Learning genetic networks from gene expression data

Dirk Husmeier
Biomathematics & Statistics Scotland (BioSS)
JCMB, The King's Buildings, Edinburgh EH9 3JZ
United Kingdom
http://www.bioss.ac.uk/∼dirk

Paradigm Shift in Molecular Biology

Paradigm Shift in Molecular Biology

Pre-Genomic

- Reductionist (DNA or RNA or protein)
- Generally qualitative, non-numeric
- Hypothesis driven

Paradigm Shift in Molecular Biology

Pre-Genomic

- Reductionist (DNA or RNA or protein)
- Generally qualitative, non-numeric
- Hypothesis driven

Post-Genomic

- Holistic, systems approach: DNA and RNA and protein
- Quantitative, highly numeric
- Data driven

Paradigm Shift in Molecular Biology

Pre-Genomic

- Reductionist (DNA or RNA or protein)
- Generally qualitative, non-numeric
- Hypothesis driven

Post-Genomic

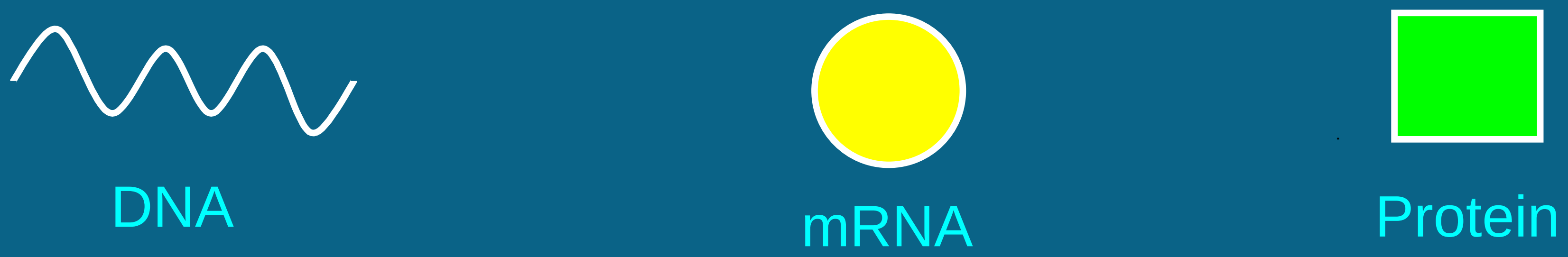
- Holistic, systems approach: DNA and RNA and protein
- Quantitative, highly numeric
- Data driven

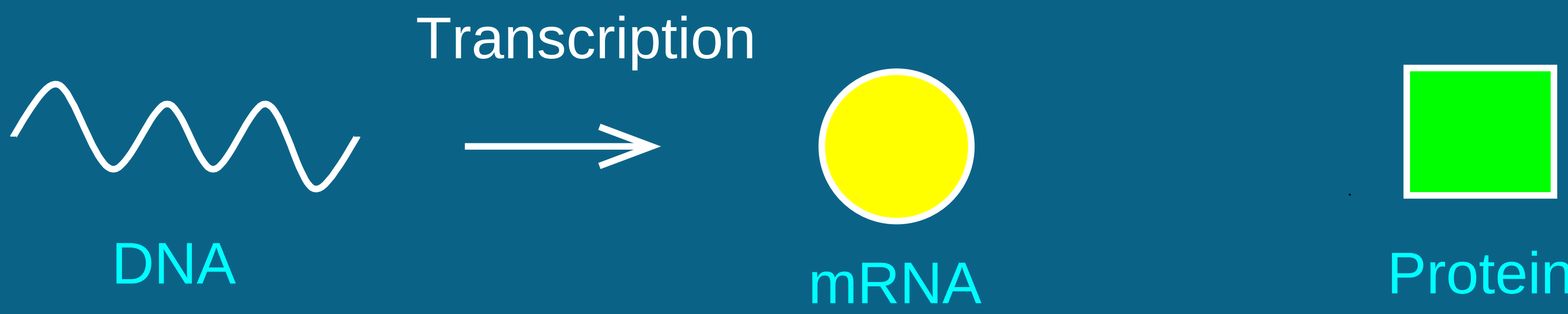
$\Longrightarrow$ Need for **machine learning** and **statistics**

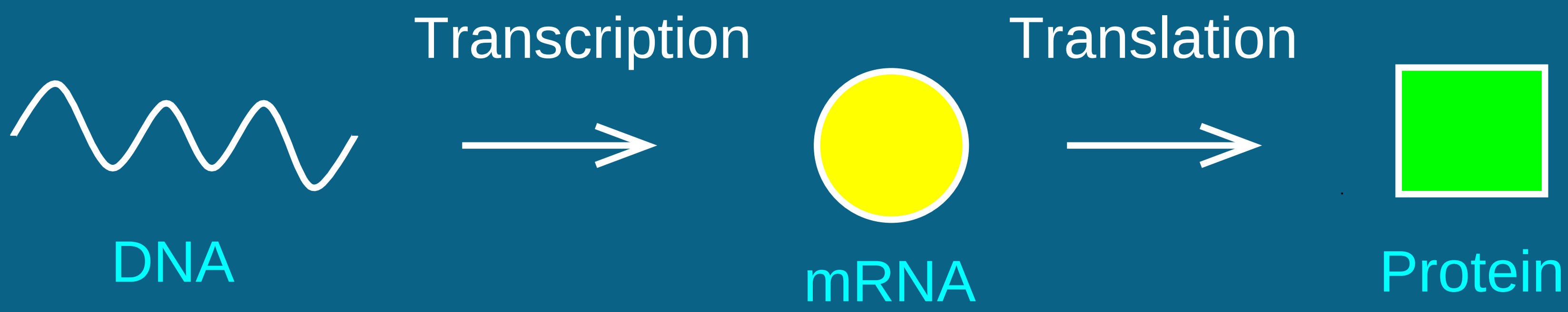
Inferring genetic networks

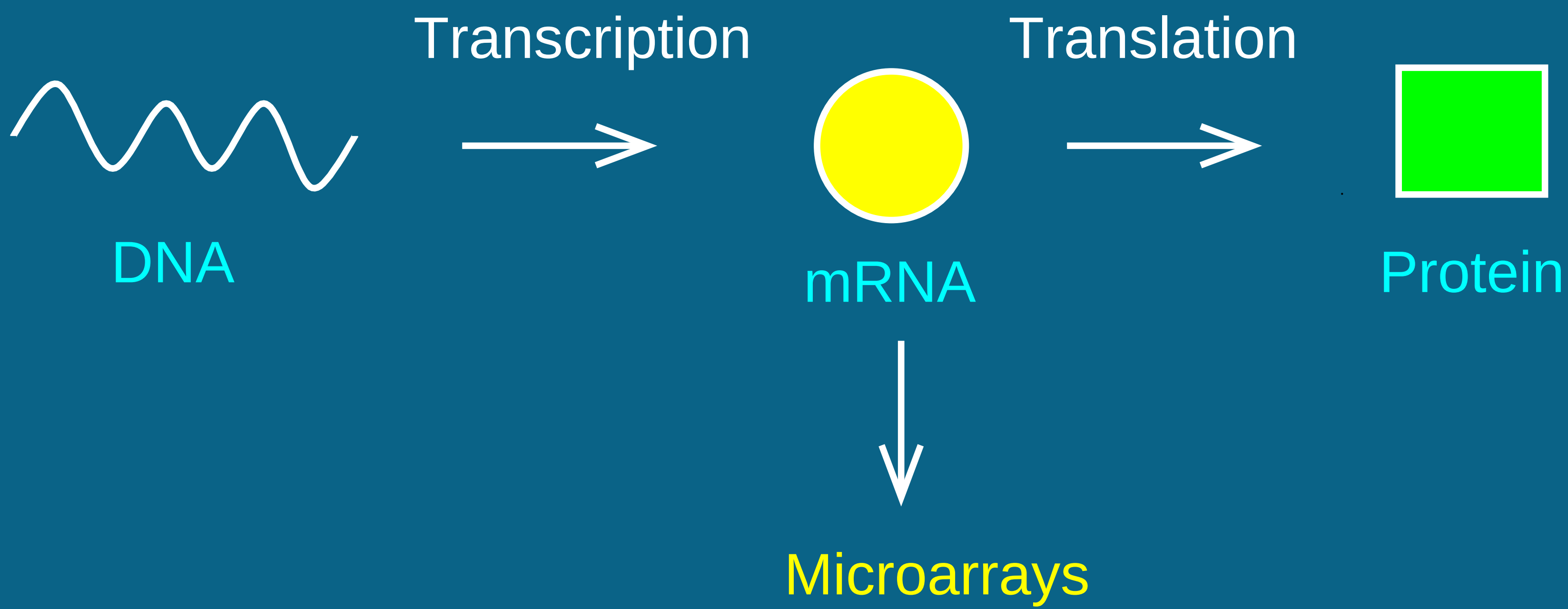
from

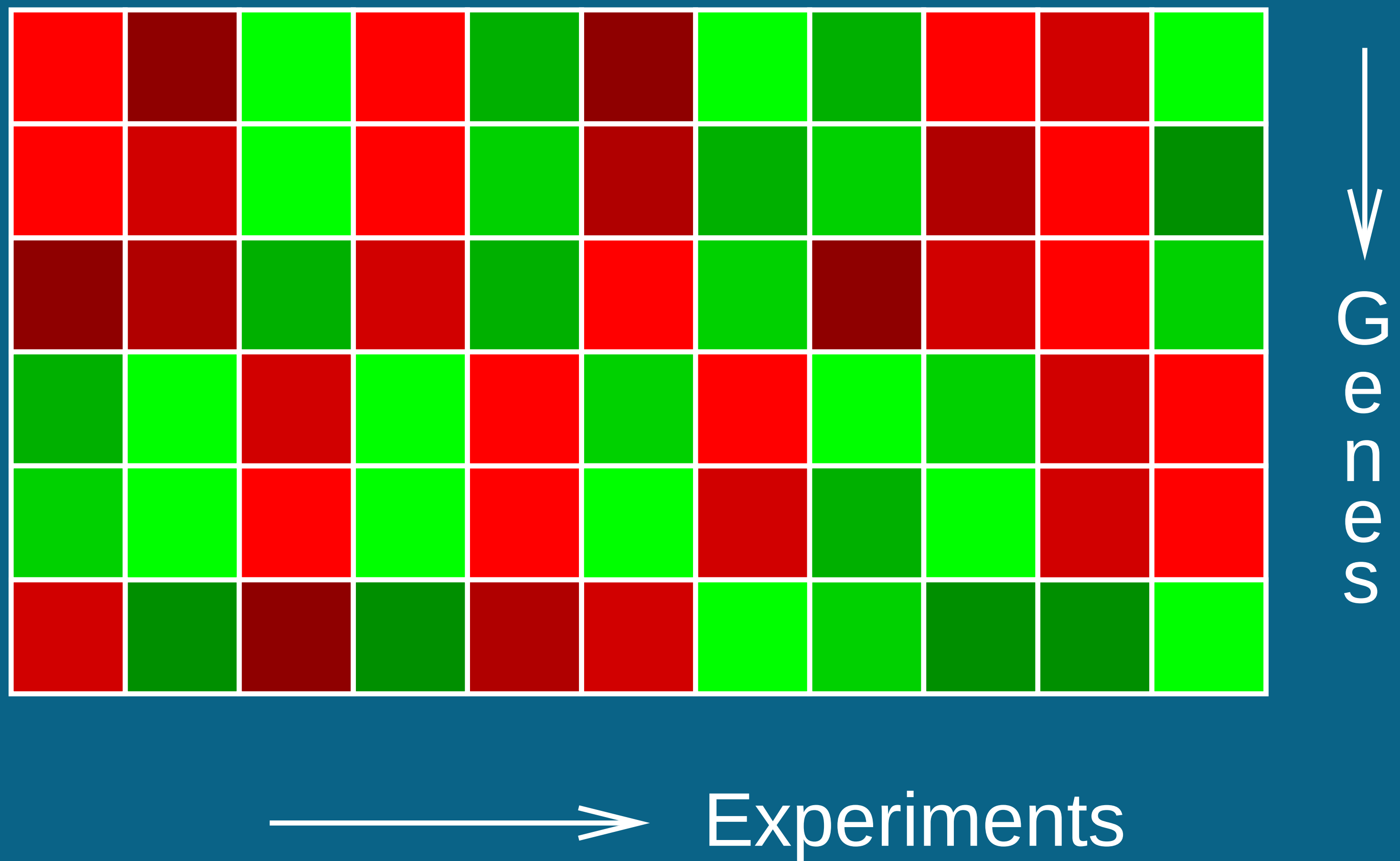
microarray gene expression data

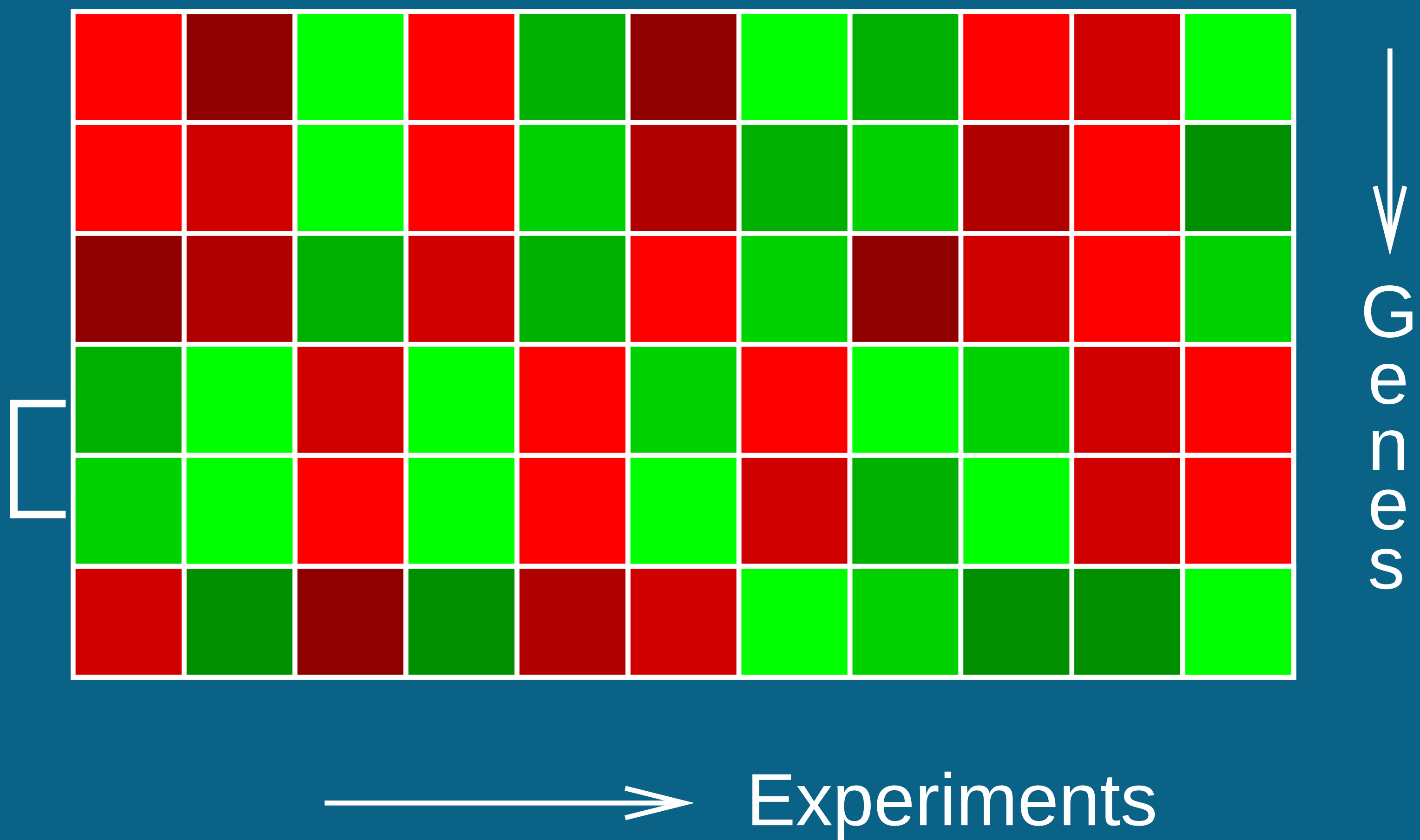

DNA
mRNA
Protein

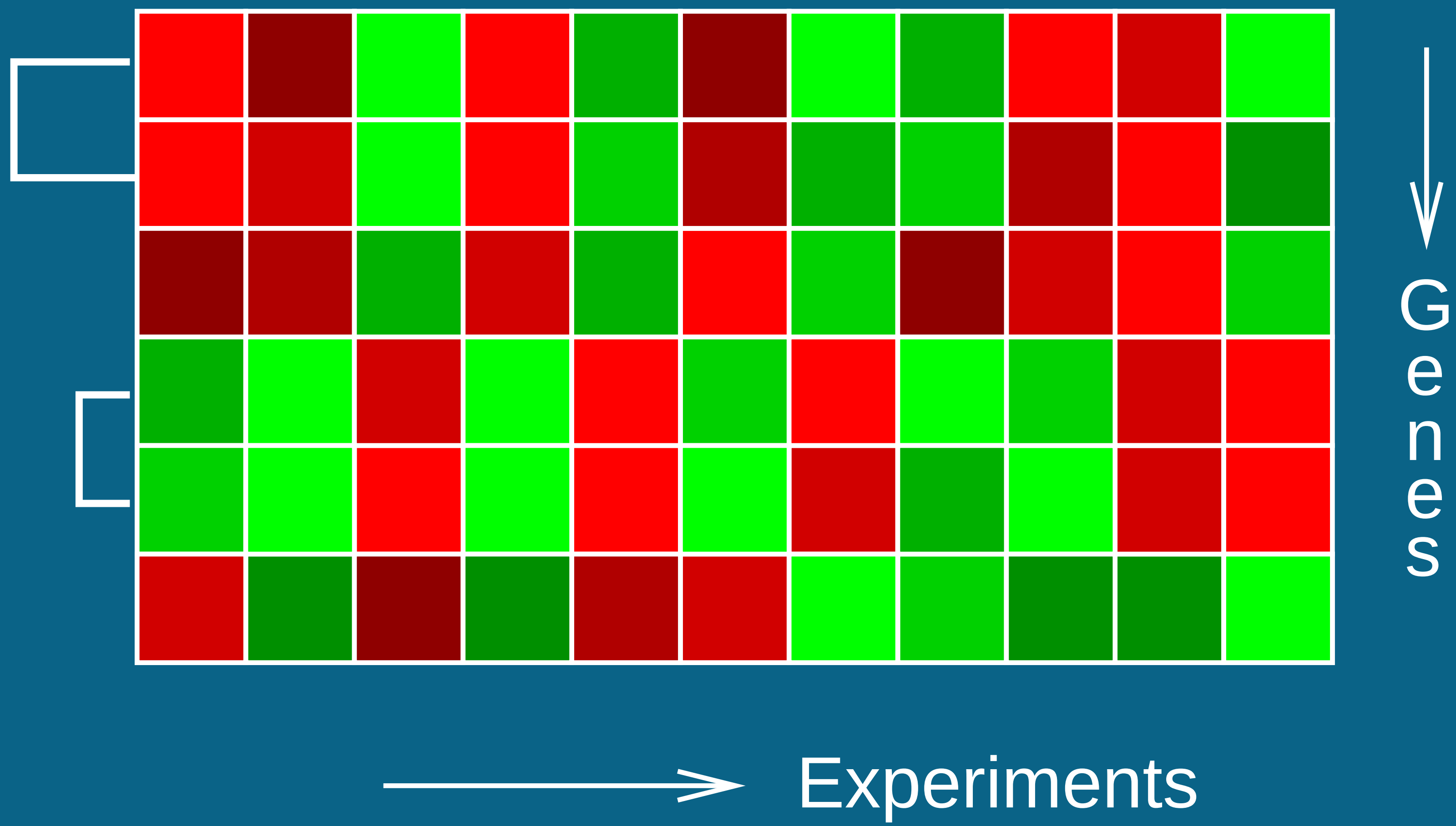

Transcription
DNA
mRNA
Protein

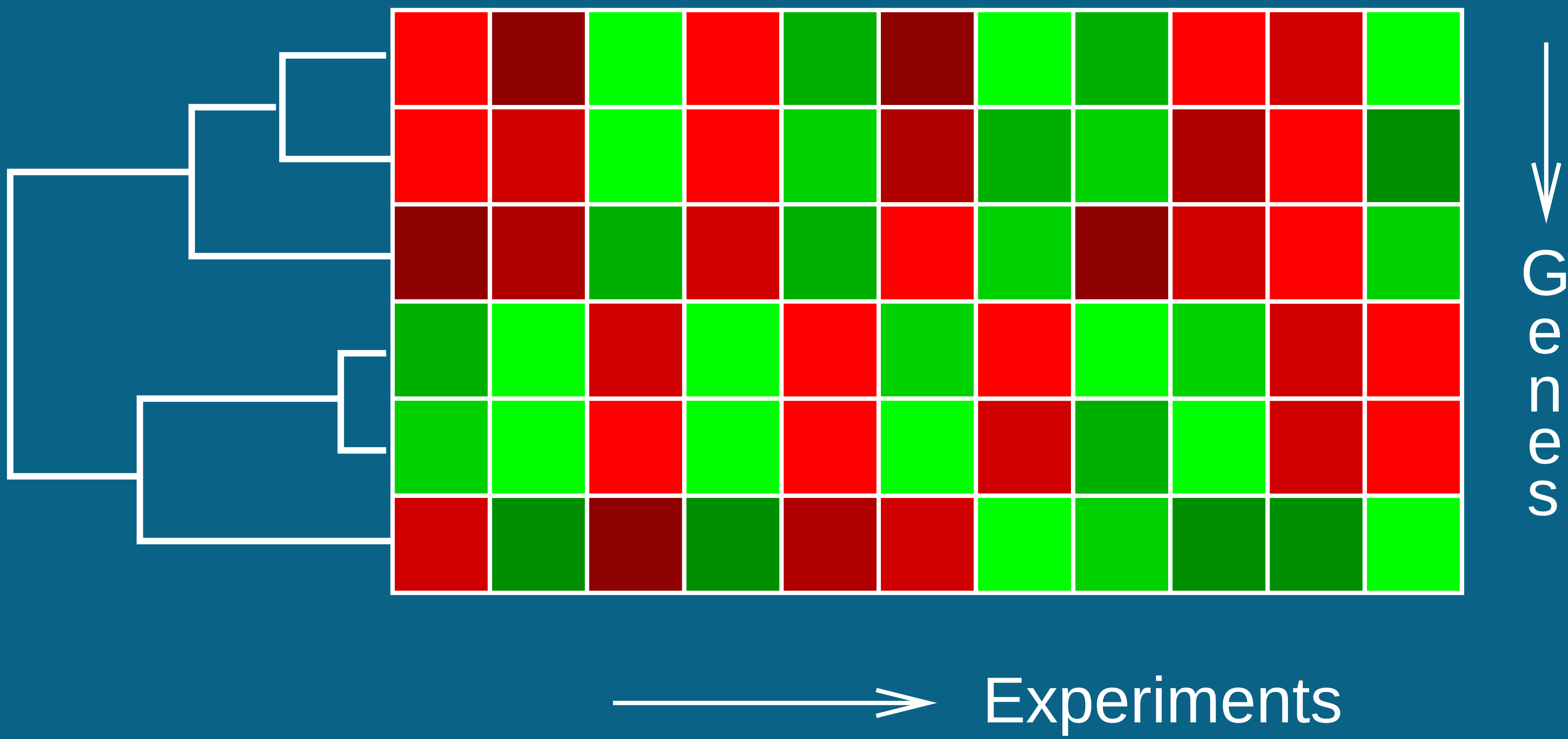

Transcription
Translation
DNA
mRNA
Protein

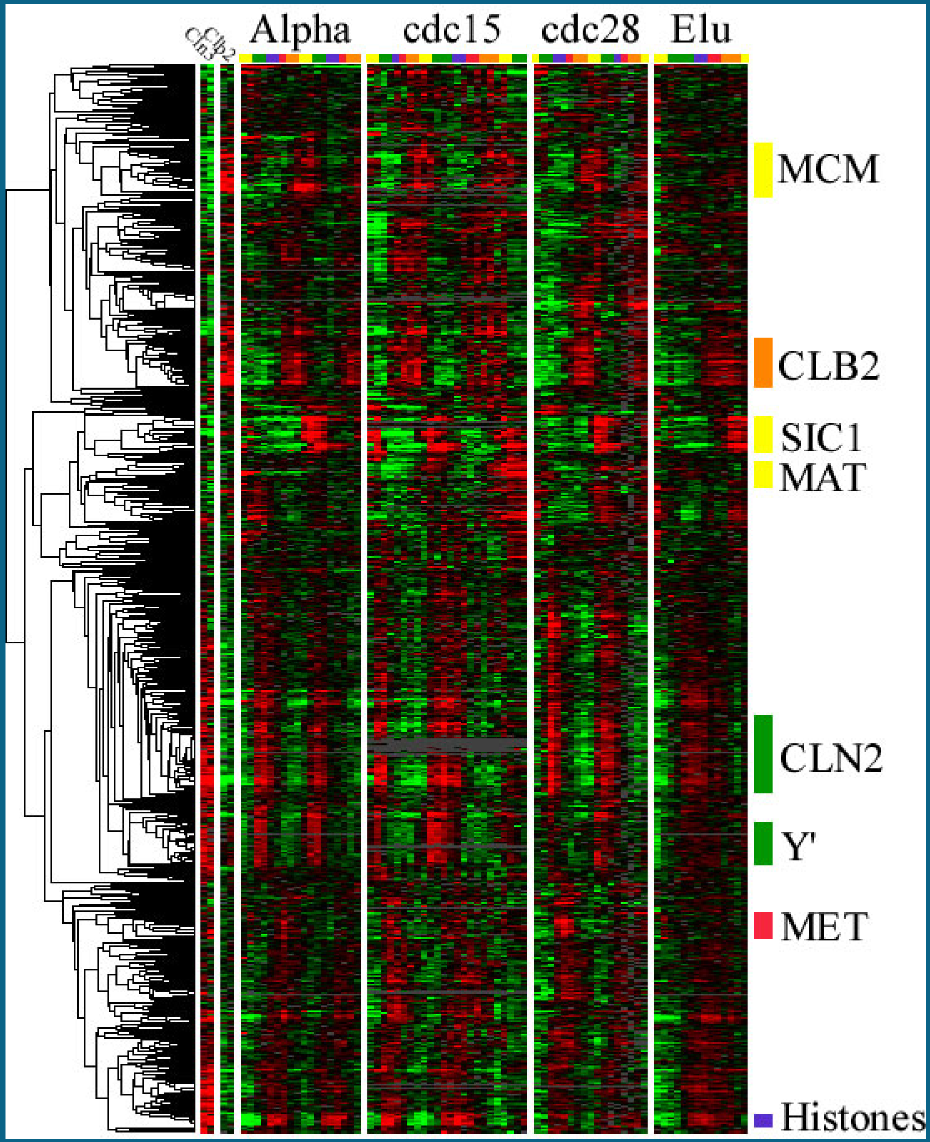

Transcription
Translation
DNA
mRNA
Protein
Microarrays


Genes
Experiments


Genes
Experiments


Genes
Experiments


Genes
Experiments



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

Advantage of clustering

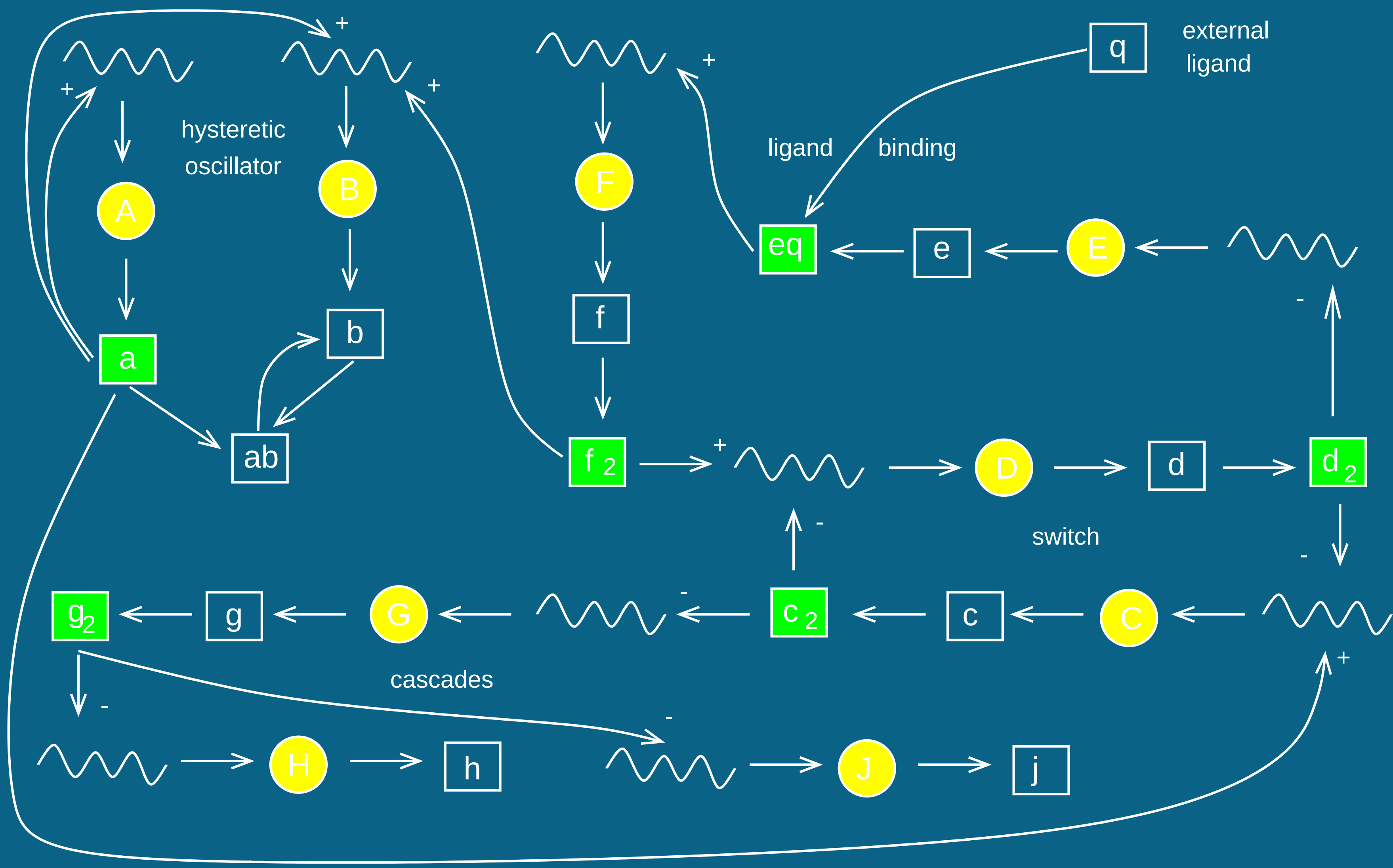
Fast, computationally cheap

Advantage of clustering

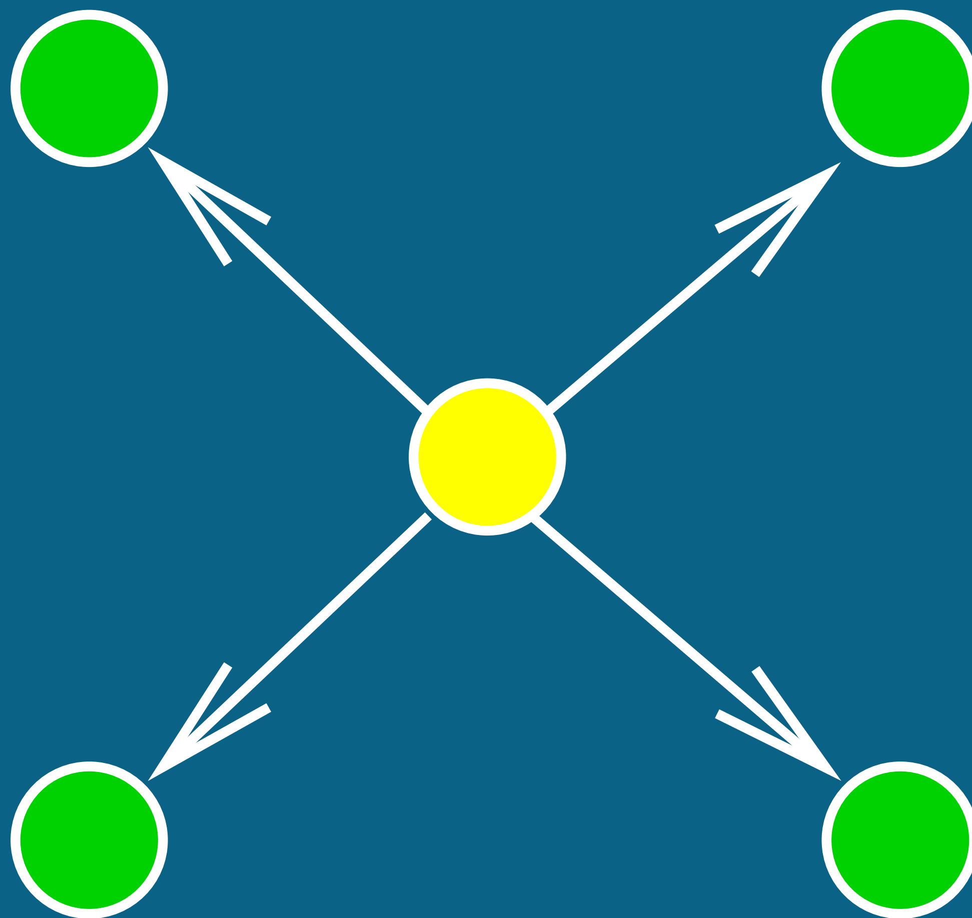
Fast, computationally cheap

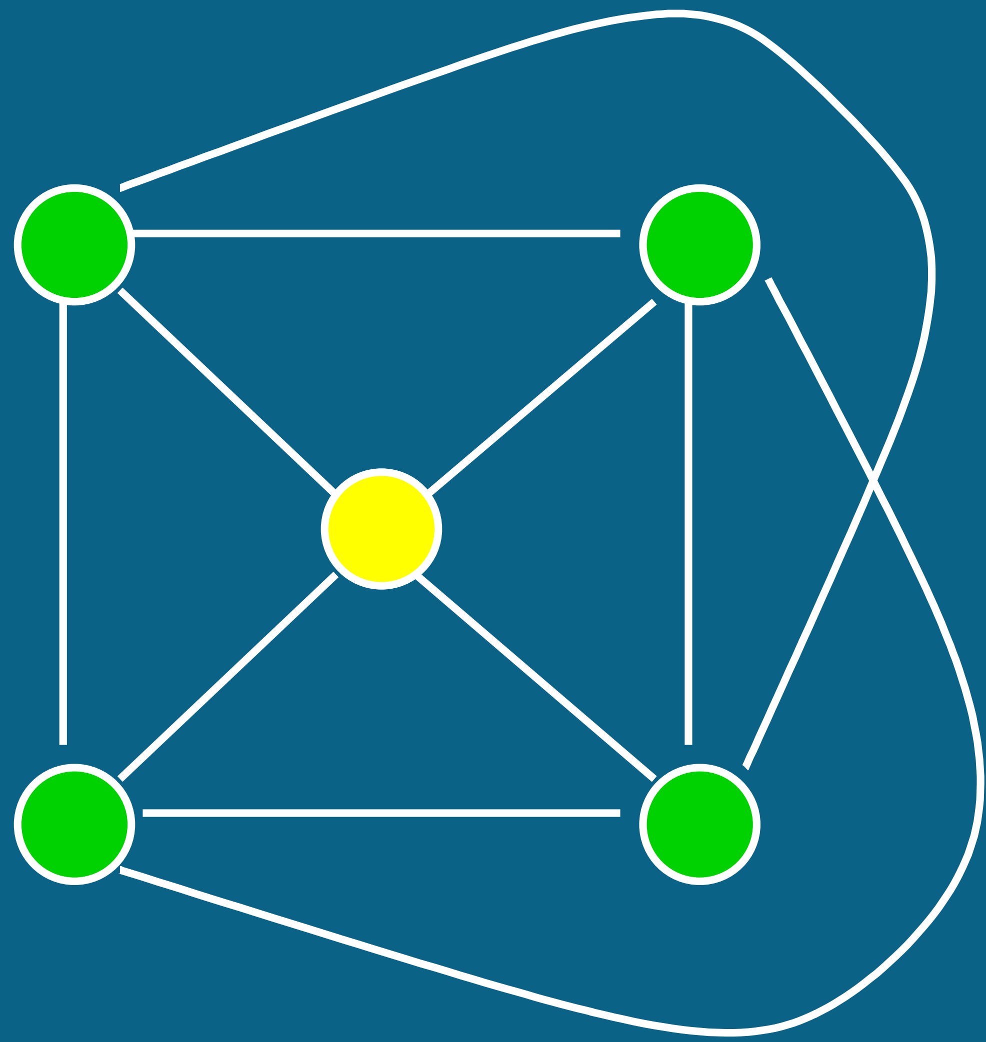
Shortcoming of clustering

It is **NOT** reverse engineering



+
+
+
+
hysteretic
oscillator
A
B
F
external
ligand
q
ligand
binding
eq
e
E
-
a
b
f
ab
f 2
+
D
d
d 2
-
switch
-
g2
g
G
-
c 2
c
C
+
cascades
-
-
H
h
J
j





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

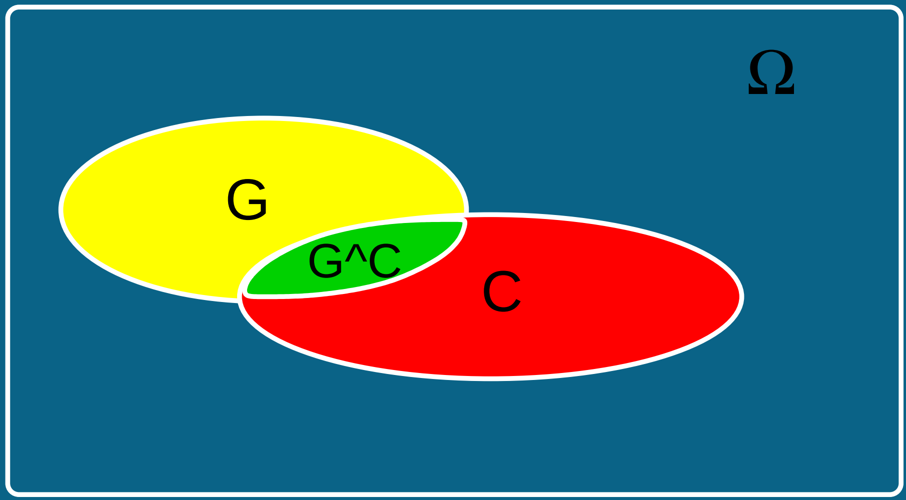
- Recapitulation: Bayesian networks
- Reverse engineering:
  Learning networks from data
- Application to the yeast cell cycle
- Estimating the accuracy of inference

Outline of the talk

- **Recapitulation: Bayesian networks**
- Reverse engineering:
  Learning networks from data
- Application to the yeast cell cycle
- Estimating the accuracy of inference

Revision: Bayes' Rule

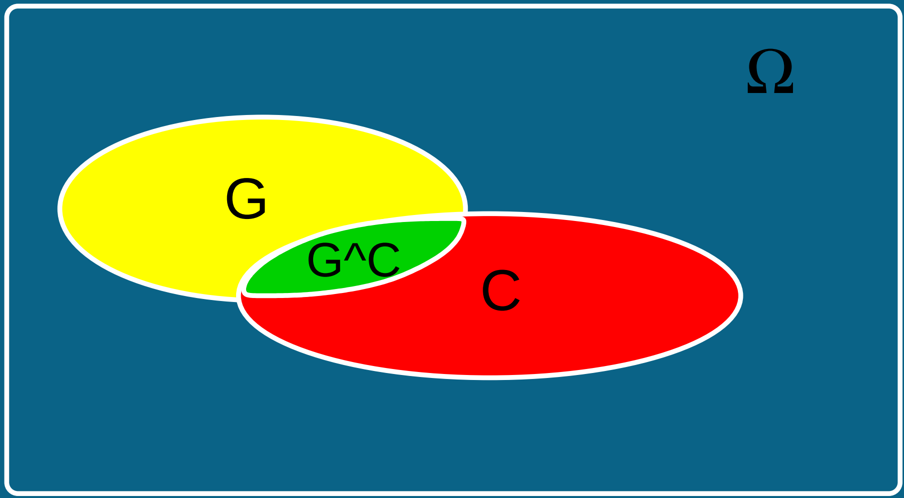
G: A certain gene is over-expressed
C: A patient is suffering from cancer



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

Revision: Bayes' Rule

G: A certain gene is over-expressed
C: A patient is suffering from cancer

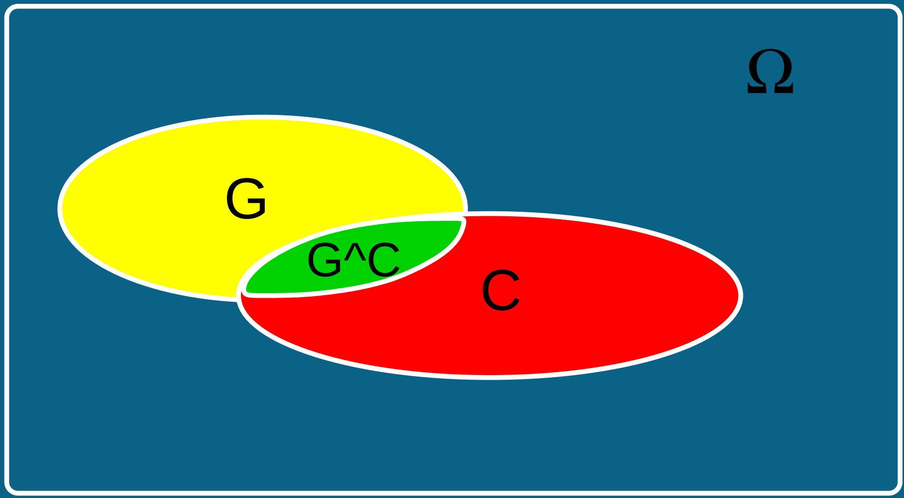


$$P(G|C) = \frac{P(G,C)}{P(C)} \qquad P(C|G) = \frac{P(G,C)}{P(G)}$$

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

Revision: Bayes' Rule

G: A certain gene is over-expressed
C: A patient is suffering from cancer



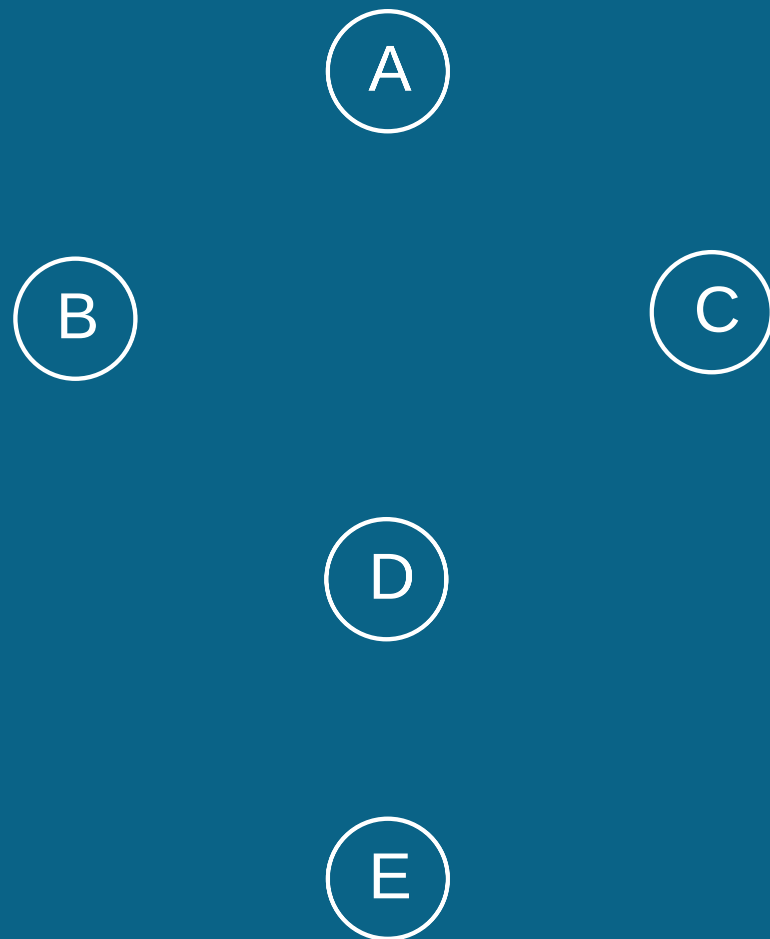
$$P(G|C) = \frac{P(G,C)}{P(C)} \qquad P(C|G) = \frac{P(G,C)}{P(G)}$$

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

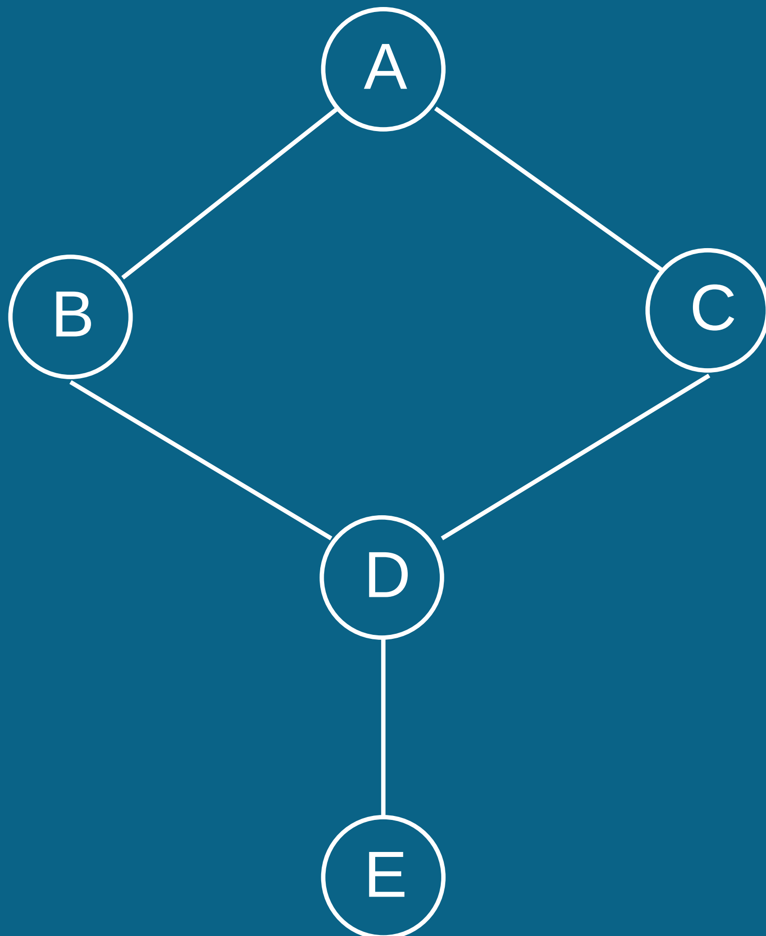
$$P(C|G) = \frac{P(G|C)P(C)}{P(G)}$$

Introduction to probabilistic graphical models

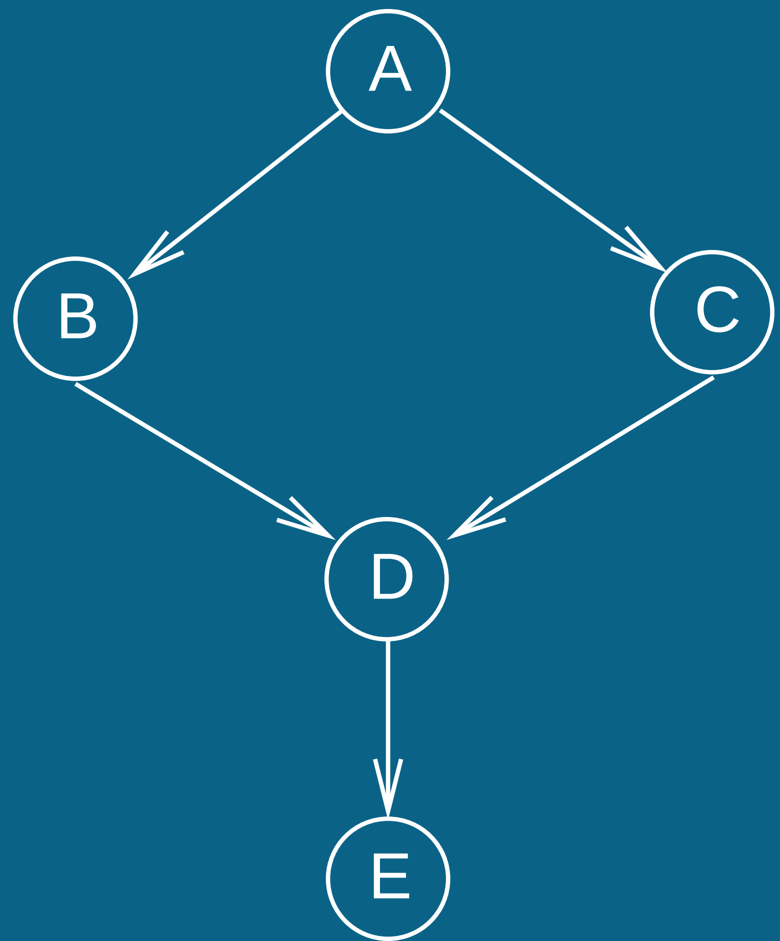
- Nodes: Random variables
- Edges or arcs indicate conditional dependencies
- For random variables $\{A_1, \ldots, A_n\}$: $$P(A_1, \ldots, A_n) = \prod_i P(A_i | \mathrm{parent}[A_i])$$



