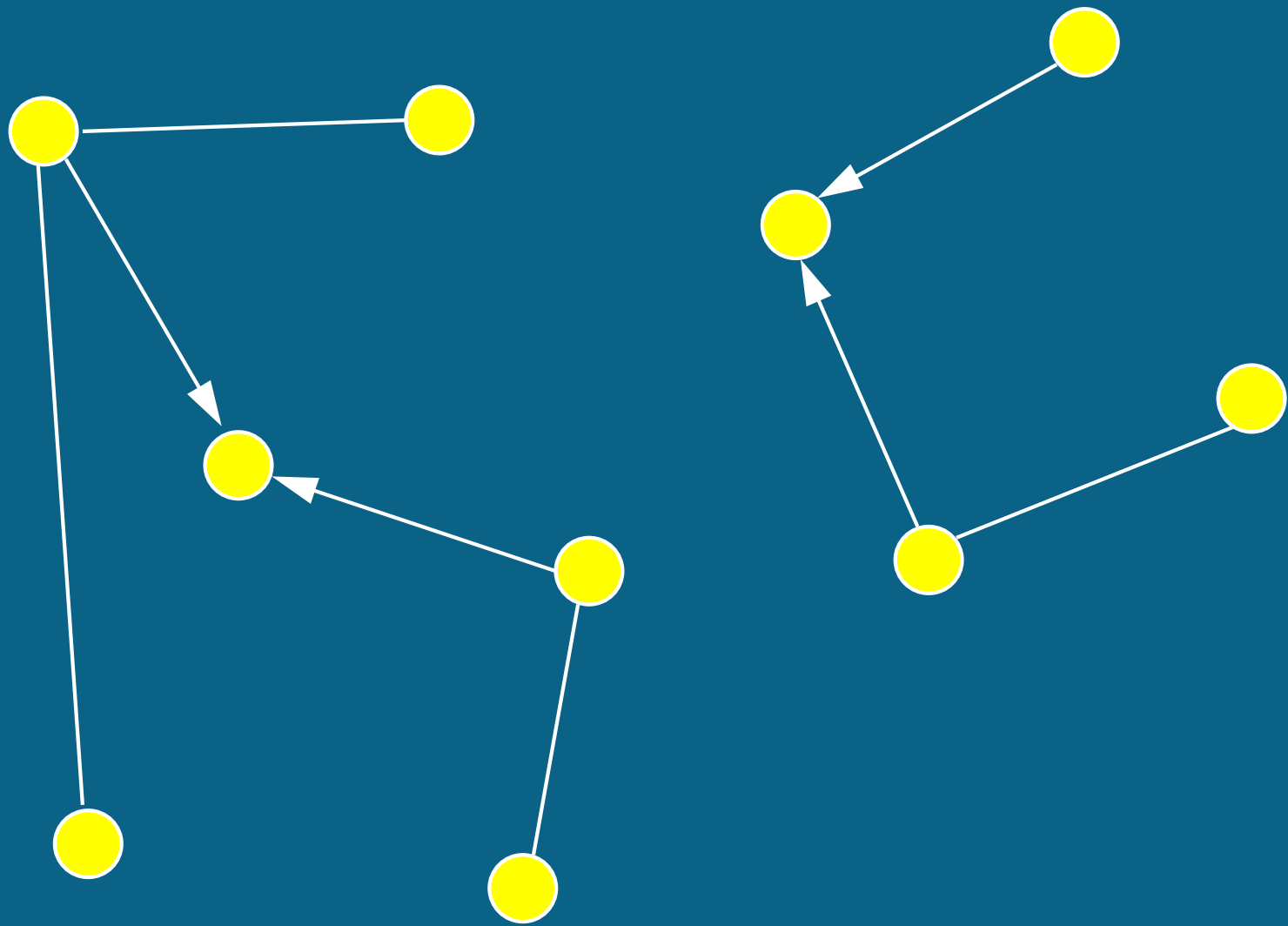




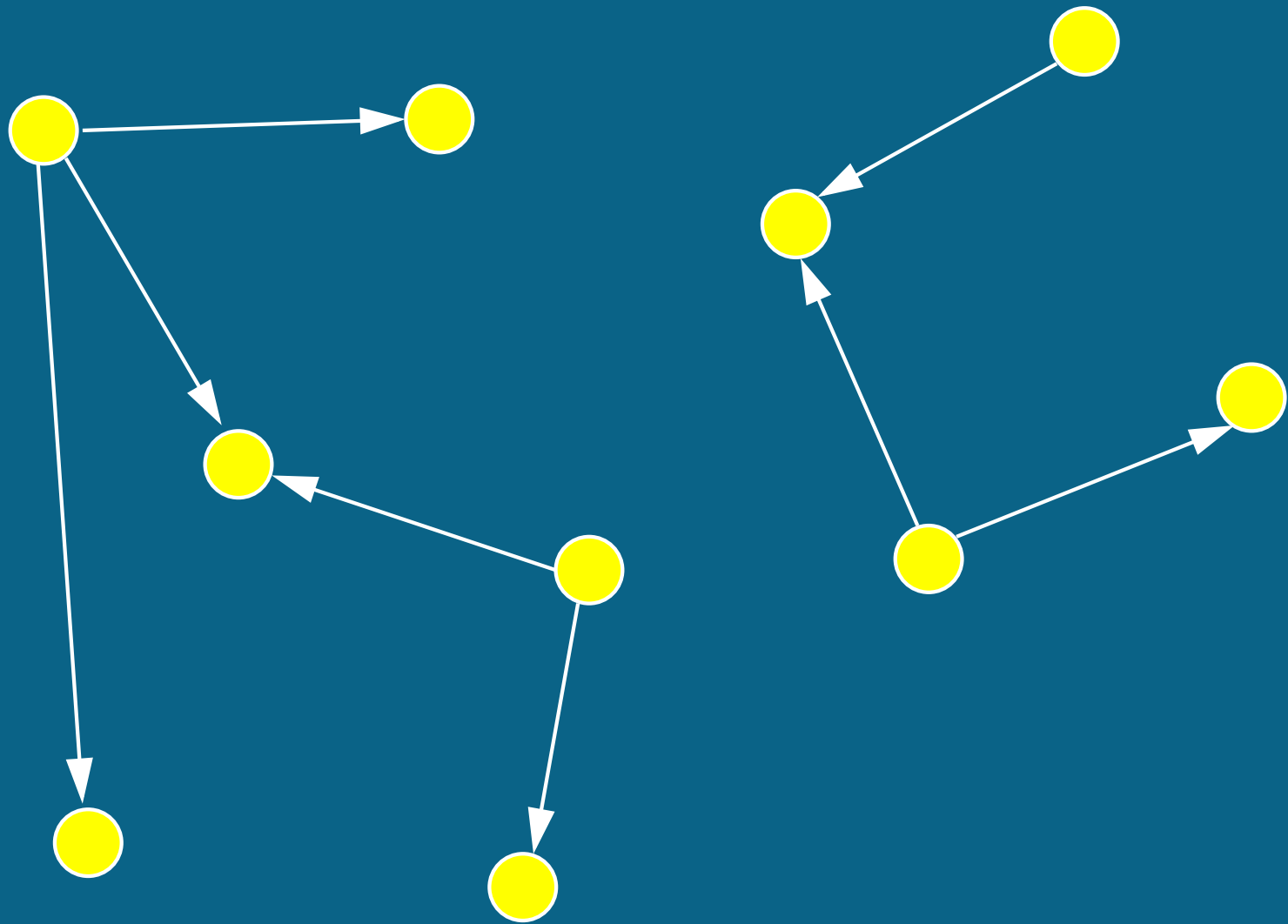