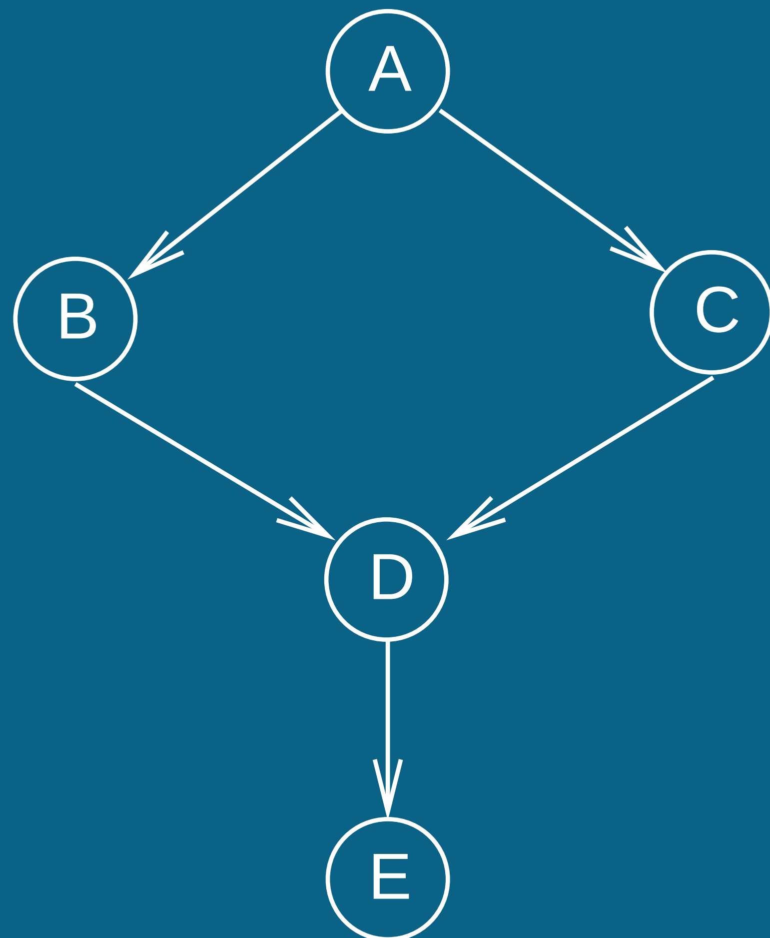
Nodes



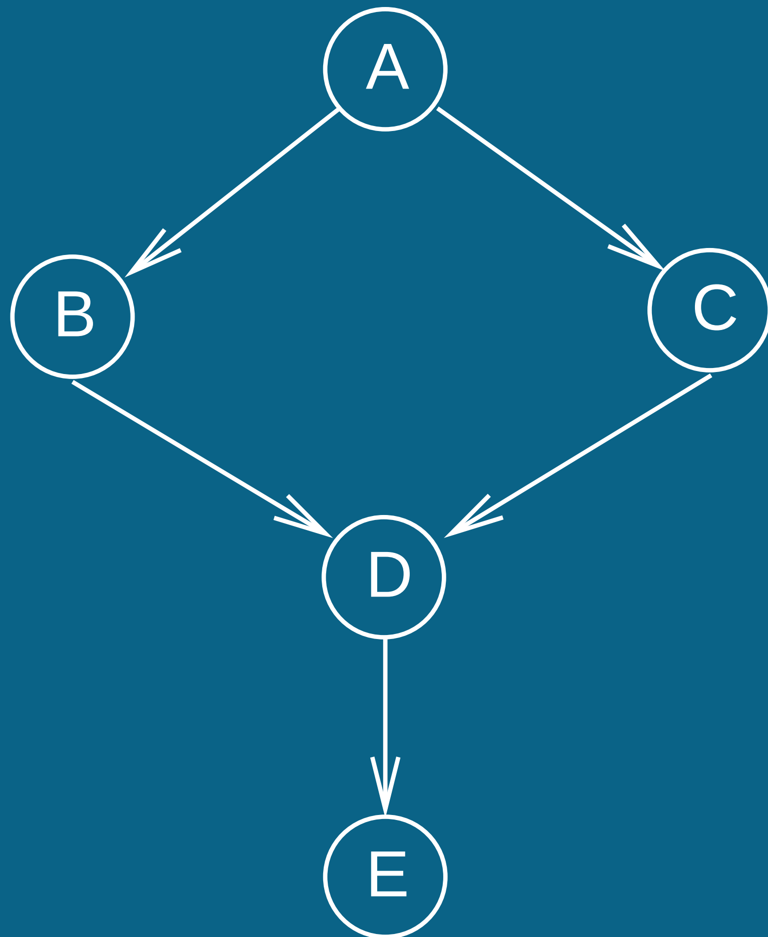
Edges



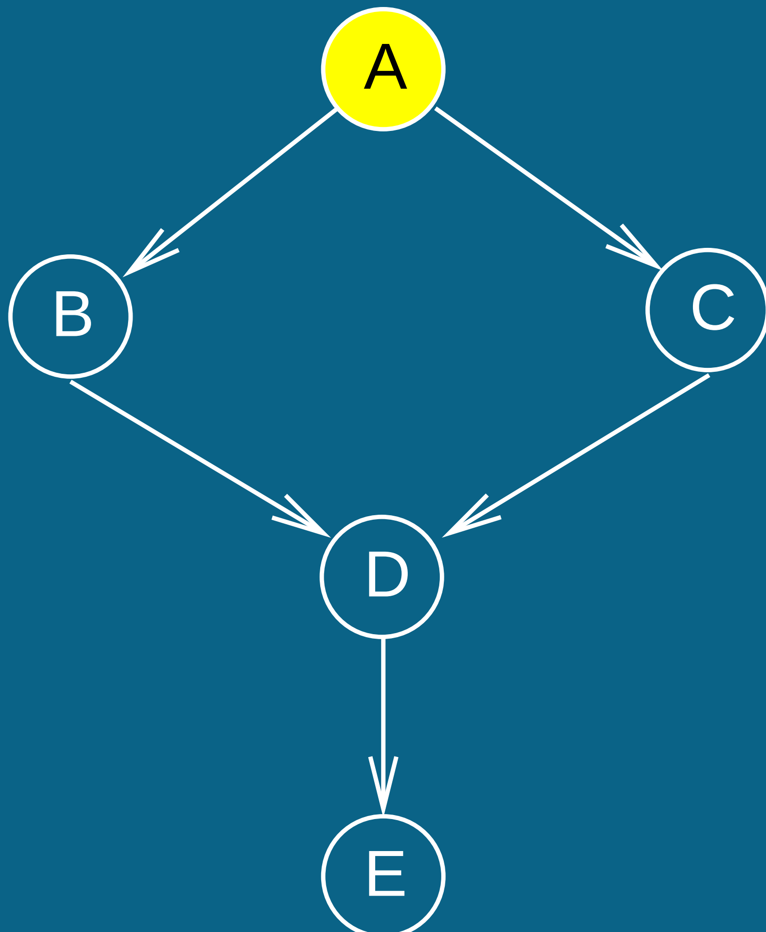
Edges = directed



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

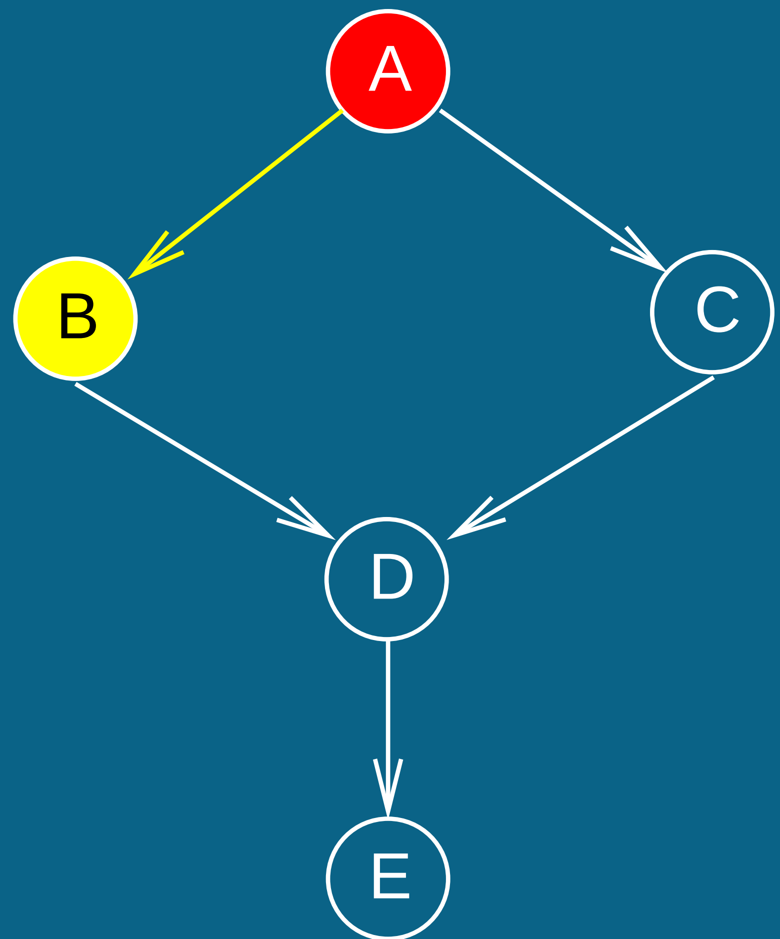


$$P(A, B, C, D, E) =$$



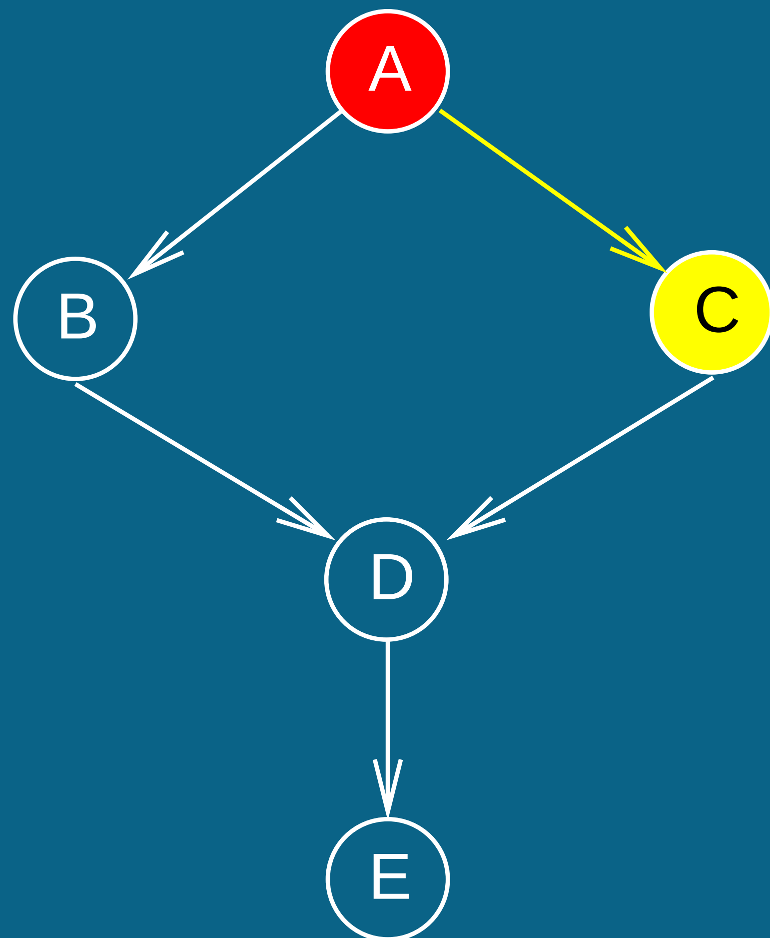
$$P(A, B, C, D, E) =$$

$$P(A)$$



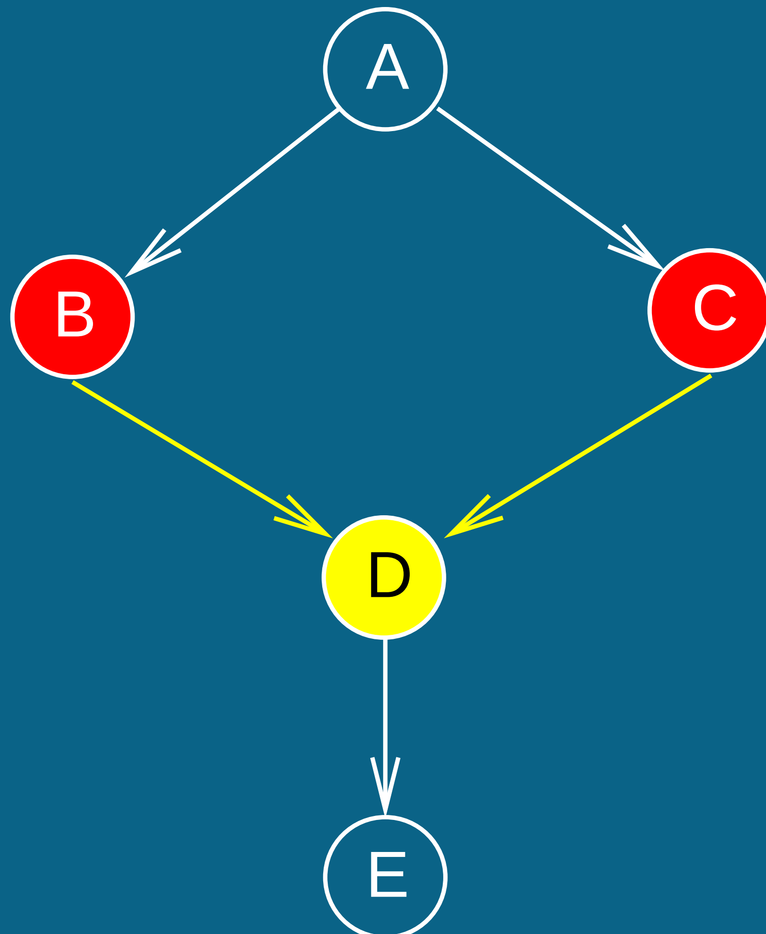
$$P(A, B, C, D, E) =$$

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



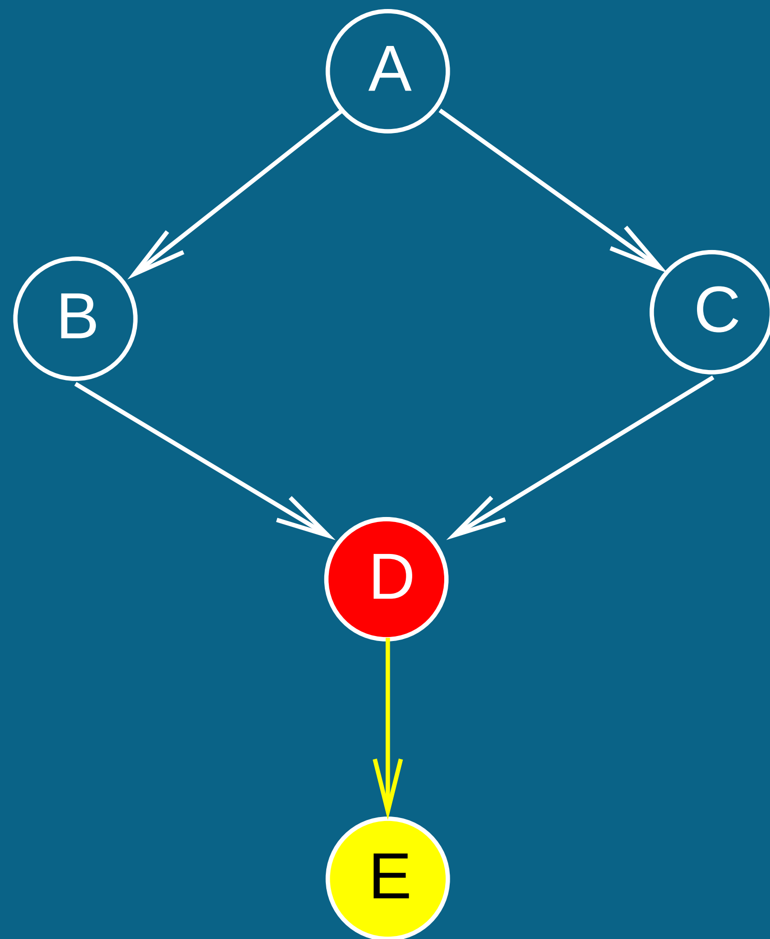
$$P(A, B, C, D, E) =$$

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



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

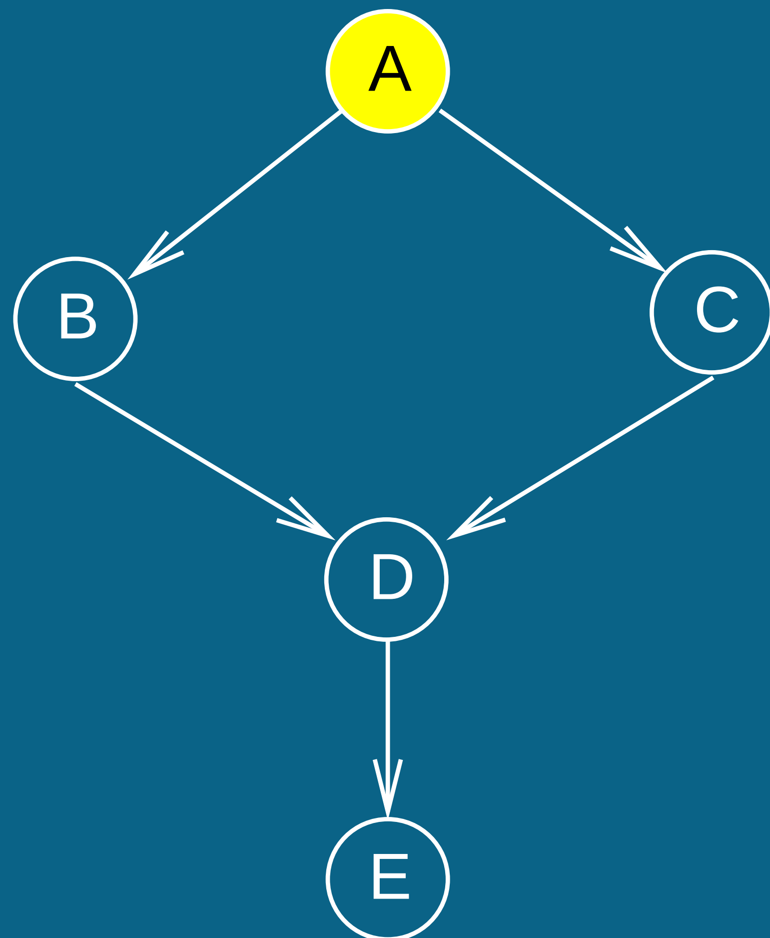
$$P(A)P(B|A)P(C|A)P(D|B, C)$$



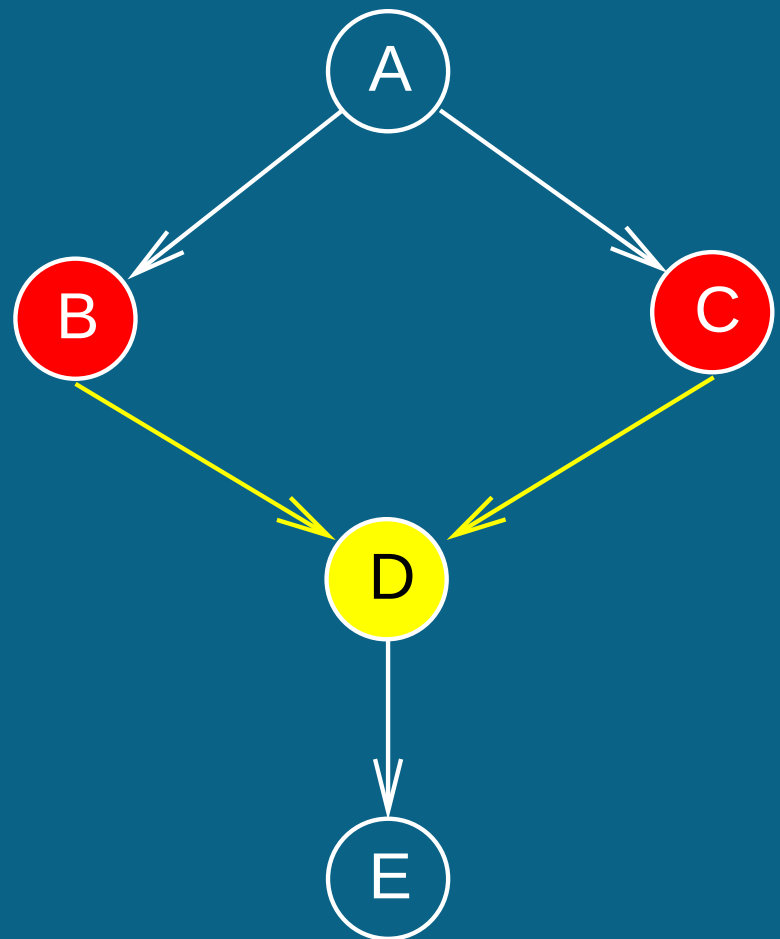
$$P(A, B, C, D, E) =$$

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

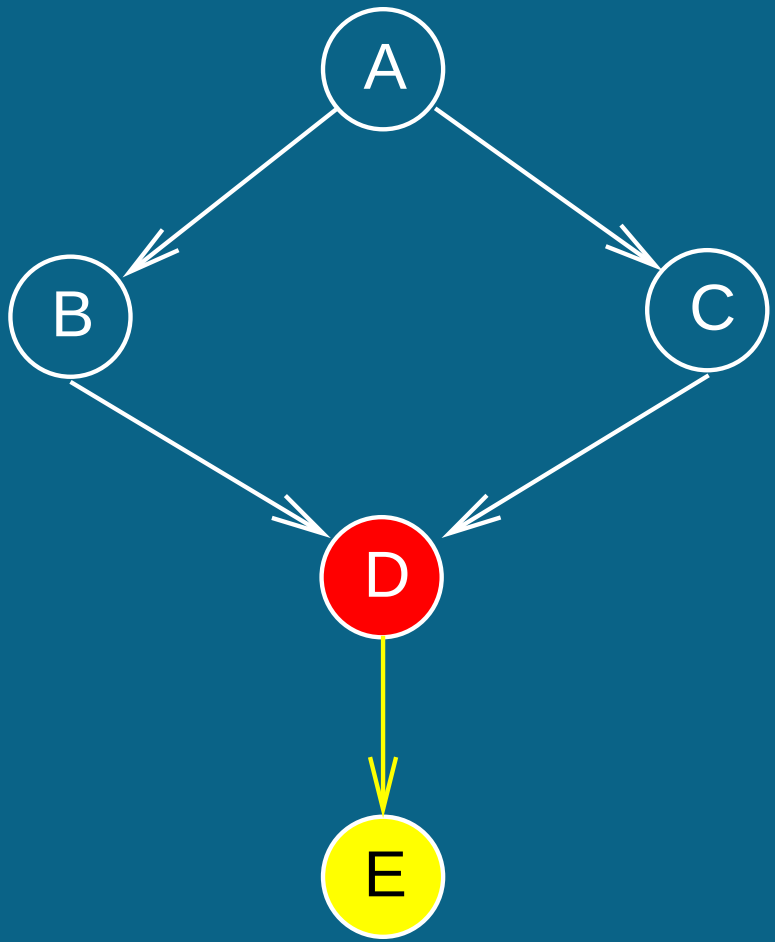
Biological interpretation



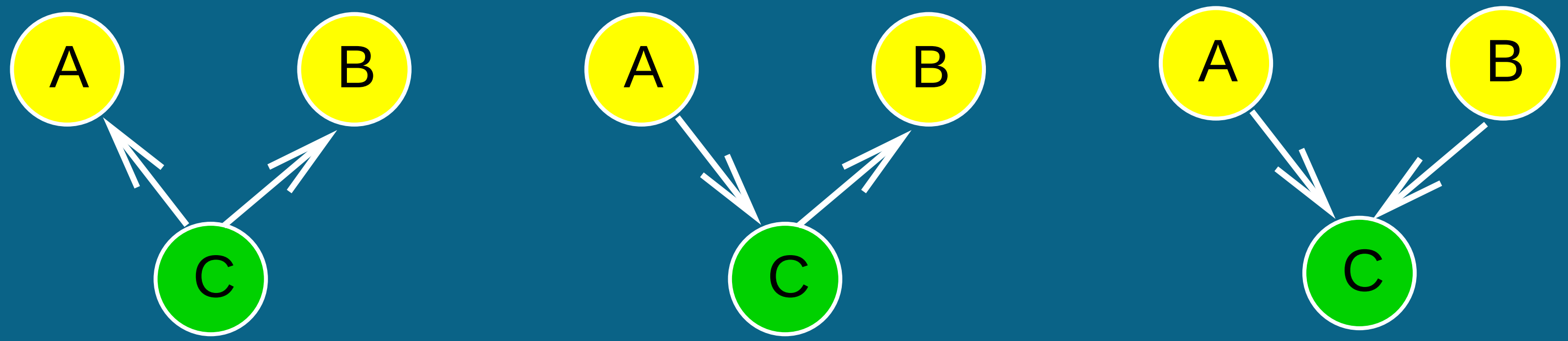
Initiation of cell (sub-)cycle



Co-regulation

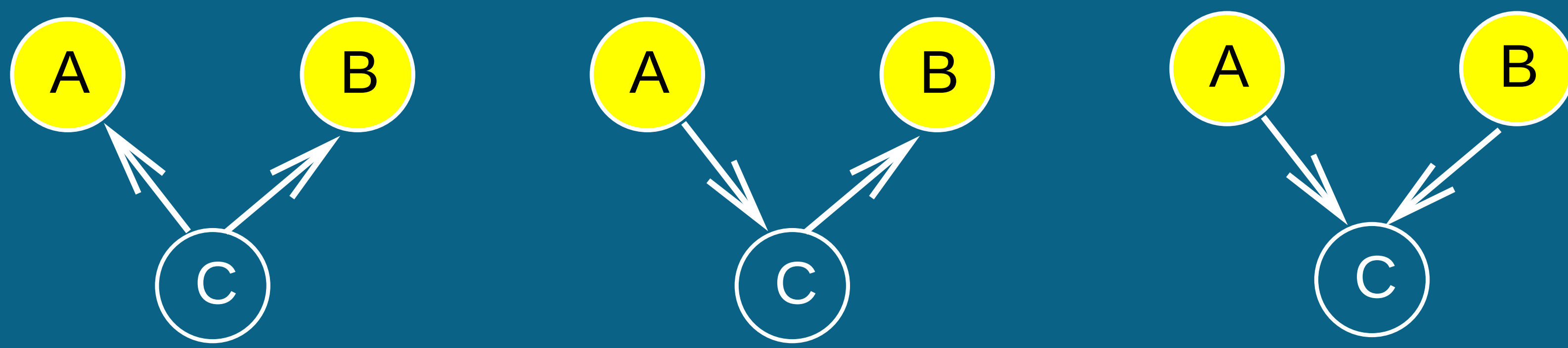


Mediation



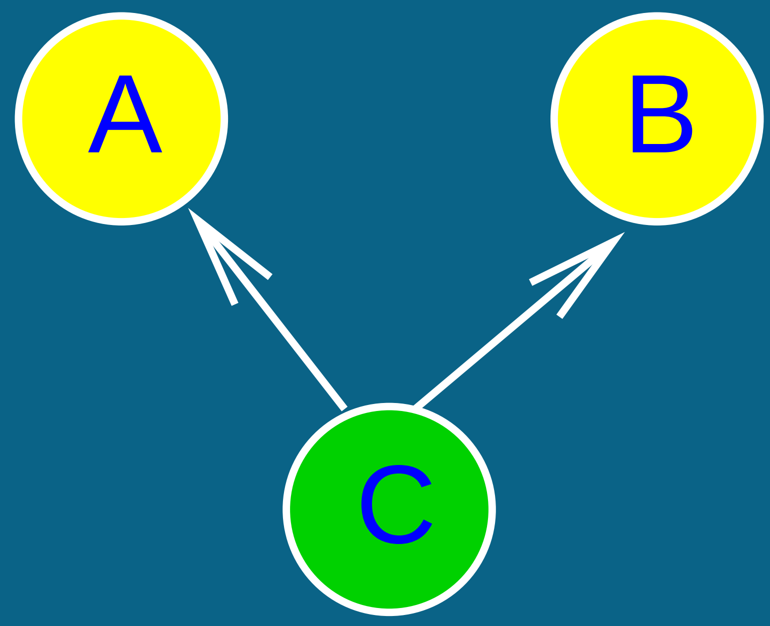
When are A and B conditionally independent given C?

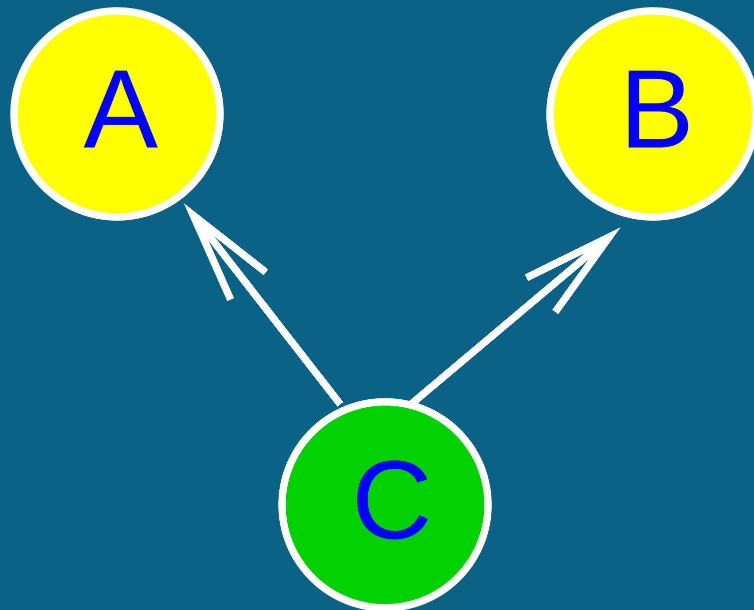
$$P(A, B|C) = P(A|C)P(B|C)$$



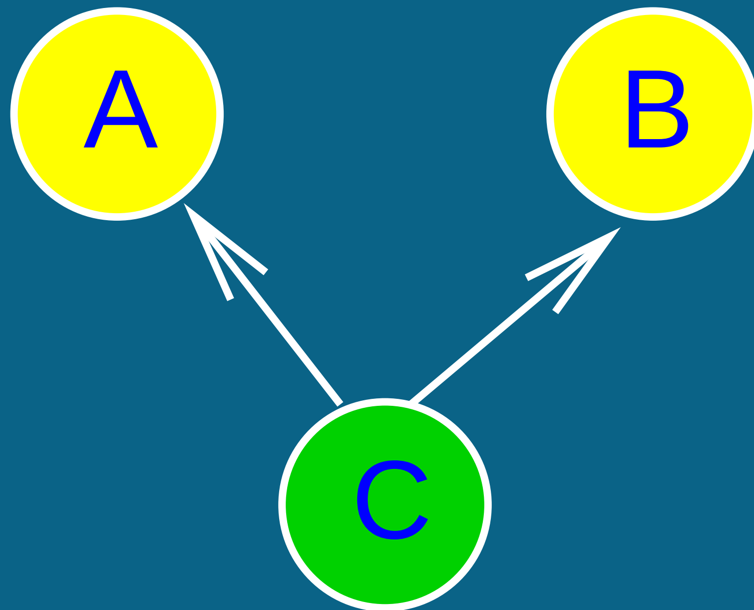
When are A and B marginally independent?

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


A
B
C

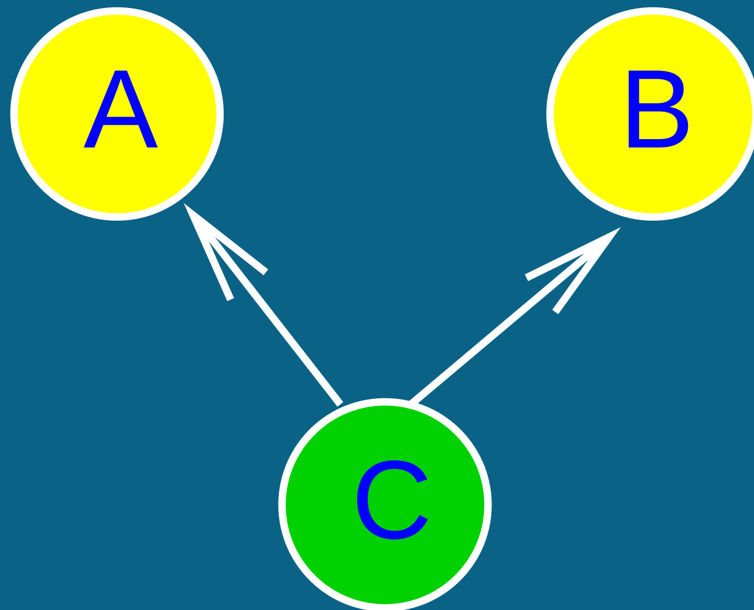


$$P(A, B, C) = P(A|C)P(B|C)P(C)$$



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

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

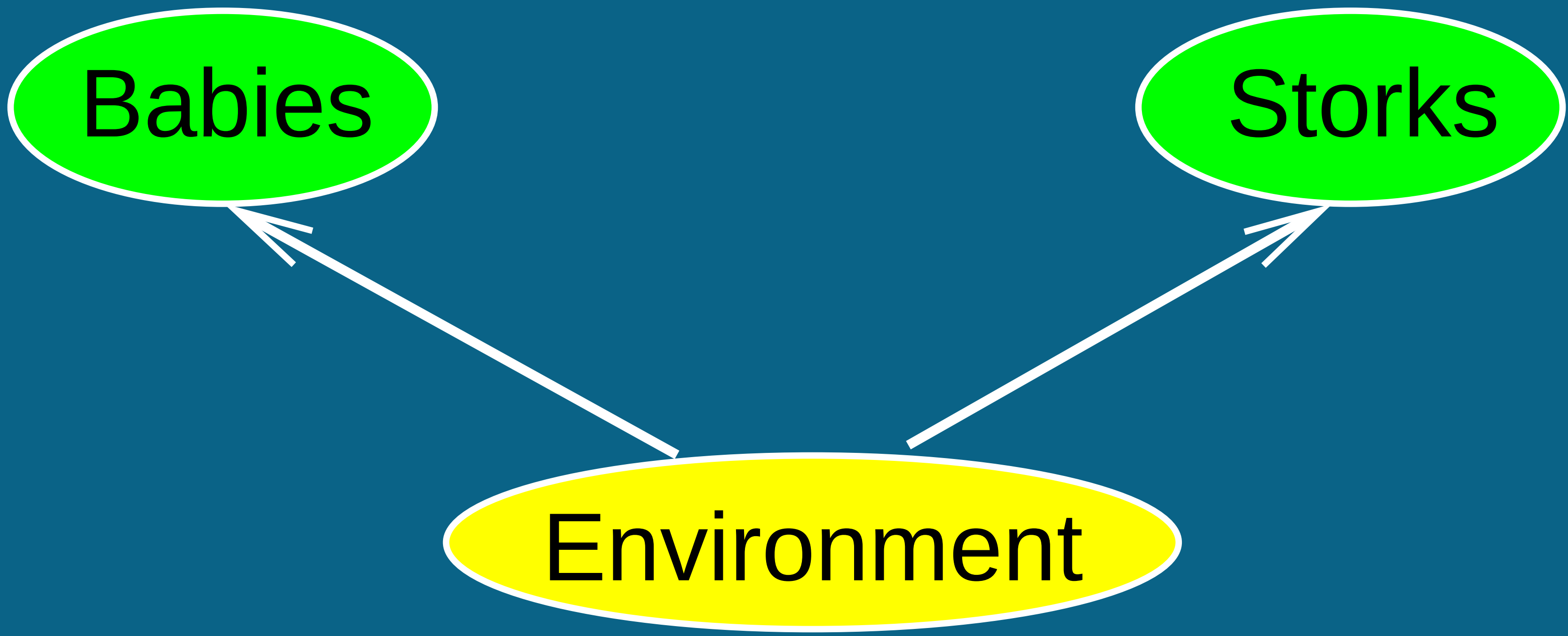


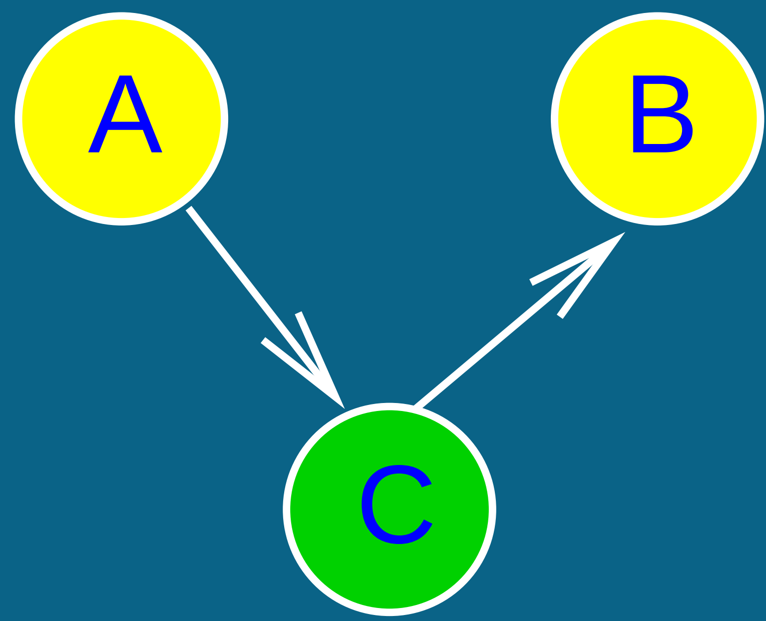
$$P(A,B,C) = P(A|C)P(B|C)P(C)$$

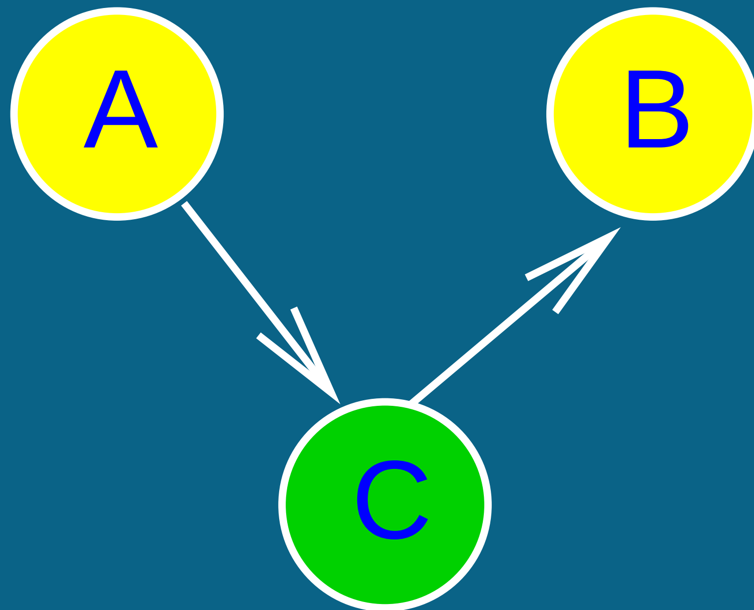
$$P(A,B|C) = \frac{P(A,B,C)}{P(C)} = P(A|C)P(B|C)$$

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

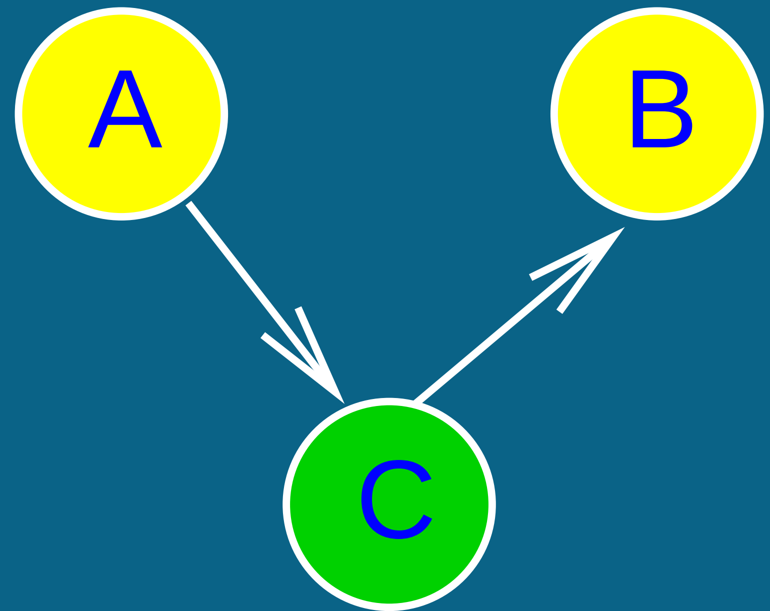

Babies
Storks


Babies
Storks
Environment


A
B
C

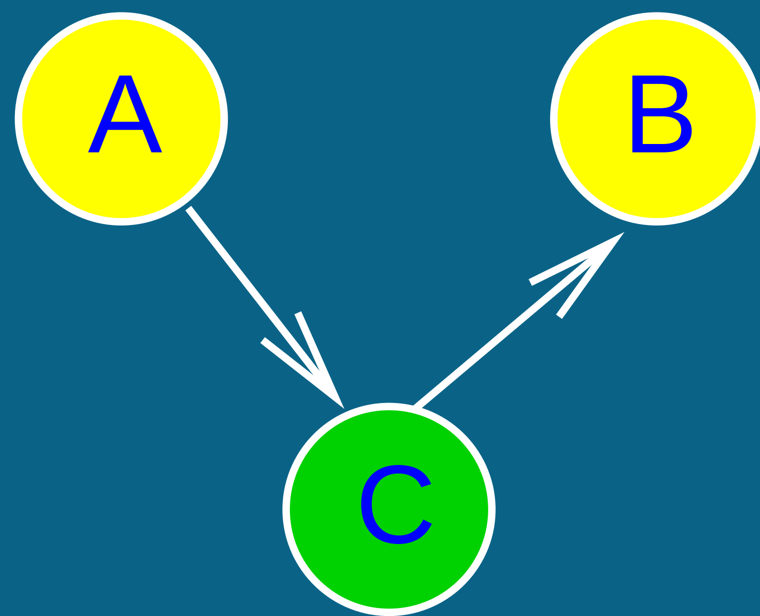


$$P(A, B, C) = P(B|C)P(C|A)P(A)$$



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

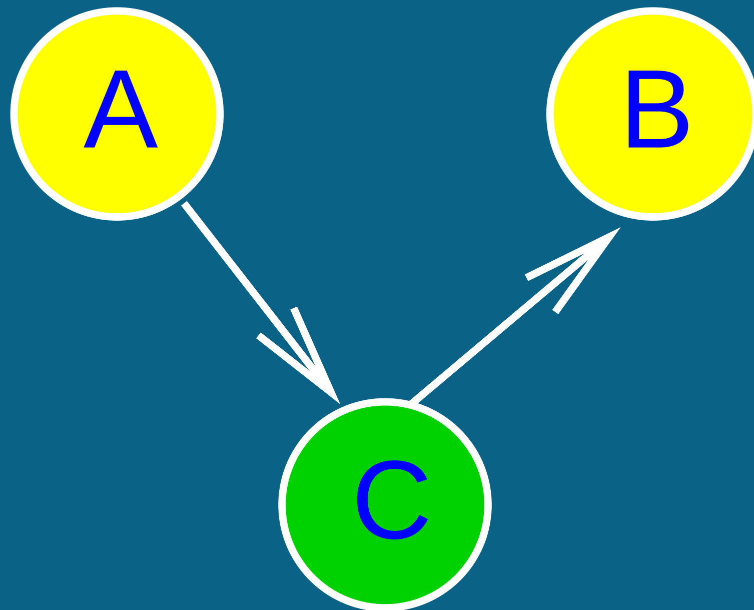
$$P(A,B|C) = \frac{P(A,B,C)}{P(C)} = P(B|C)\frac{P(C|A)P(A)}{P(C)}$$



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

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

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



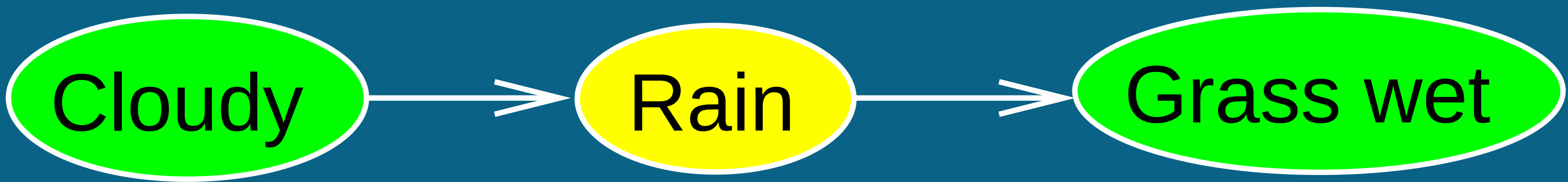
$$P(A,B,C) = P(B|C)P(C|A)P(A)$$

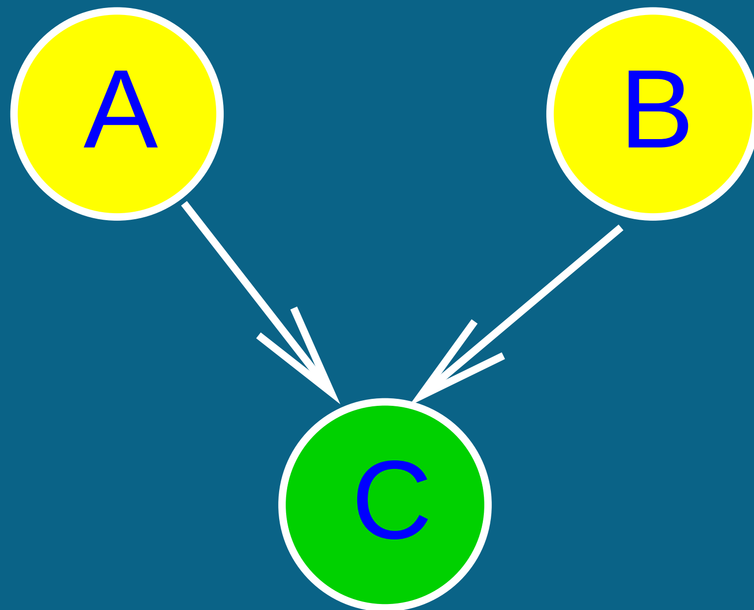
$$P(A,B|C) = \frac{P(A,B,C)}{P(C)} = P(B|C)\frac{P(C|A)P(A)}{P(C)}$$

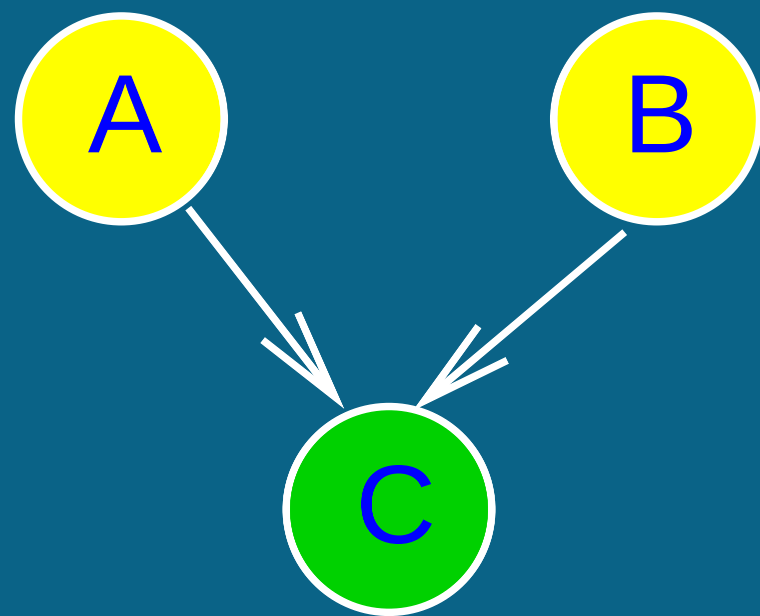
$$P(A,B|C) = P(B|C)P(A|C)$$

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

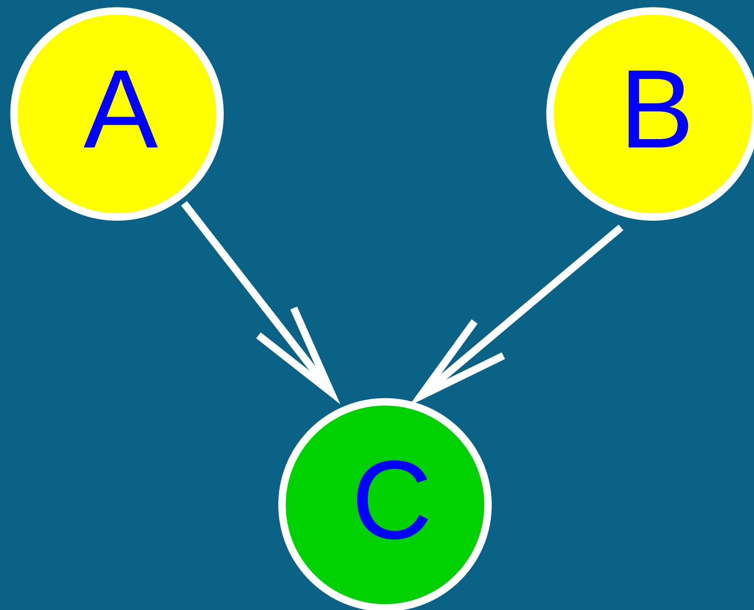

Cloudy
Rain
Grass wet


Cloudy
Rain
Grass wet


A
B
C

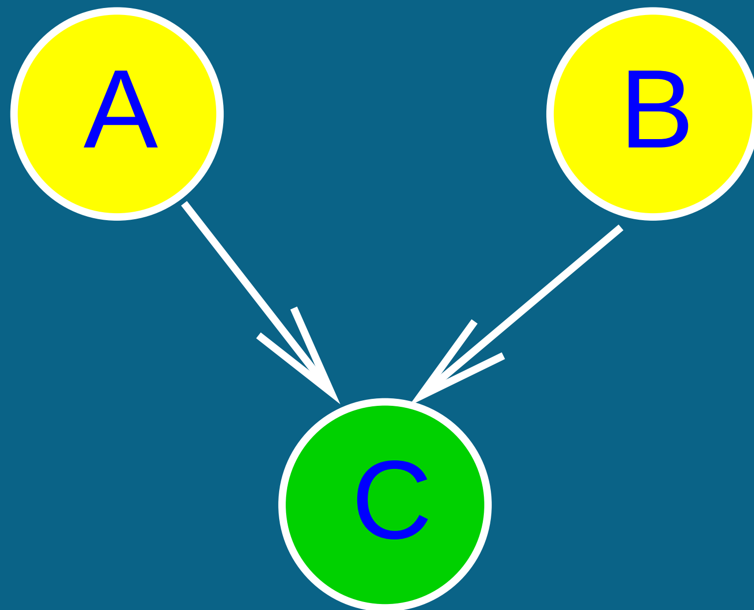


$$P(A, B, C) = P(C|A, B)P(A)P(B)$$



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

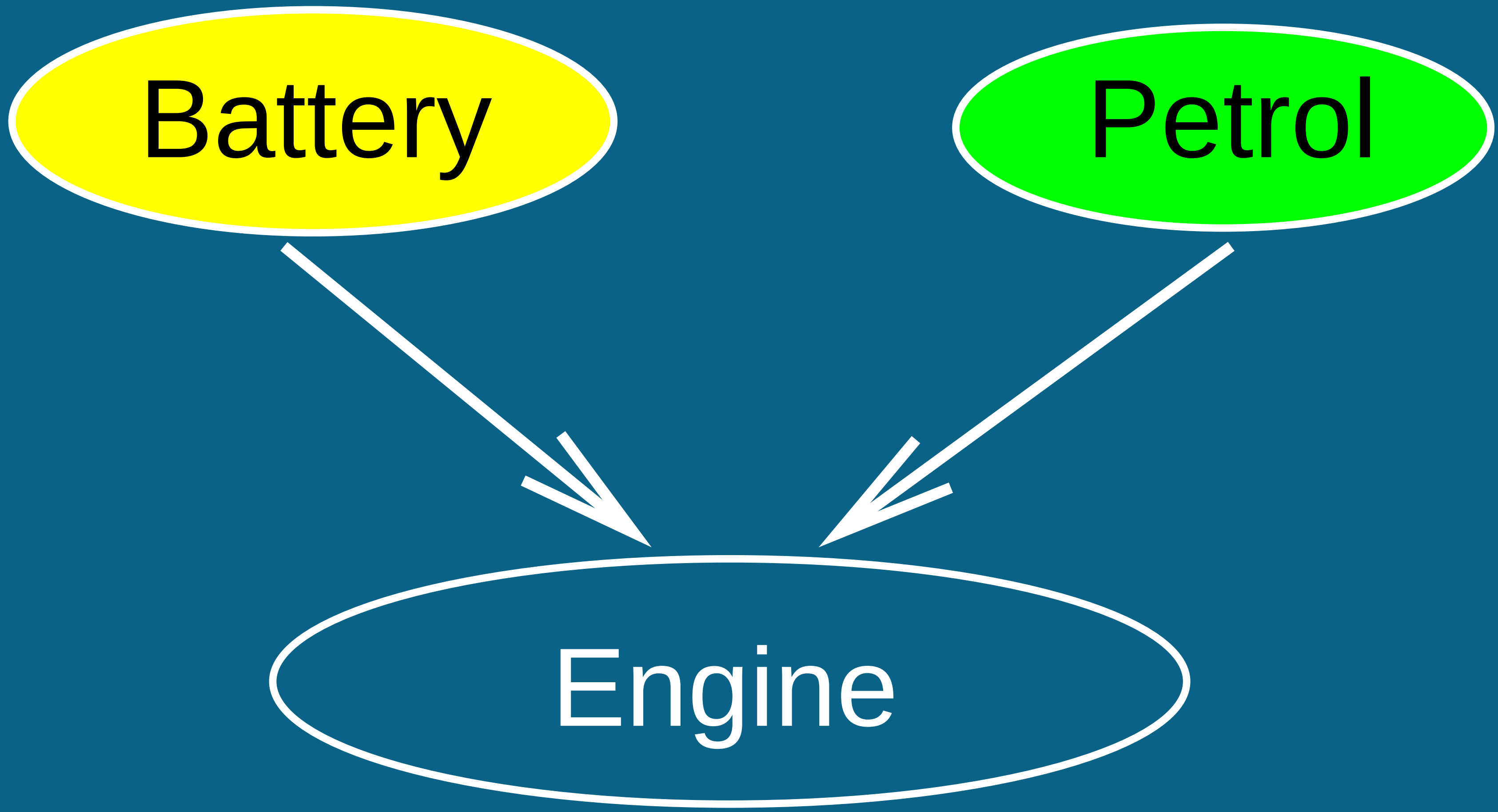
$$P(A,B) = \sum_{C} P(A,B,C) = P(A)P(B)$$

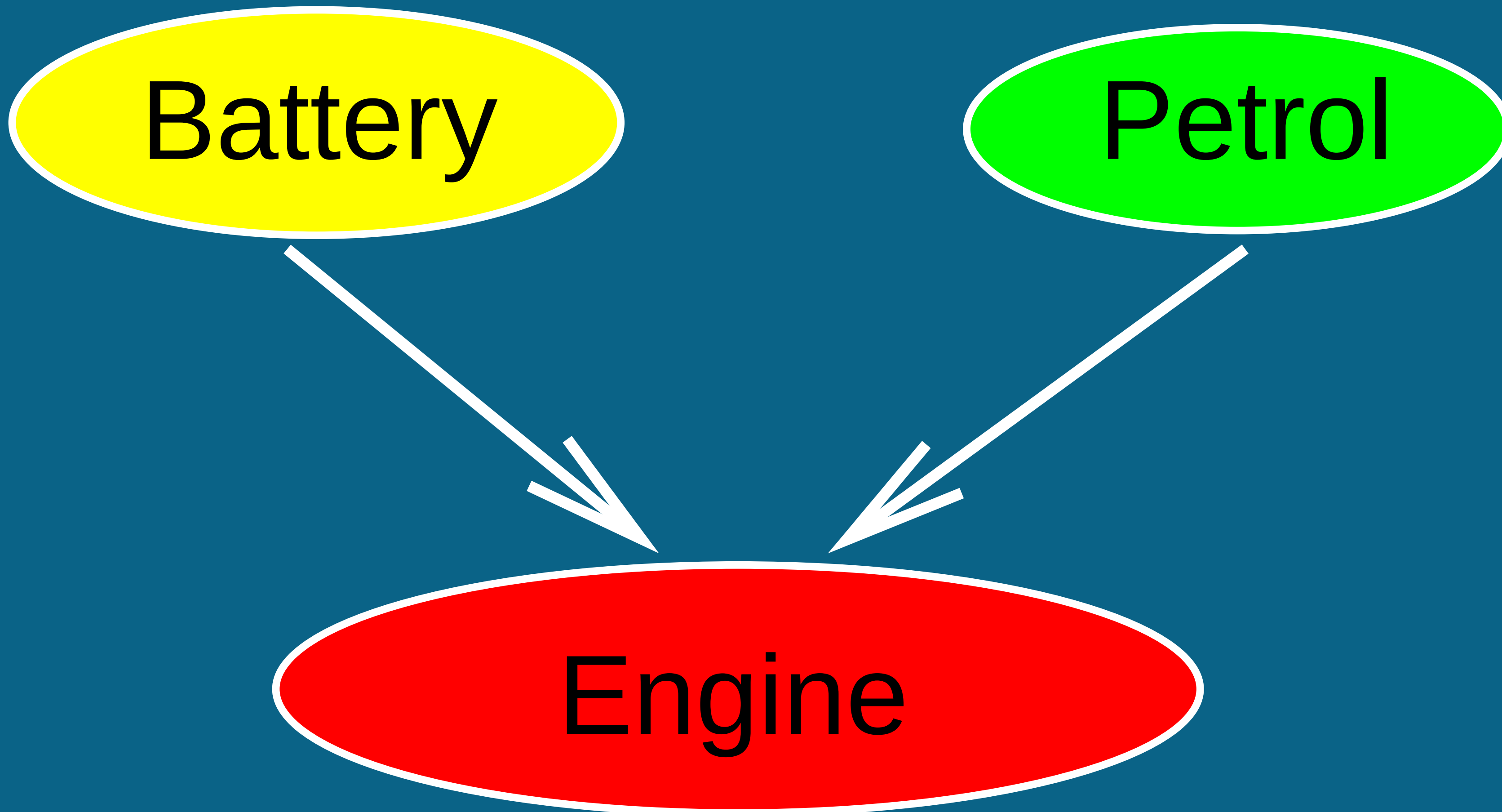


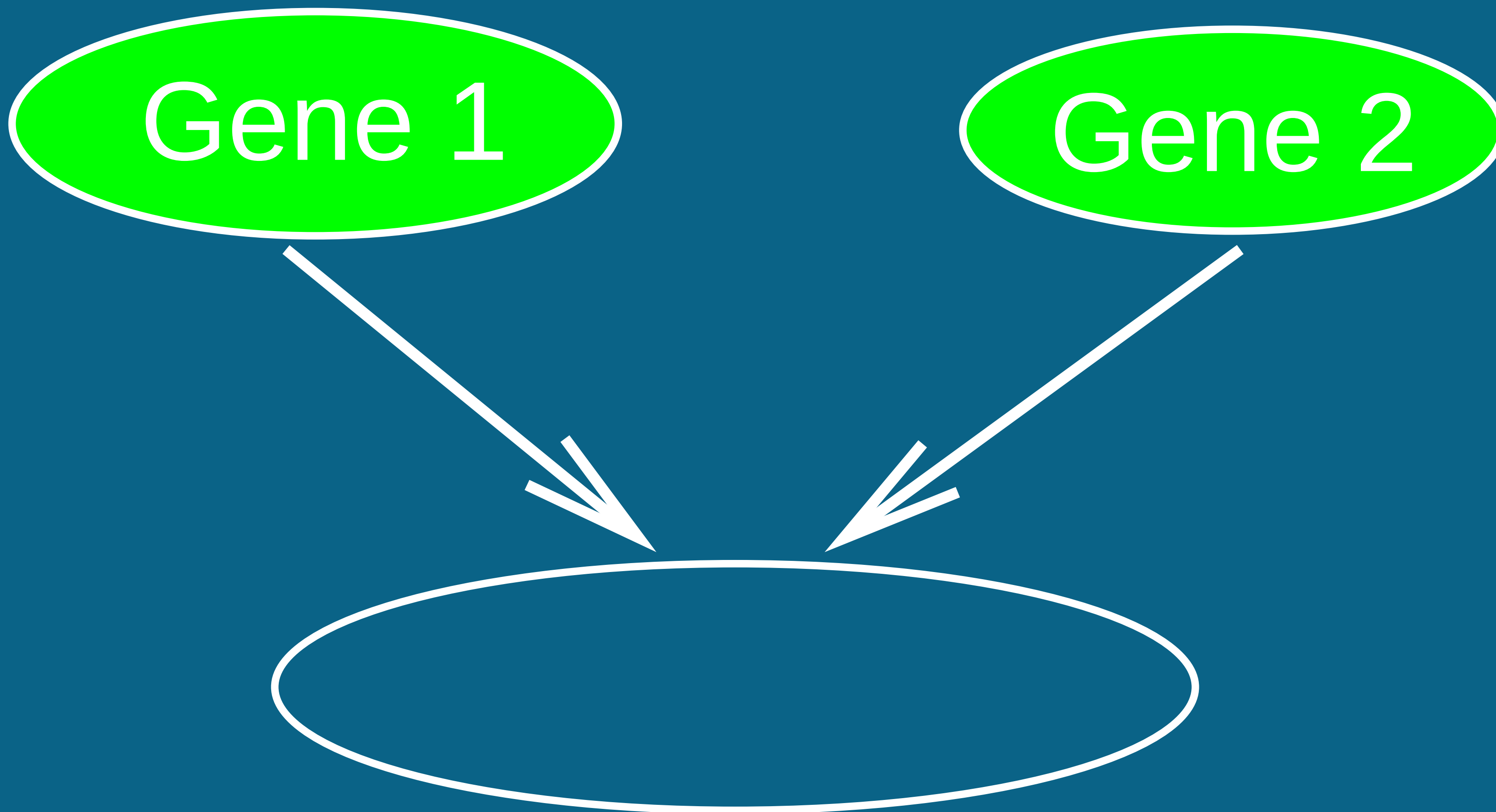
$$P(A,B,C) = P(C|A,B)P(A)P(B)$$

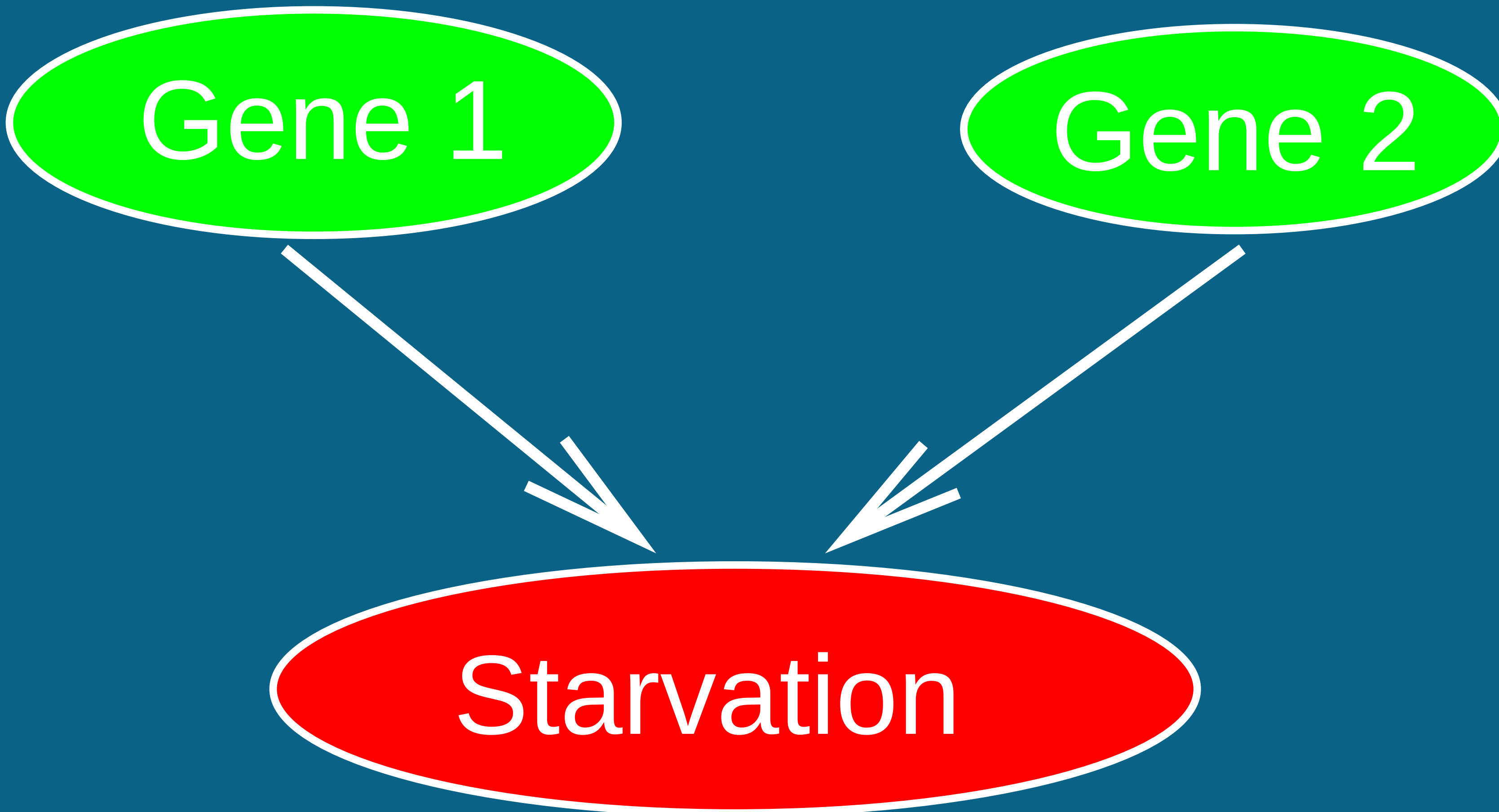
$$P(A,B) = \sum_C P(A,B,C) = P(A)P(B)$$

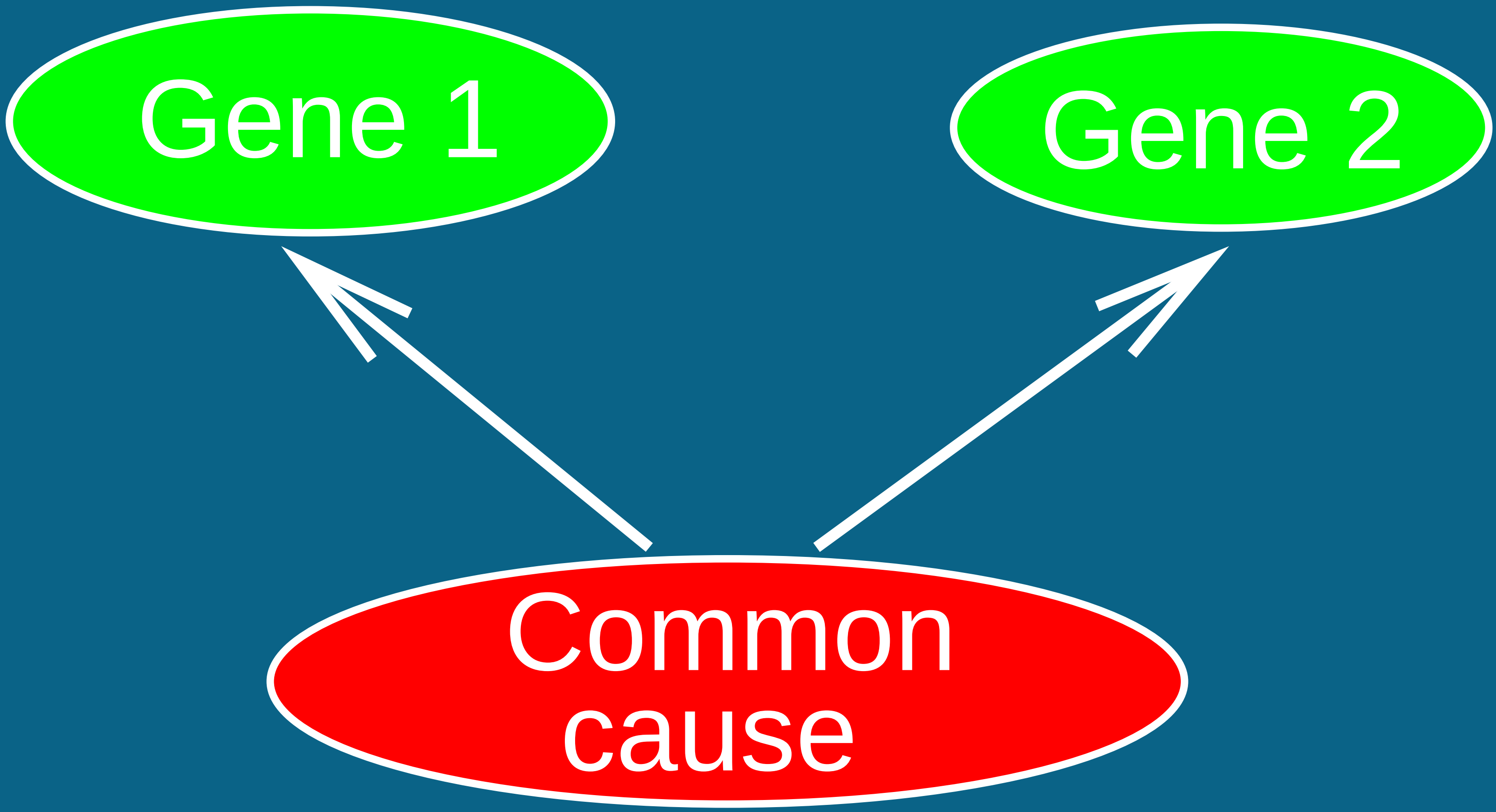
But: $P(A,B|C) \neq P(A|C)P(B|C)$

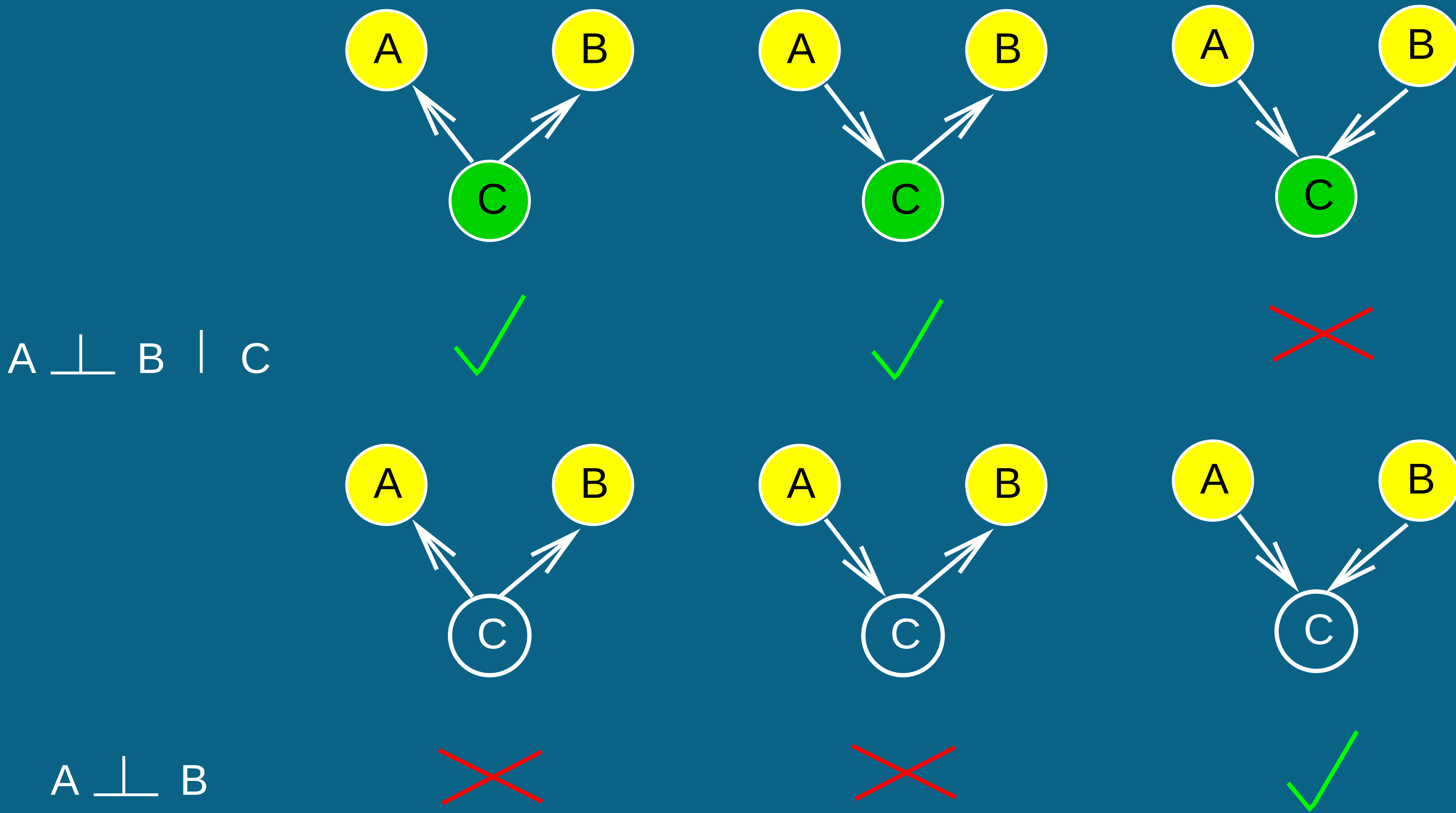

Battery
Petrol
Engine


Battery
Petrol
Engine


Gene 1
Gene 2


Gene 1
Gene 2
Starvation


Gene 1
Gene 2
Common cause

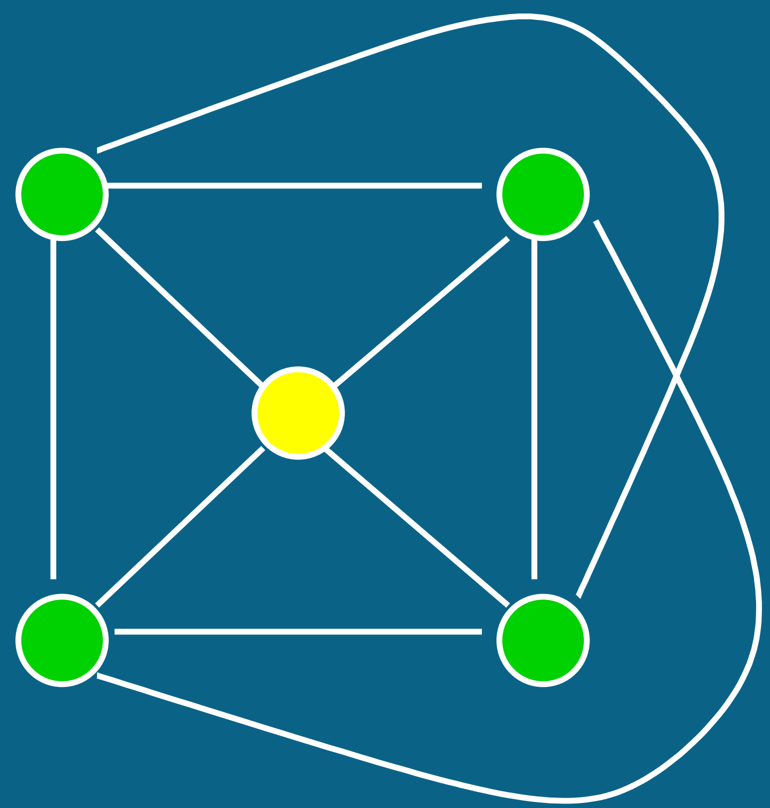

A
B
C
A ⊥ B | C
A ⊥ B

Biological example

Yeast cell cycle

Clustering

Spellman et al. 1998
Molecular Biology of the Cell 9 (12) :3273-97



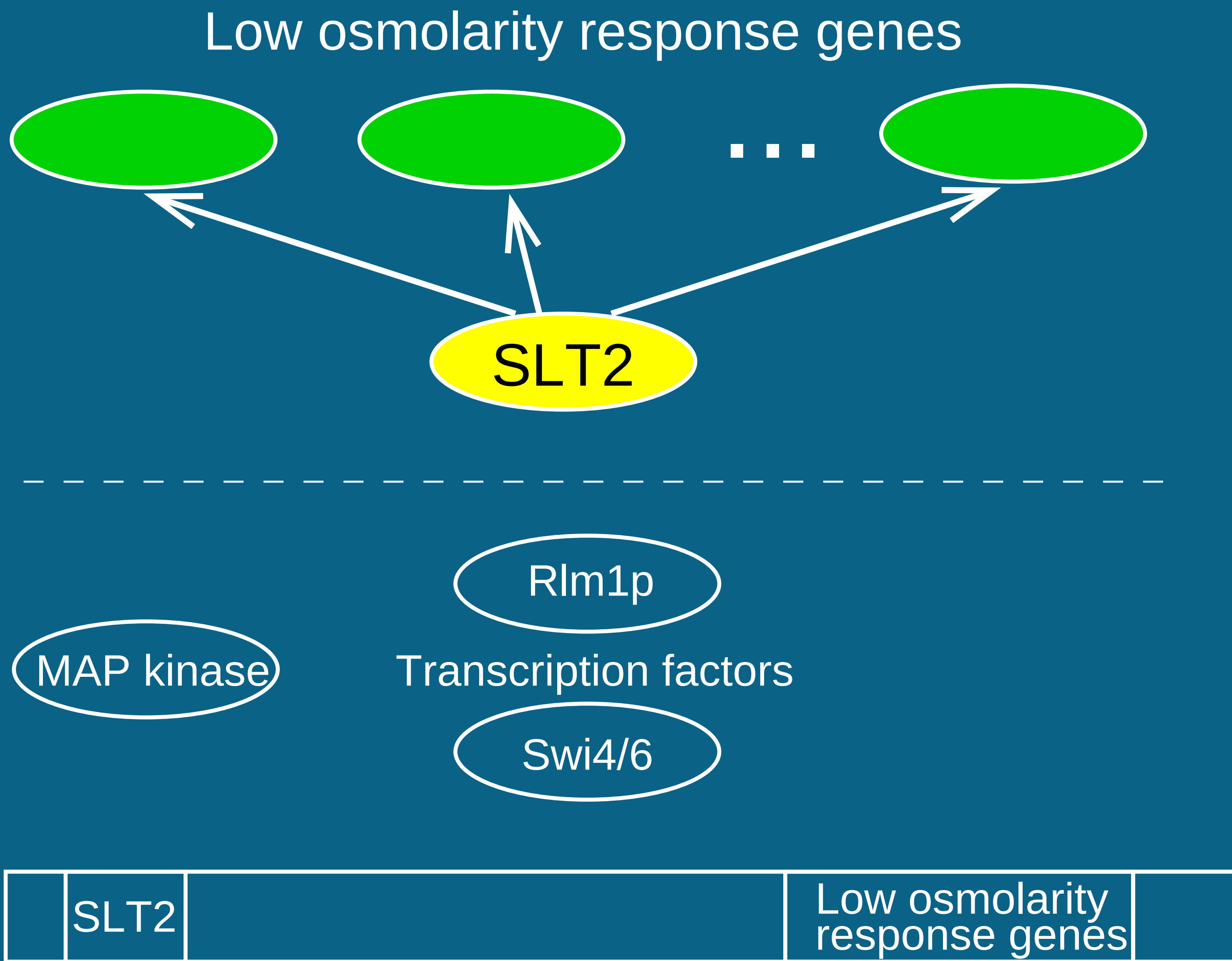
SLT2 clusters with low-osmolarity response genes

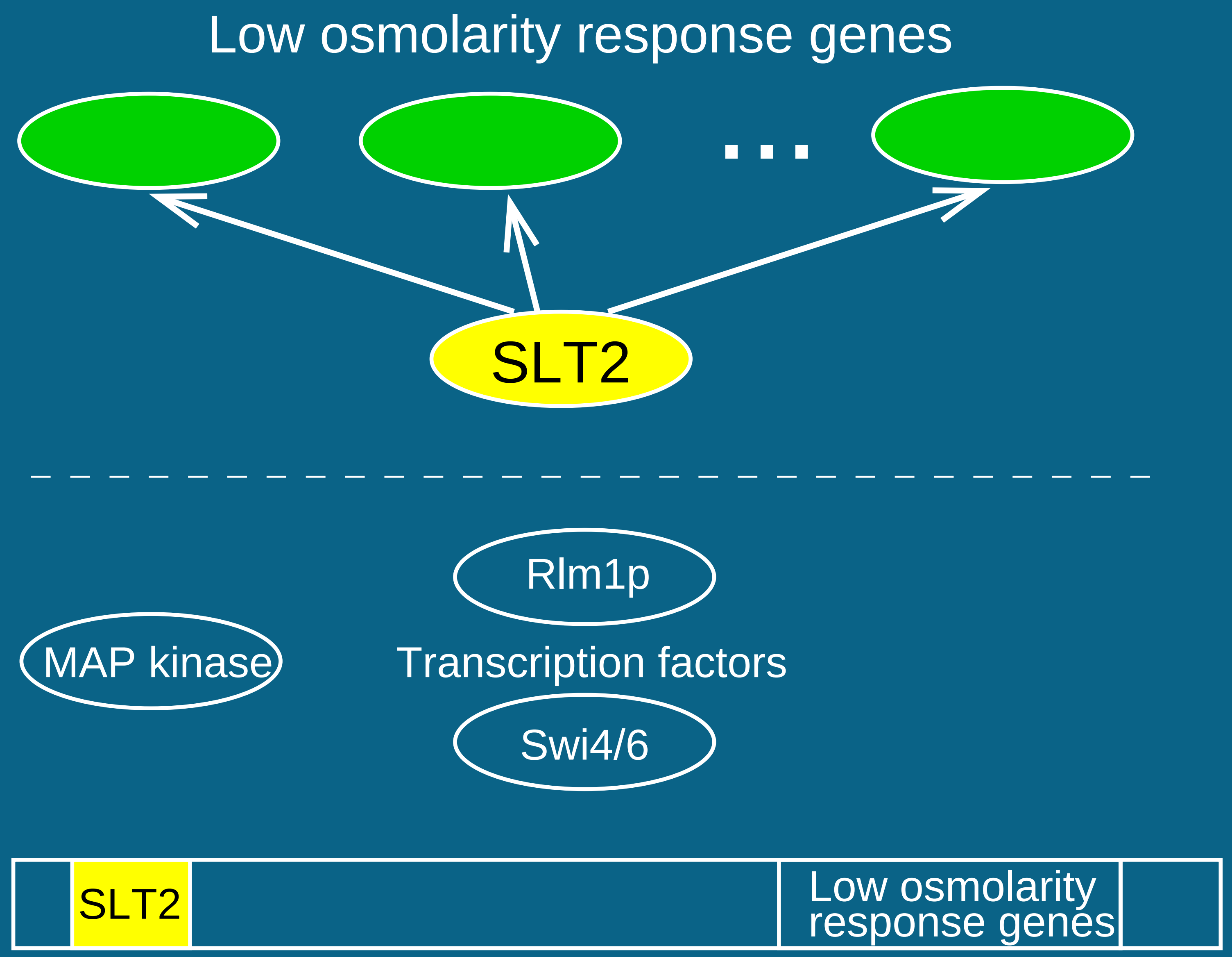
Biological example

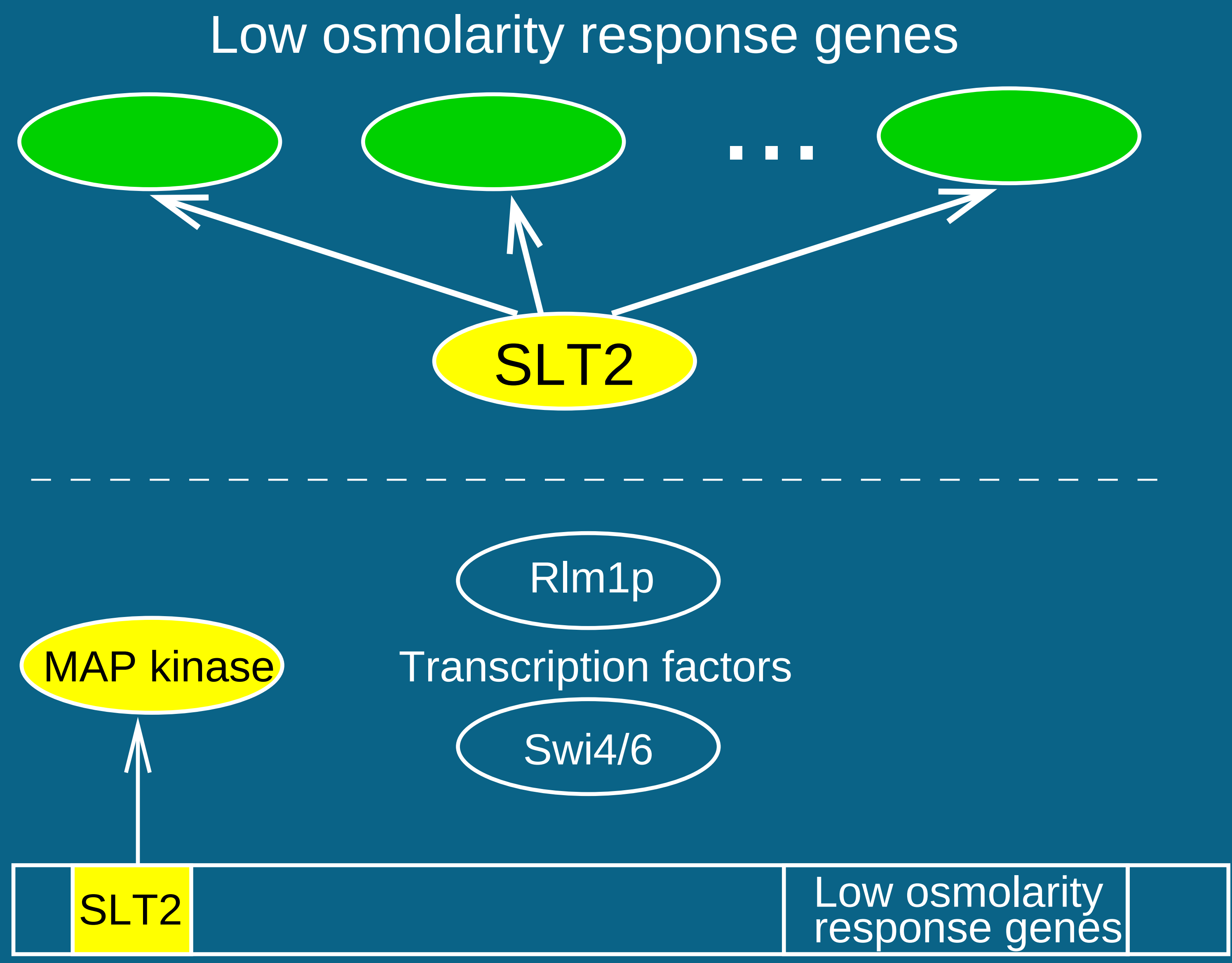
Yeast cell cycle

Nir Friedman et al. (2000)

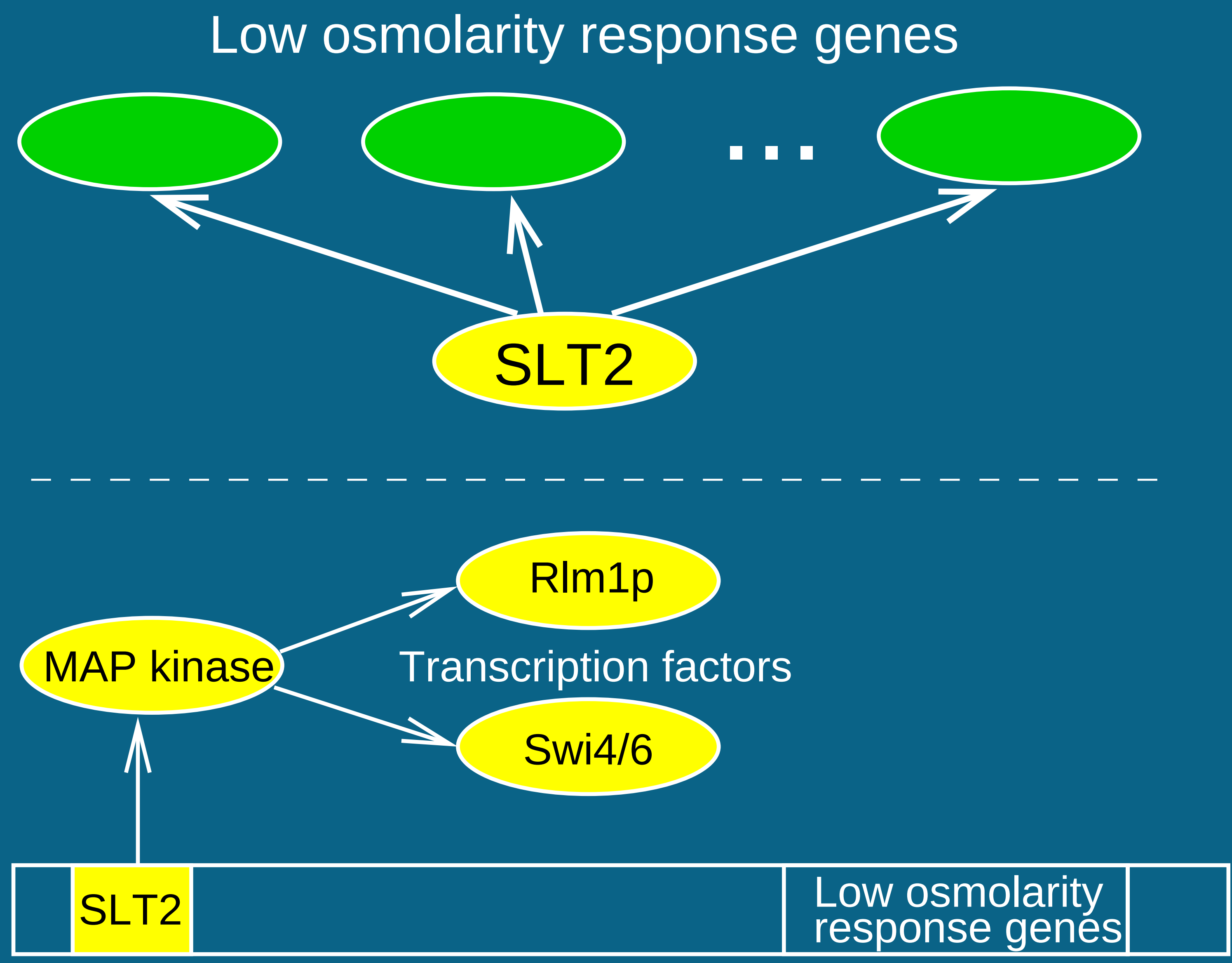
Journal of Computational Biology 7: 601-620

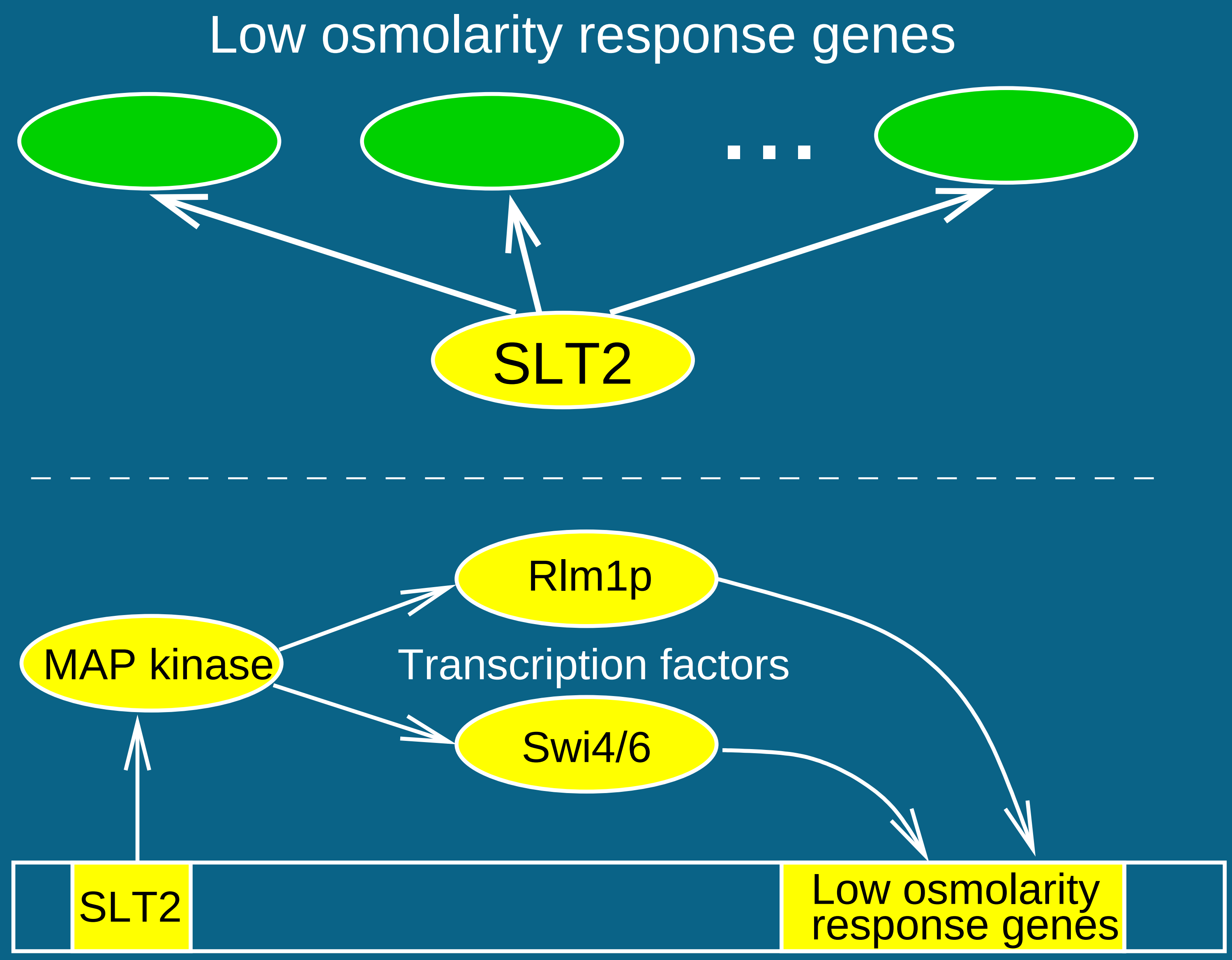

Low osmolarity response genes
...
SLT2
Rlm1p
MAP kinase
Transcription factors
Swi4/6
SLT2
Low osmolarity response genes


Low osmolarity response genes
...
SLT2
Rlm1p
MAP kinase
Transcription factors
Swi4/6
SLT2
Low osmolarity
response genes



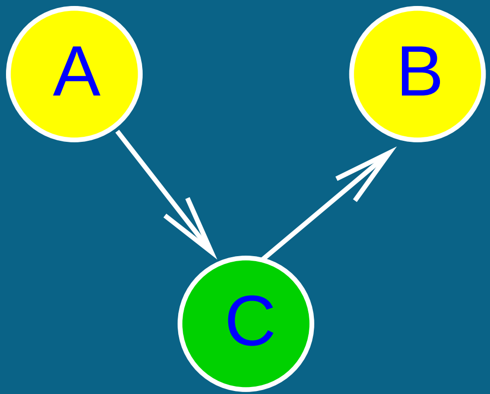
Low osmolarity response genes
...
SLT2
Rlm1p
MAP kinase
Transcription factors
Swi4/6
SLT2
Low osmolarity
response genes


Low osmolarity response genes
...
SLT2
Rlm1p
MAP kinase
Transcription factors
Swi4/6
SLT2
Low osmolarity
response genes


Low osmolarity response genes
...
SLT2
Rlm1p
MAP kinase
Transcription factors
Swi4/6
SLT2
Low osmolarity
response genes

Can we learn causal relationships
from conditional dependencies ?

Can we learn causal relationships from conditional dependencies ?



Bayesian network:
A node is independent of its nondescendants, given its parents.

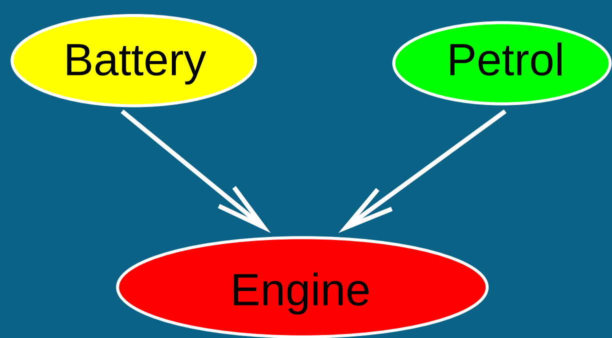
Causal network:
A node is independent of its earlier causes, given its immediate causes.

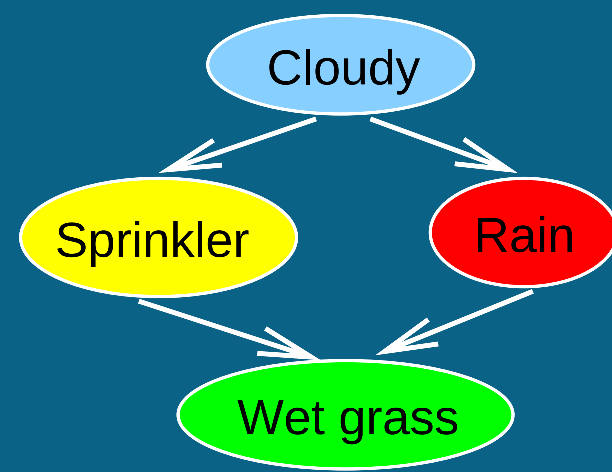
Intuitive notion

Causal explanation of a dependency model

=

Conditional independence facts entailed by the corresponding DAG



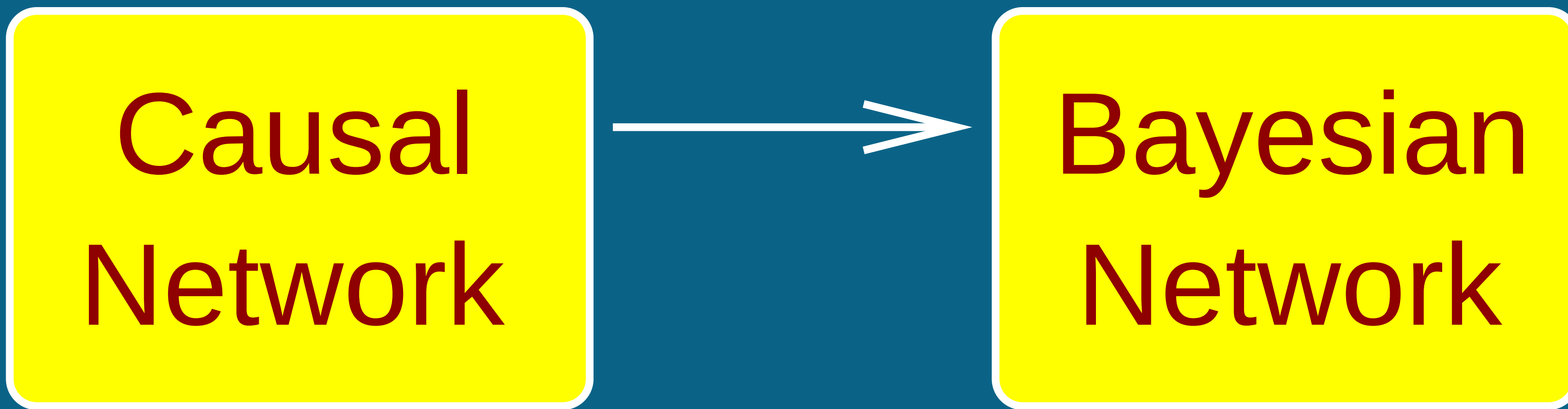


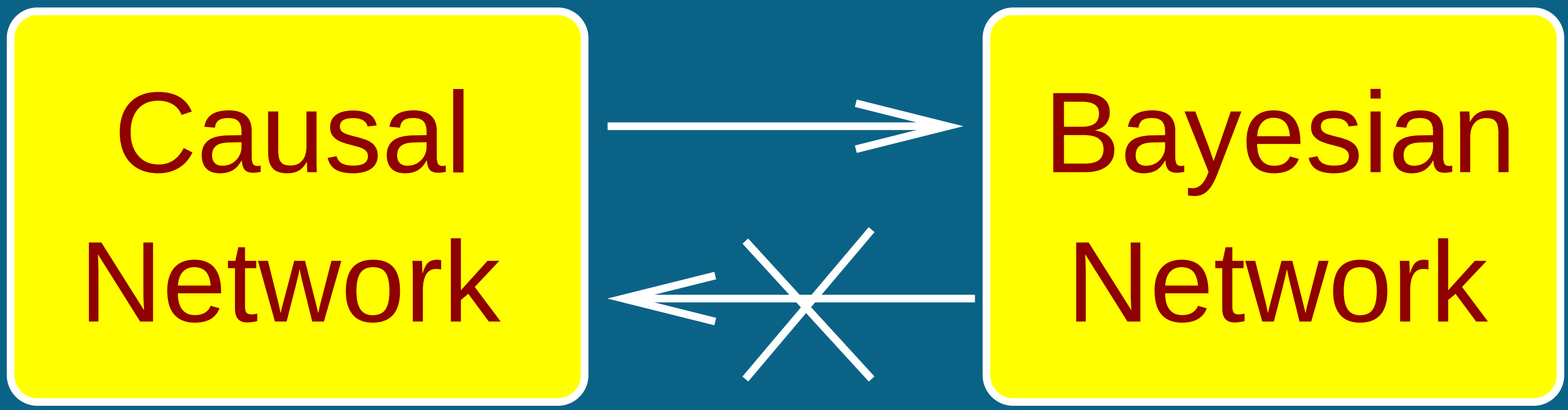
Causal Markov assumption

A domain with causal relationships given by a graph $\mathcal{G}$ satisfies the conditional independencies of the corresponding Bayesian network.

Causal Network

Bayesian Network

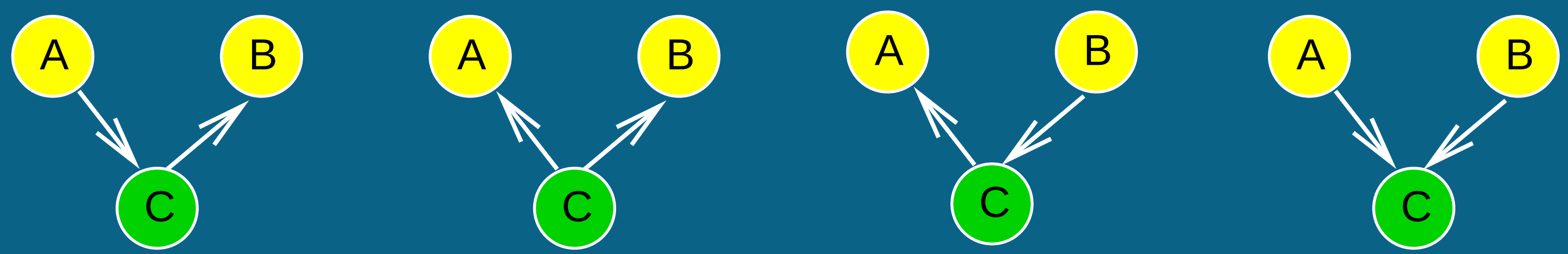

Causal Network
Bayesian Network

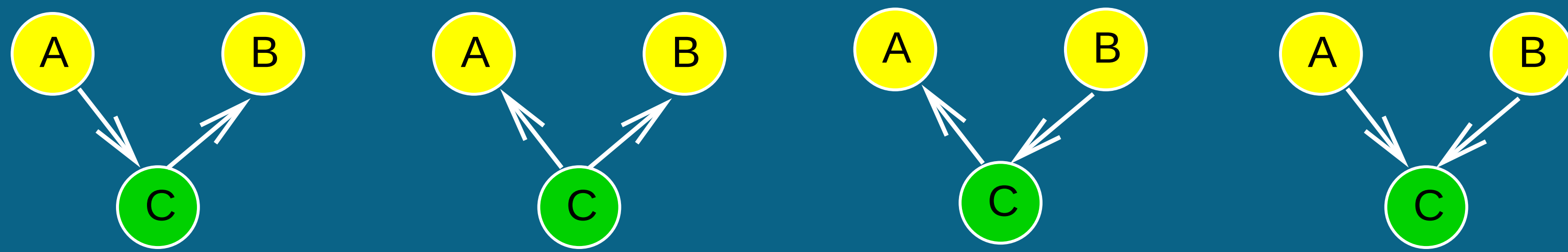

Causal Network
Bayesian Network

Equivalence classes

and

PDAGs


A
B
C
A
B
C
A
B
C
A
B
C



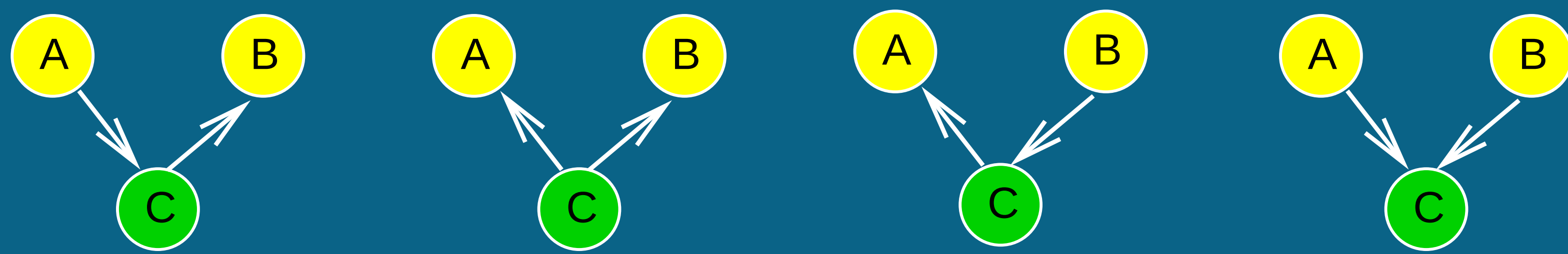
P(A,B,C) =

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)



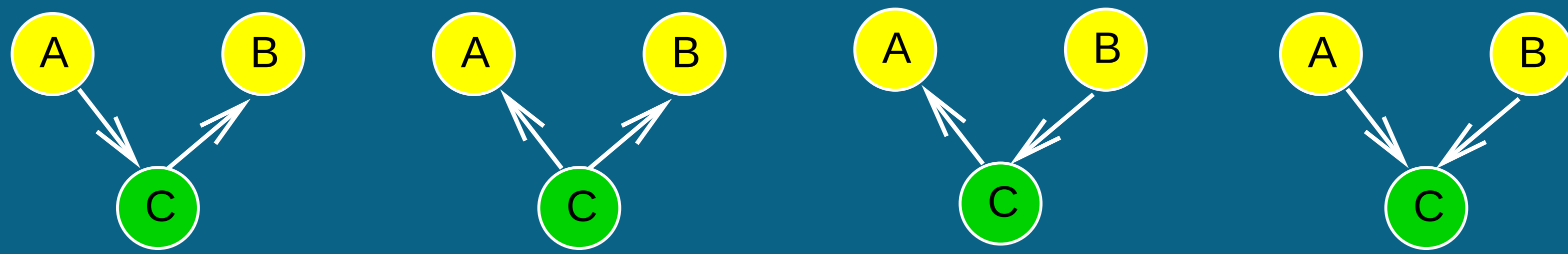
$$P(A,B,C) =$$

$P(B|C)\ P(C|A)\ P(A)$ — where $P(C|A)\ P(A) = P(A|C)\ P(C)$

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

$P(A|C)\ P(C|B)\ P(B)$ — where $P(C|B)\ P(B) = P(B|C)\ P(C)$

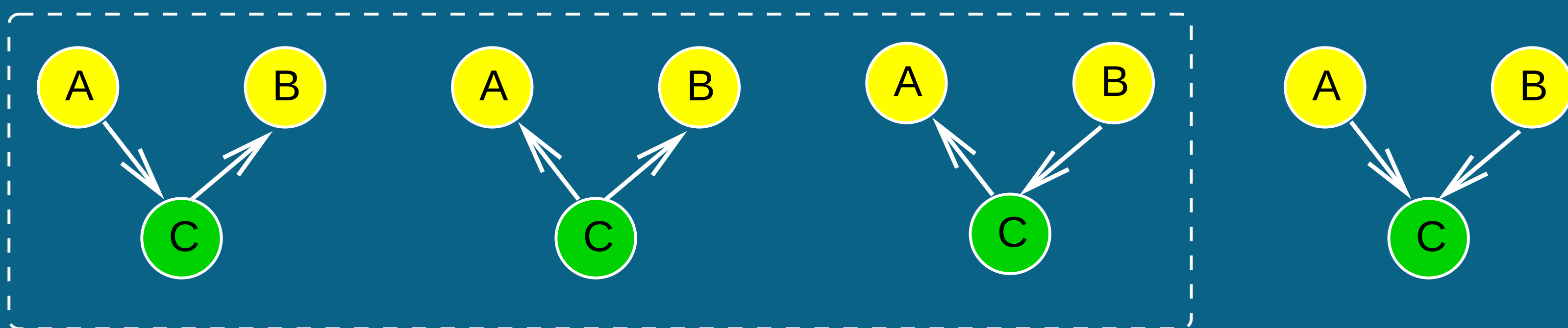
$P(C|A,B)\ P(A)\ P(B)$

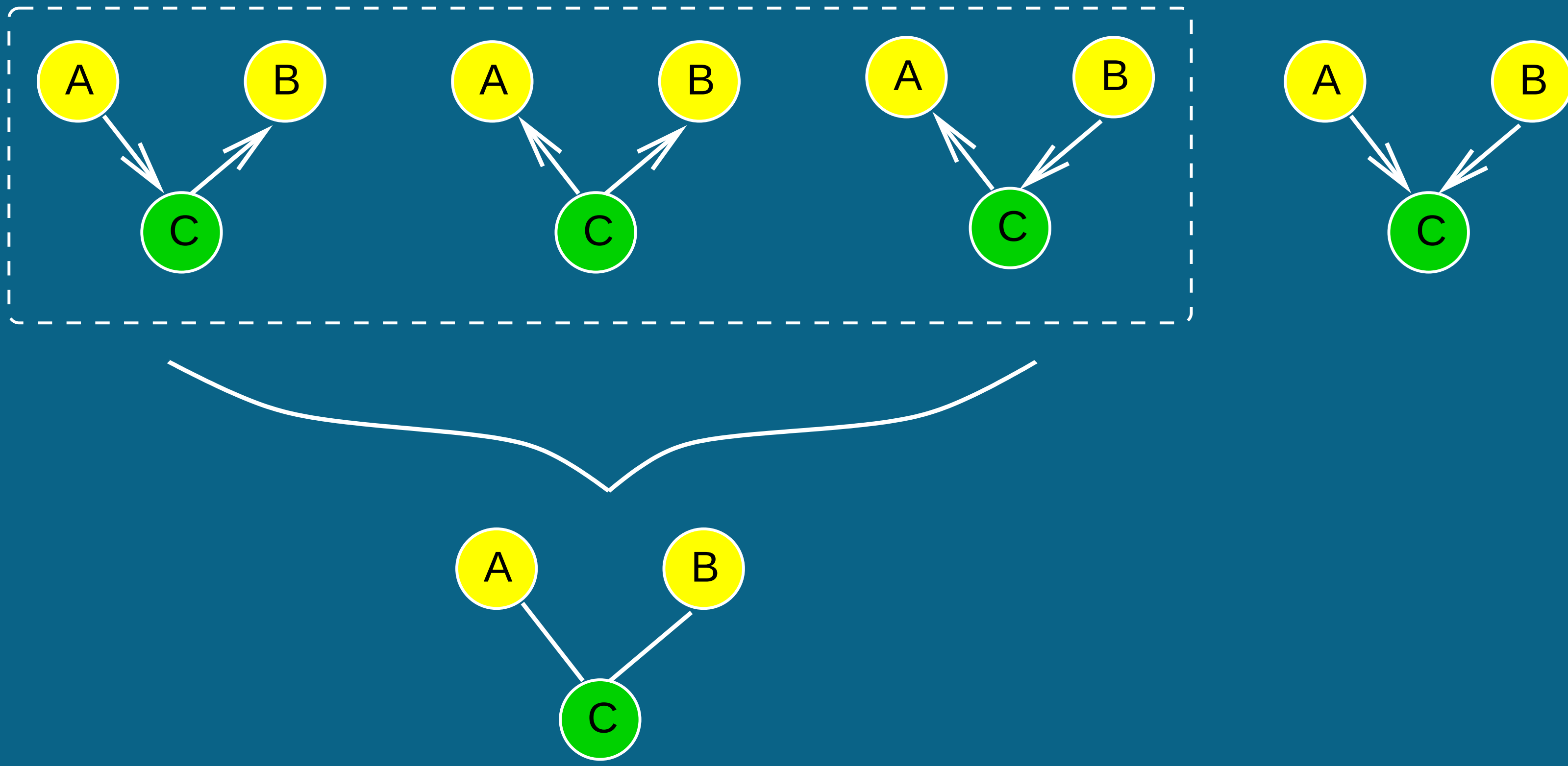


P(A,B,C) =

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)

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

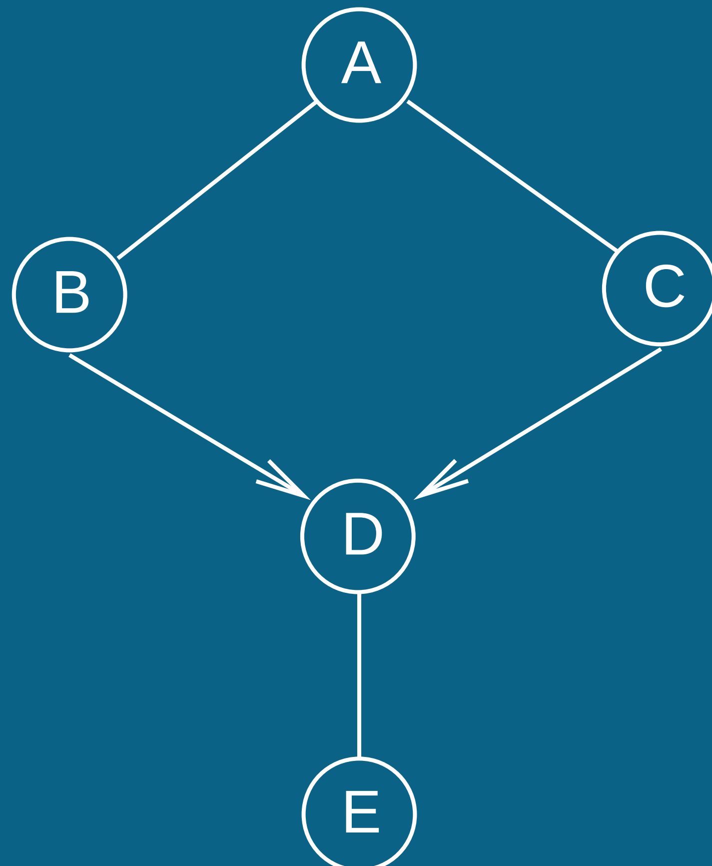




- Two DAGs are equivalent iff they have the same skeleton (= the underlying undirected graph) and the same v-structure.