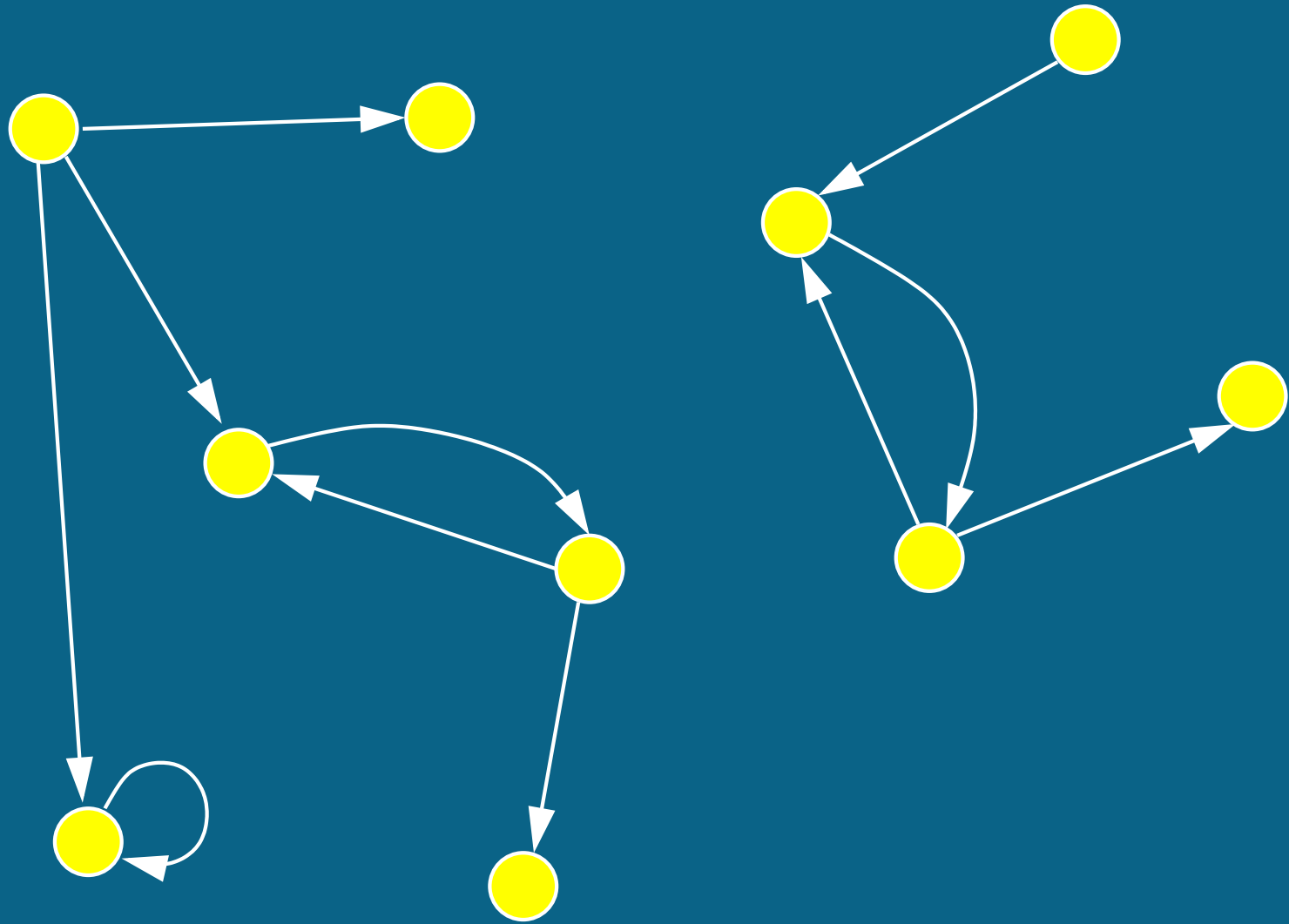
- We cannot distinguish between **equivalent networks** (same score).
- Loss of clues about **causal interactions**.