- v-structure: Converging directed edges into the same node without an edge between the parents.

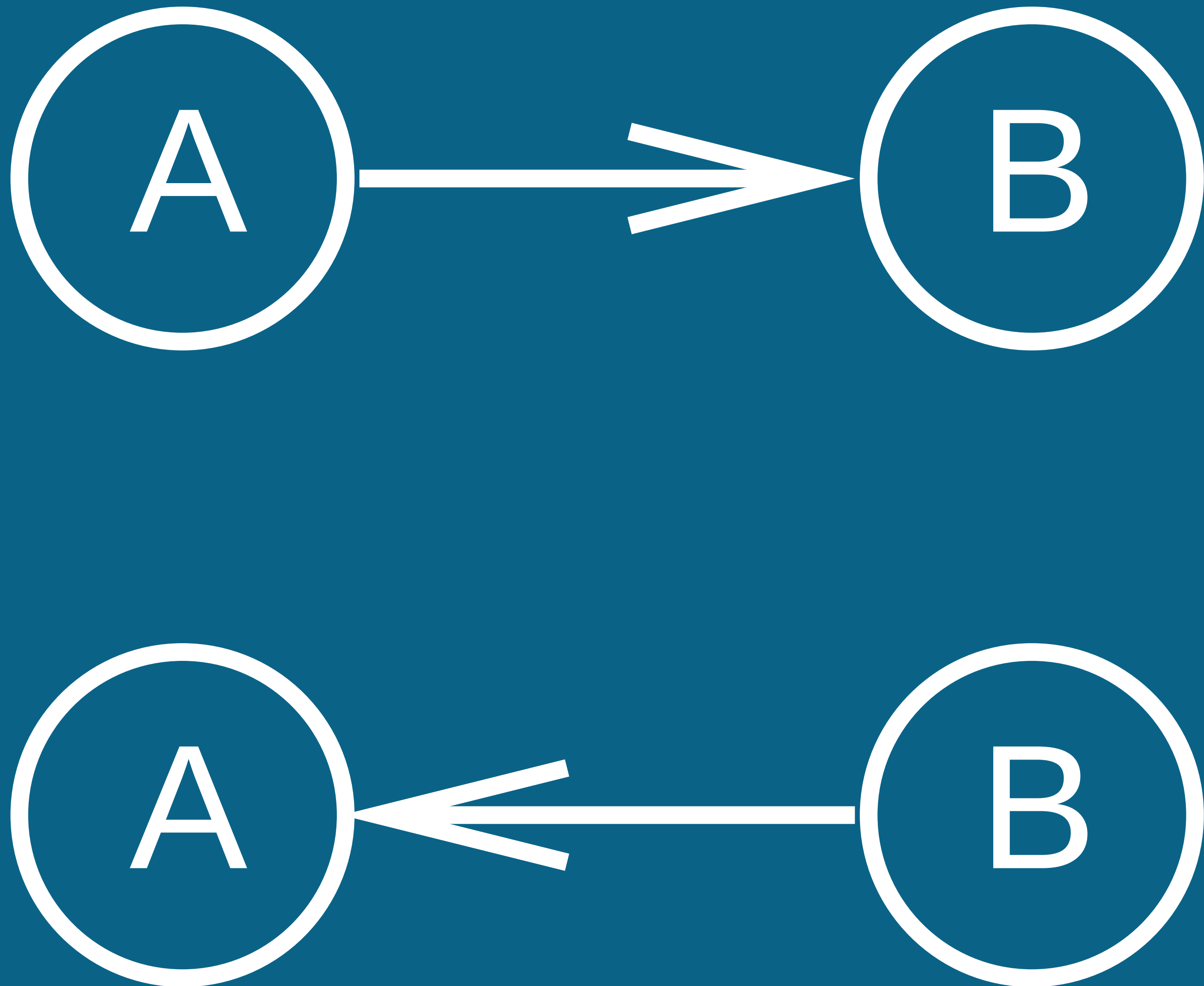
An equivalence class of DAGs can be represented by a PDAG (partially directed acyclic graph).

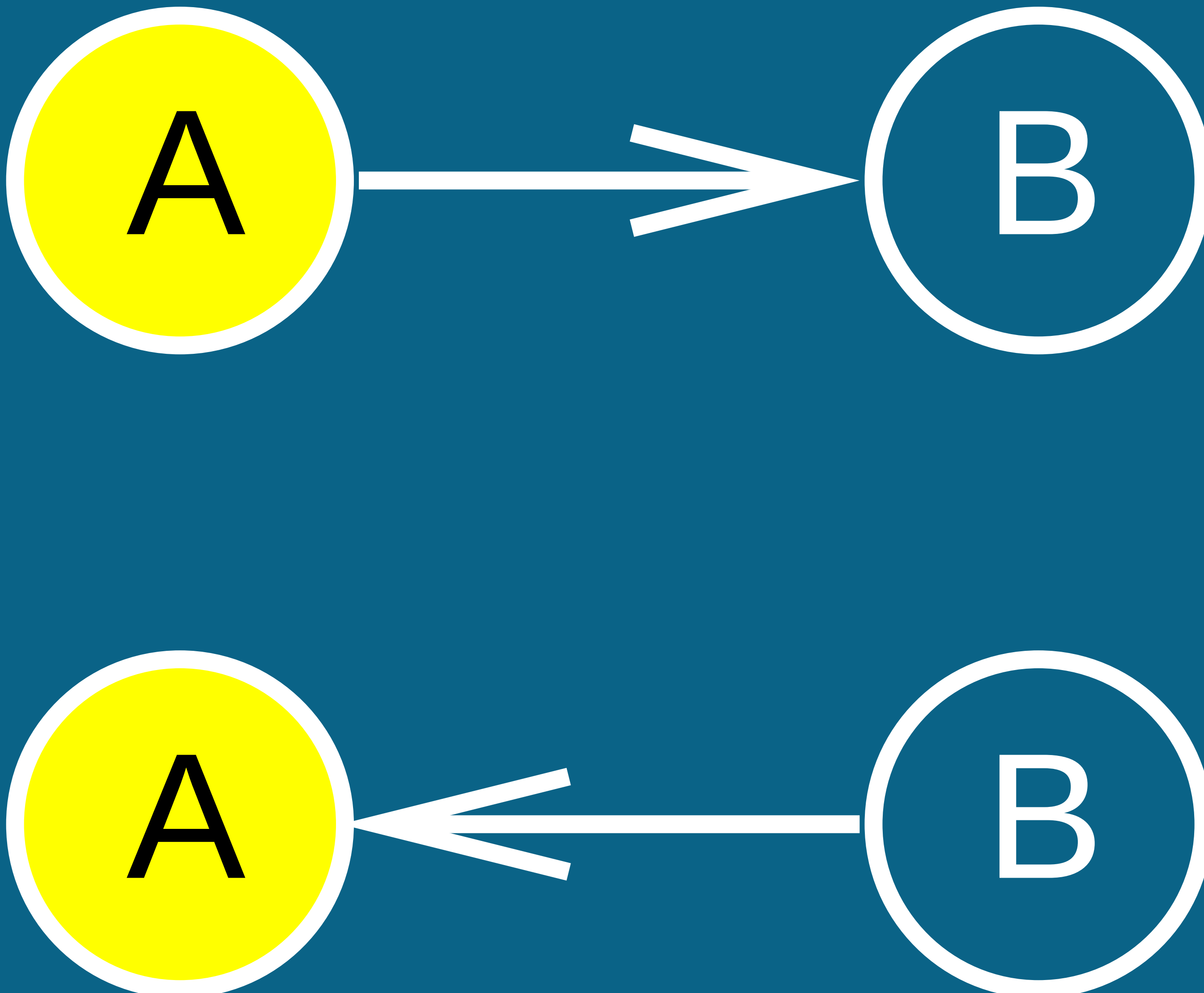


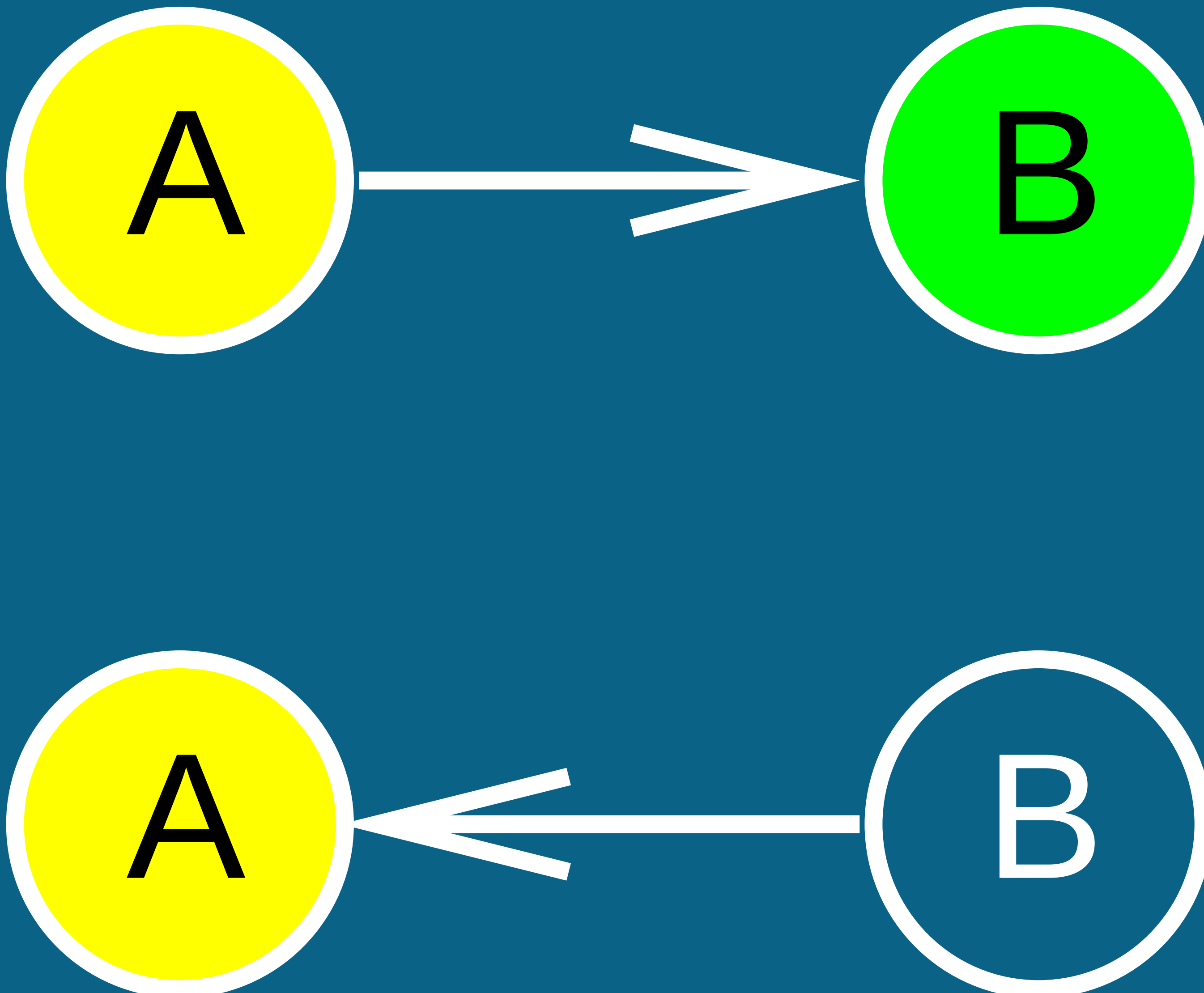
We can only learn PDAGs from the data!

- Observation: A passive measurement of the domain of interest.
- Intervention: Setting the values of some variables using forces outside the causal model, e.g., gene knockout or over-expression

- Observation: A passive measurement of the domain of interest.
- Intervention: Setting the values of some variables using forces outside the causal model, e.g., gene knockout or over-expression
- Interventions can destroy the symmetry within an equivalent class.


A
B
A
B


A
B
A
B


A
B
A
B

Learning with interventions

No intervention:

$$P(D|M) = \prod_i P(X_i|Pa(X_i))$$

Two models $\mathcal{M}$, $\mathcal{M}'$ with the same score are structure equivalent.

Learning with interventions

No intervention:

$$P(D|M) = \prod_i P(X_i|Pa(X_i))$$

Two models $\mathcal{M}$, $\mathcal{M}'$ with the same score are structure equivalent.

Int: Set of interventions $\longrightarrow$ Modified score:

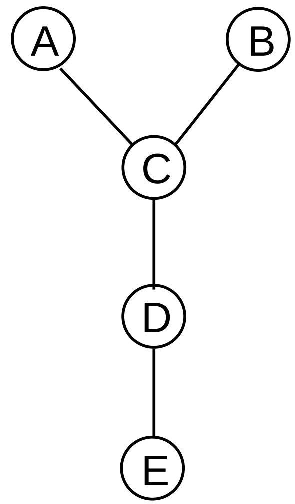
$$P(D|M) = \prod_{i, X_i \notin Int} P(X_i|Pa(X_i))$$

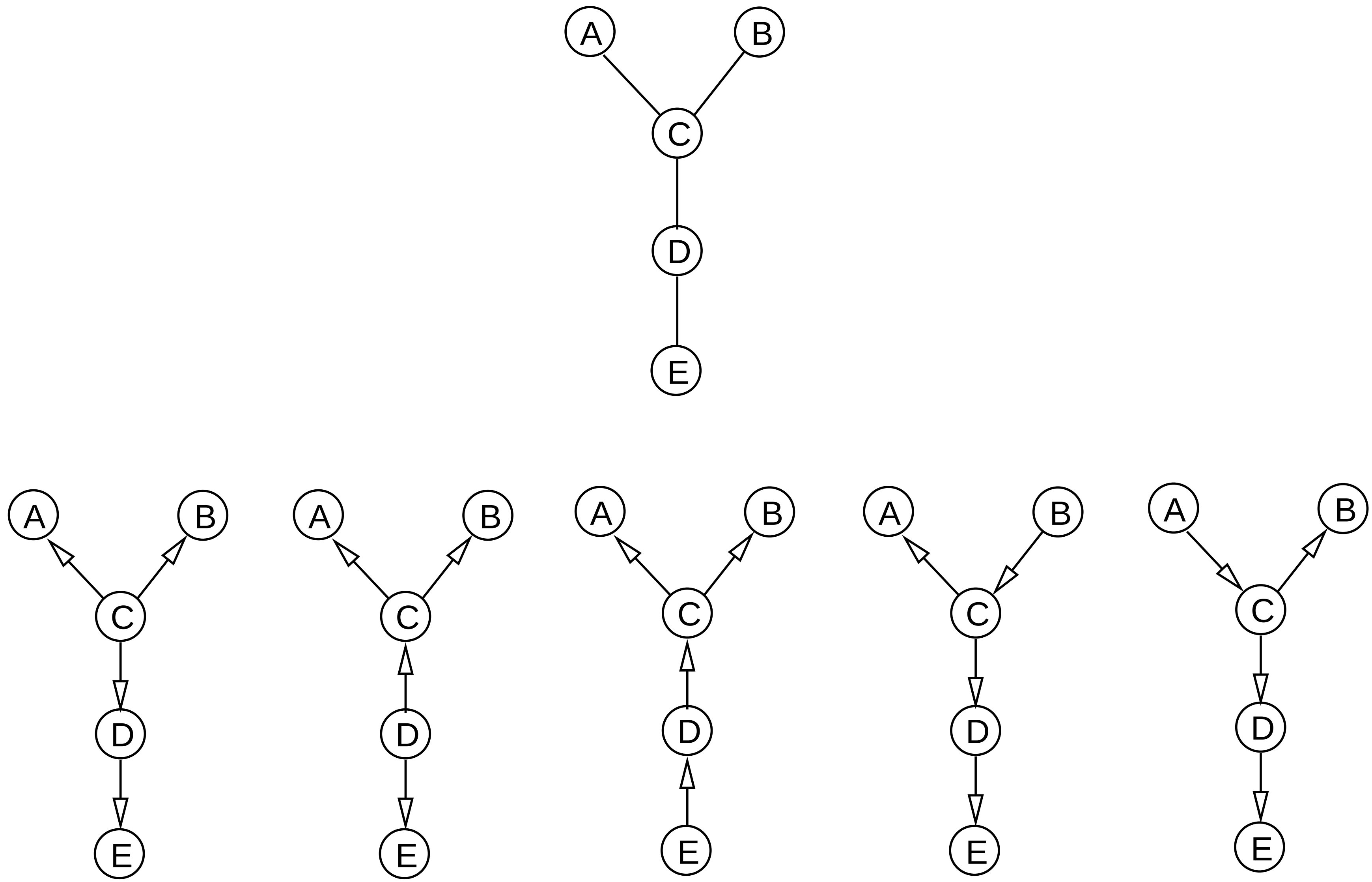
This score is no longer structure equivalent.

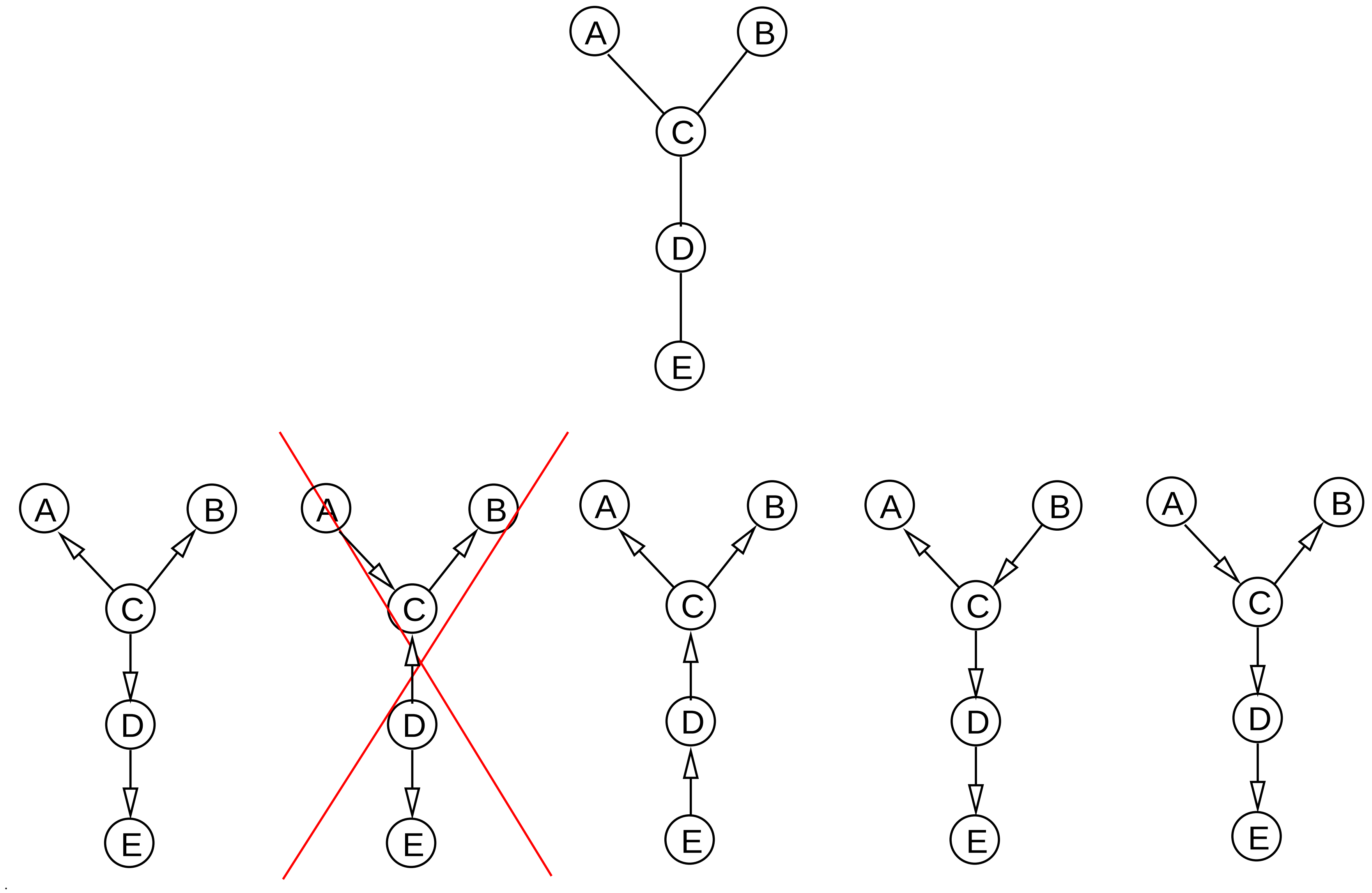
$\mathcal{M}$, $\mathcal{M}'$ are intervention equivalent, if they have the same score given the data and the interventions.

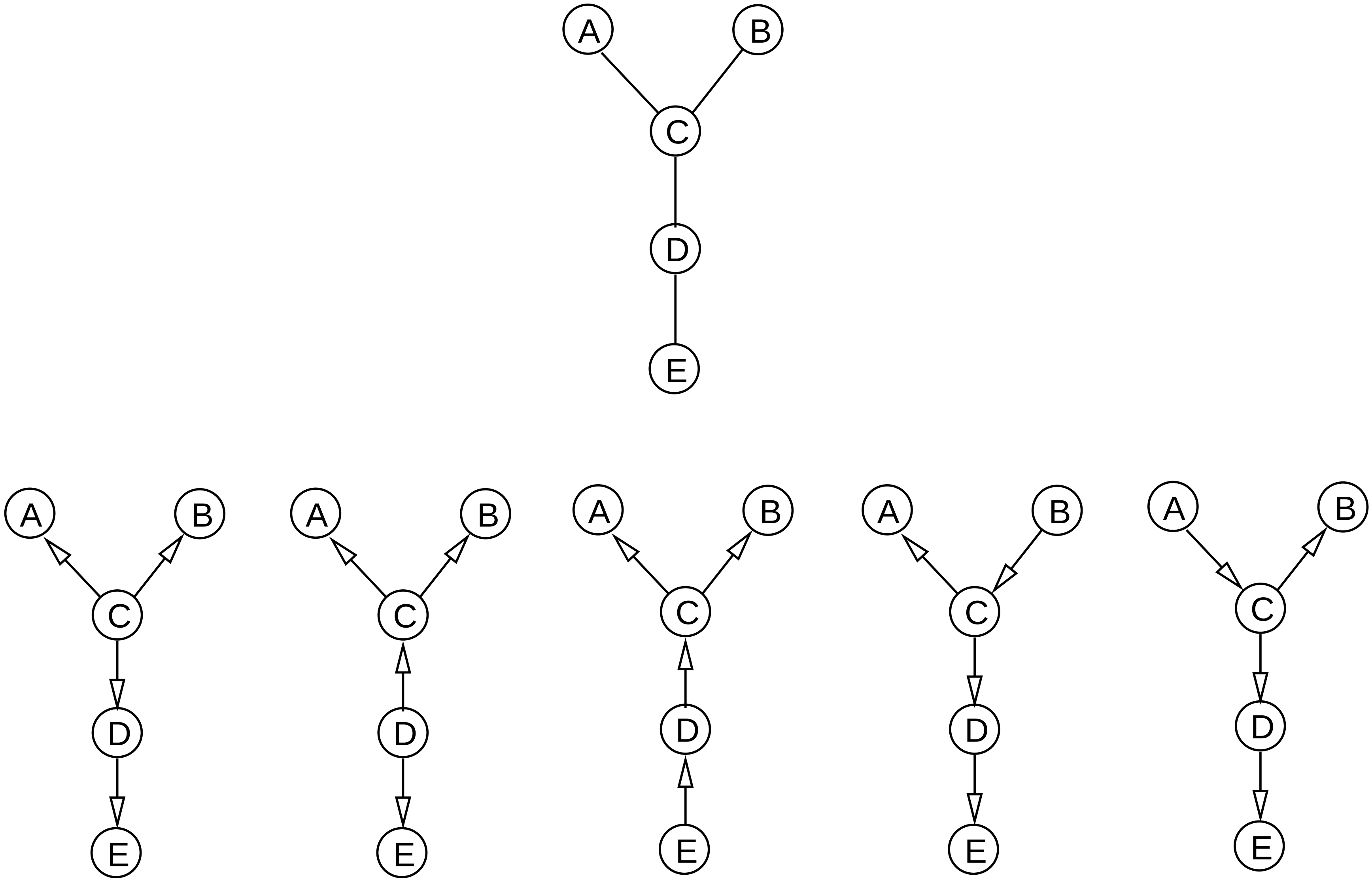
$$\{\mathcal{M}|\text{intervention equivalent}\} \subset \{\mathcal{M}|\text{structure equivalent}\}$$

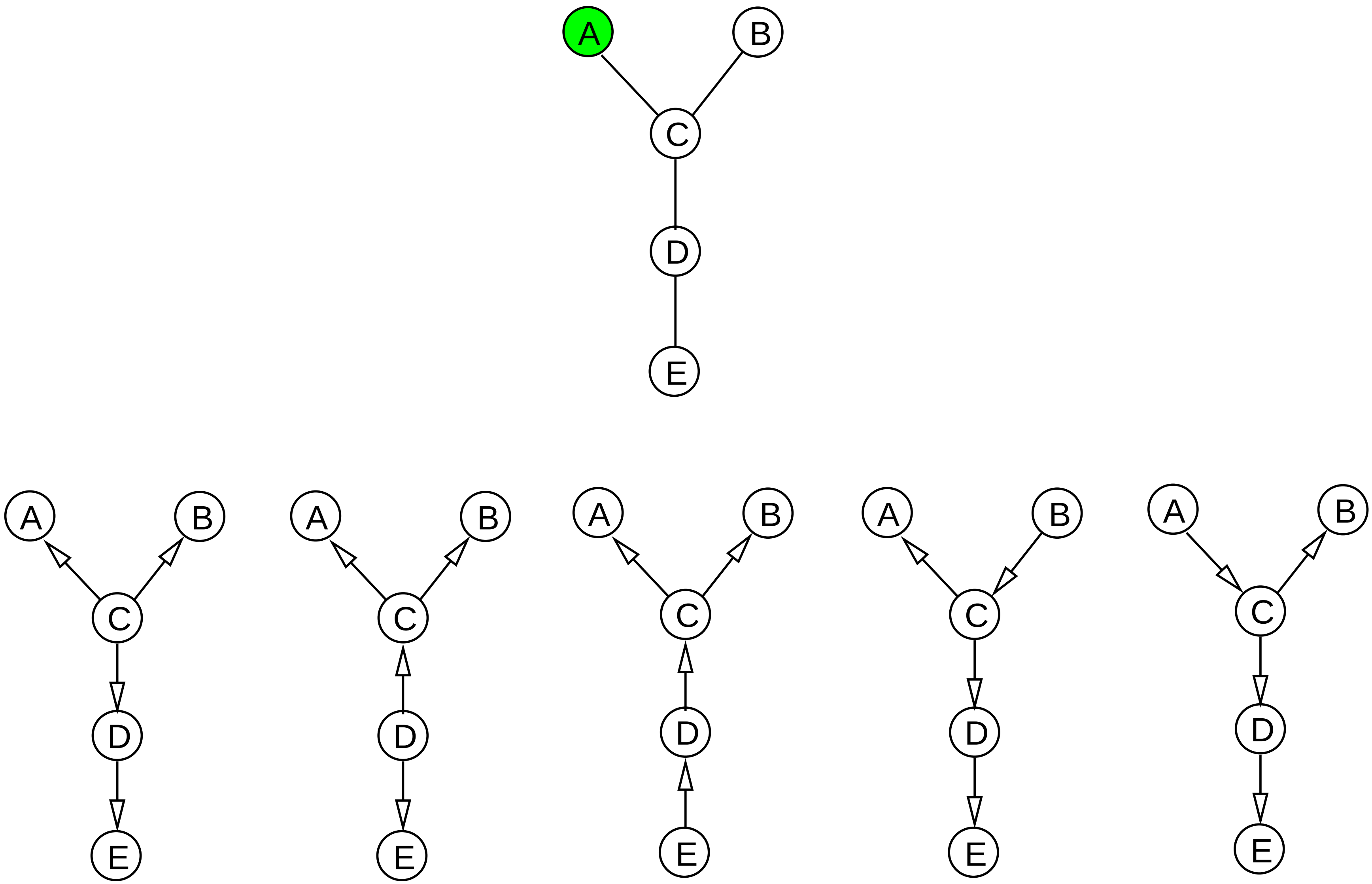
Result of interventions: More edges in the PDAG will be directed, including all edges entering or leaving an intervened variable.

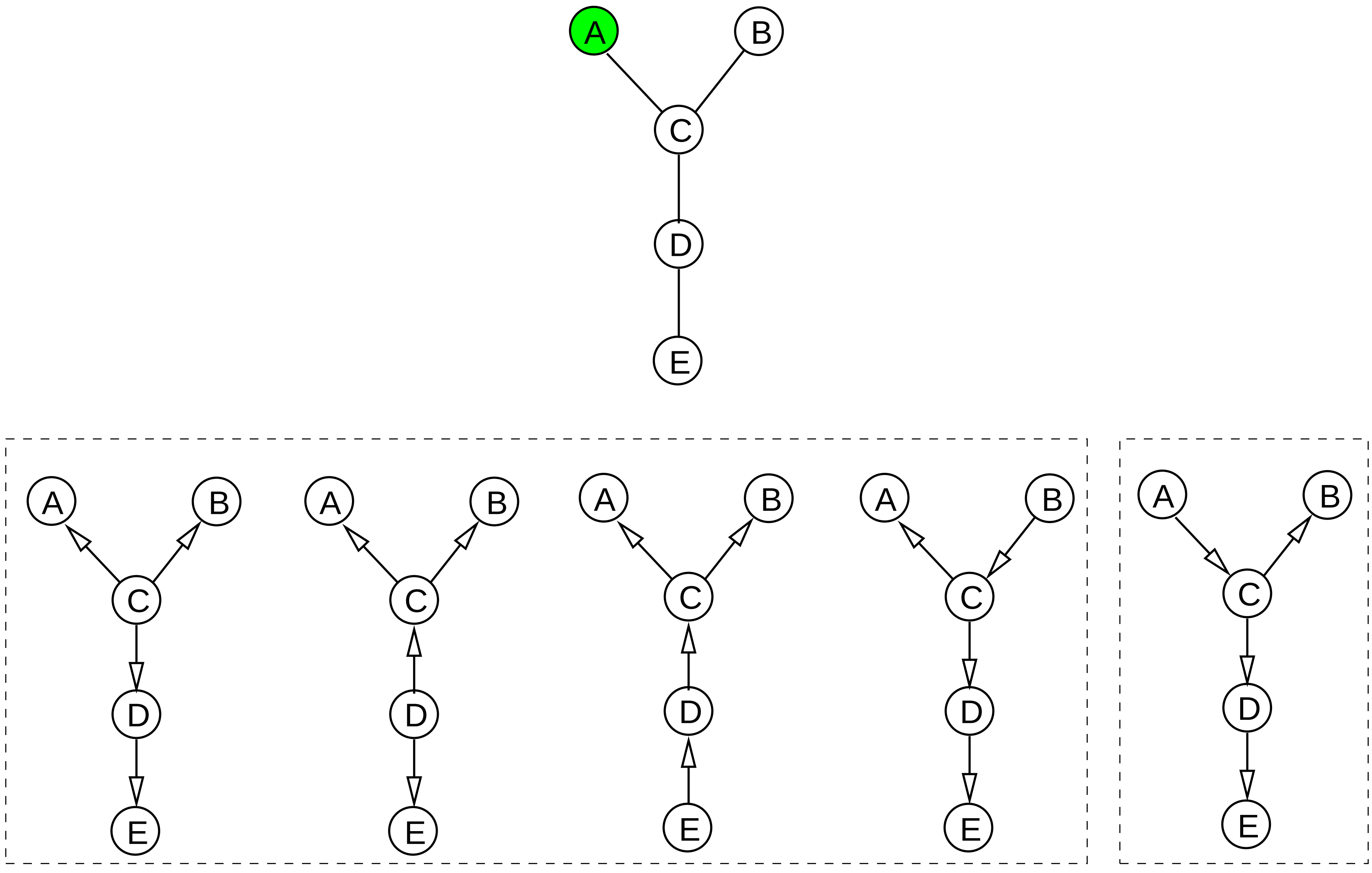

A
B
C
D
E


A
B
C
D
E

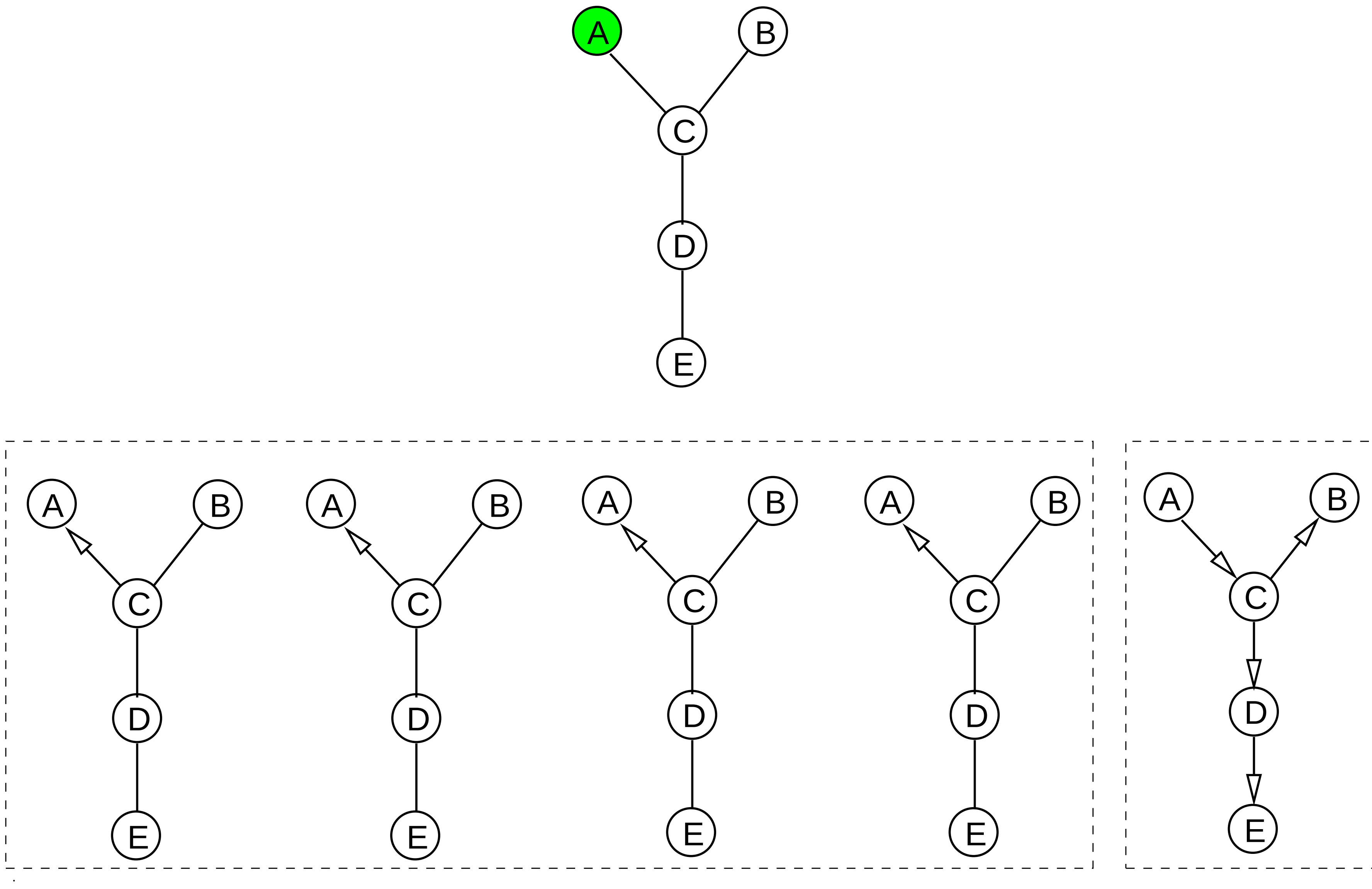

A
B
C
D
E

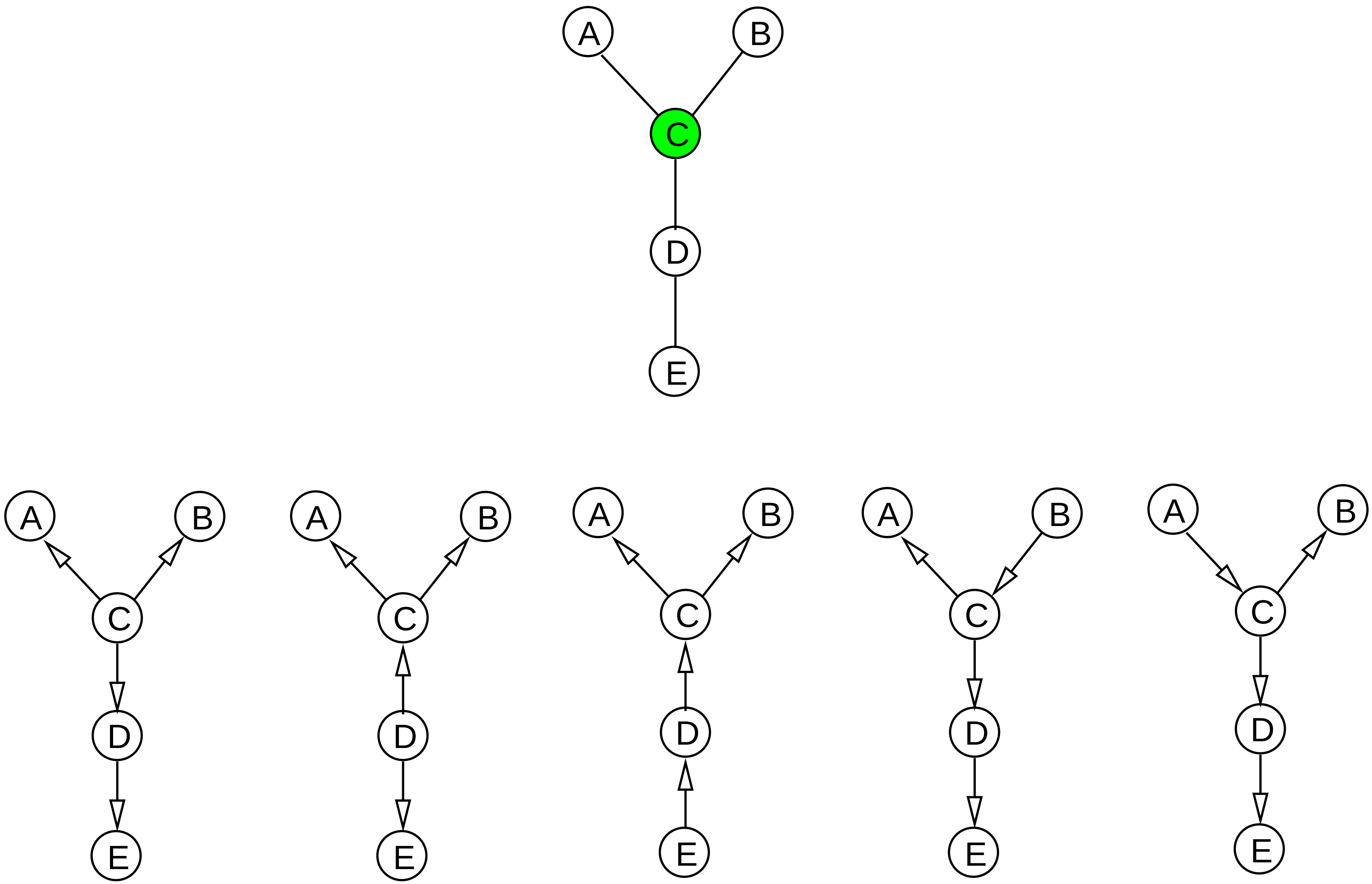

A
B
C
D
E

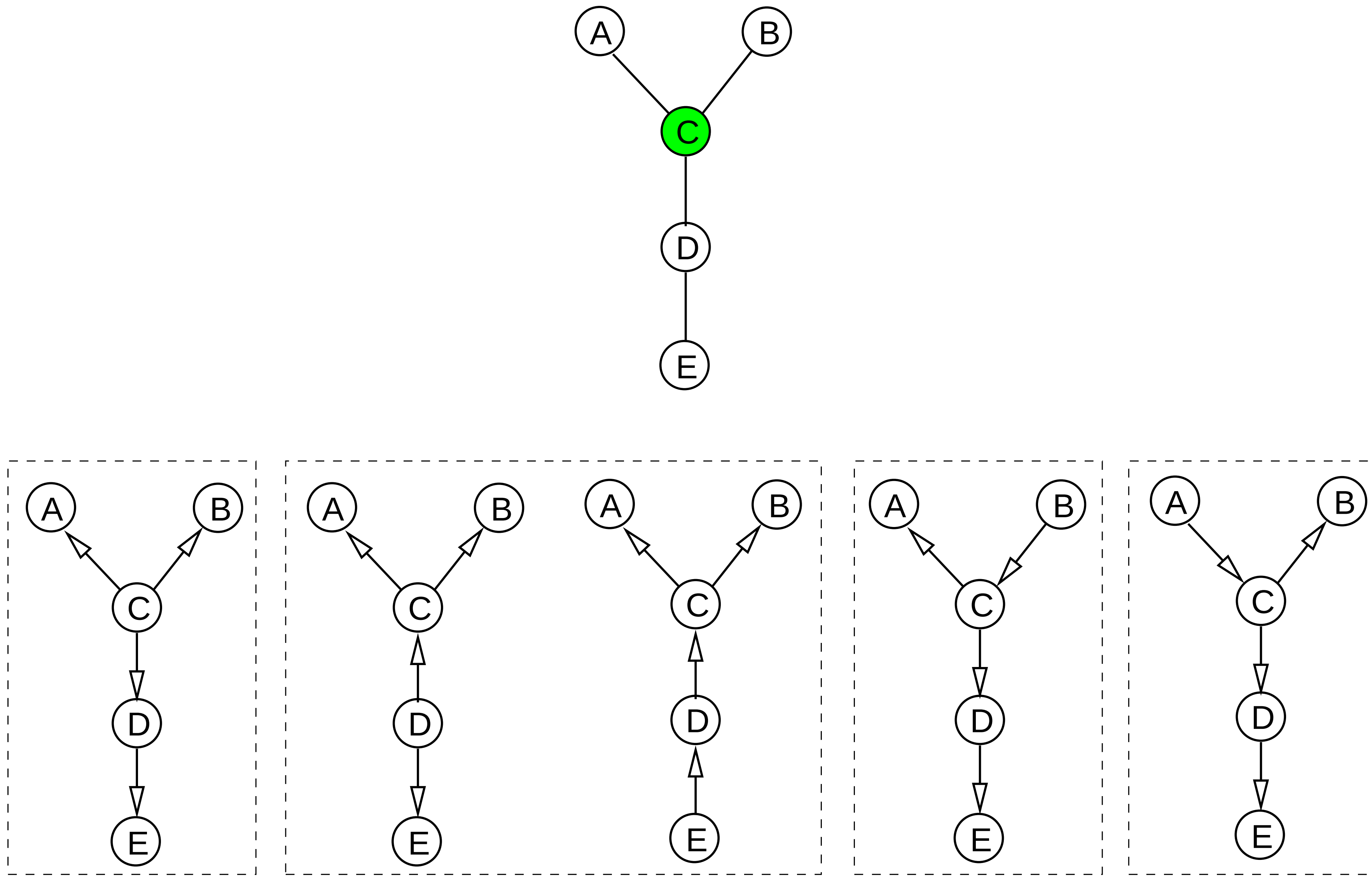

A
B
C
D
E
A
B
C
D
E
A
B
C
D
E
A
B
C
D
E
A
B
C
D
E
A
B
C
D
E



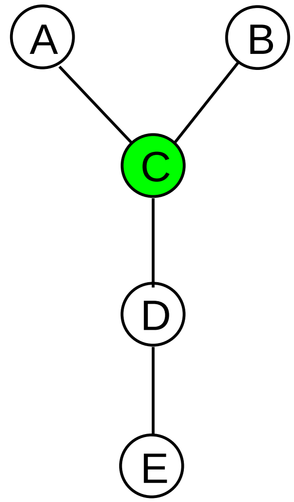
A
B
C
D
E

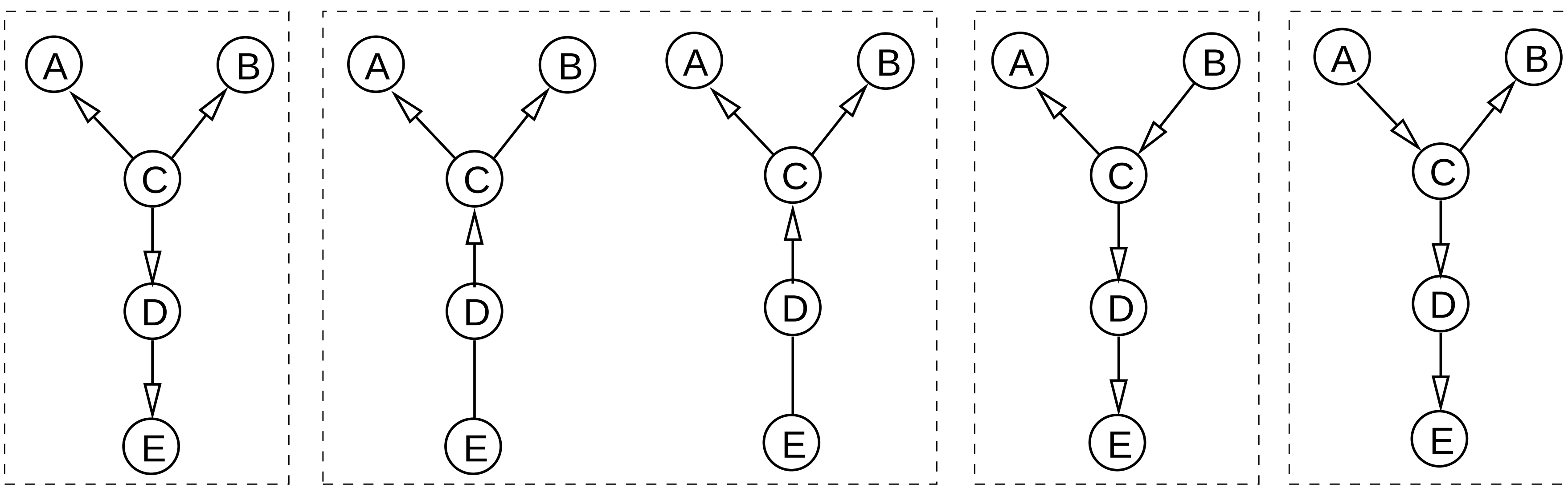

A
B
C
D
E


A
B
C
D
E



A
B
C
D
E

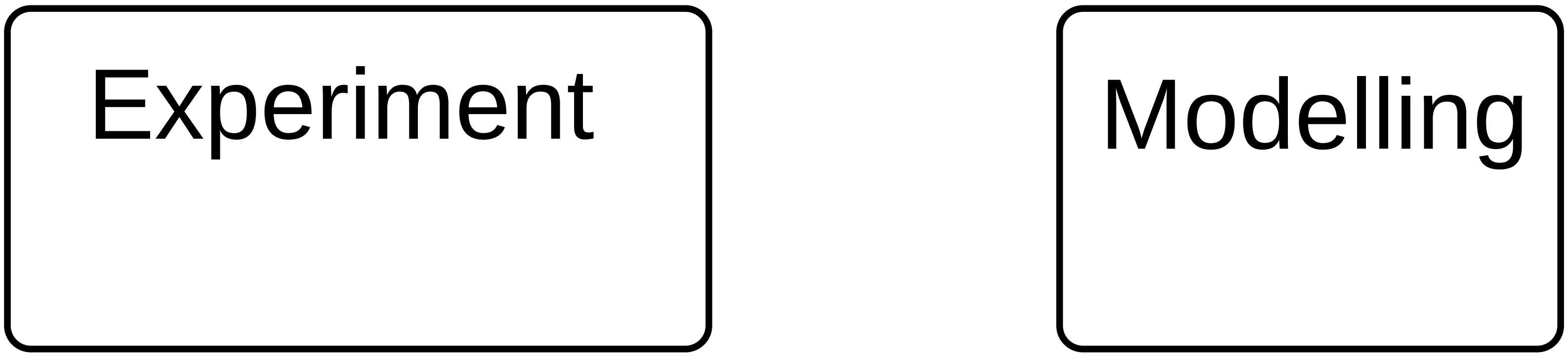

A
B
C
D
E

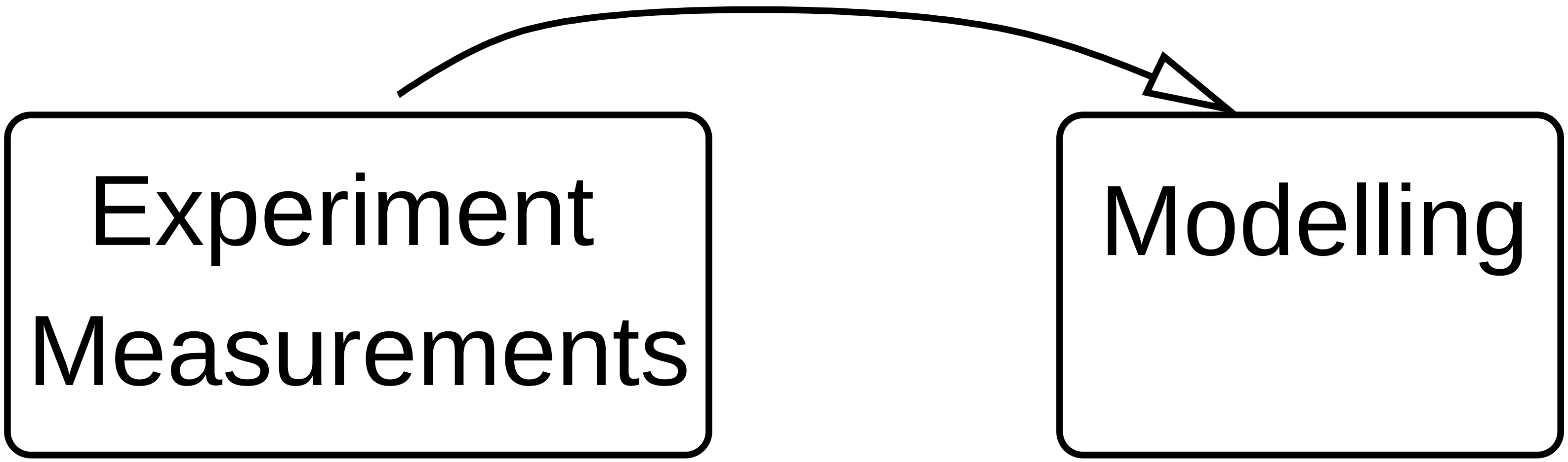

A
B
C
D
E

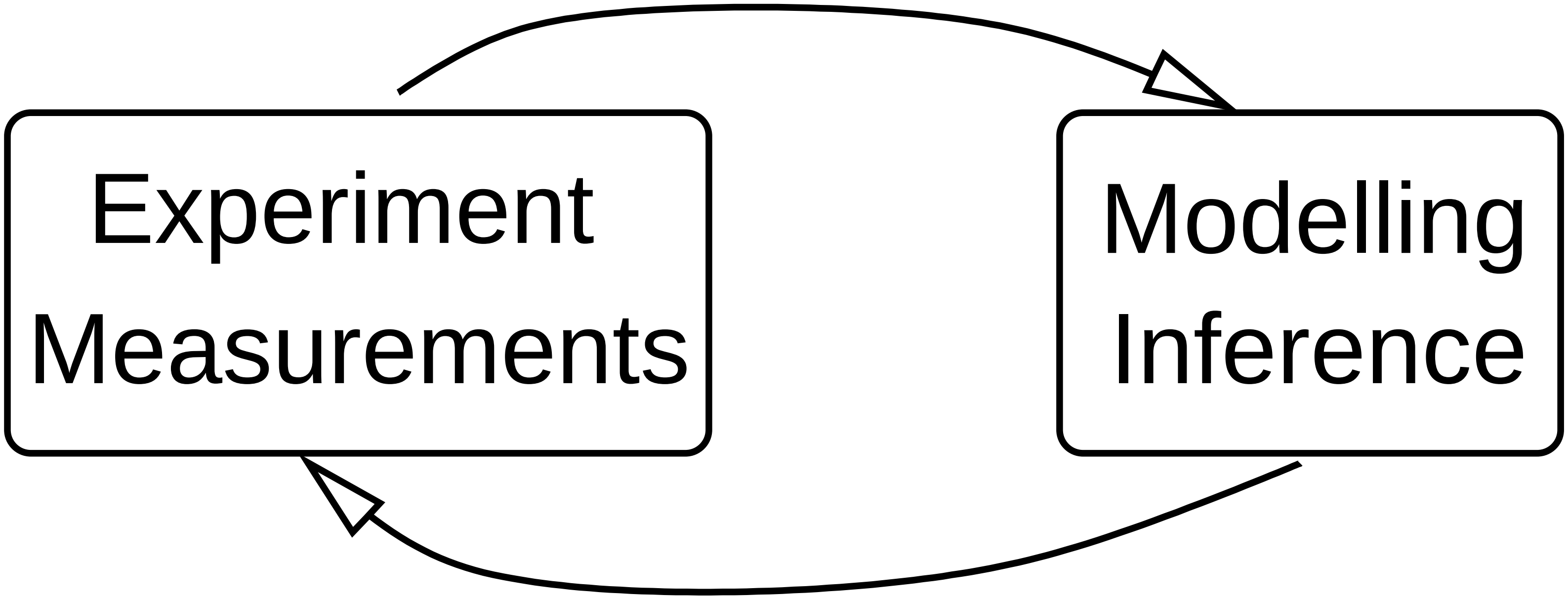
Active learning

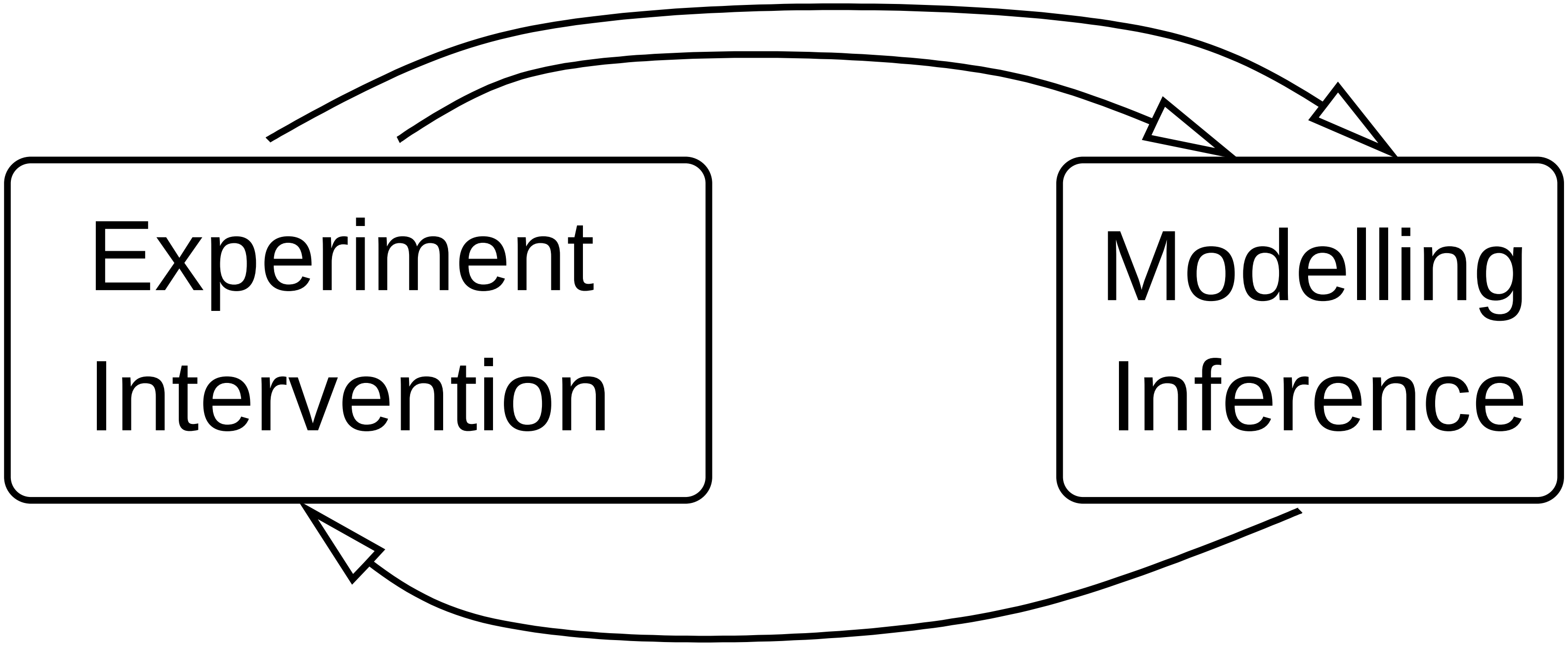
Based on preliminary inference:

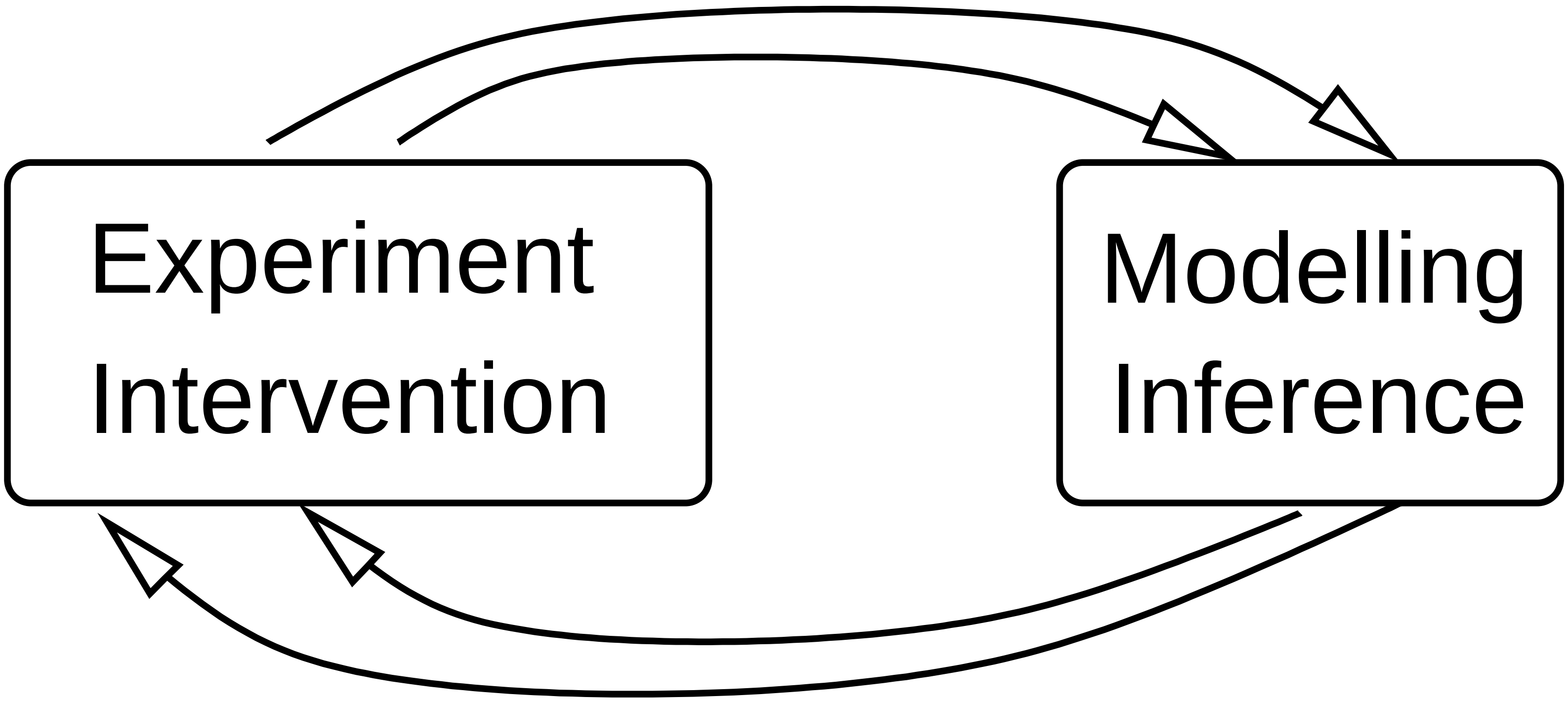
Predict the intervention that maximizes the information content of the expected response.


Experiment
Modelling


Experiment
Measurements
Modelling


Experiment
Measurements
Modelling
Inference


Experiment
Intervention
Modelling
Inference


Experiment
Intervention
Modelling
Inference

Reconstructing gene networks by passive and active Bayesian learning

PhD thesis

Iosifina Pournara

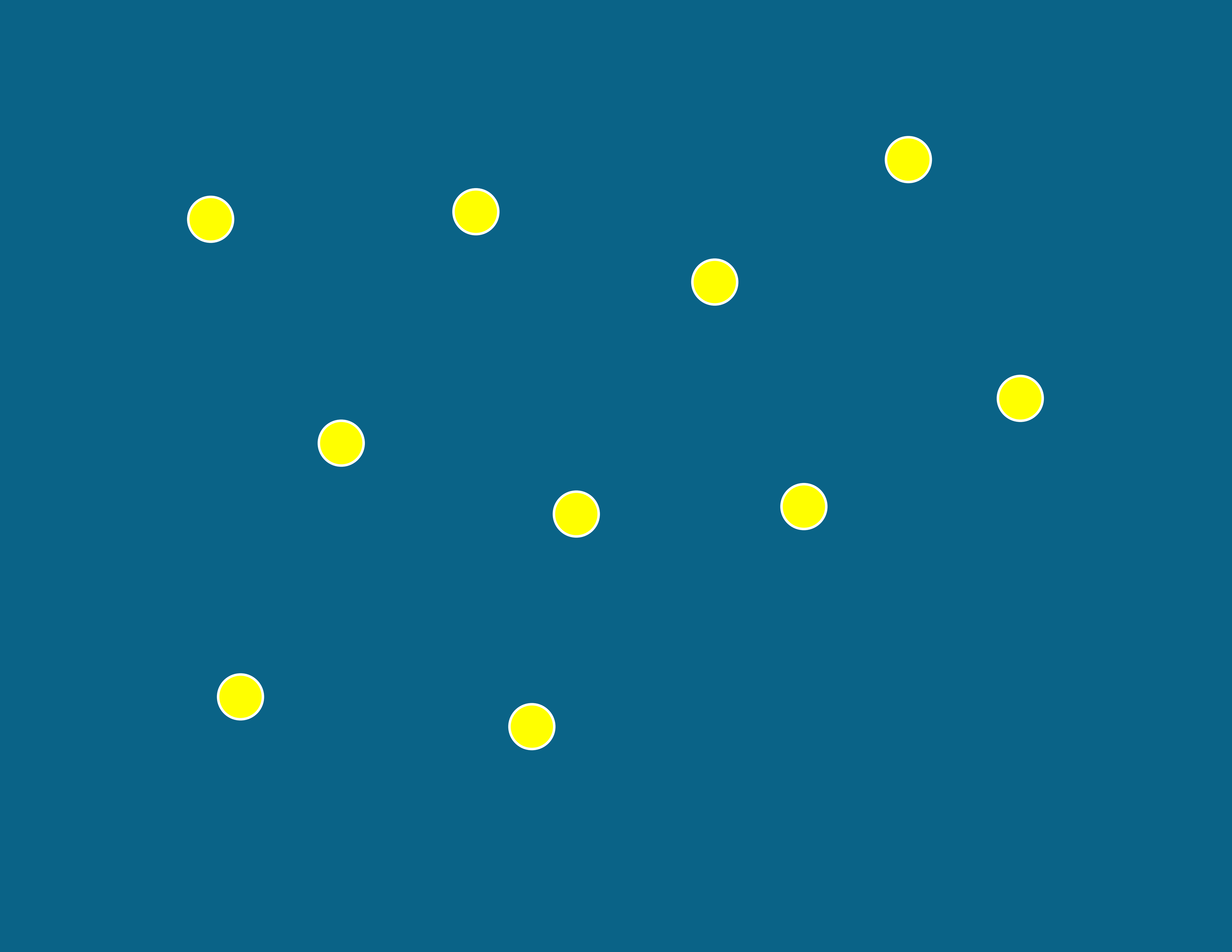
Birbeck College London, January 2005

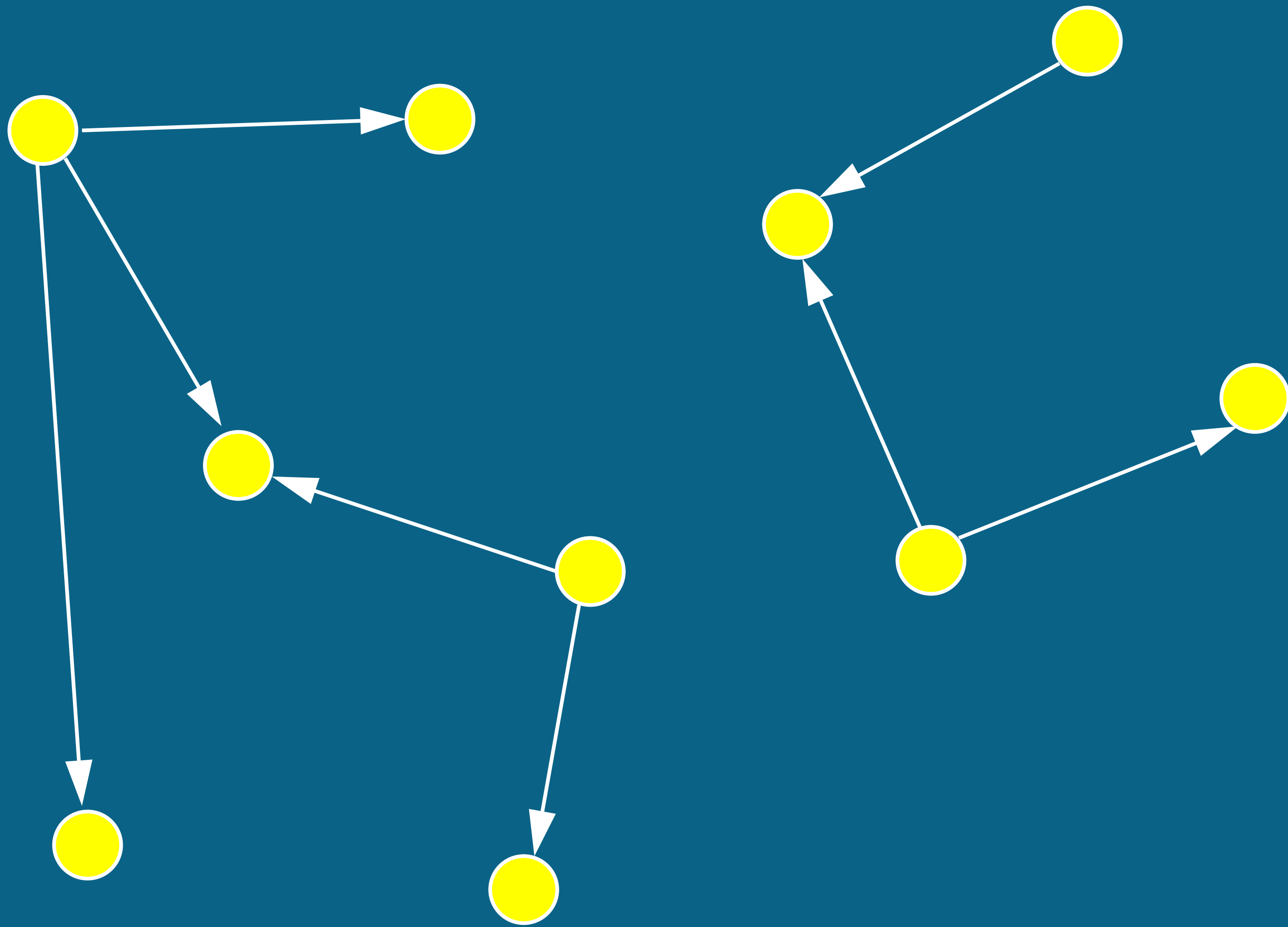
Dynamic Bayesian Networks

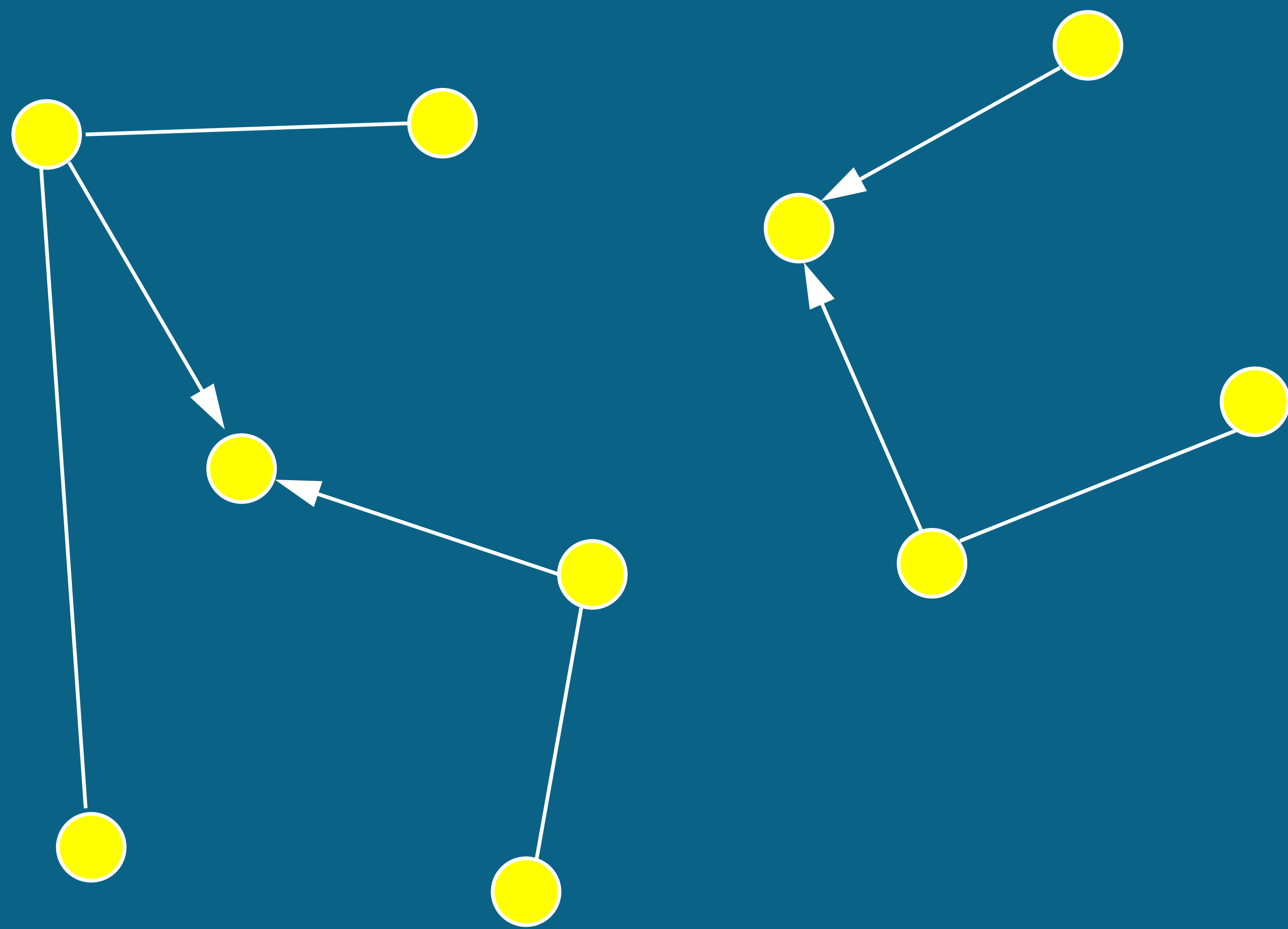
Dynamic Bayesian Networks

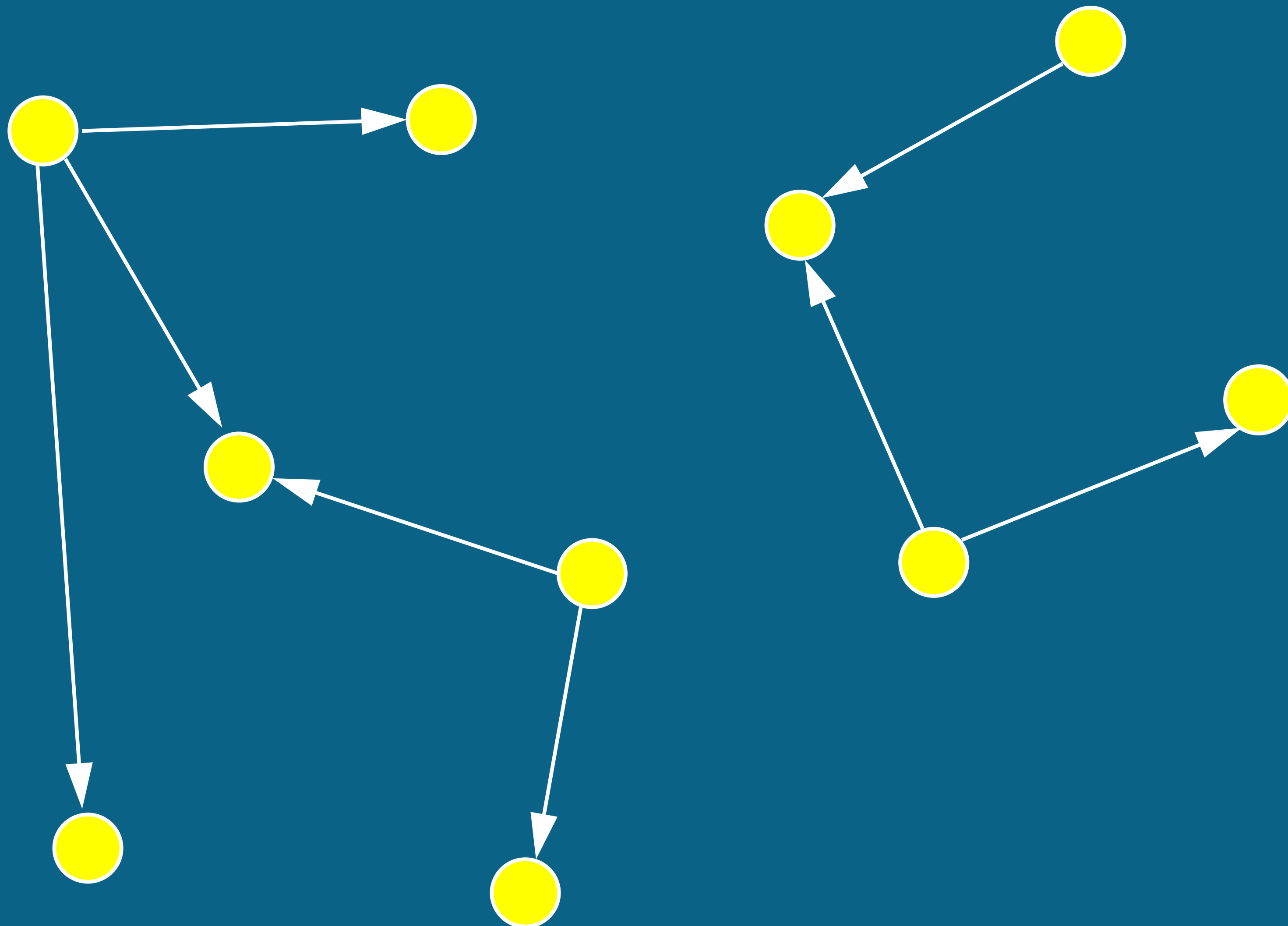
Symmetry breaking by the direction of time:
Cause precedes its effect

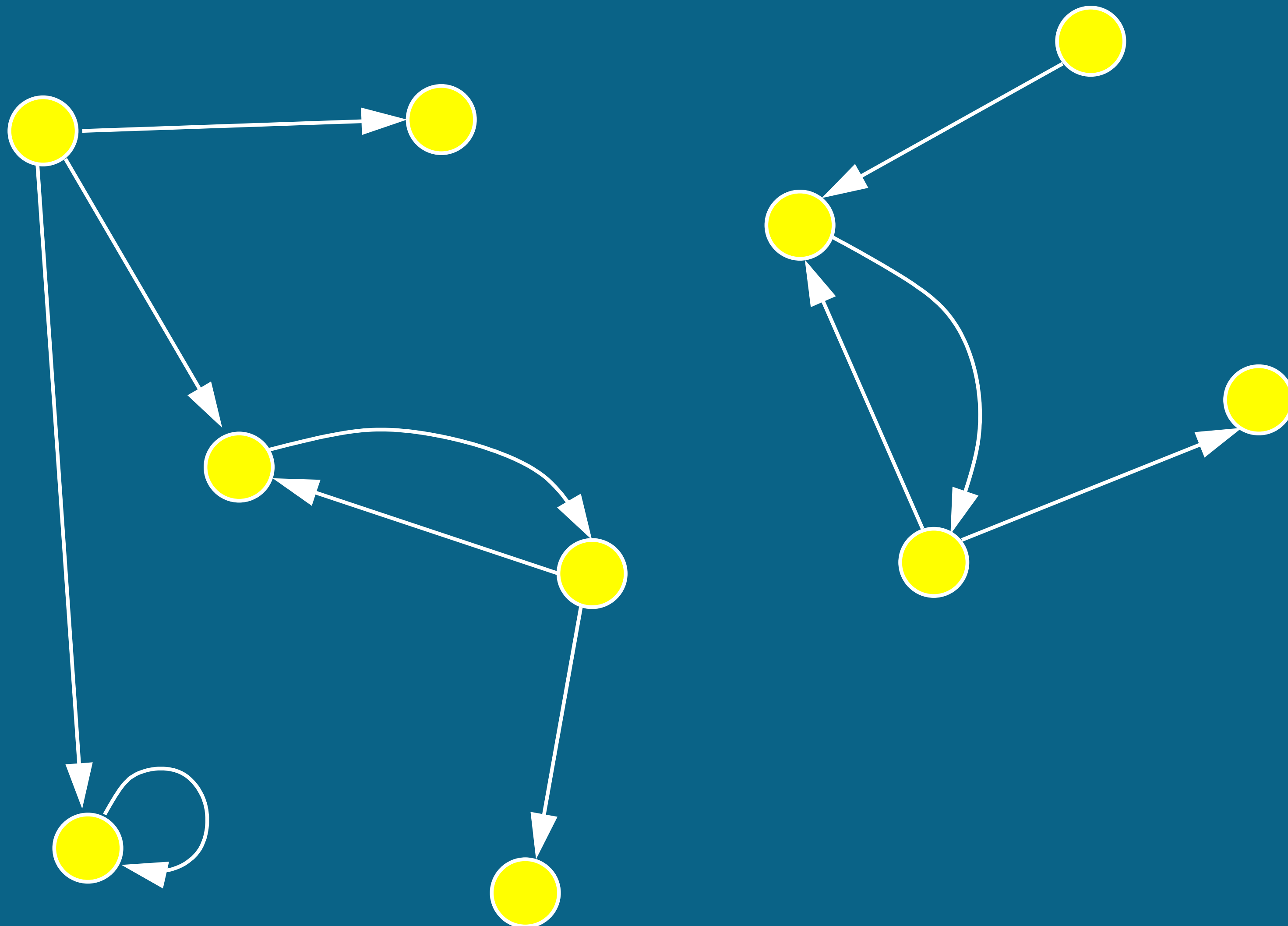
Modelling recurrent structures and
feedback loops

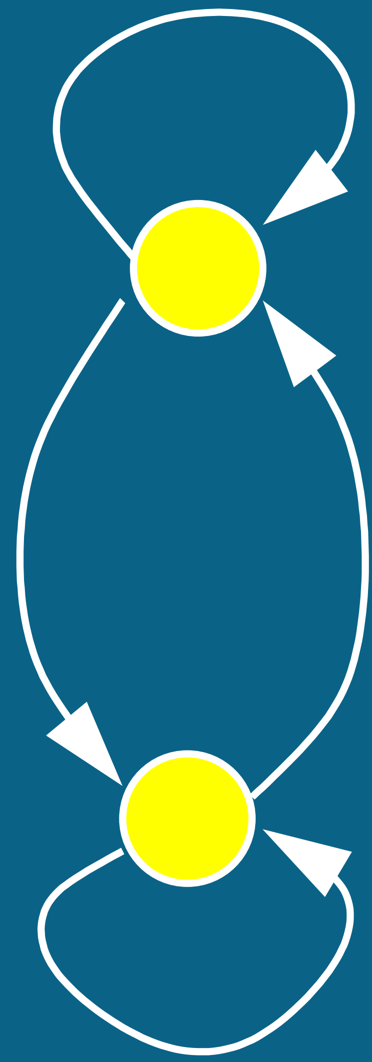


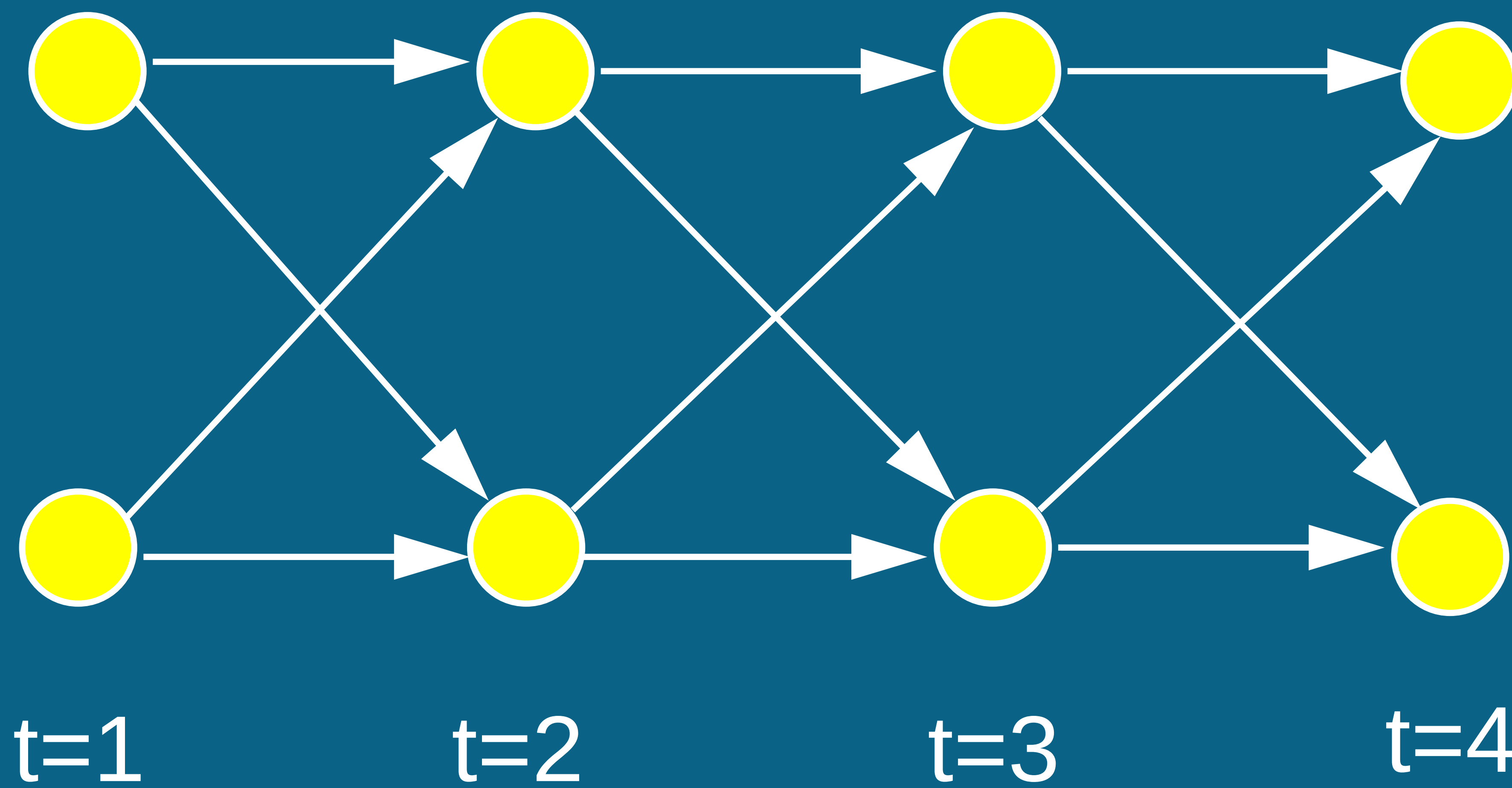











t=1
t=2
t=3
t=4

Outline of the talk

- Recapitulation: Bayesian networks
- **Reverse engineering:
  Learning networks from data**
- Application to the yeast cell cycle
- Estimating the accuracy of inference

Learning from data

Learning from data

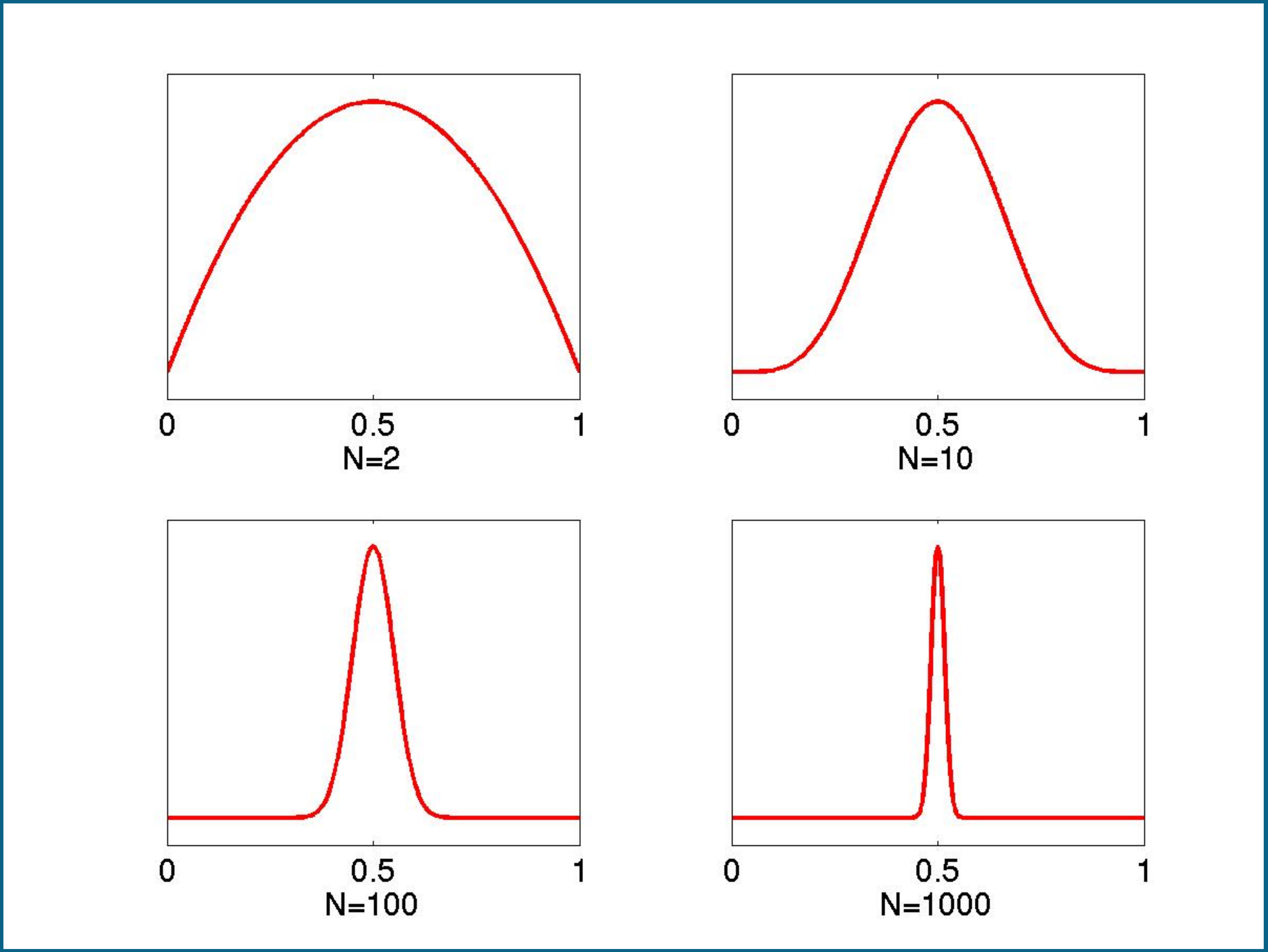
Heads   Tails

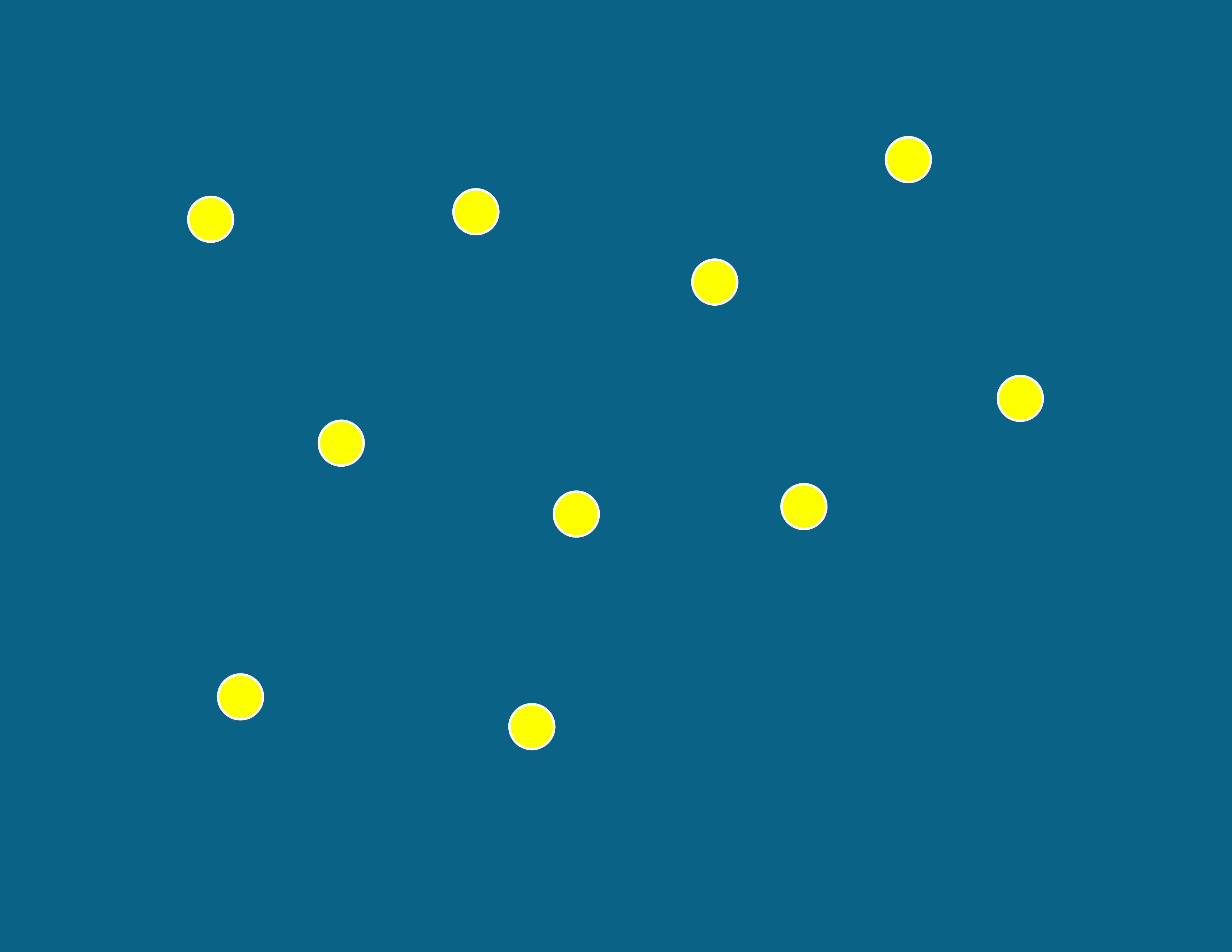
Probability θ $1-\theta$

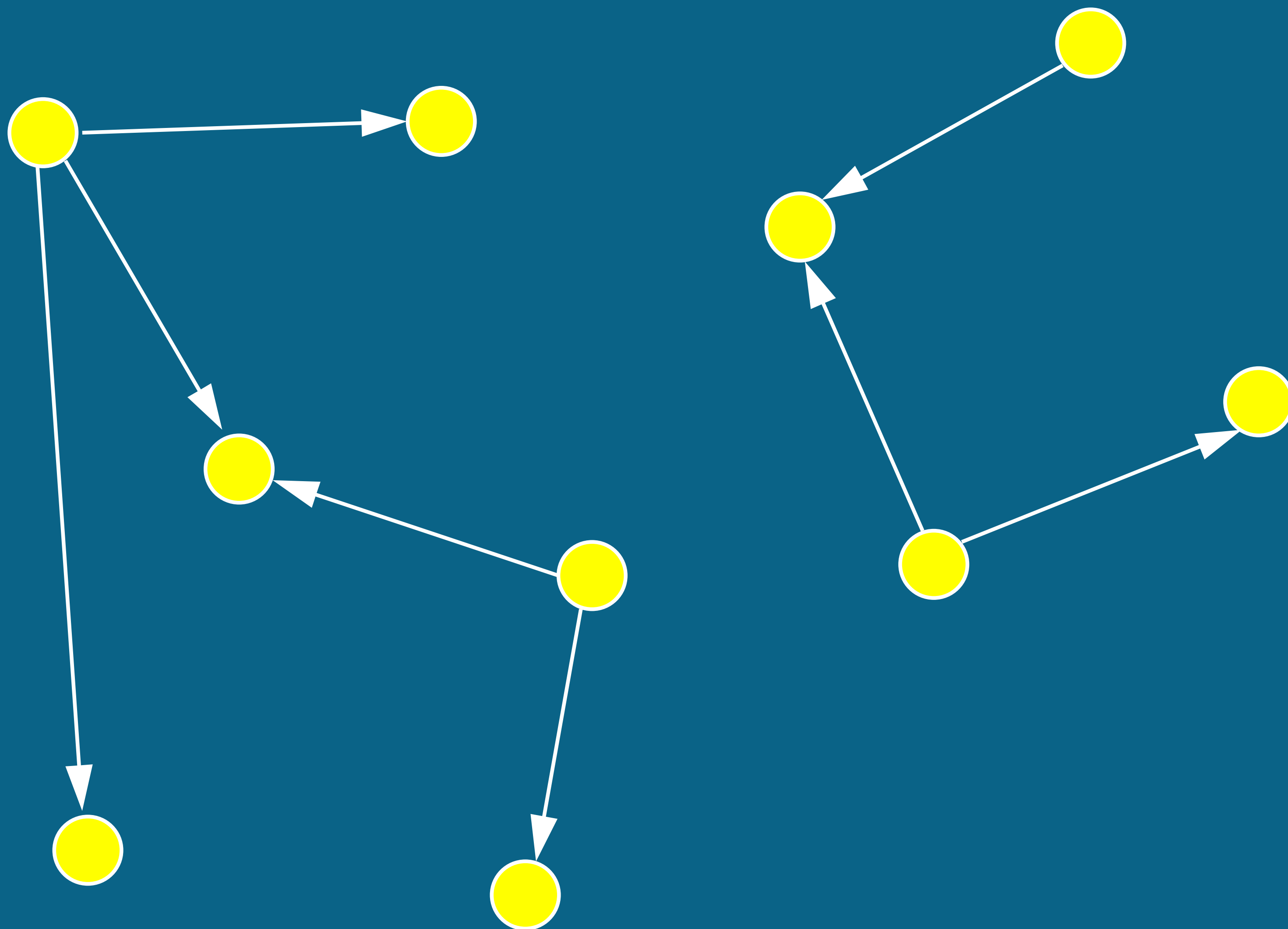
Data

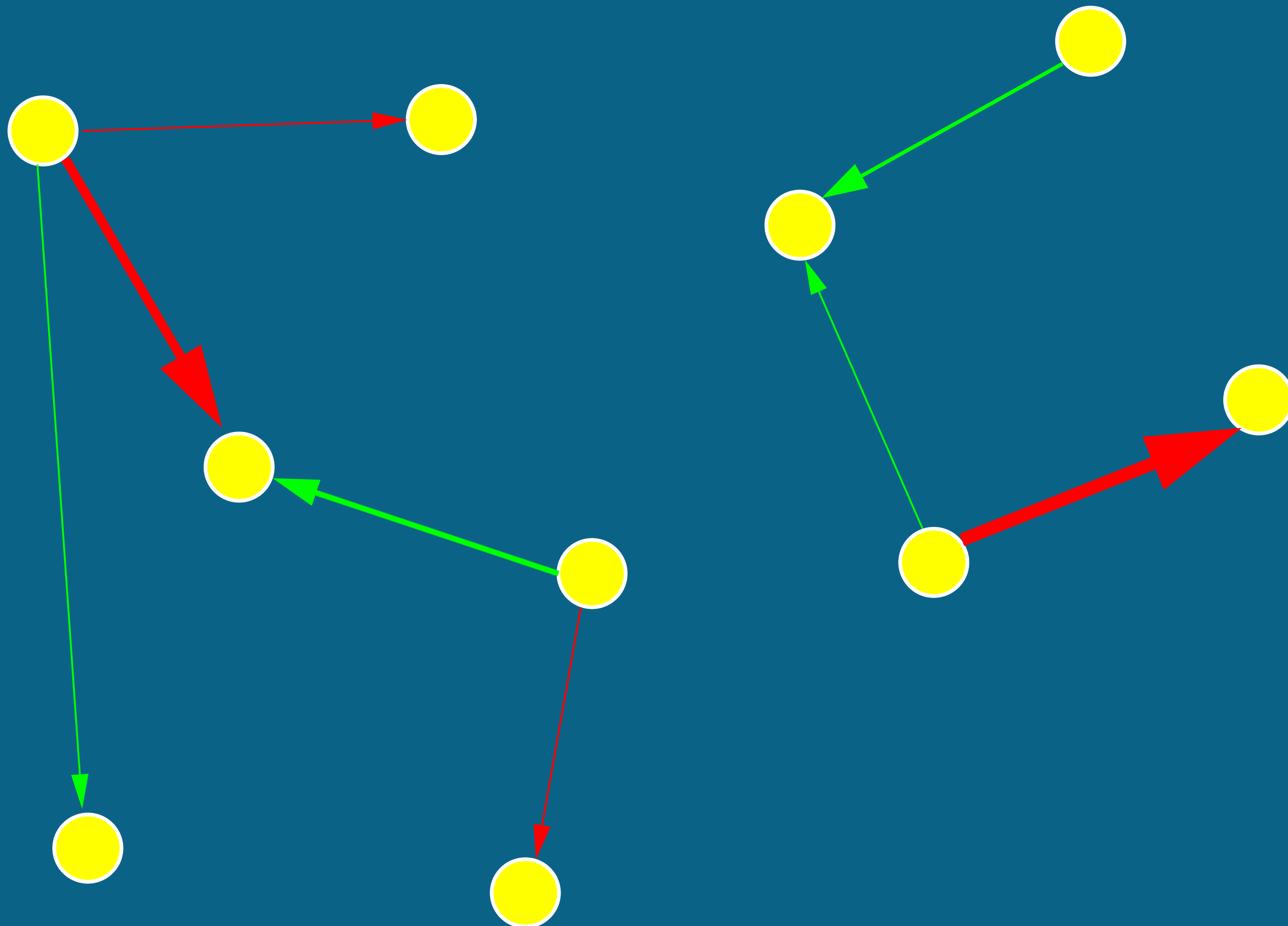
N tosses, k observations of "heads"

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









Classical learning paradigm

Classical learning paradigm

Find the best network structure M:

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

Classical learning paradigm

Find the best network structure M:

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

Find the best parameters θ^*

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

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 Dirchlet 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}

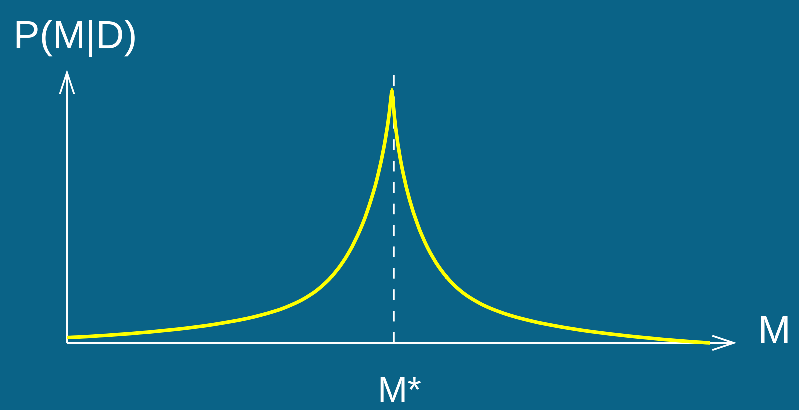
$\longrightarrow$ Optimization problem intractable for large N of nodes

Naive approach

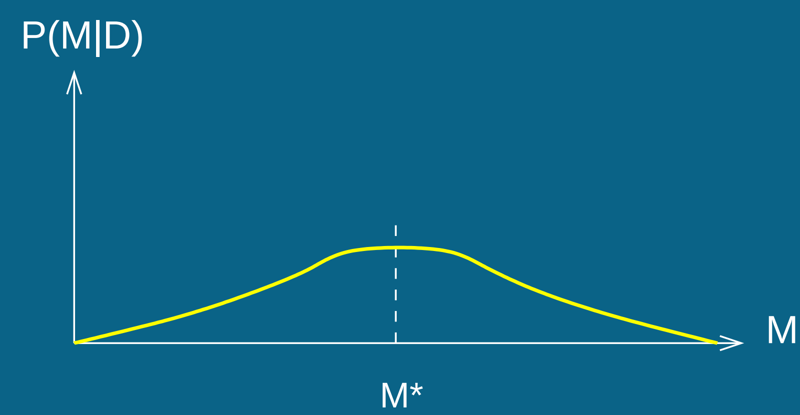
- Compute $P(M|D)$ for all possible network structures M .
- Select network structure M^* that maximizes $P(M|D)$

Problem 2:

Data are sparse $\longrightarrow$ Intrinsic uncertainty of inference



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



Small data set D:
Intrinsic uncertainty about M*

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

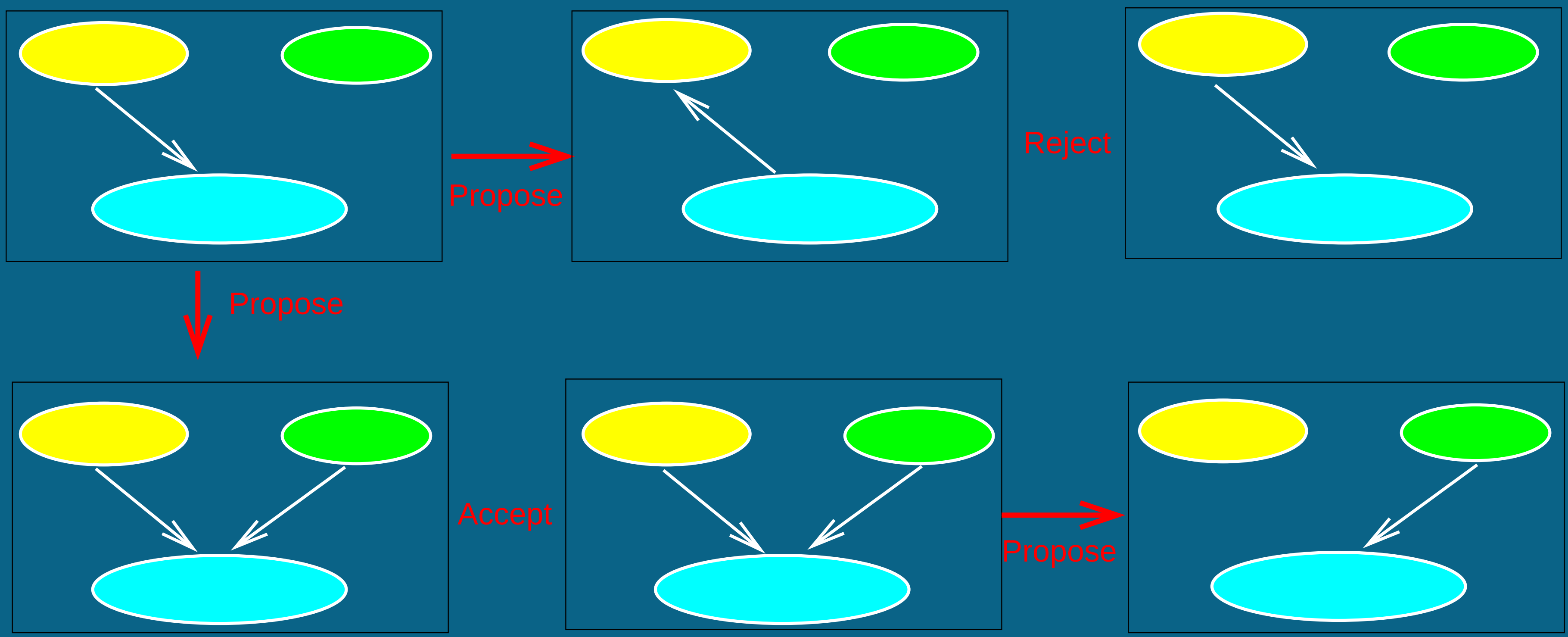
$$P(M_k|D) = \frac{P(D|M_k)P(M_k)}{\sum_i P(D|M_i)P(M_i)}$$

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

Markov chain Monte Carlo (MCMC):

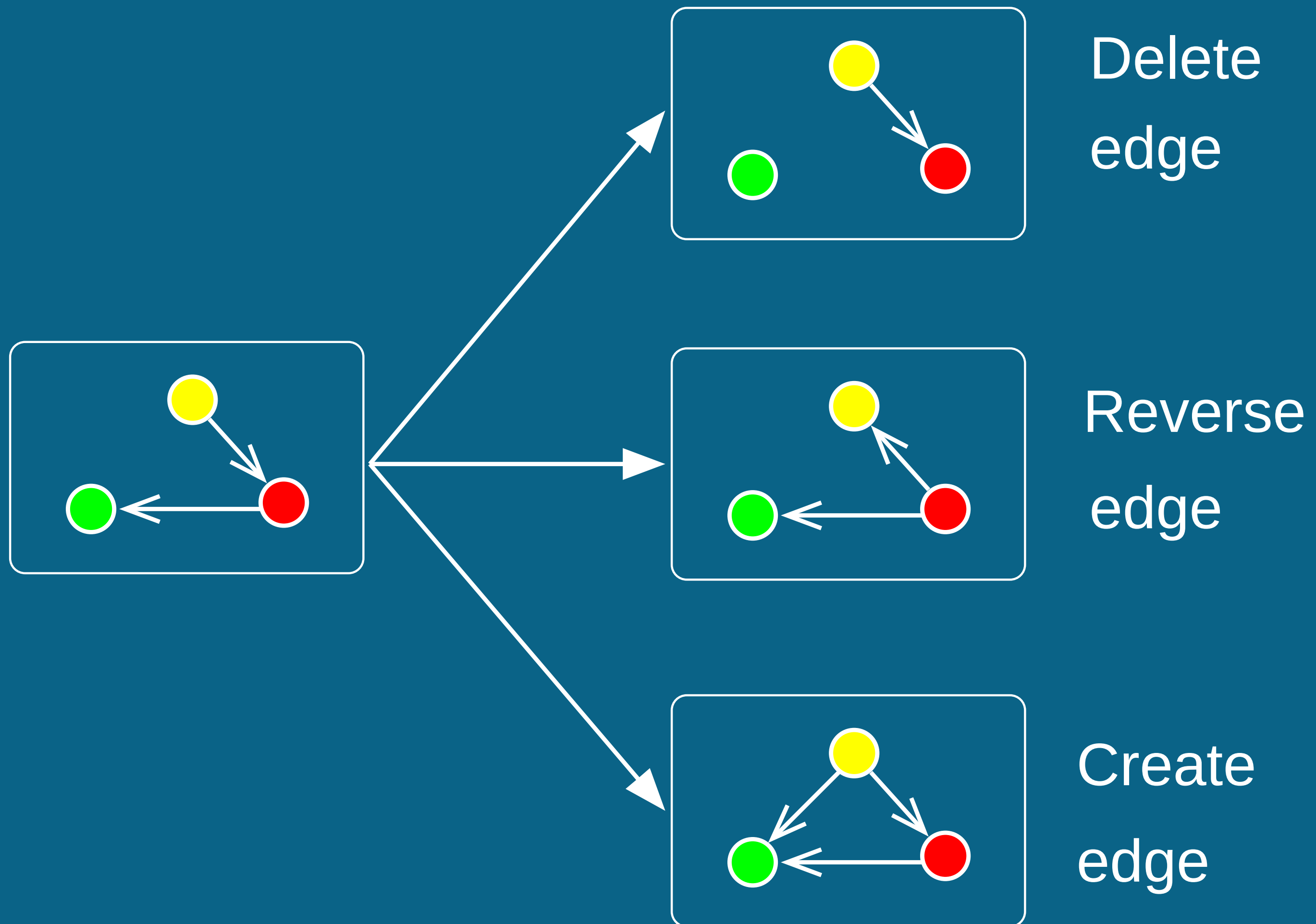
- Proposal move: Given network M_{old}, propose a new network M_{new} with probability $Q(M_{new}|M_{old})$.
- Acceptance/Rejection: Accept this new network with probability $\min\left\{1, \frac{P(M_{new}|D)}{P(M_{old}|D)} \times \frac{Q(M_{old}|M_{new})}{Q(M_{new}|M_{old})}\right\}$

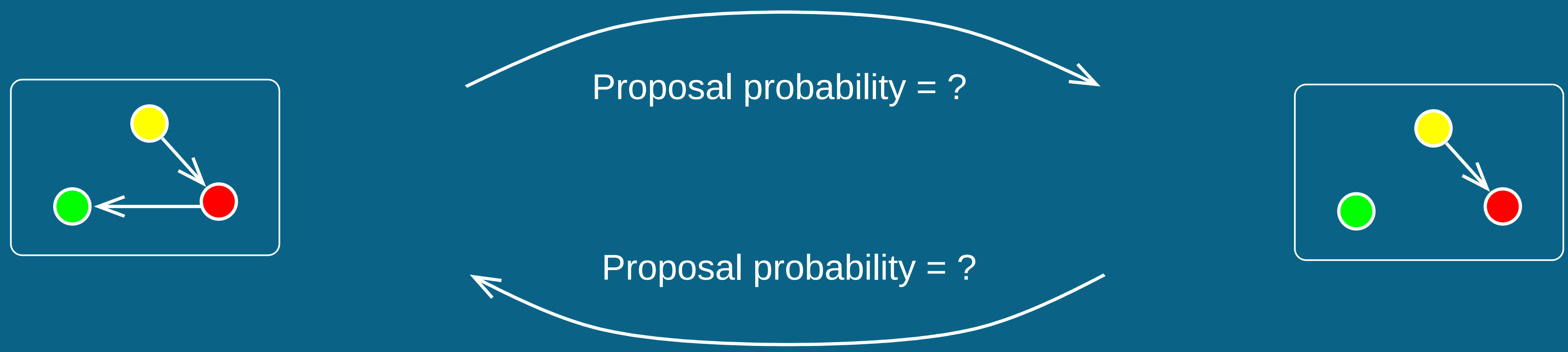
Markov chain Monte Carlo (MCMC)

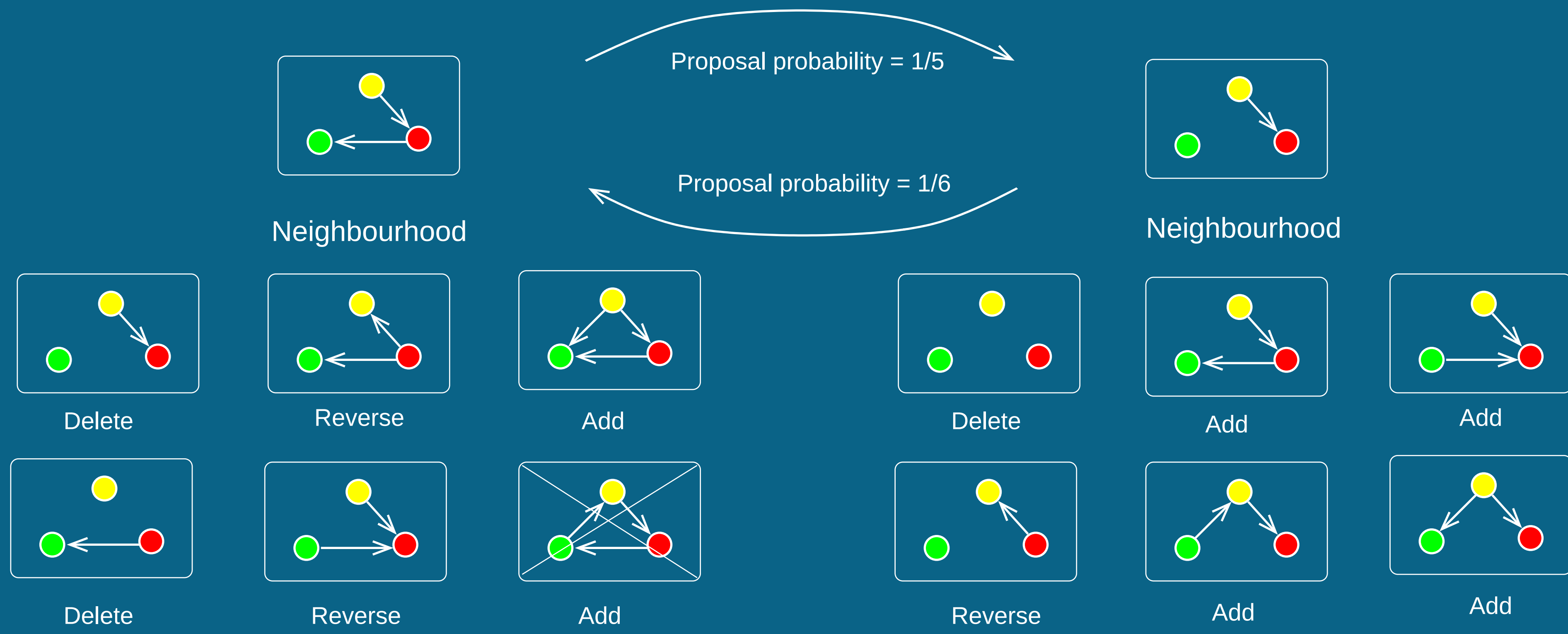


Accept move with probability: $\min\left\{1, \frac{P(D|M_{new})P(M_{new})}{P(D|M_{old})P(M_{old})} \times \frac{Q(M_{old}|M_{new})}{Q(M_{new}|M_{old})}\right\}$

MCMC moves




Proposal probability = ?
Proposal probability = ?