Feedback

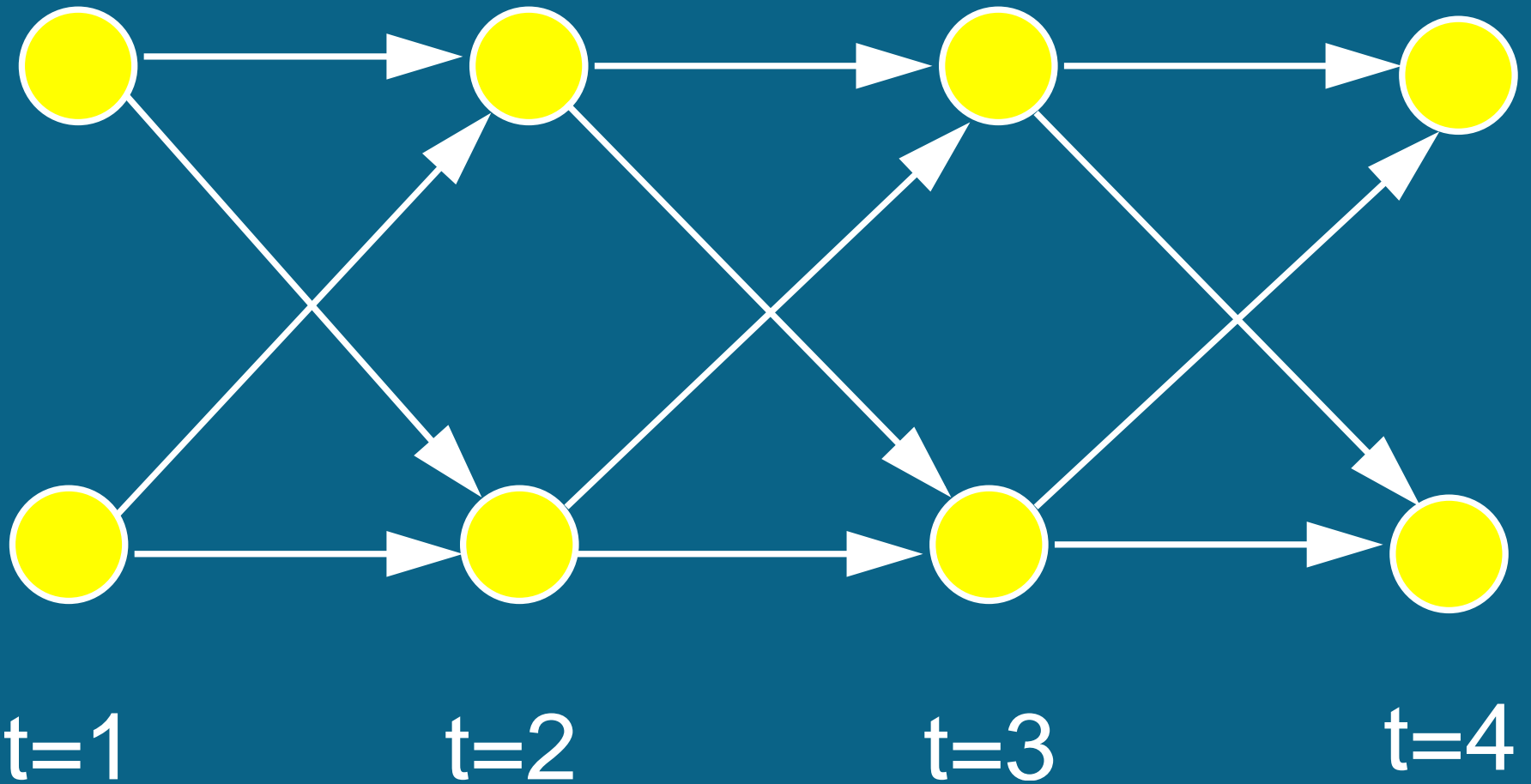




Causality and feedback

Dynamic Bayesian networks

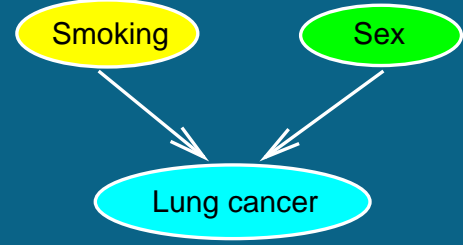
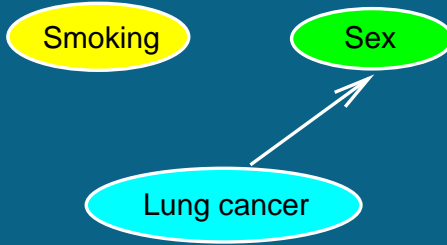
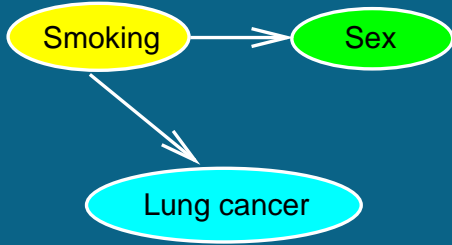
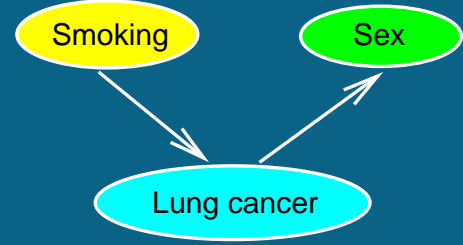
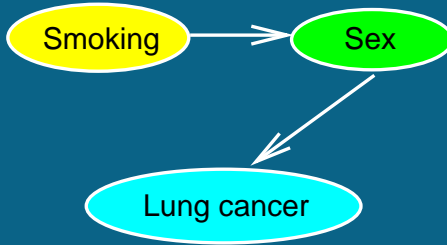
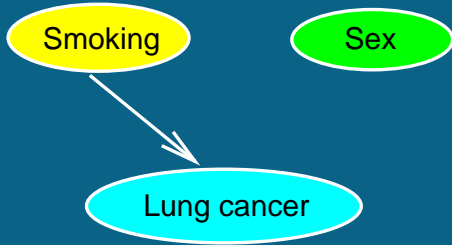