Proposal probability = 1/5
Proposal probability = 1/6
Neighbourhood
Neighbourhood
Delete
Reverse
Add
Delete
Add
Add
Delete
Reverse
Add
Reverse
Add
Add

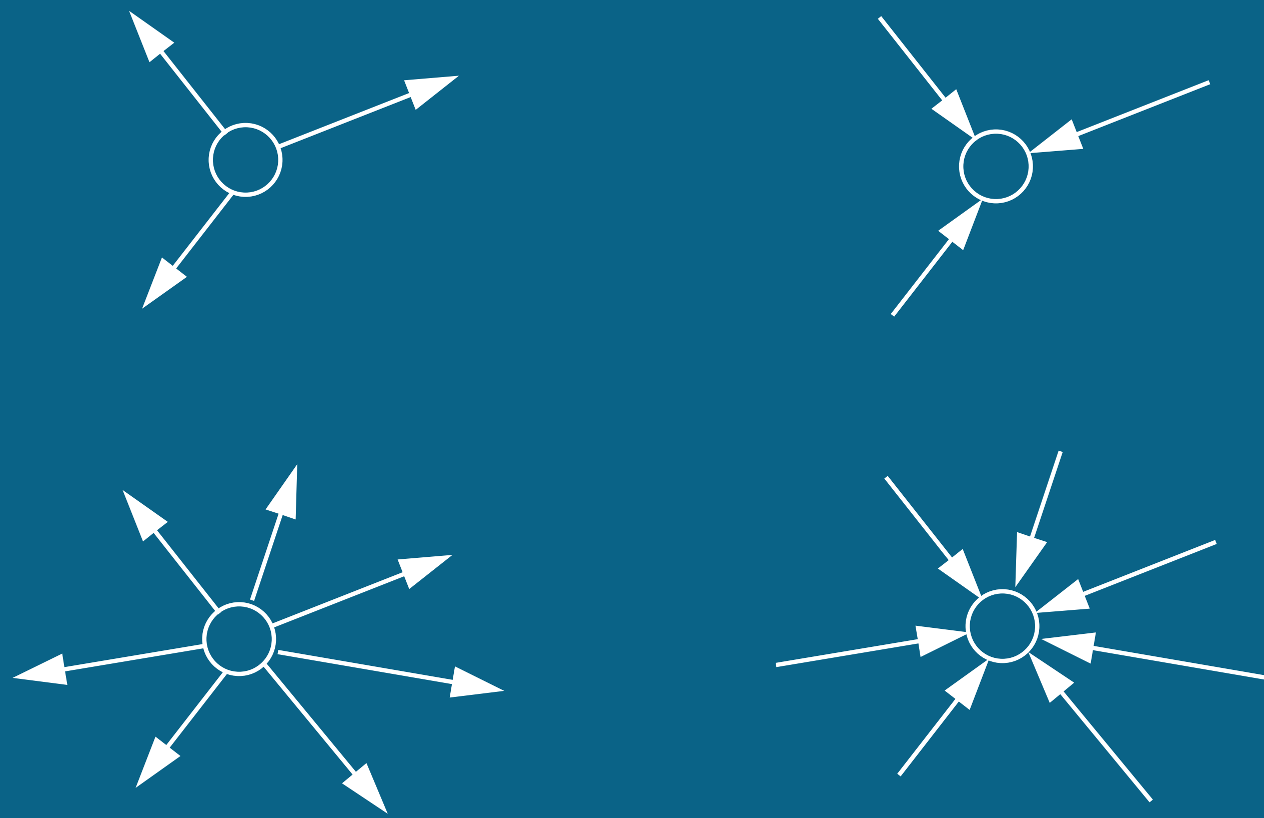
Bayes rule:

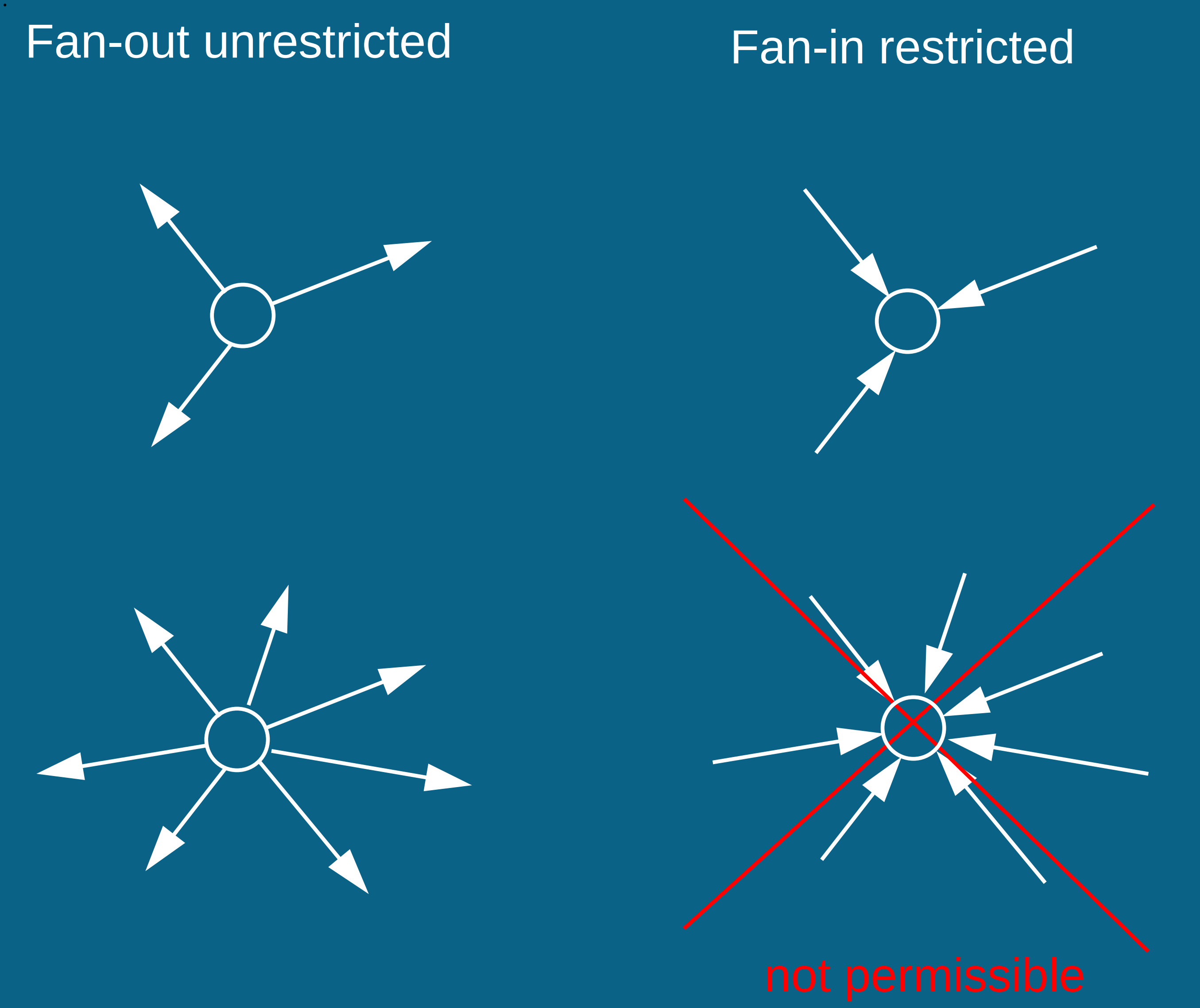
$$P(\mathcal{M}|D) \propto P(D|\mathcal{M})P(\mathcal{M})$$

Bayes rule:

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

$$P(D|M) = \prod_i P(X_i|Pa(X_i))$$




Fan-out unrestricted
Fan-in restricted
not permissible

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

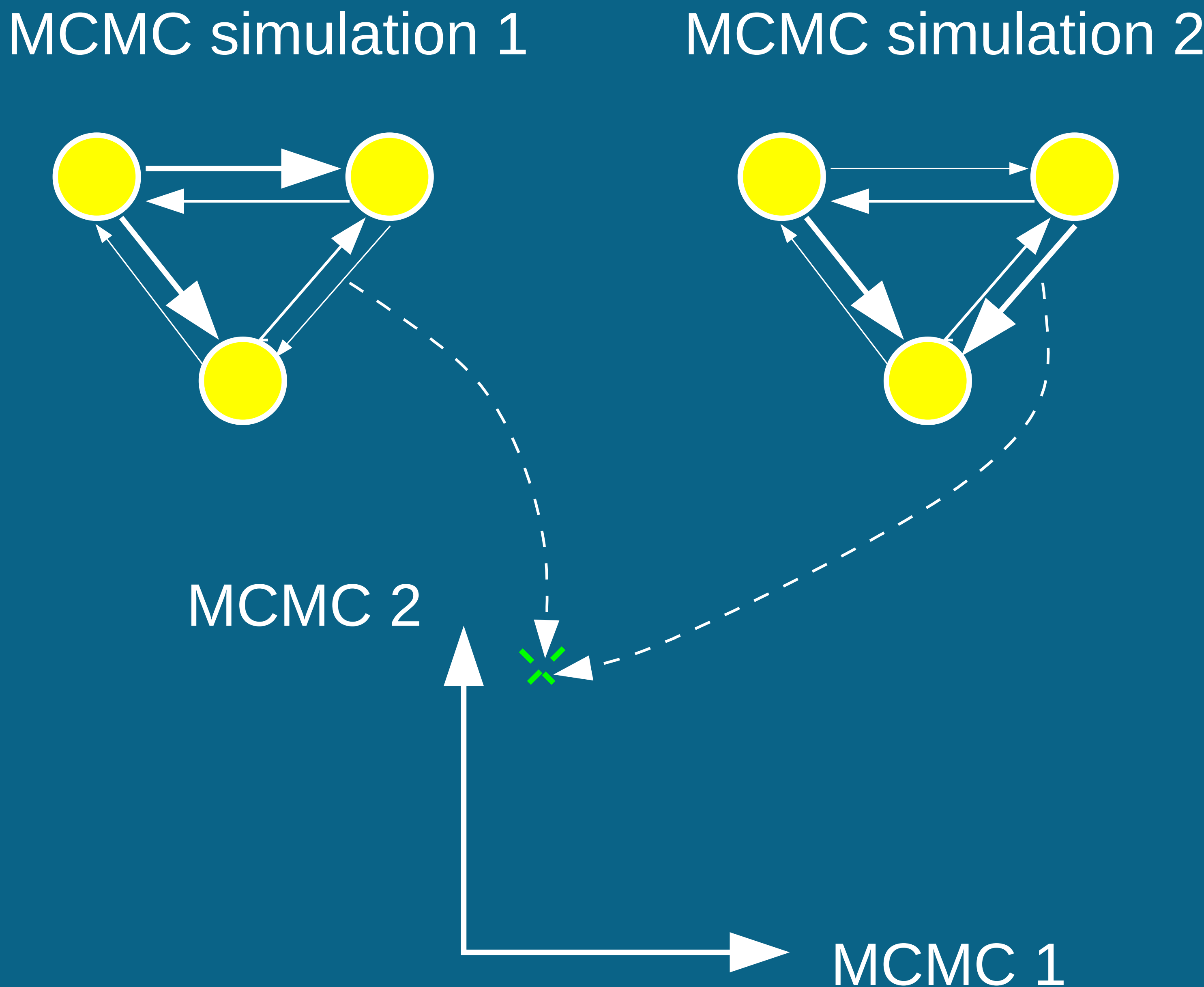
Burn-in: T MCMC steps

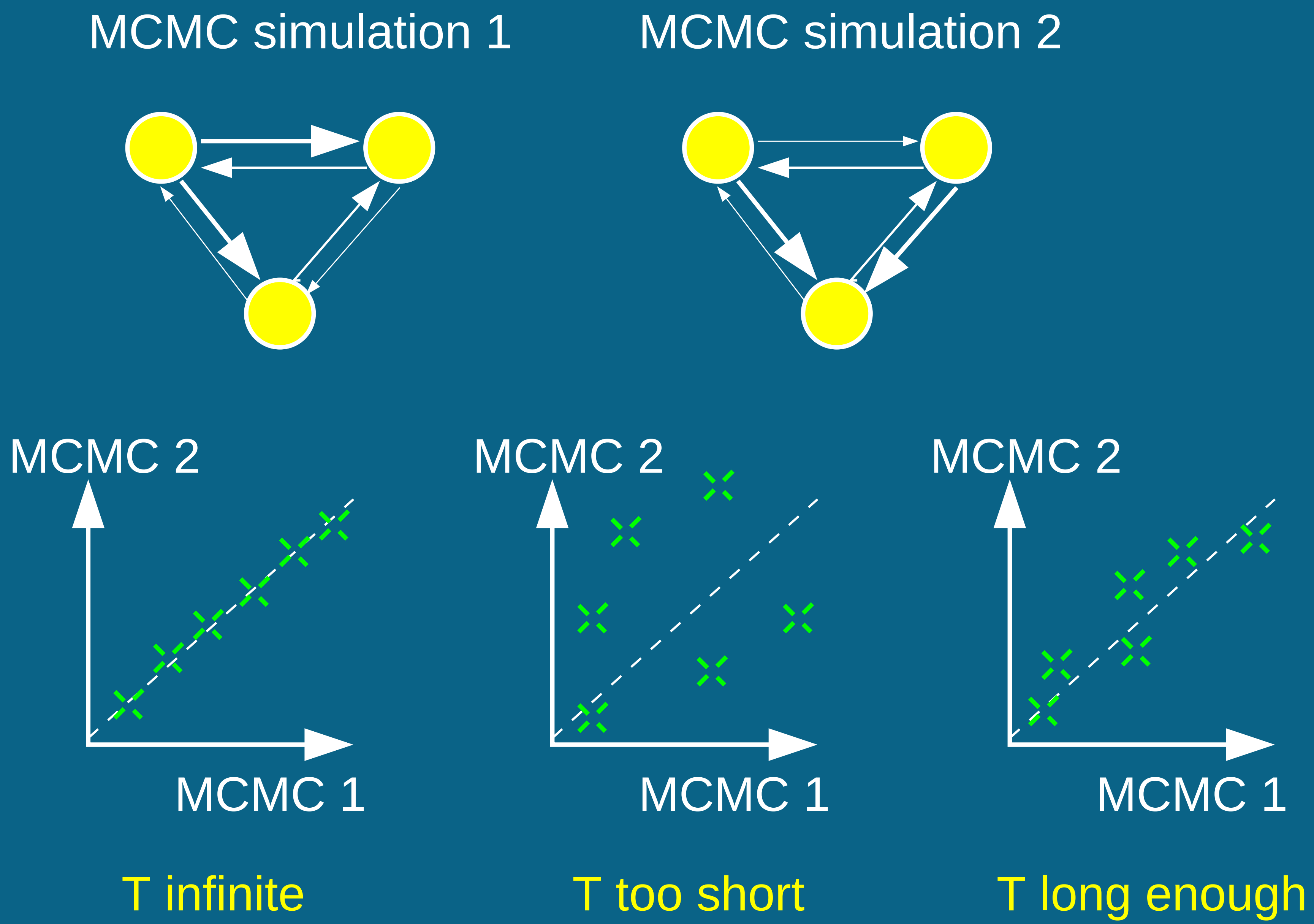
Sampling: T MCMC steps

How large do we have to choose T?

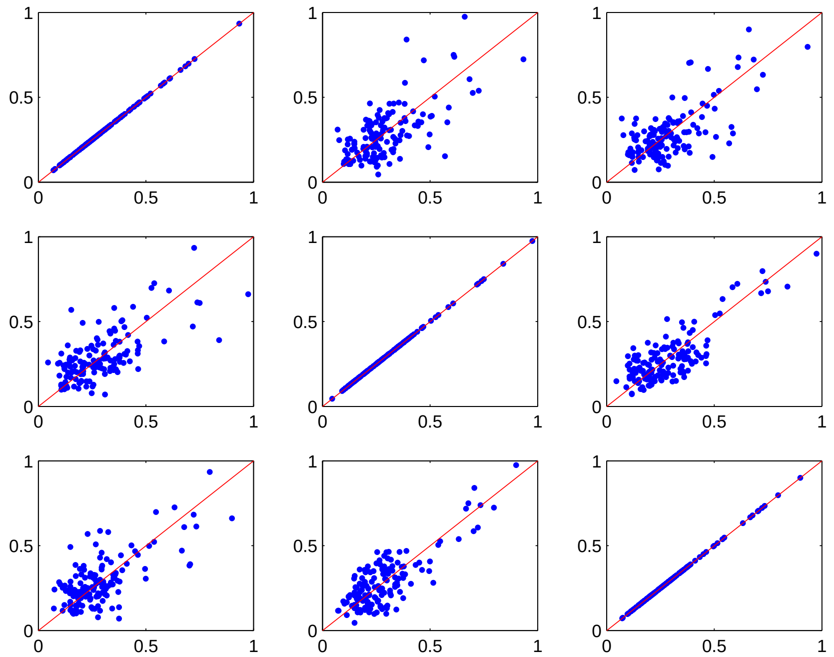
Repeat MCMC simulations from different initializations

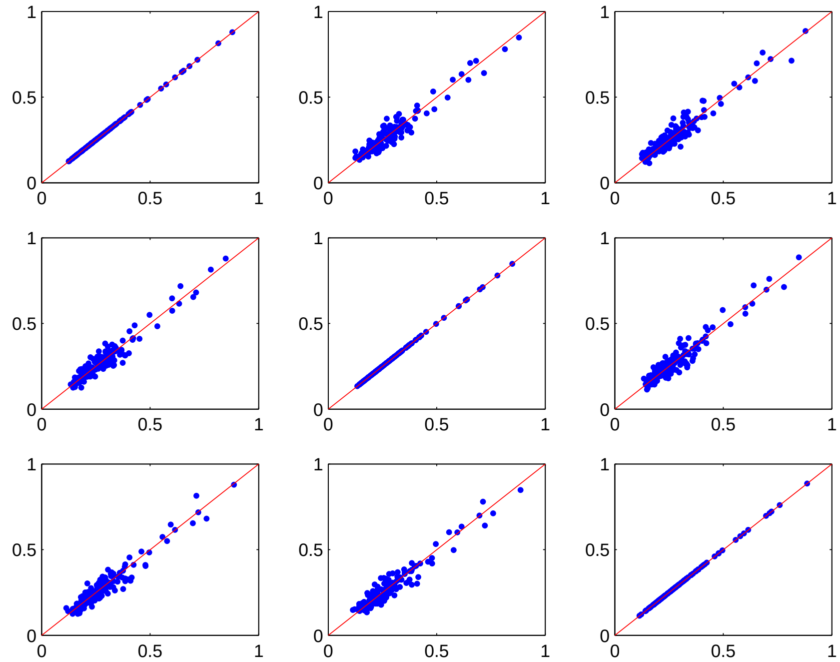
Scatter plots of posterior probabilities of edges


MCMC simulation 1
MCMC simulation 2
MCMC 2
MCMC 1


MCMC simulation 1
MCMC simulation 2
MCMC 2
MCMC 1
T infinite
MCMC 2
MCMC 1
T too short
MCMC 2
MCMC 1
T long enough

MCMC steps: 10,000 (left) versus 100,000 (right)

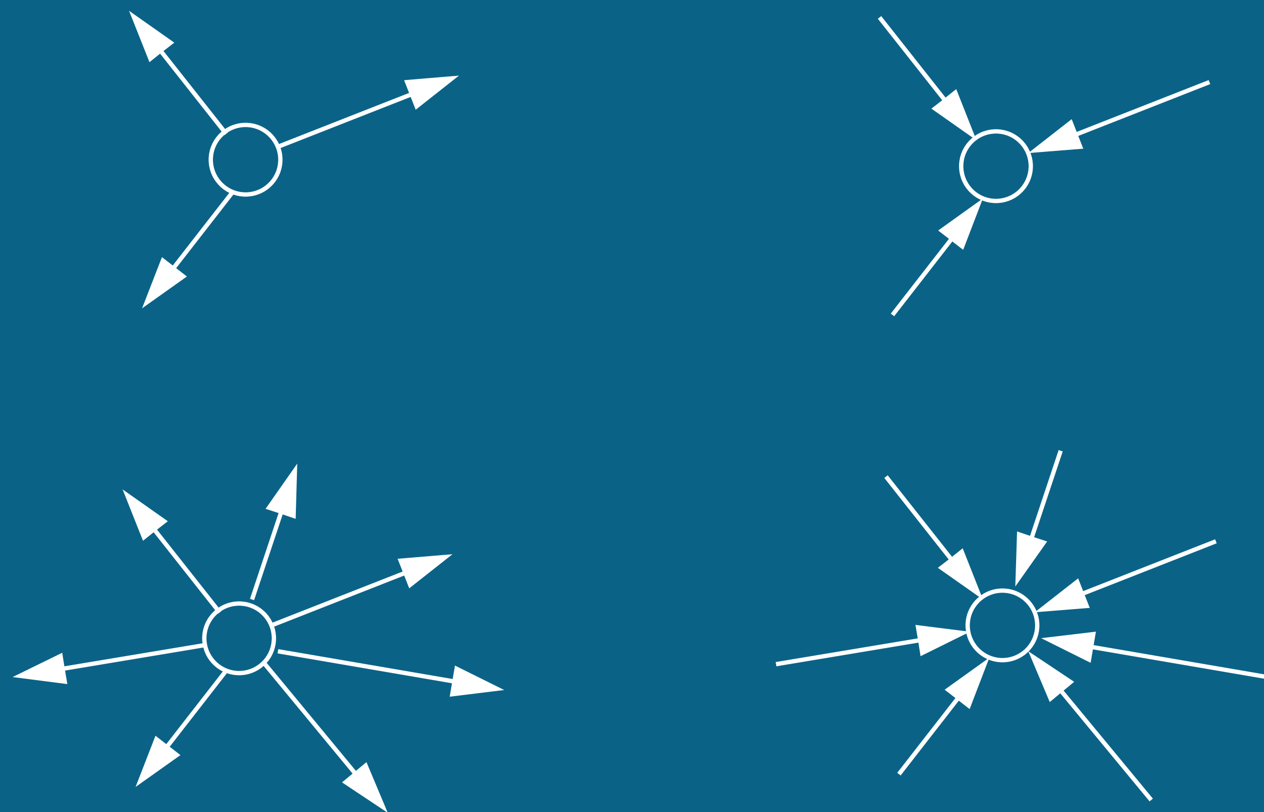


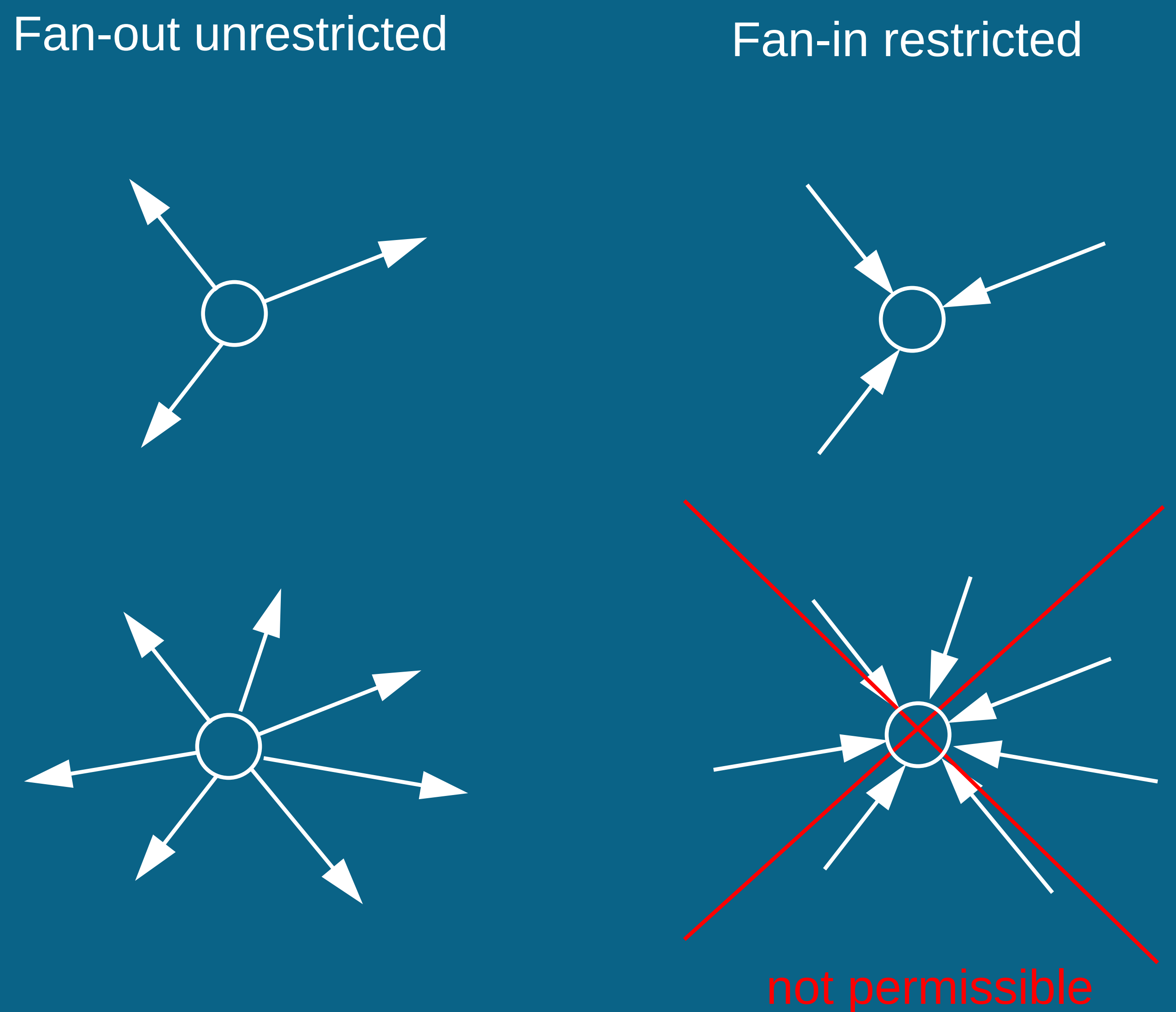


Bayes rule:

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

The structure prior $P(\mathcal{M})$




Fan-out unrestricted
Fan-in restricted
not permissible

$P(\mathcal{M})$ uniform over structures

or

$P(\mathcal{M})$ uniform over cardinalities of parent sets

$P(\mathcal{M})$ uniform over structures

or

$P(\mathcal{M})$ uniform over cardinalities of parent sets

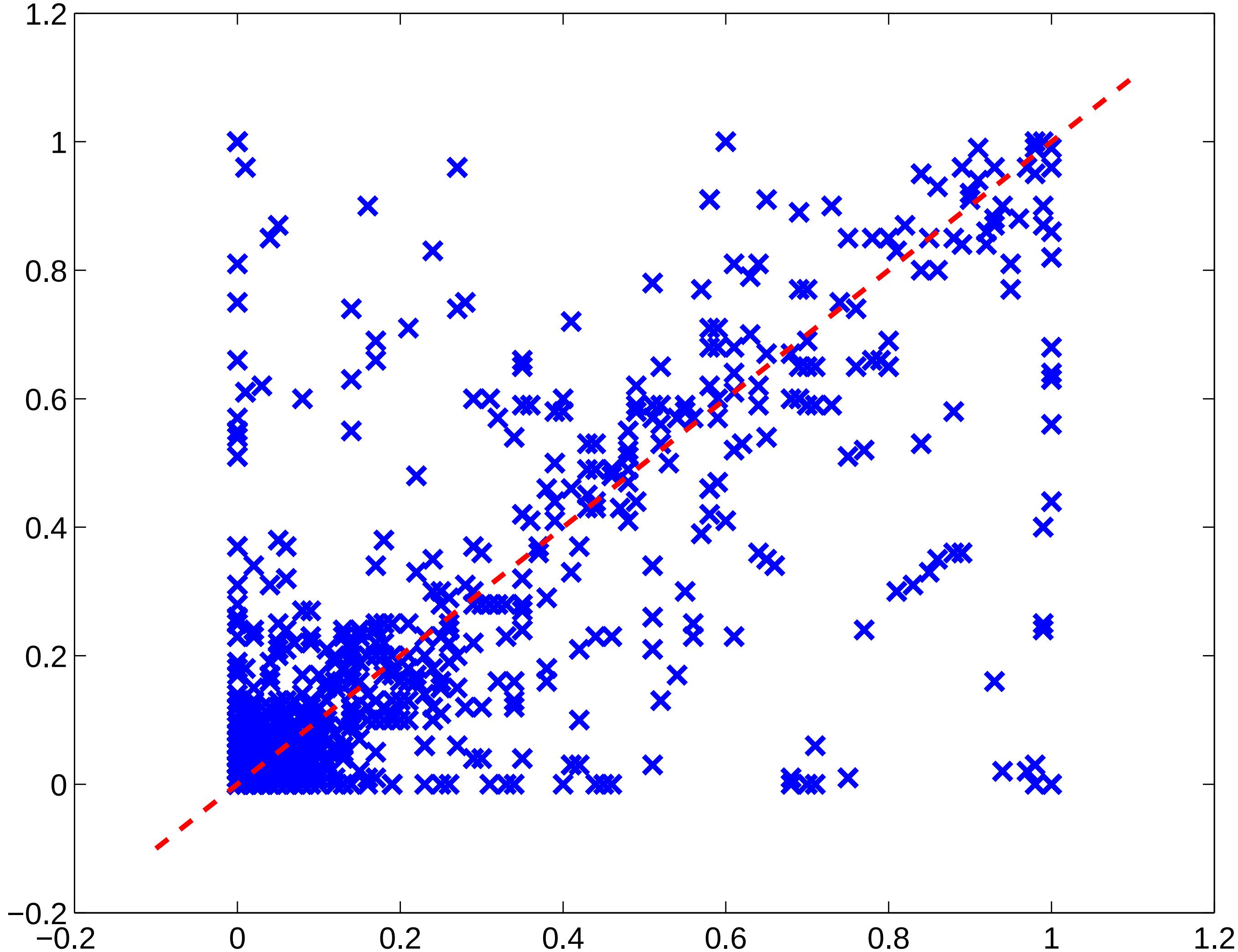
$$\propto \binom{\text{Number of nodes}}{\text{Cardinality}}^{-1}$$

Study with Marco Grzegorczyk and Wolfgang Urfer (Dortmund University)

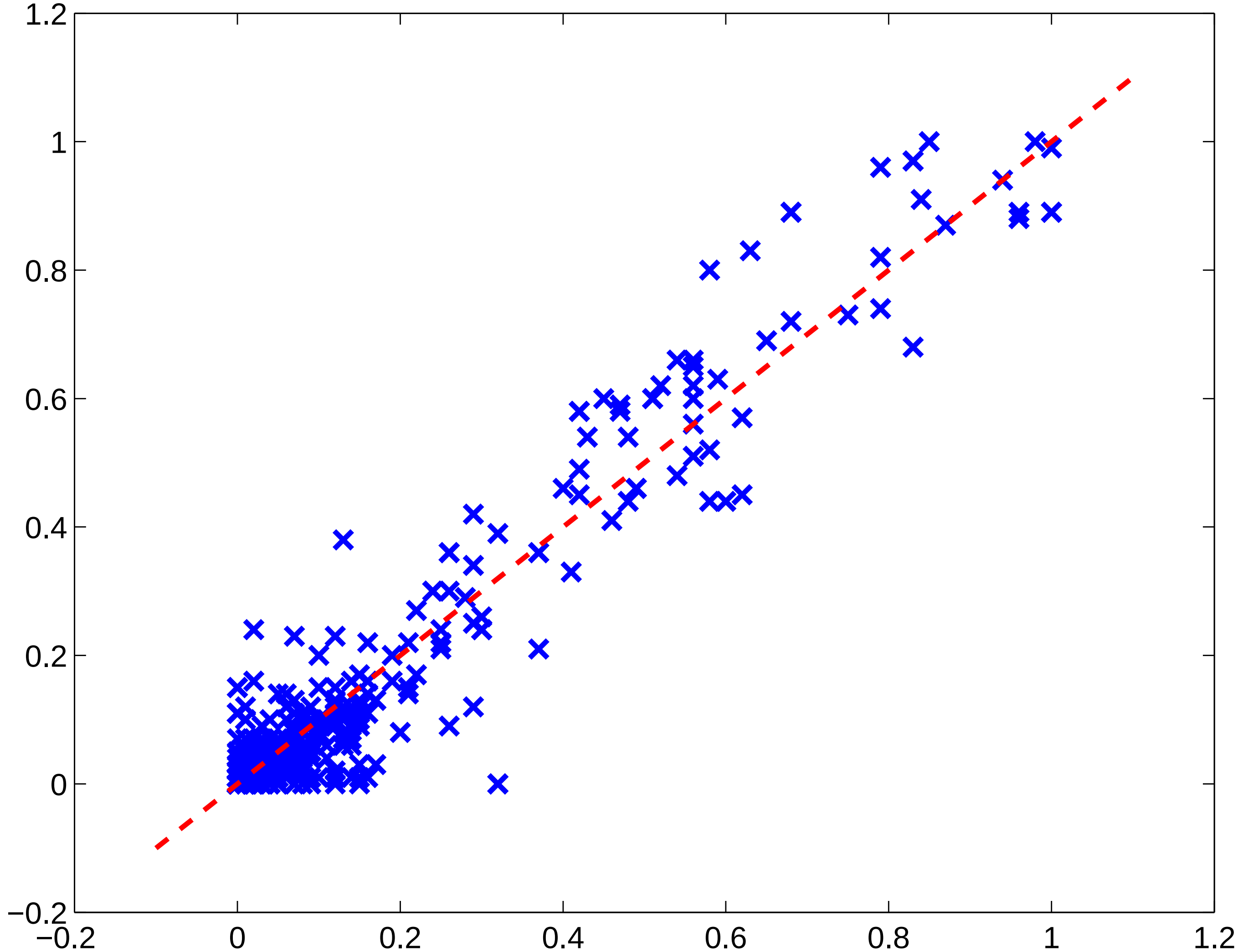
Carcinogenesis in kidney

- 100 genes
- 60 kidney cancer patients
- MCMC: 100,000 Metropolis-Hastings steps

$P(\mathcal{M})$ uniform over structures

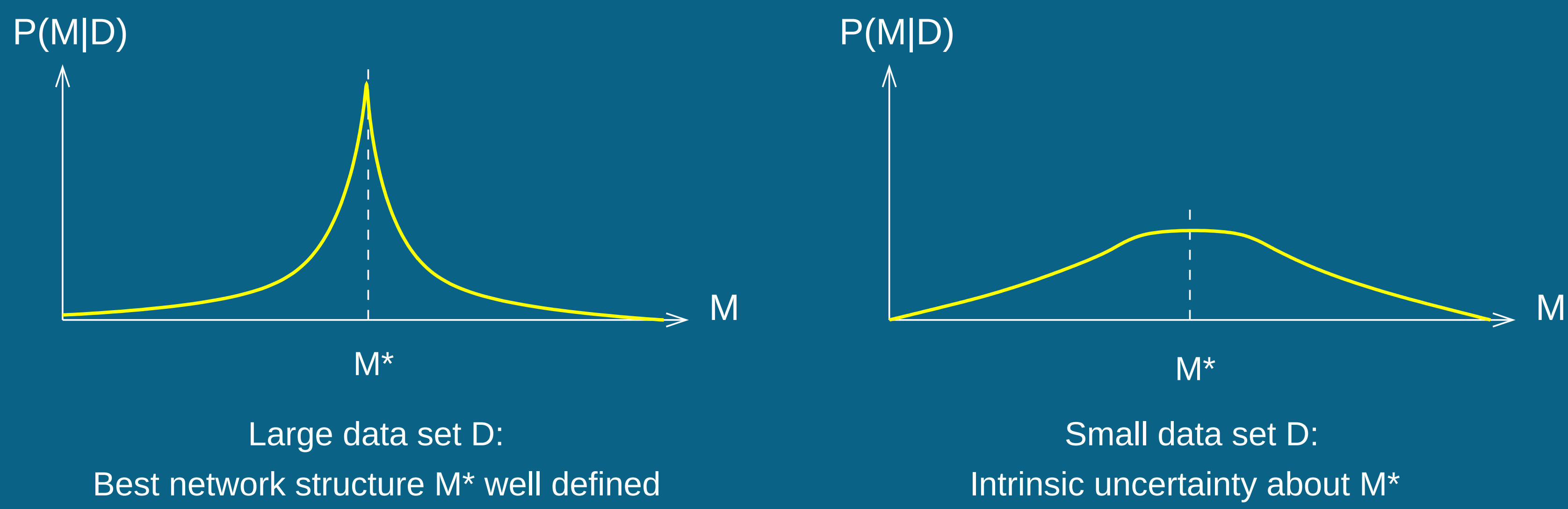


$P(\mathcal{M})$ uniform over cardinalities



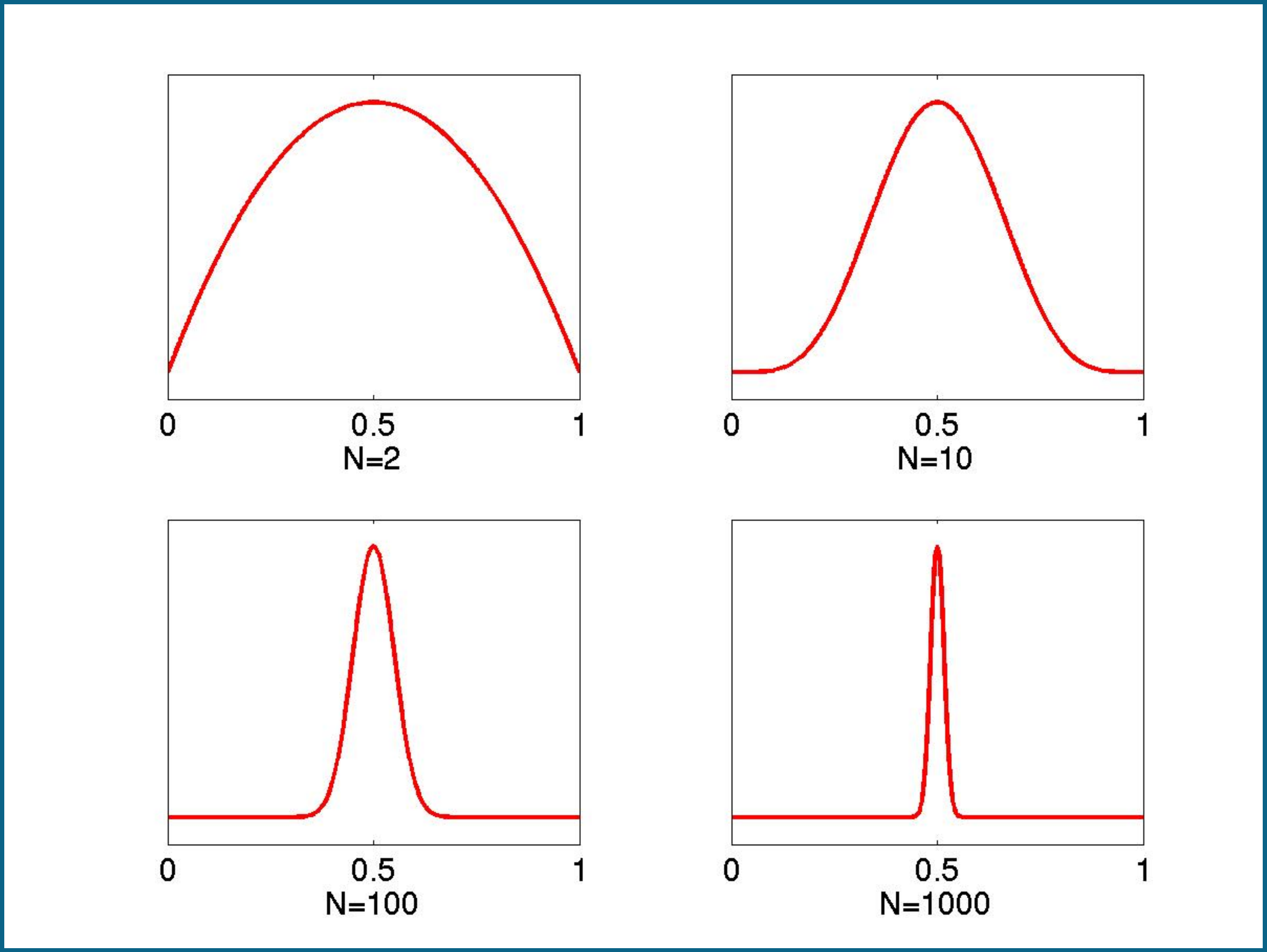
Problem: Statistical significance of the networks

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



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

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



Solution: Focus on features and subnetworks

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?

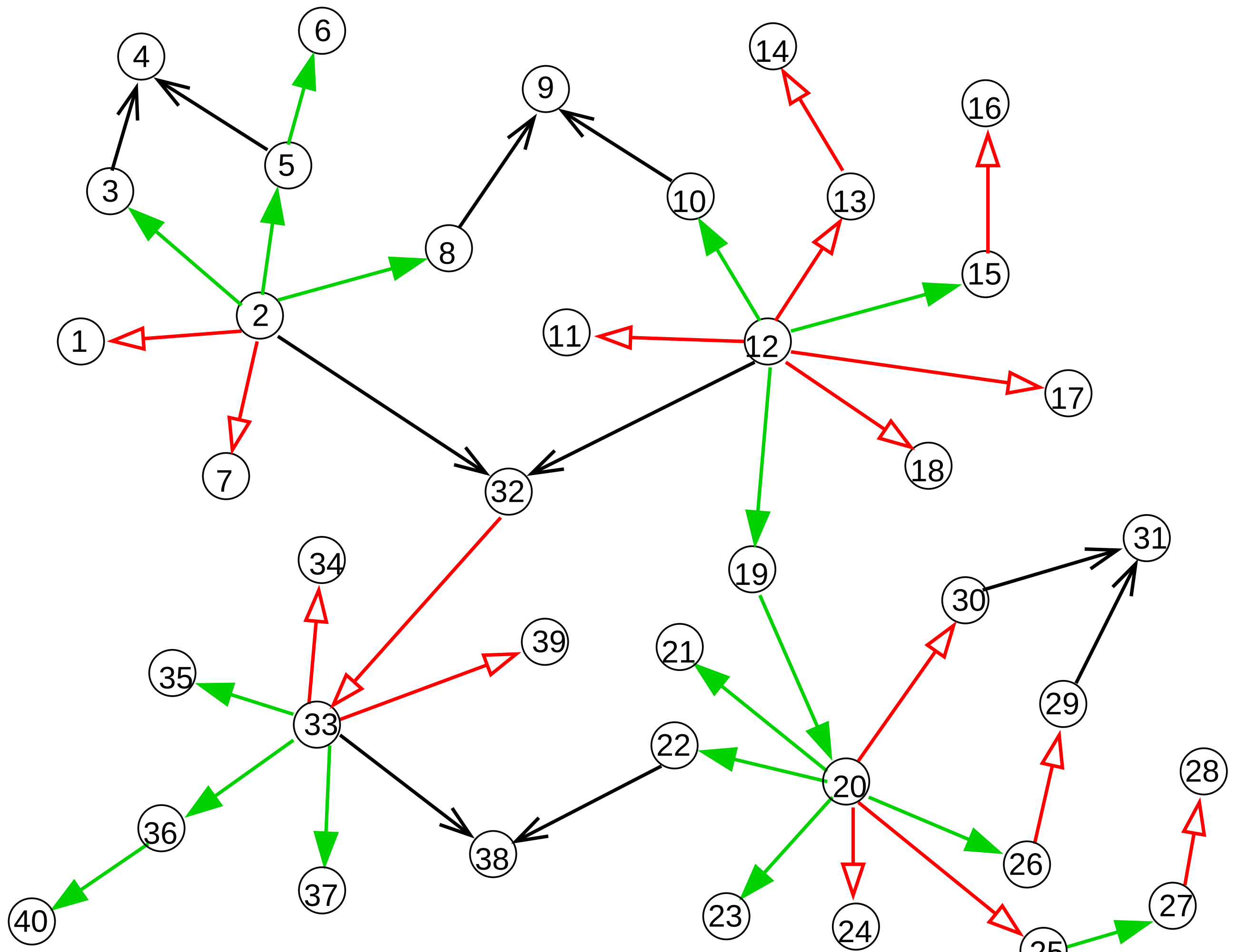
$$f(M) = \begin{cases} 1 \text{ if } M \text{ satisfies the feature} \\ 0 \text{ otherwise} \end{cases}$$

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

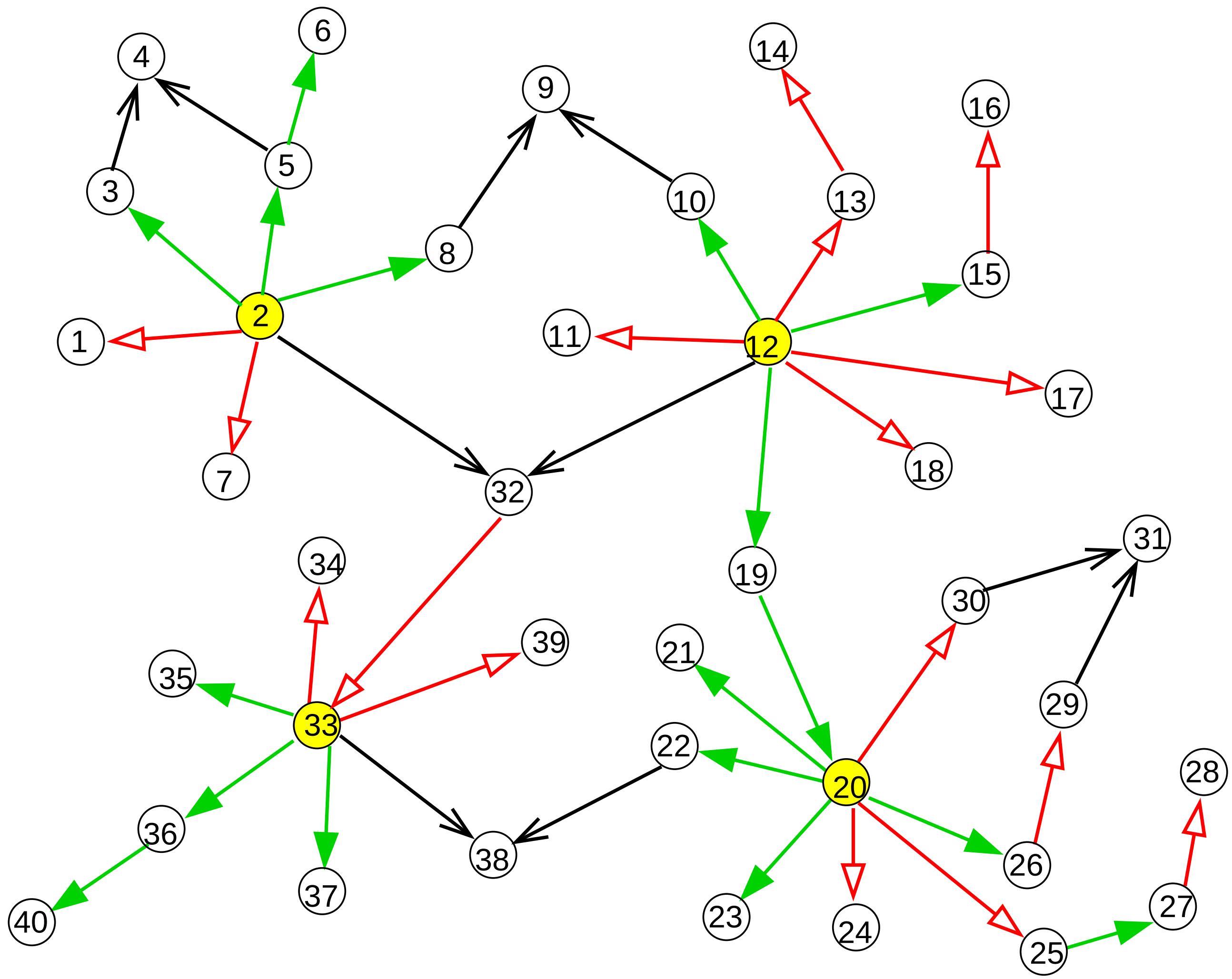
Approximate this sum with MCMC: $P(f|D) = \frac{1}{T}\sum_{i=1}^{T} f(M_i)$

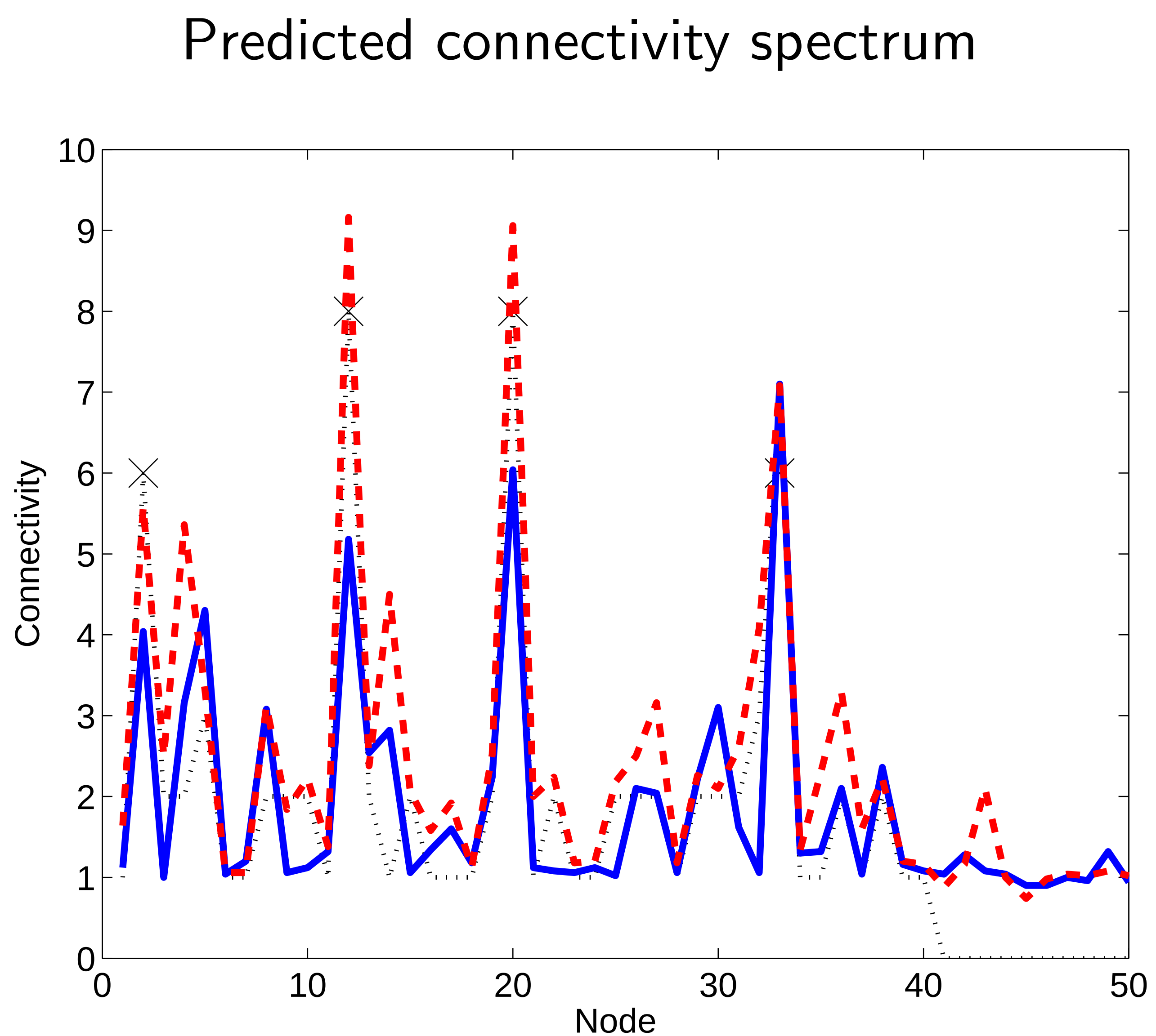
where $\{M_i\}$ is a sample from the posterior obtained with MCMC.