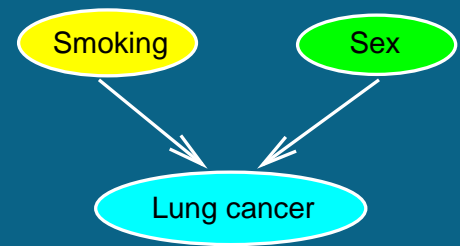
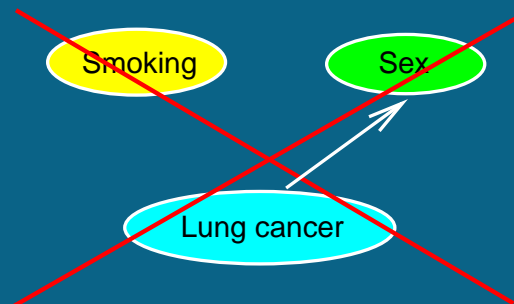
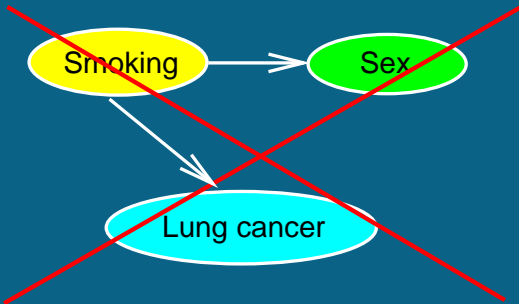
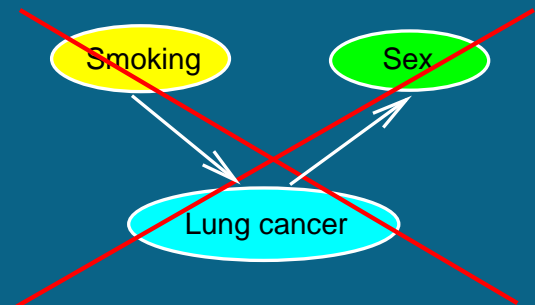
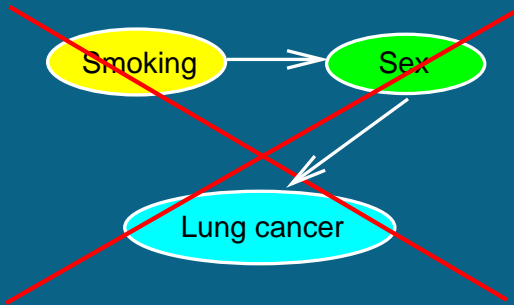
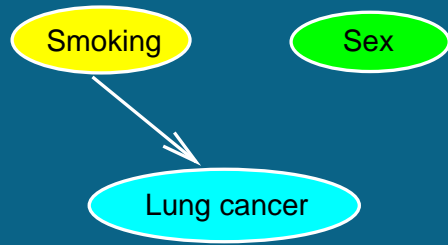


Make use of **biological prior knowledge** (from data bases, e.g. GO annotations) to **reduce the uncertainty of inference**.

$$P(\mathcal{M}|D) \propto P(D|\mathcal{M})P(\mathcal{M})$$

Use an **informative prior** $P(\mathcal{M})$.





- David Heckerman (1996)
Technical report: A Tutorial on Learning Bayesian Networks
- Paul Krause (1998)
Technical report: Learning Probabilistic Networks
- Friedman, Koller (2002)
Machine Learning, to appear
- Chickering (1995)
Uncertainty in Artificial Intelligence 95, 87-98
- Friedman, Linial, Nachman, Pe'er (2000)
Journal of Computational Biology 7 (3/4): 601-620
- Pe'er, Regev, Elidan, Friedman (2001)
Bioinformatics S1: 215-224
- Spellman, Sherlock, Zhang, Iyer, Anders, Eisen, Brown, Botstein, Futcher (1998)
Molecular Biology of the Cell 9 (12) :3273-97