Model network, data set size: $N = 50$



Model network, data set size: $N = 50$




Predicted connectivity spectrum
Connectivity
Node
0
1
2
3
4
5
6
7
8
9
10
0
10
20
30
40
50

Outline of the talk

- Recapitulation: Bayesian networks
- Reverse engineering:
  Learning networks from data
- **Application to the yeast cell cycle**
- Estimating the accuracy of inference

Experimental Results

- Friedman, Linial, Nachman, Pe'er (2000)
  Journal of Computational Biology 7: 601-620

- Pe'er, Regev, Elidan, Friedman (2001)
  Bioinformatics S1: 215-224

- Friedman & Koller (2003)
  Machine Learning 50: 95-126

- 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).

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.

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.

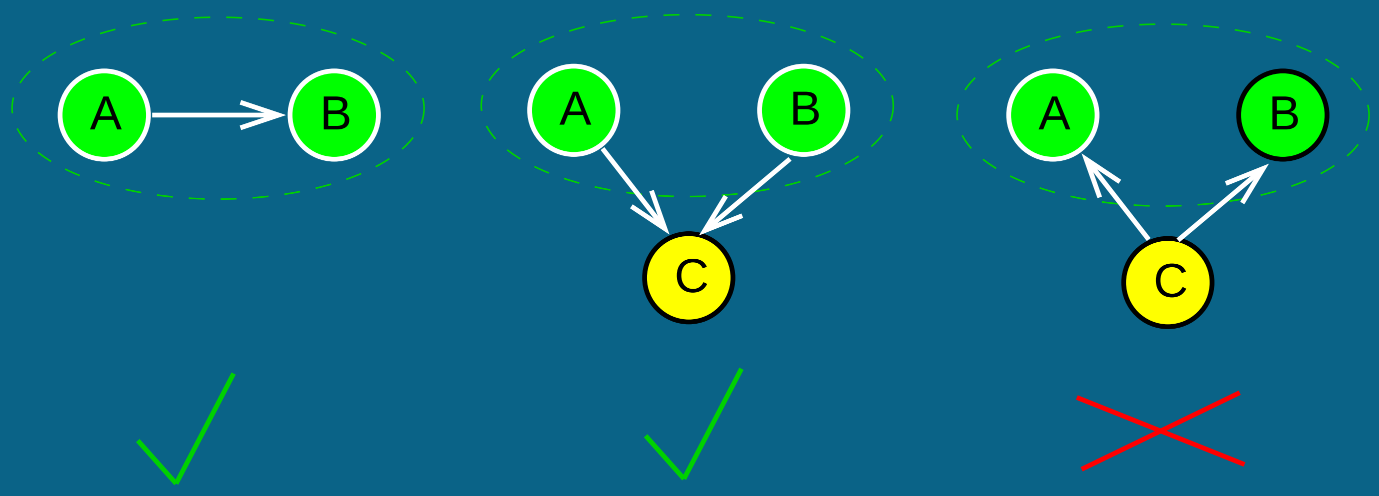
Dominant genes in the ordering relations

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 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.

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

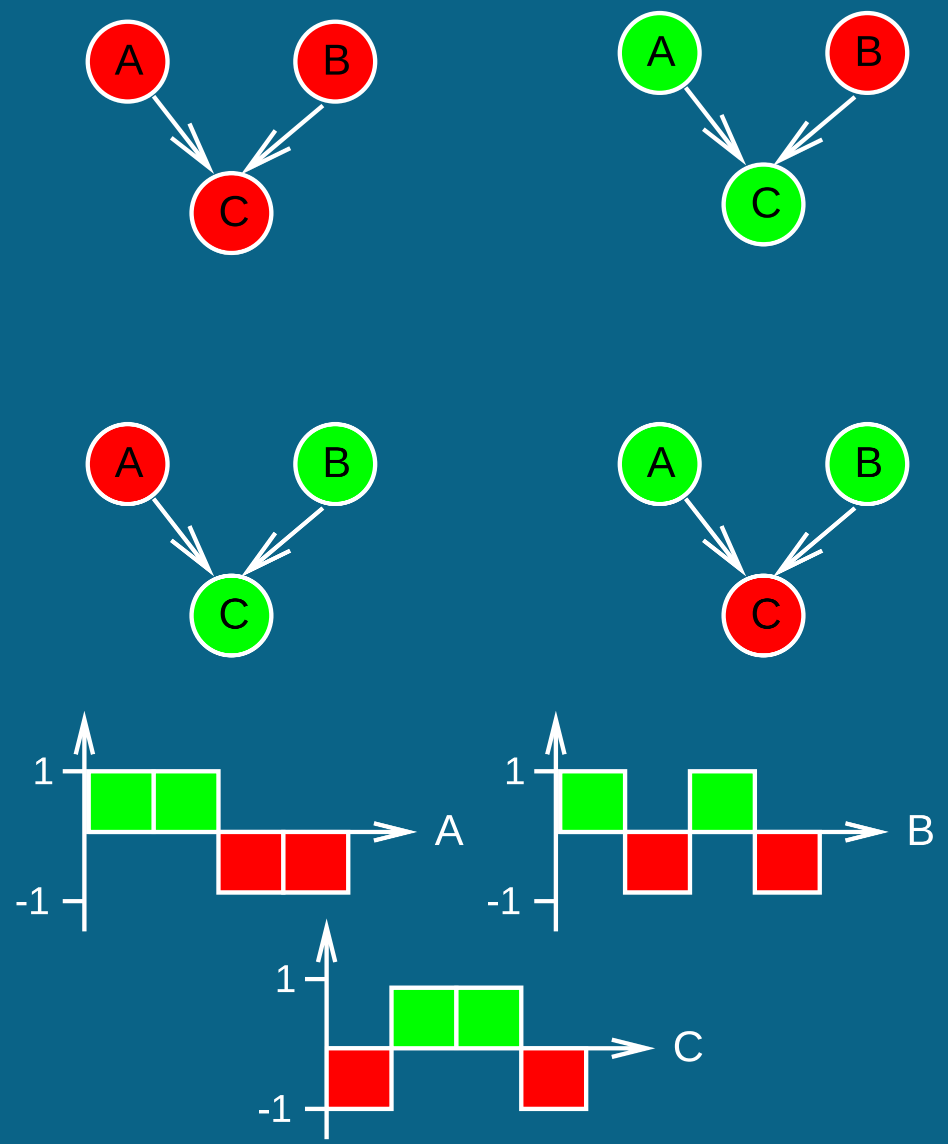
$P(X \leftrightarrow Y|D)$: Indication that genes are functionally related.

- 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

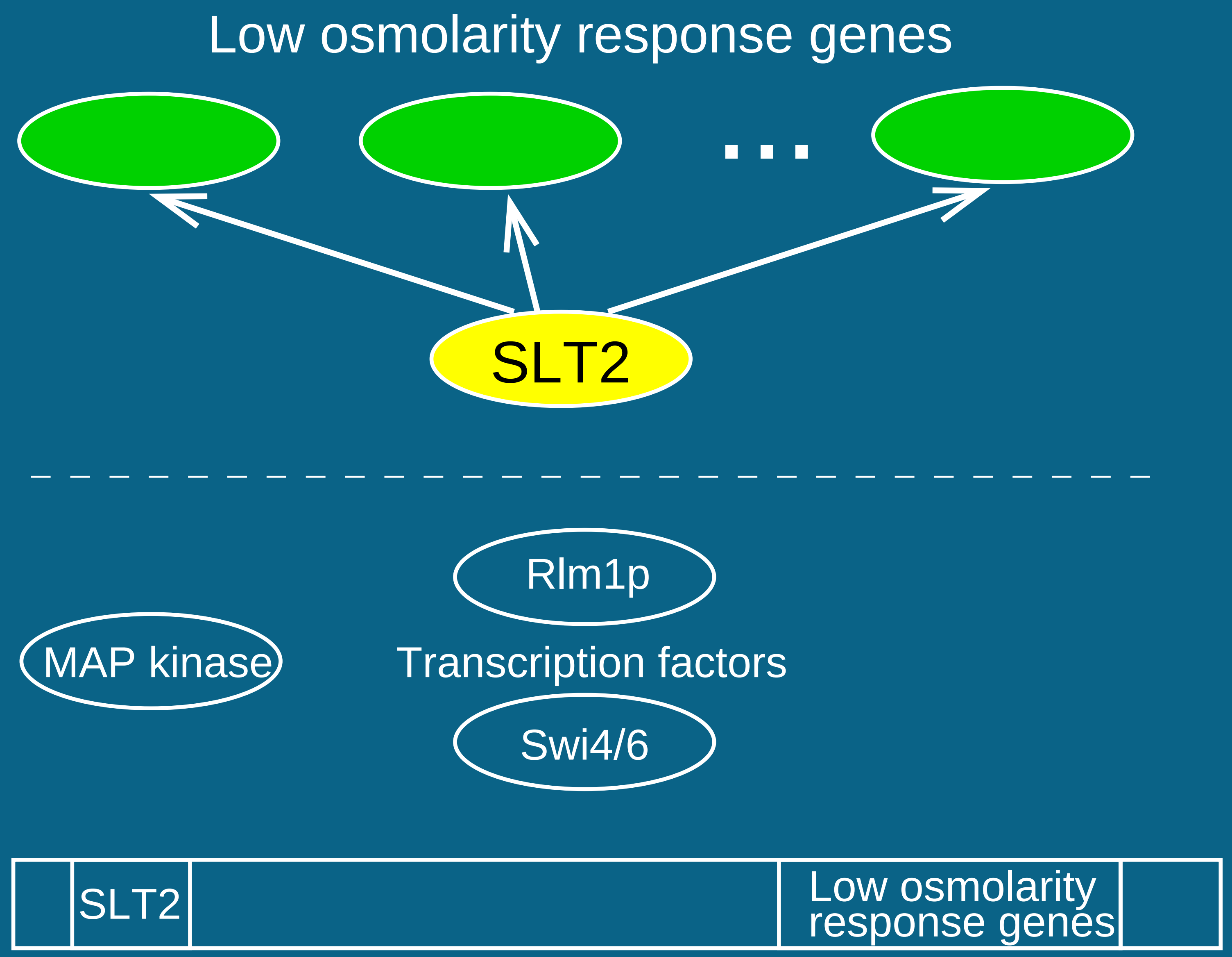
Advantage of Bayesian networks: context-specific and non-linear.

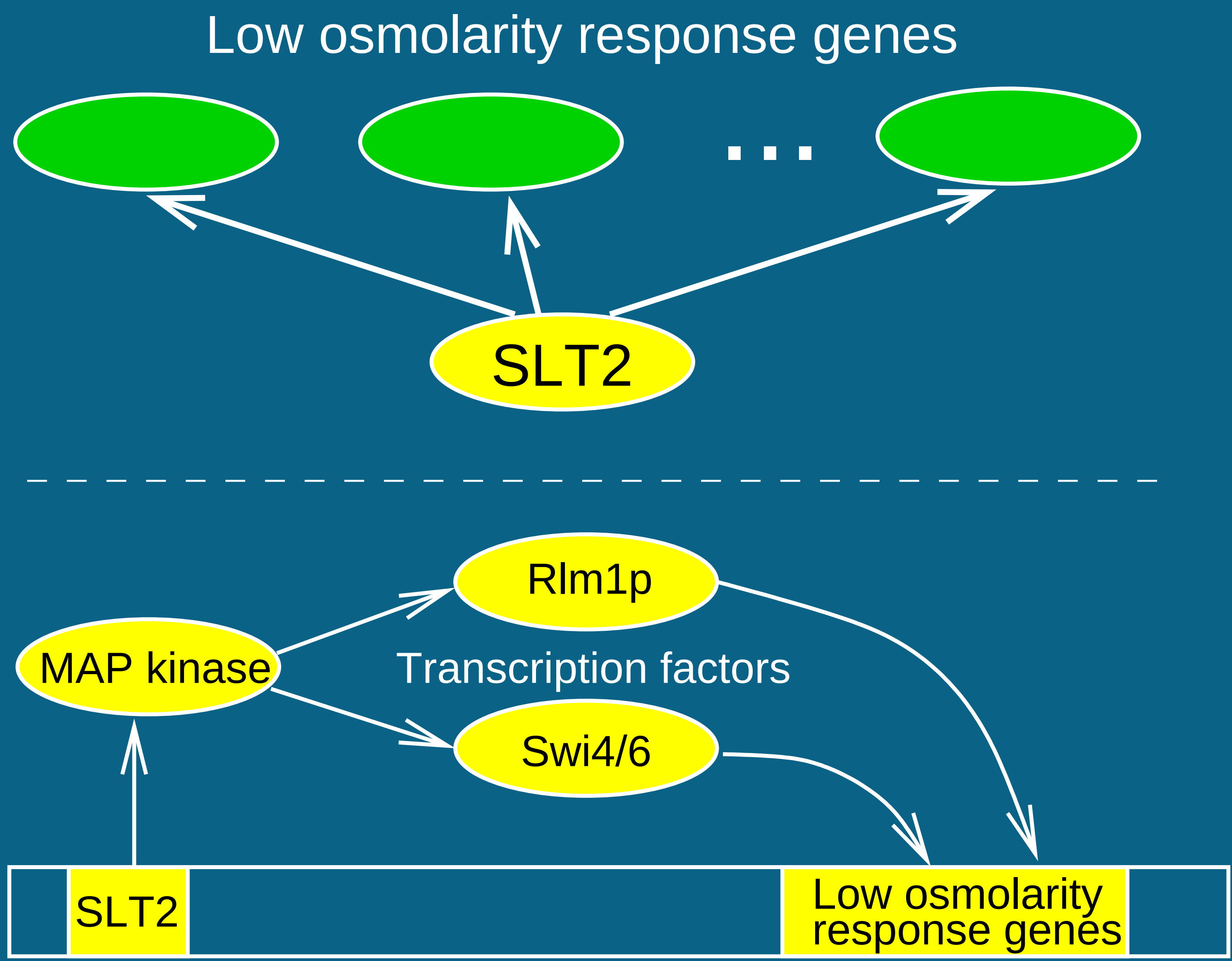

A
B
C
A
B
C
A
B
C
A
B
C
1
-1
A
1
-1
B
1
-1
C

Separator relations

and

subnetworks


Low osmolarity response genes
...
SLT2
Rlm1p
MAP kinase
Transcription factors
Swi4/6
SLT2
Low osmolarity response genes


Low osmolarity response genes
...
SLT2
Rlm1p
MAP kinase
Transcription factors
Swi4/6
SLT2
Low osmolarity
response genes

Outline of the talk

- Recapitulation: Bayesian networks
- Reverse engineering:
  Learning networks from data
- Application to the yeast cell cycle
- **Estimating the accuracy of inference**

Dirk Husmeier (2003)
Bioinformatics 19, 2271-2282

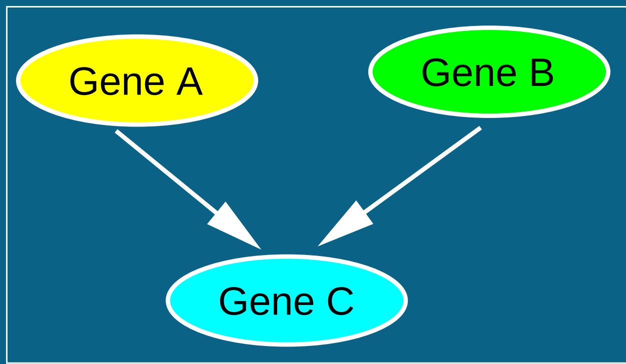
Disadvantage of **real data:**

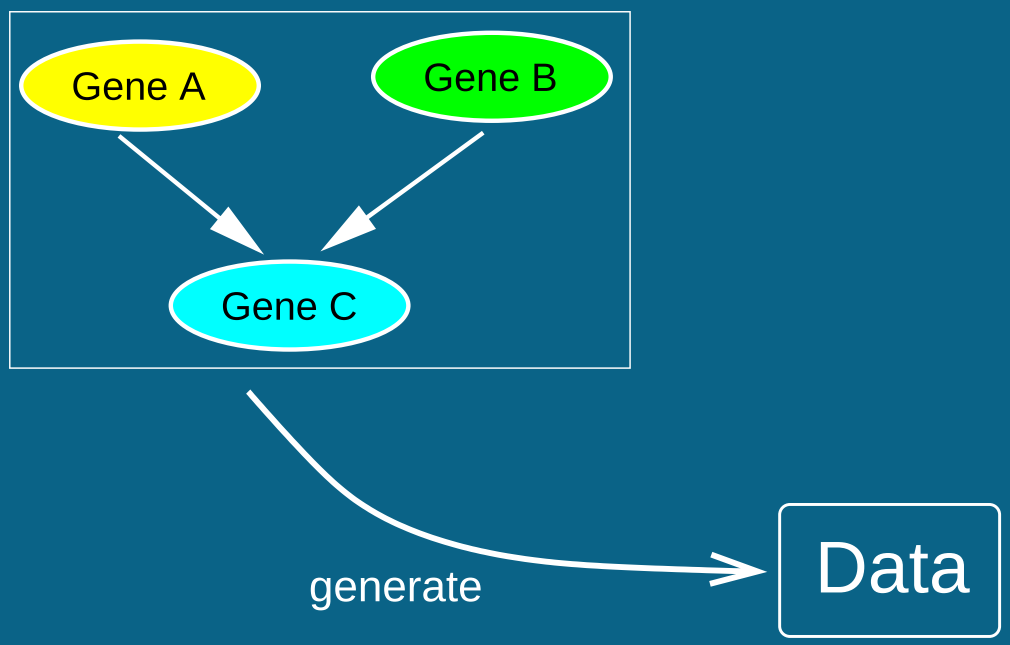
No gold standards !

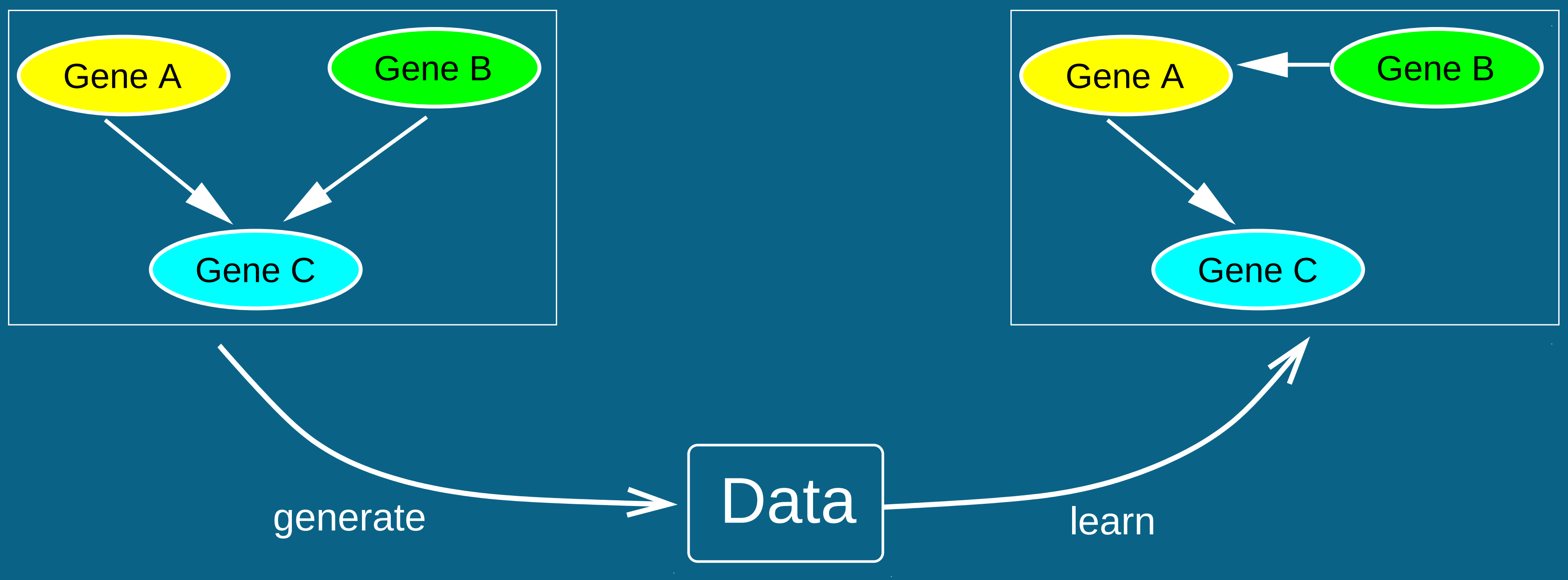
It may be possible to

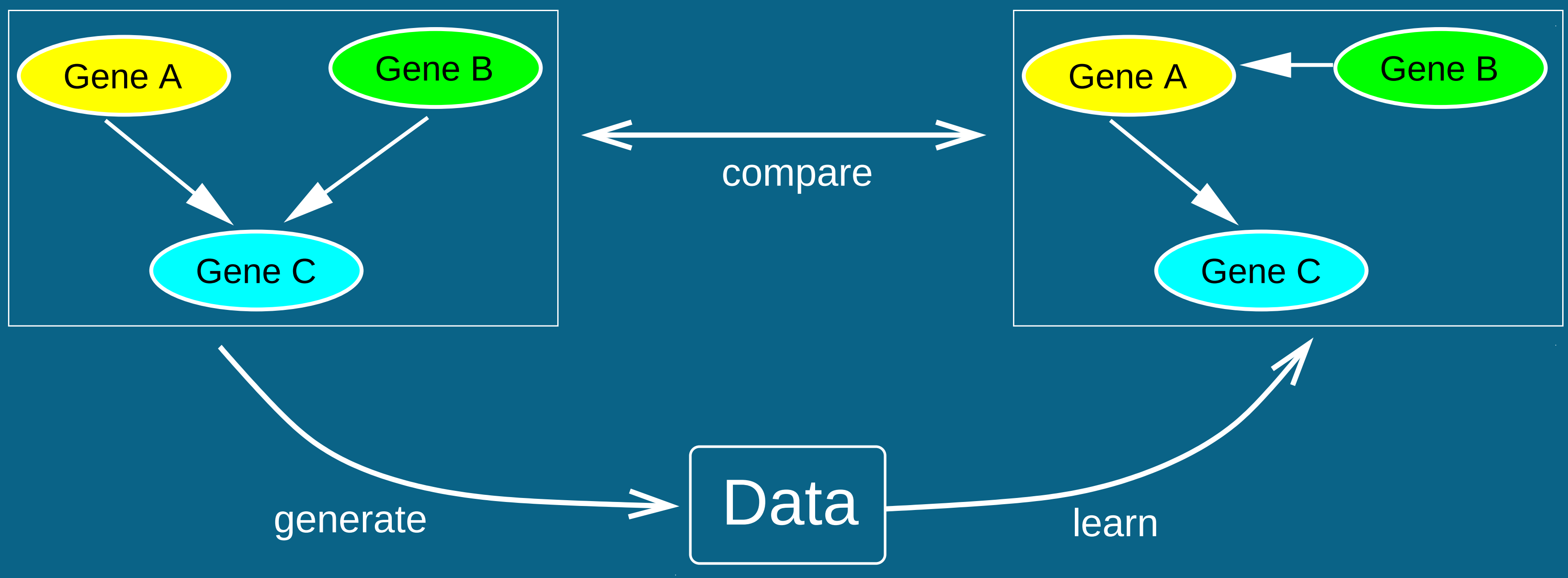
estimate the **sensitivity**,

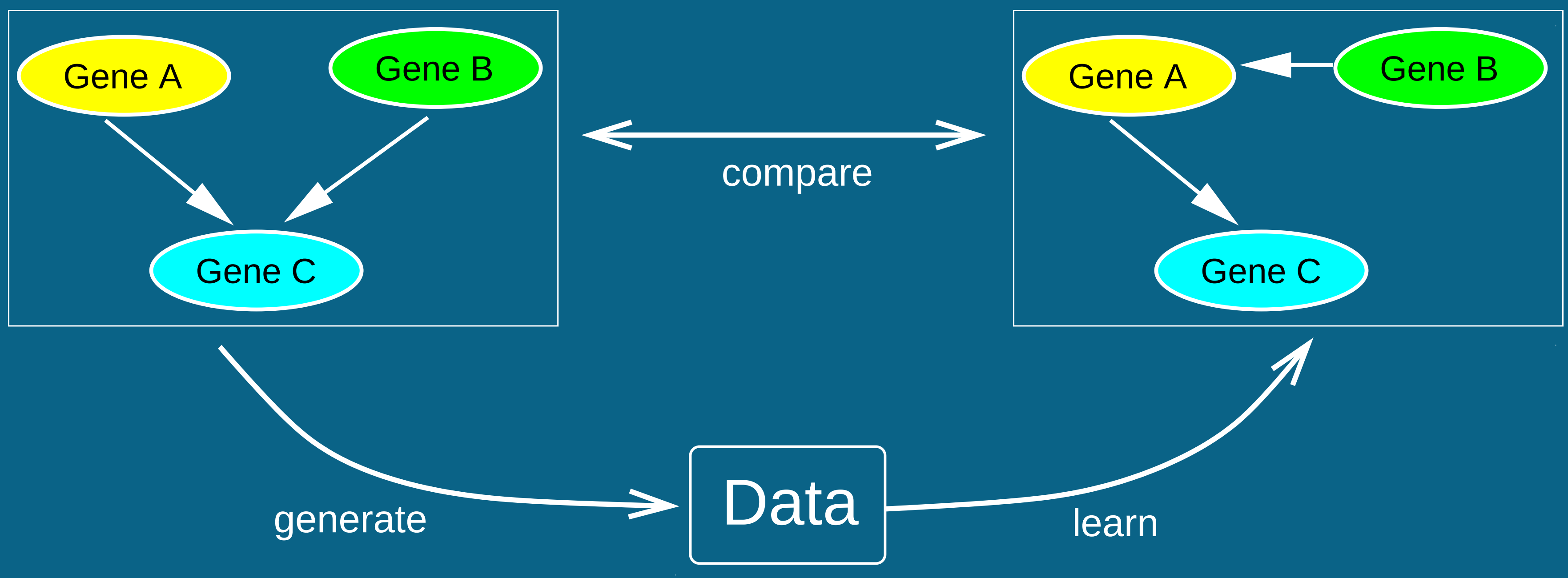
but not the **specificity**.


Gene A
Gene B
Gene C

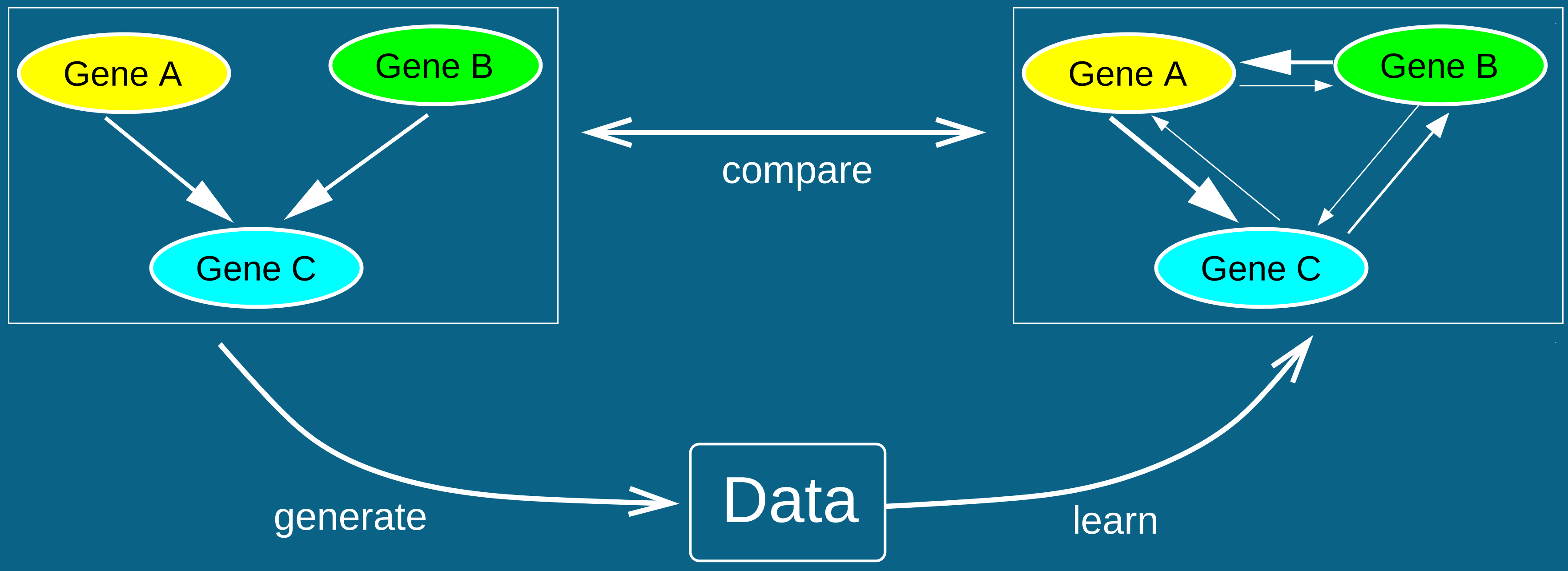

Gene A
Gene B
Gene C
generate
Data


Gene A
Gene B
Gene C
generate
Data
learn
Gene A
Gene B
Gene C

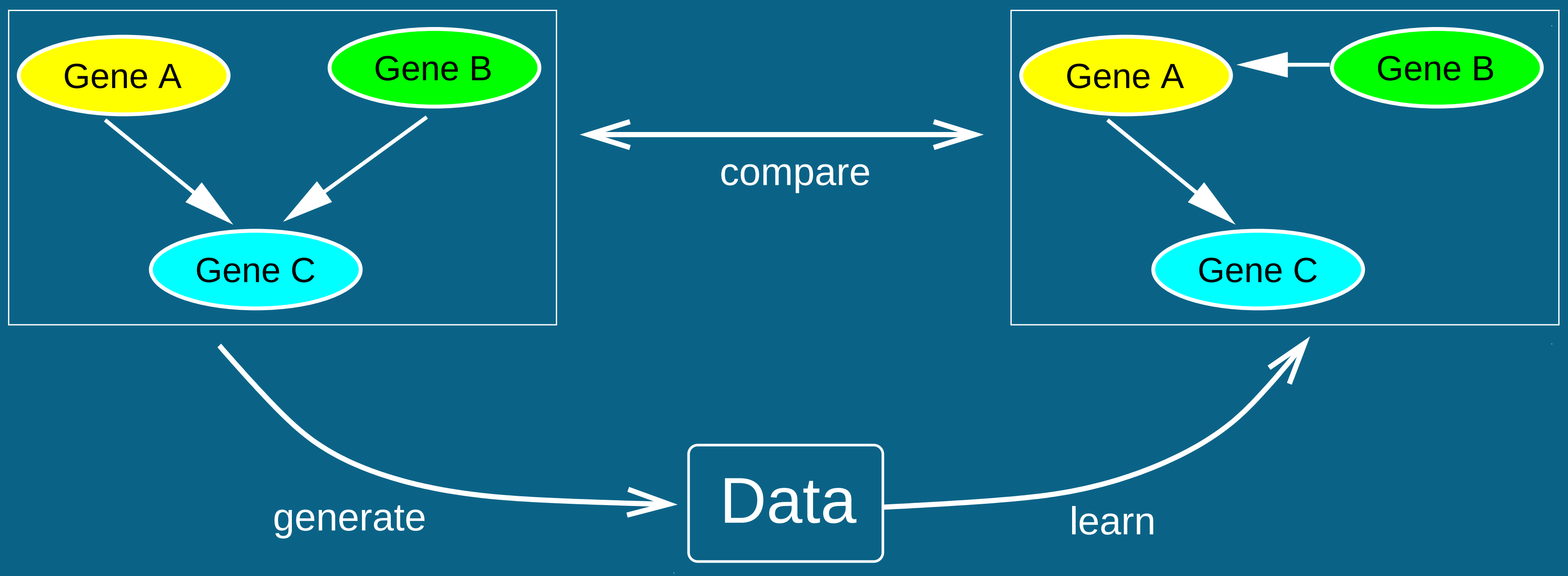

Gene A
Gene B
Gene C
compare
Gene A
Gene B
Gene C
generate
Data
learn



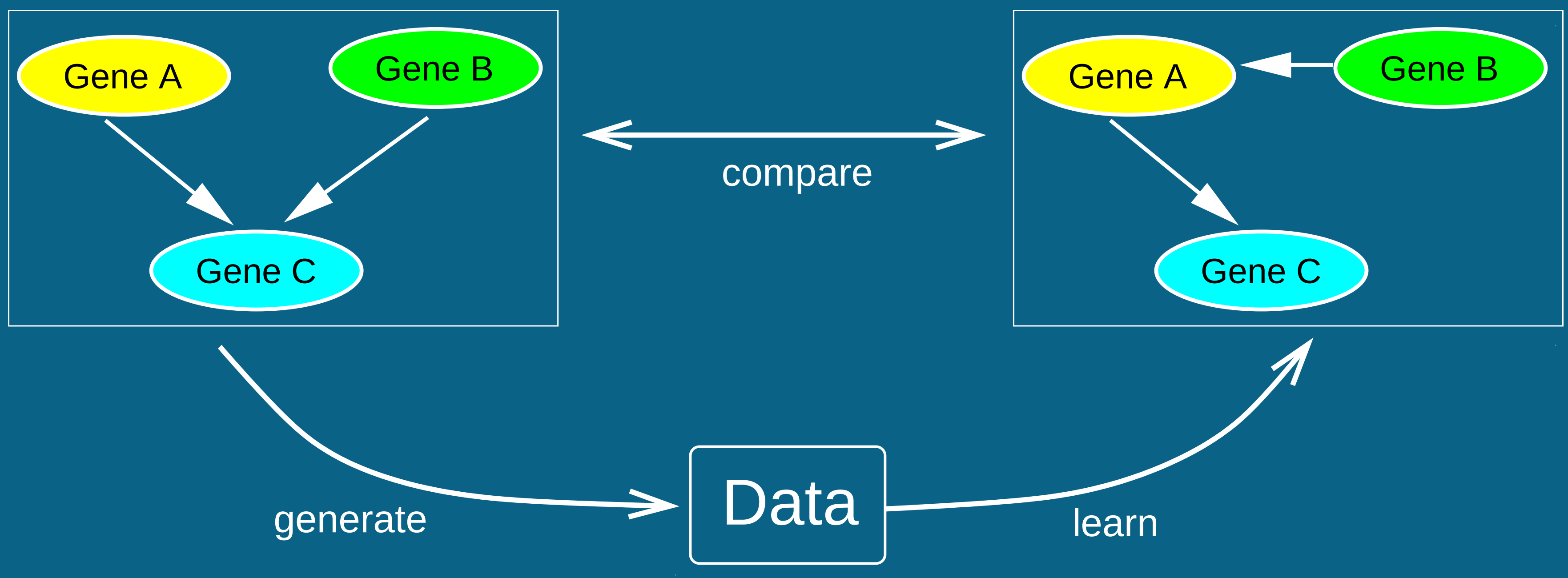
Deterministic inference



Probabilistic inference



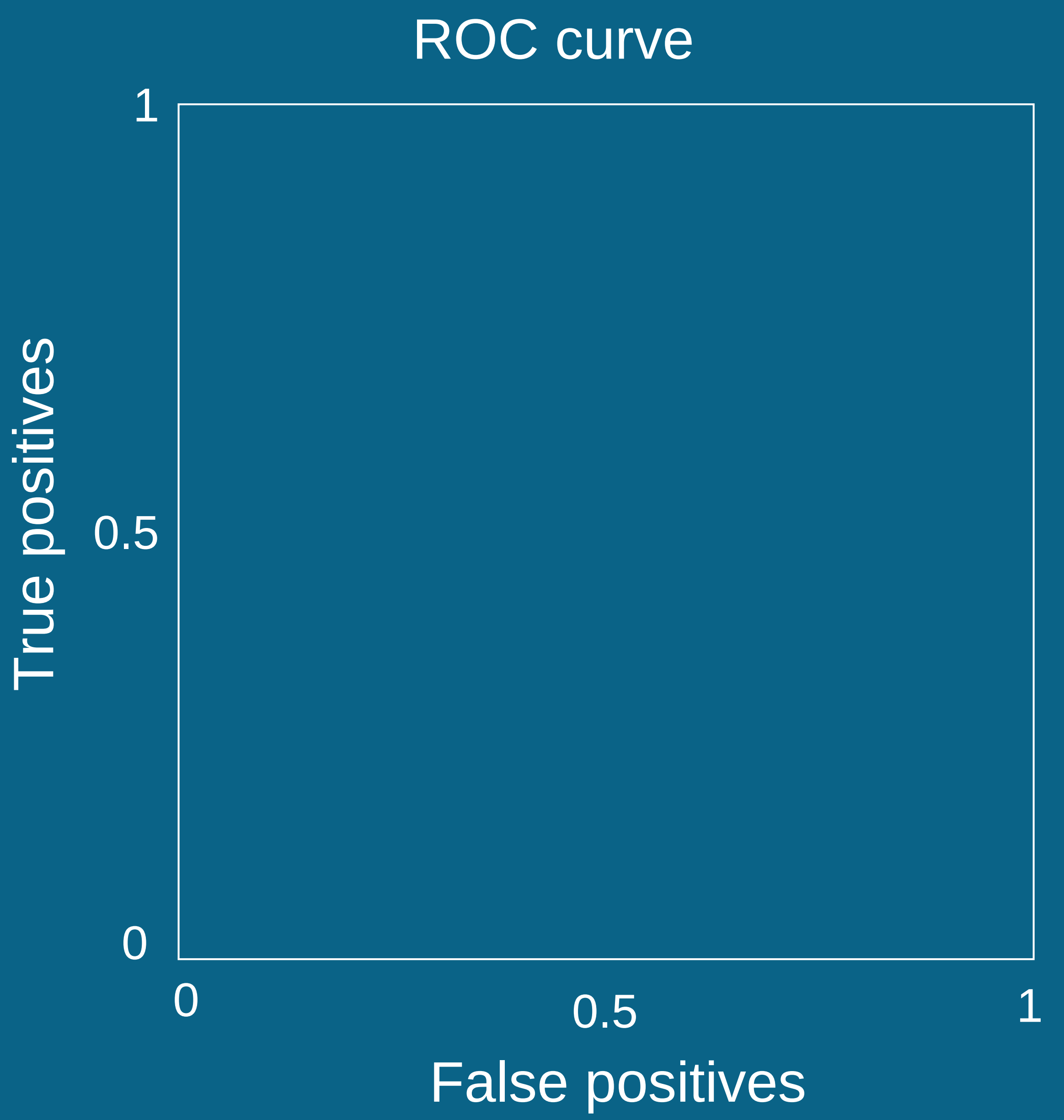
Thresholding

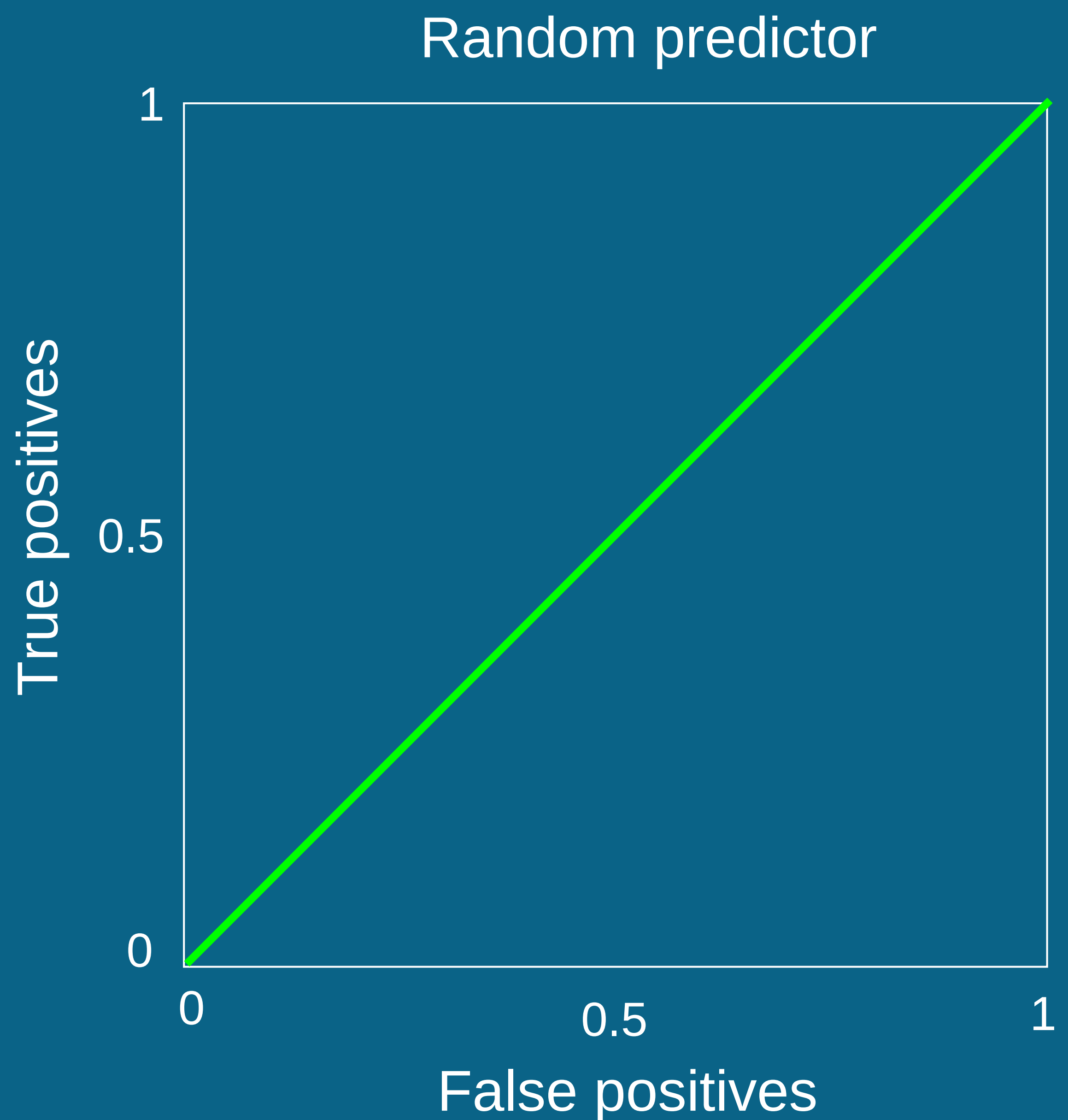


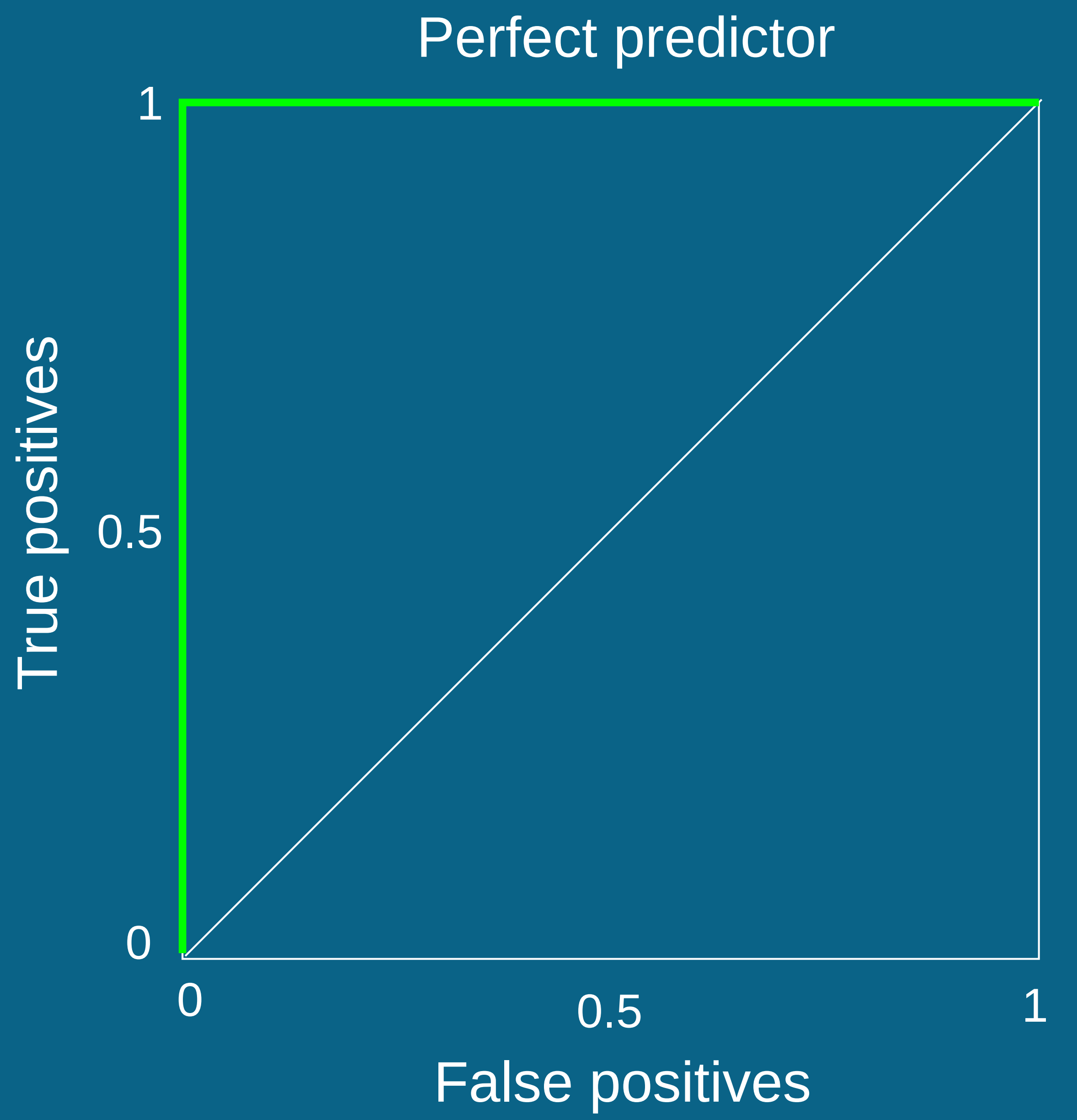
Thresholding

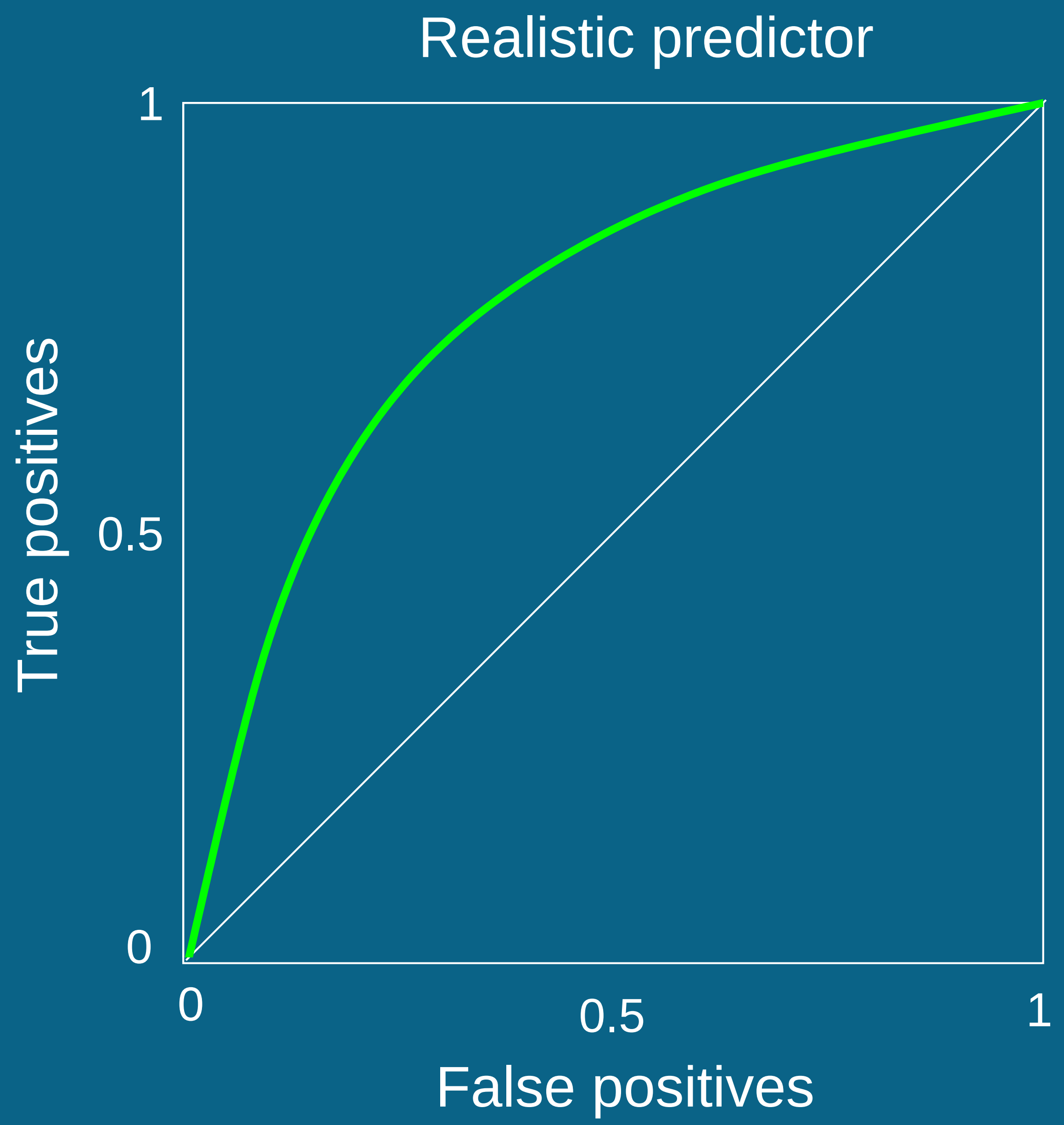
True positives

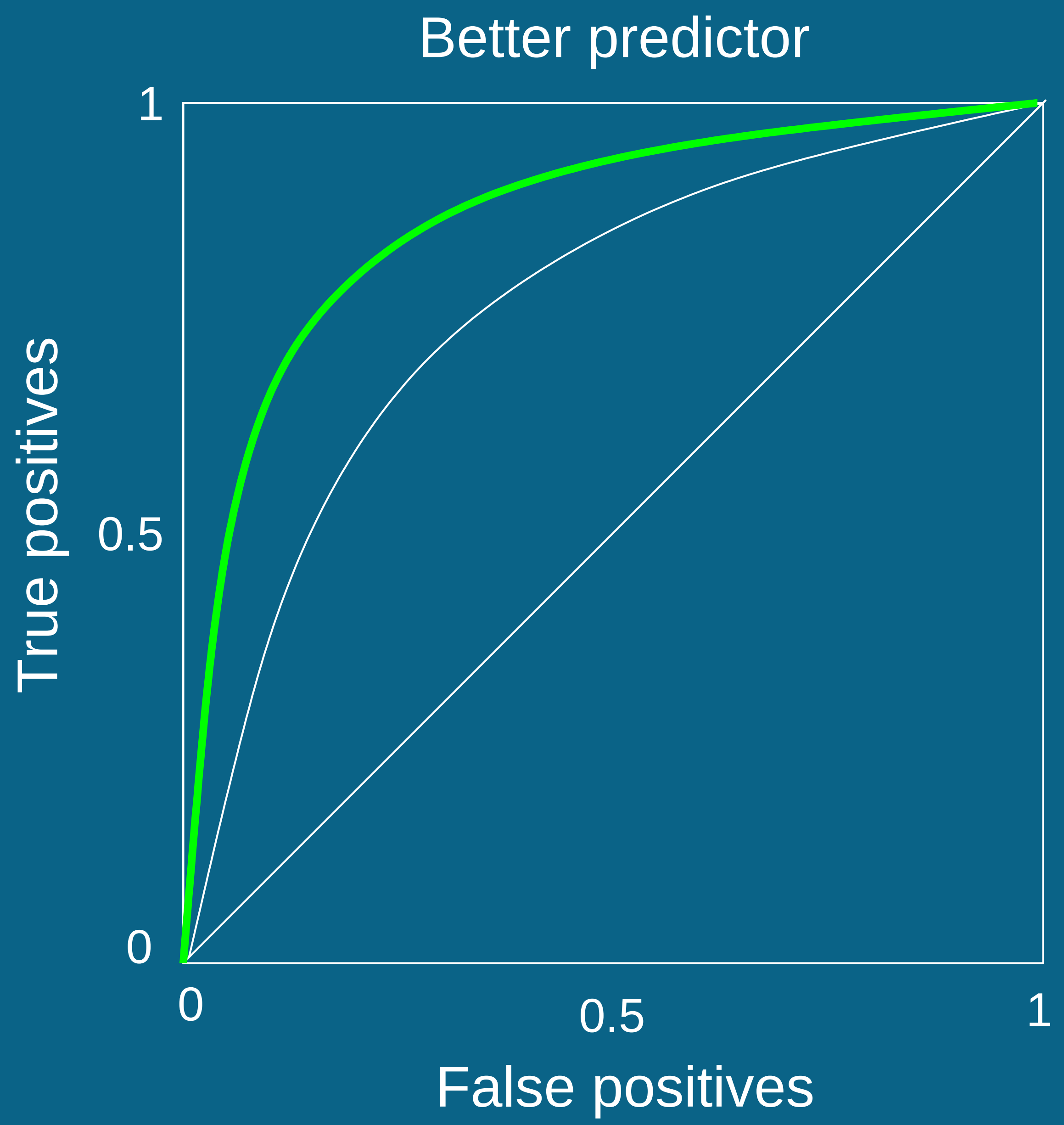
False positives

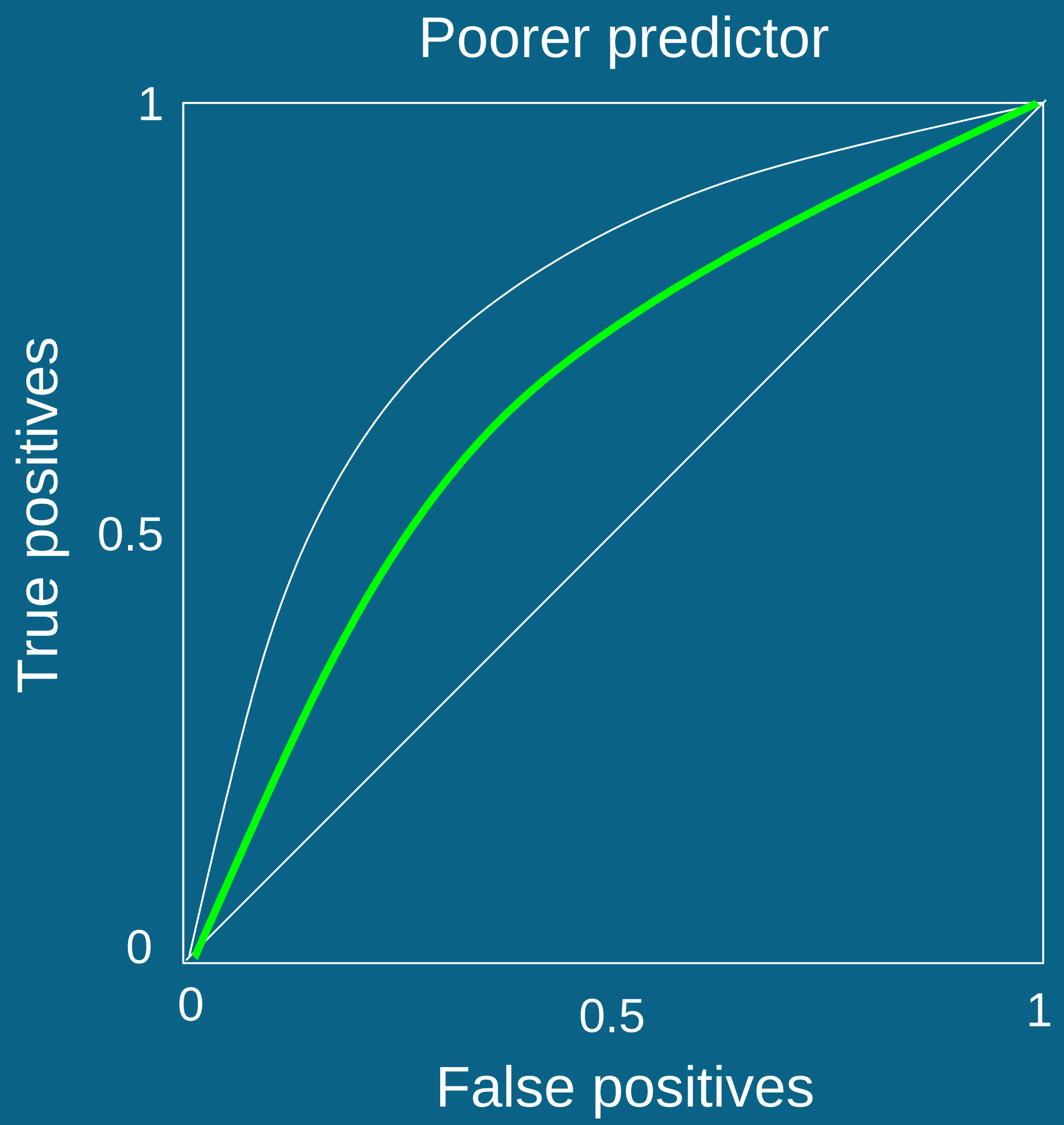

ROC curve
1
0.5
0
0
0.5
1
True positives
False positives


Random predictor
1
0.5
0
True positives
0
0.5
1
False positives


Perfect predictor
1
0.5
0
True positives
0
0.5
1
False positives

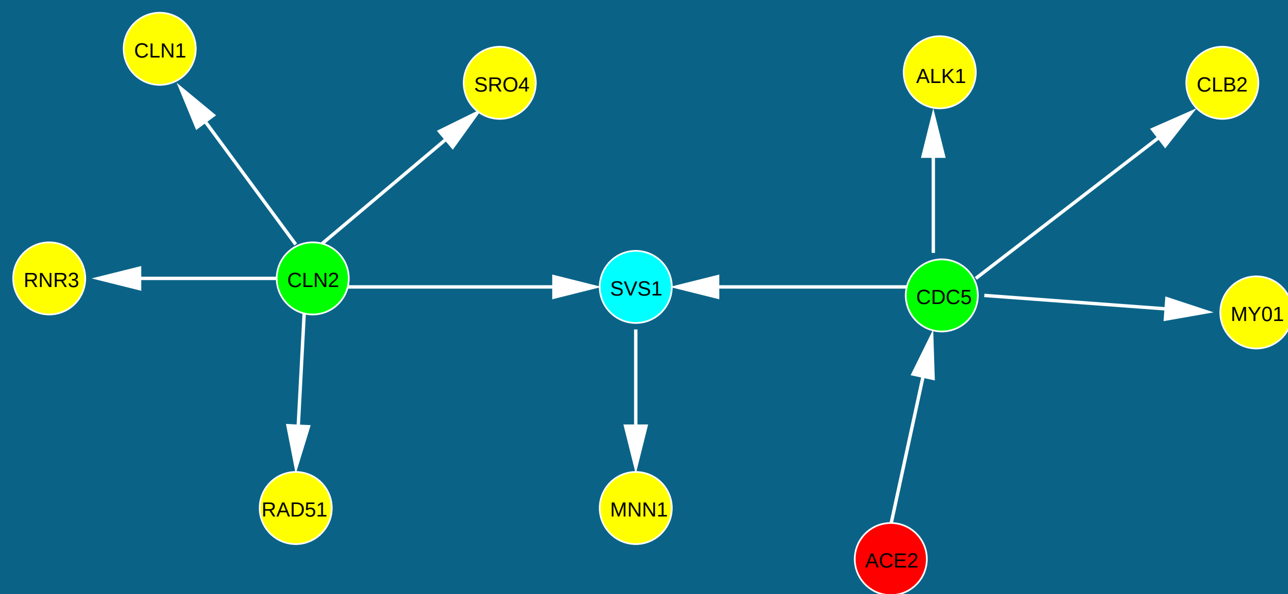

Realistic predictor
1
0.5
0
True positives
0
0.5
1
False positives


Better predictor
True positives
1
0.5
0
0
0.5
1
False positives


Poorer predictor
1
0.5
0
True positives
0
0.5
1
False positives

Data: binary

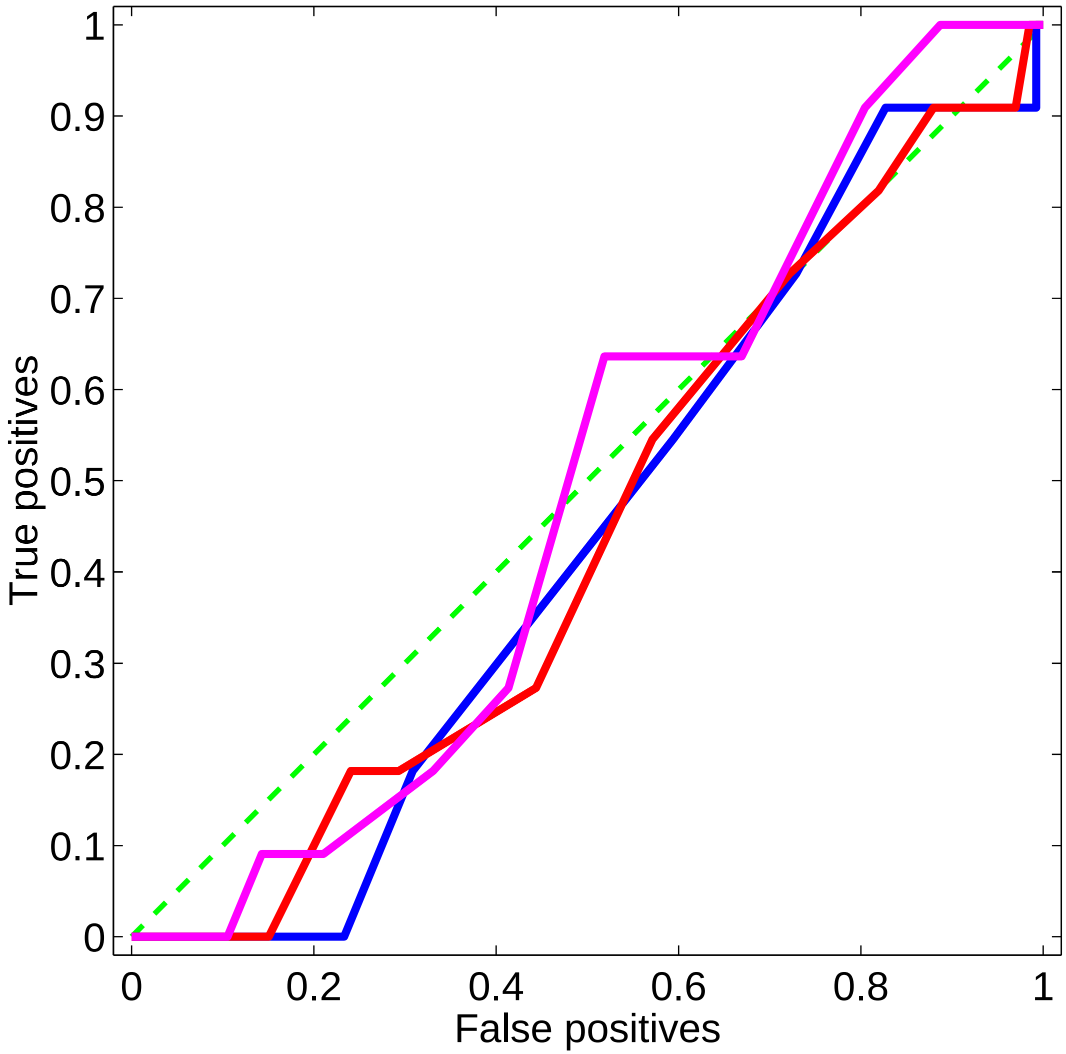
Model:



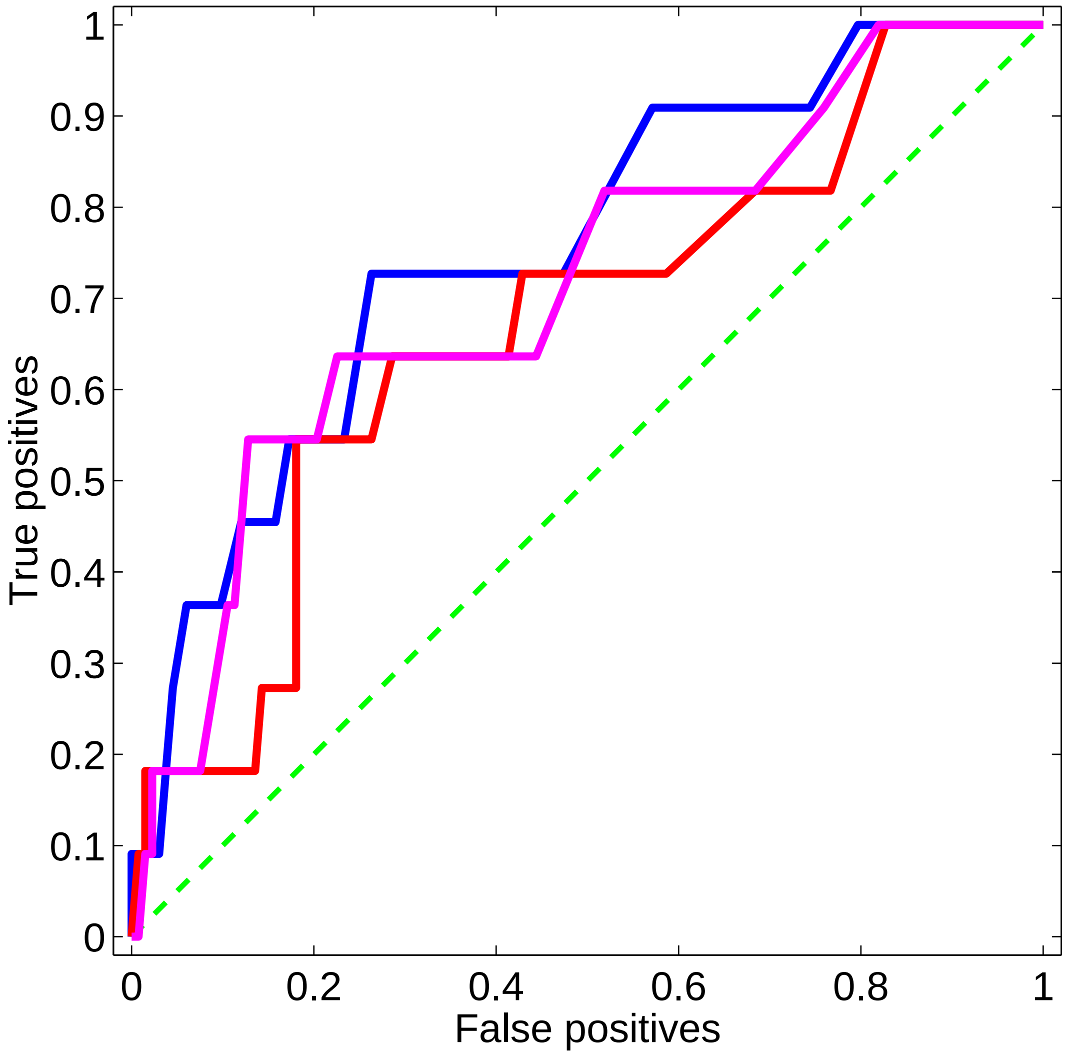
Parameters:

Noisy boolean: $P \in \{0.1, 0.9\}$

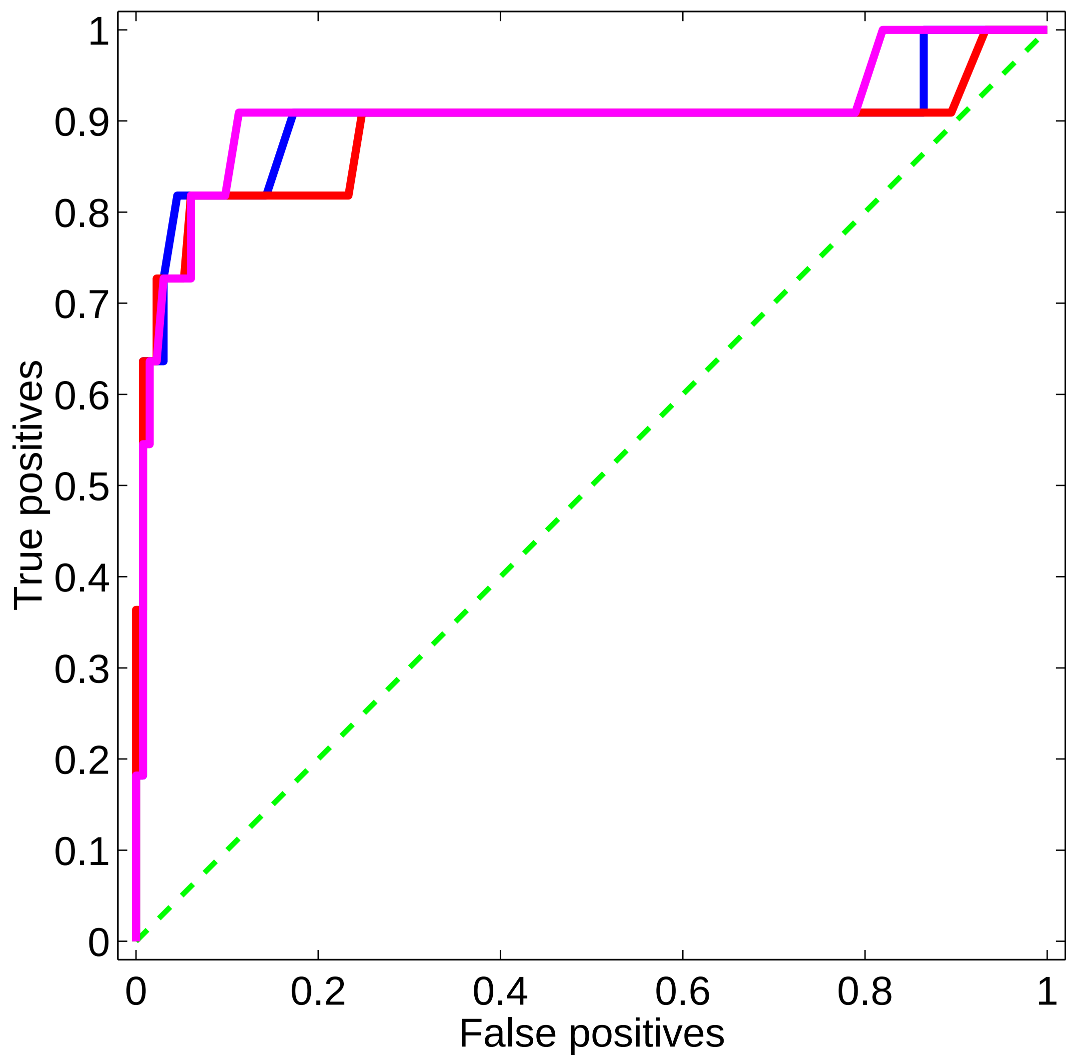
ROC curve: Sample size= 3



ROC curve: Sample size= 6

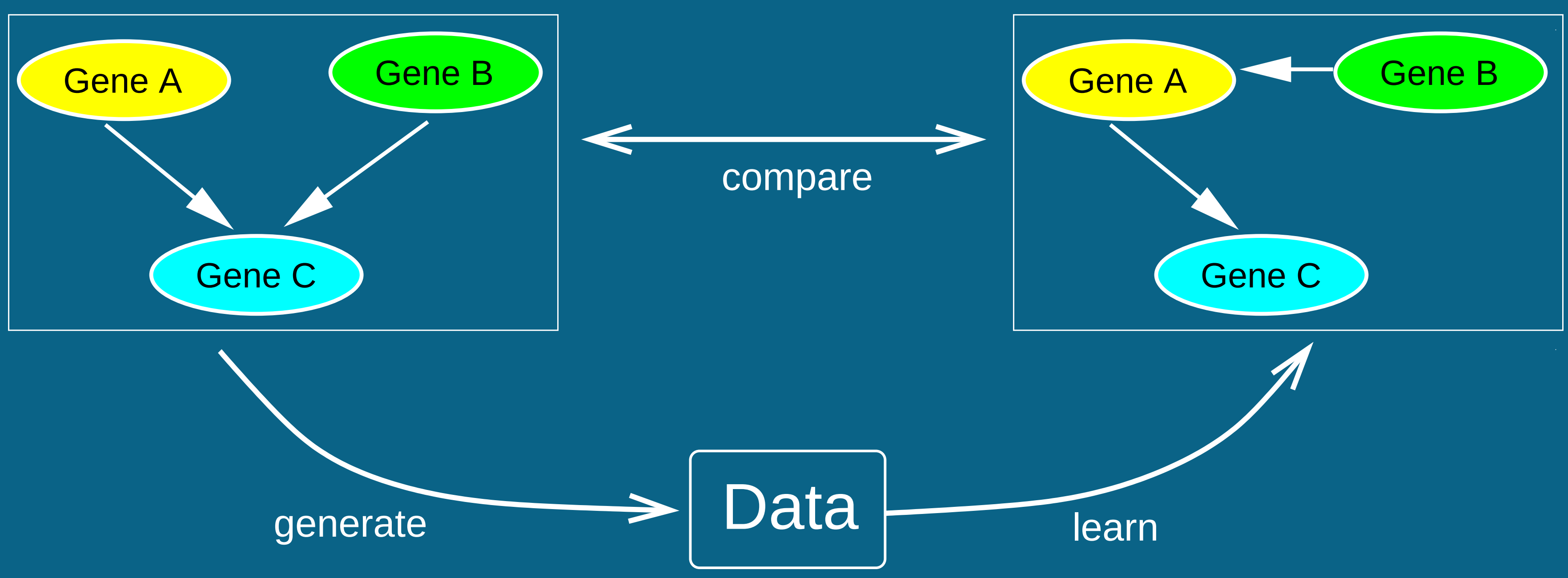


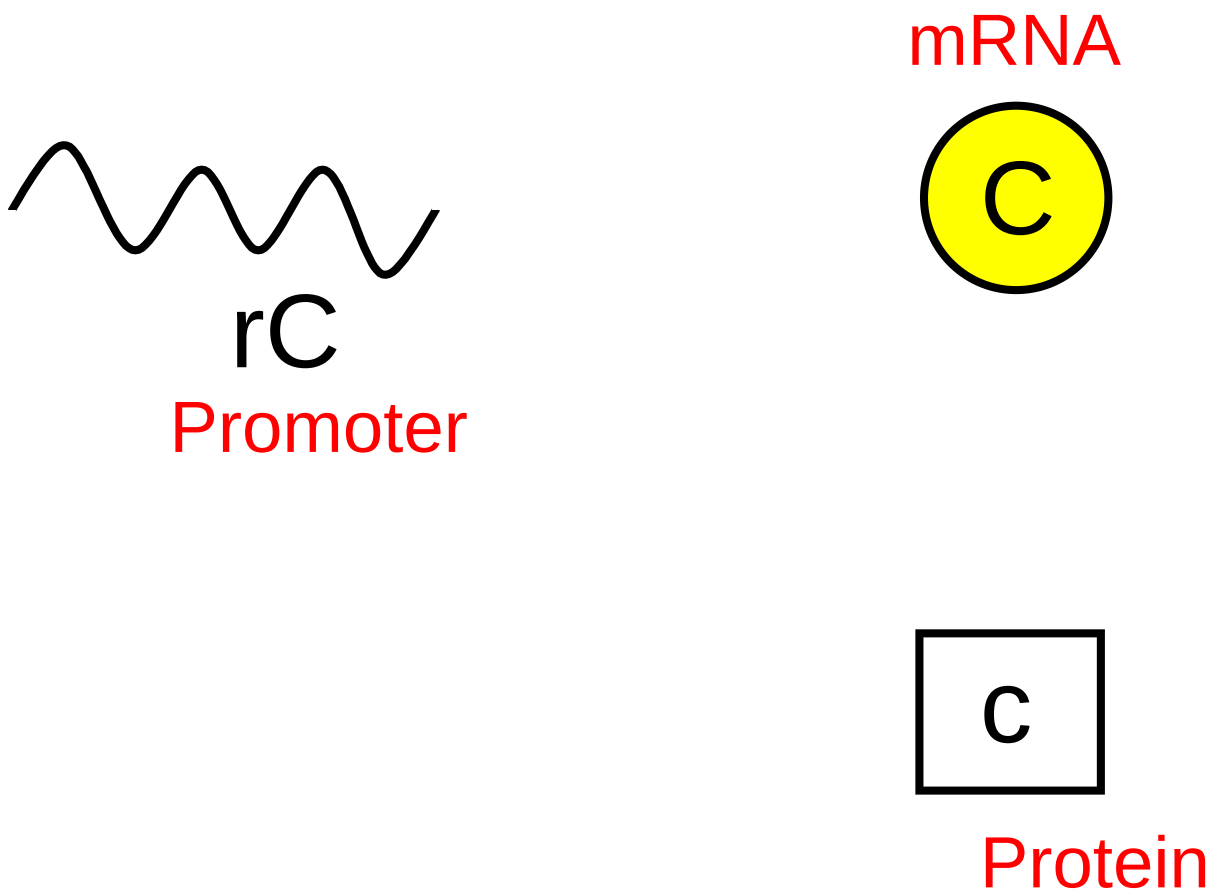
ROC curve: Sample size= 12

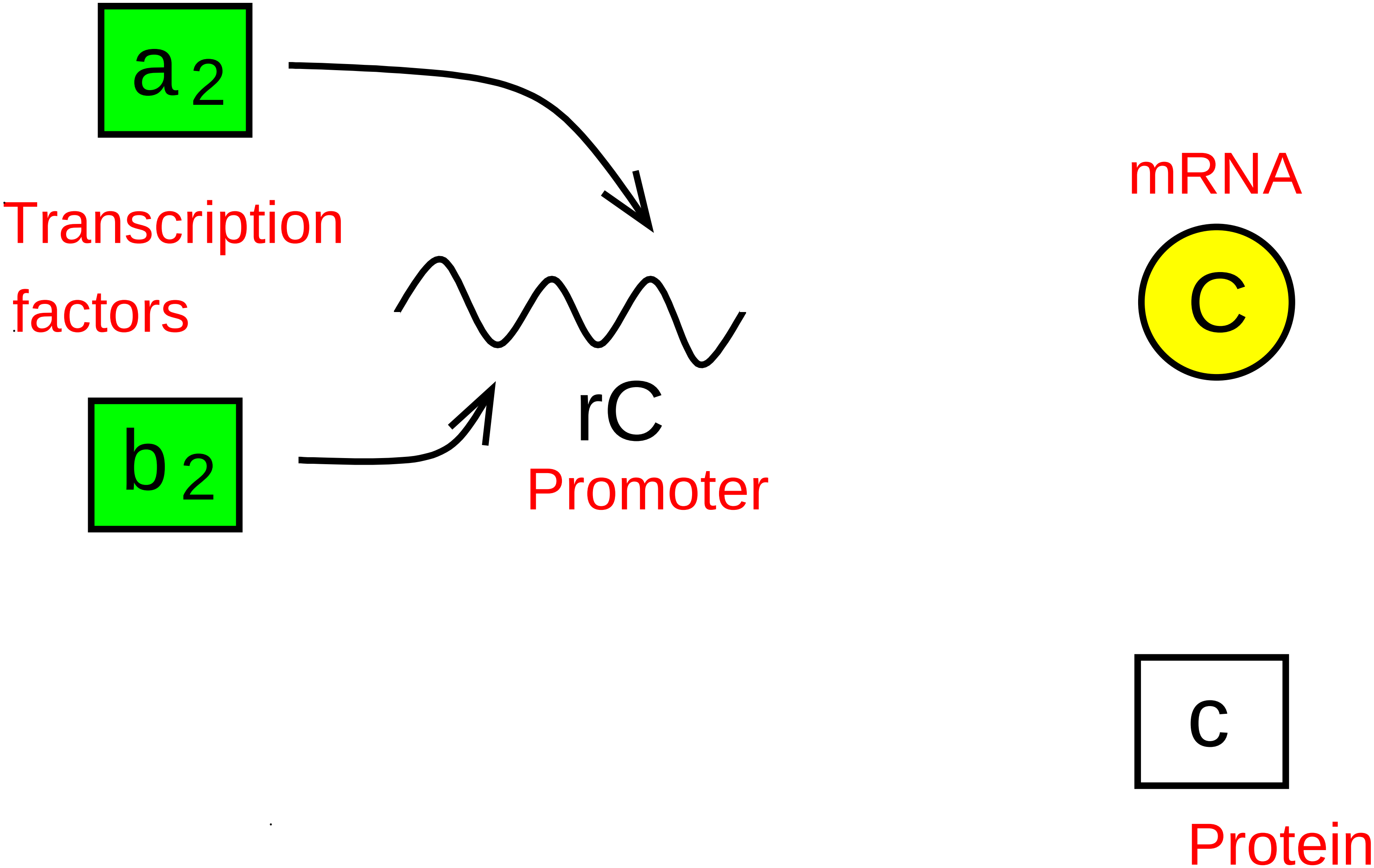


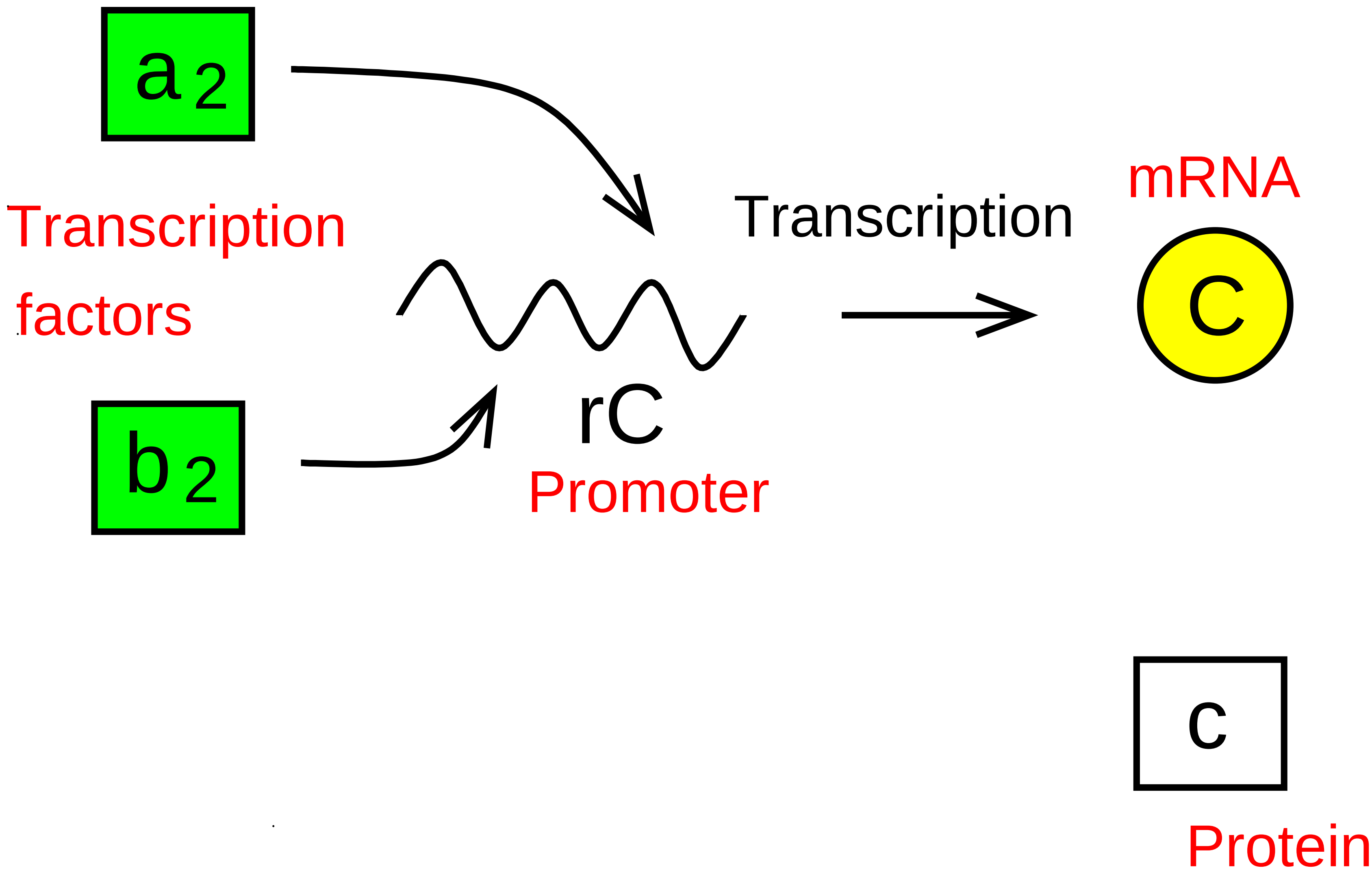
Disadvantage:

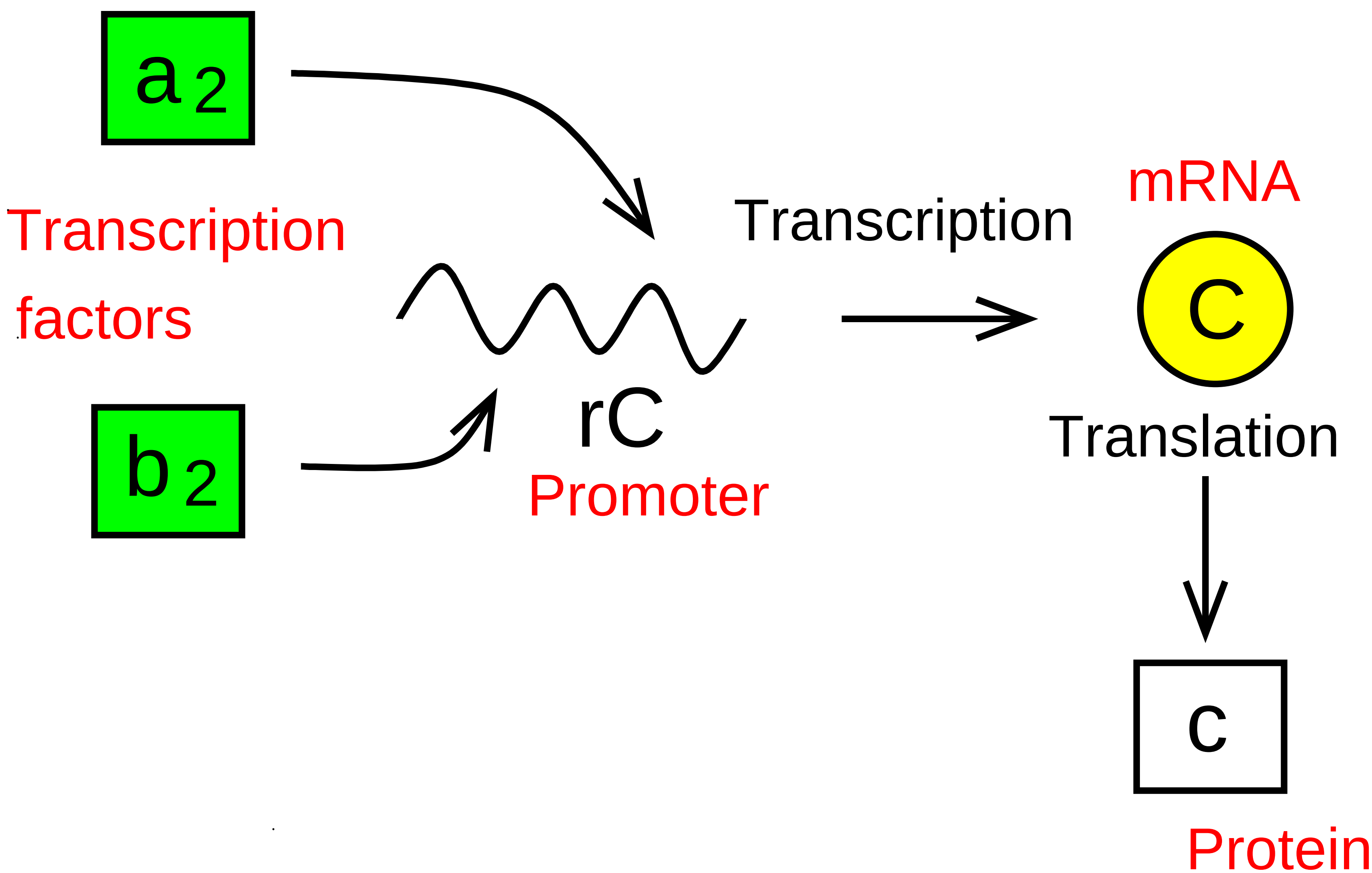
Unrealistic, no mismatch between the model used for data generation and the model used for inference.

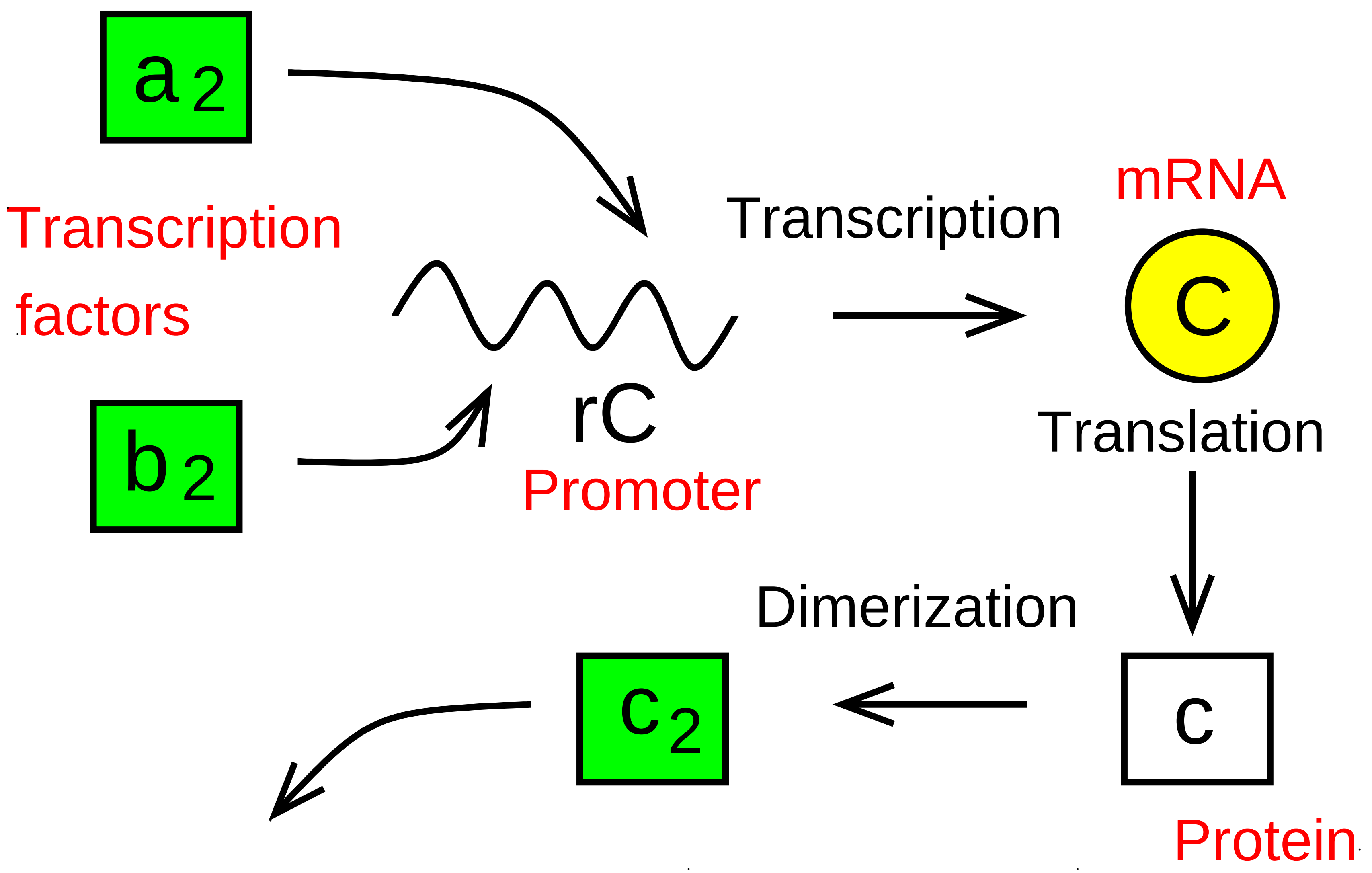



mRNA
rC
C
Promoter
c
Protein

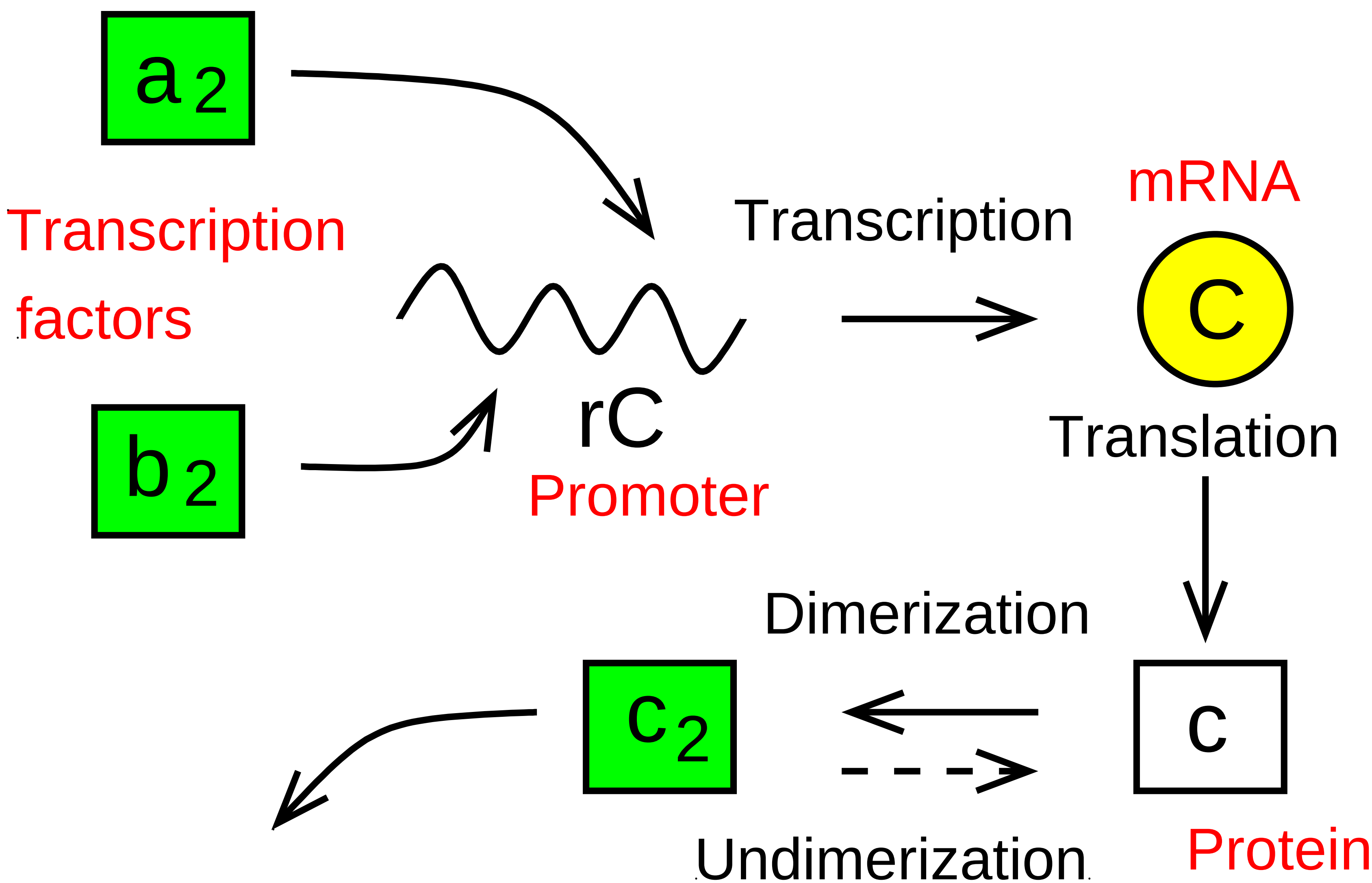

a 2
Transcription
factors
b 2
rC
Promoter
mRNA
c
c
Protein


a 2
Transcription
factors
b 2
rC
Promoter
Transcription
mRNA
C
c
Protein

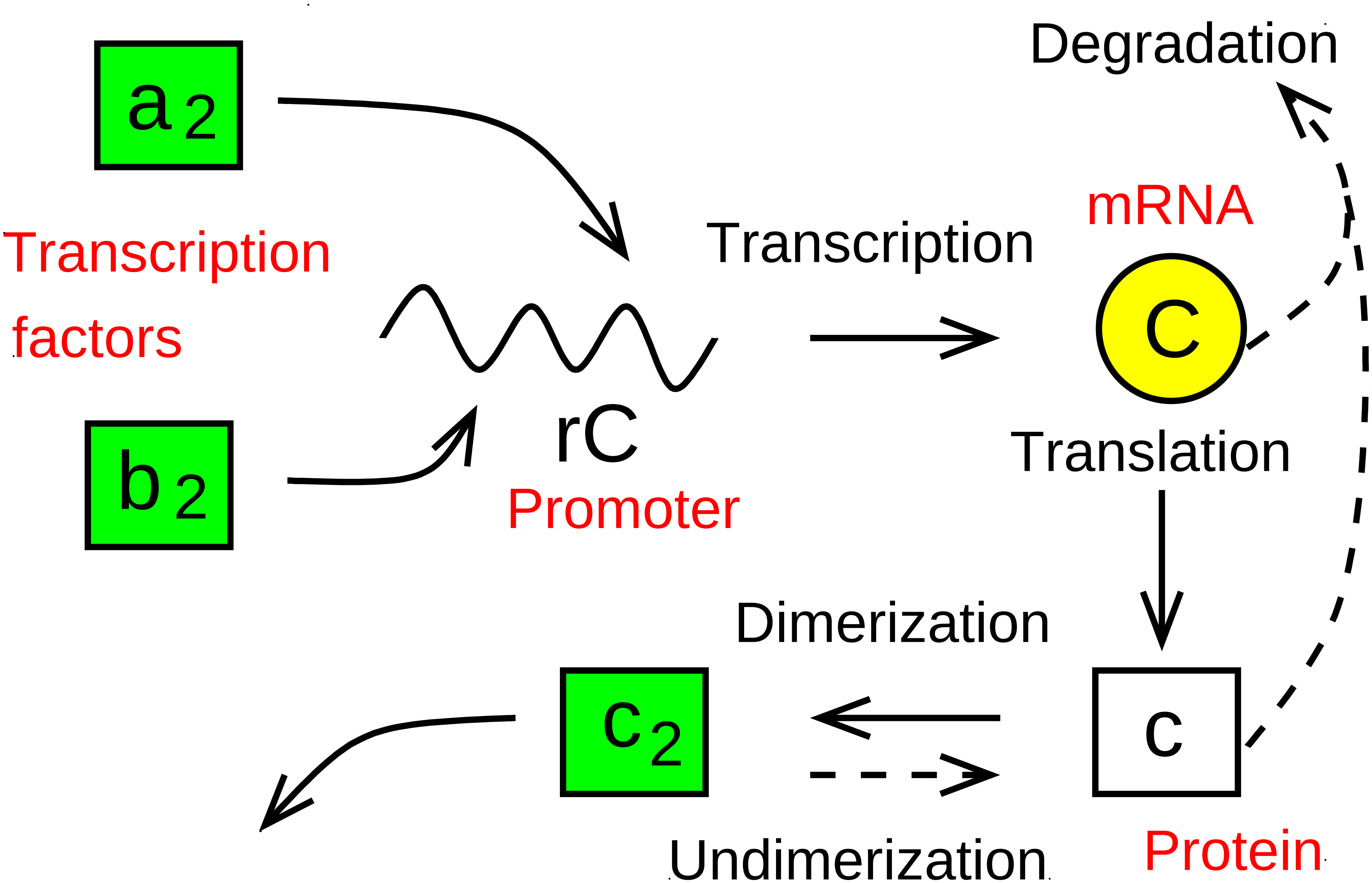

a2
Transcription
factors
b2
rC
Promoter
Transcription
mRNA
c
Translation
c
Protein



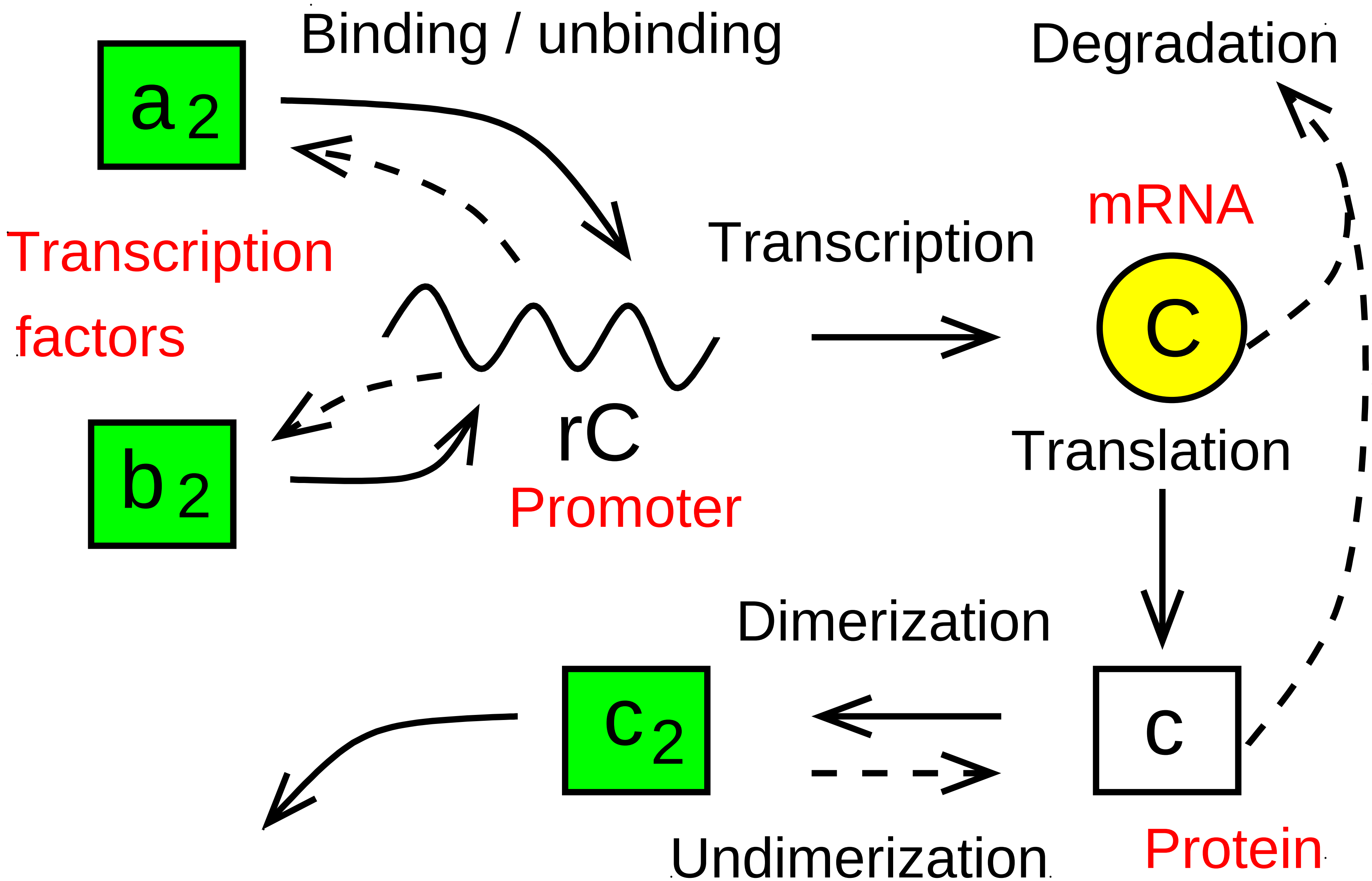
a 2
Transcription
factors
b 2
rC
Promoter
Transcription
mRNA
c
Translation
Dimerization
c 2
c
Protein



a2
Transcription factors
b2
rC
Promoter
Transcription
mRNA
c
Translation
Dimerization
c2
c
Undimerization
Protein



a2
Transcription factors
b2
rC
Promoter
Transcription
mRNA
c
Degradation
Translation
Dimerization
c2
c
Undimerization
Protein



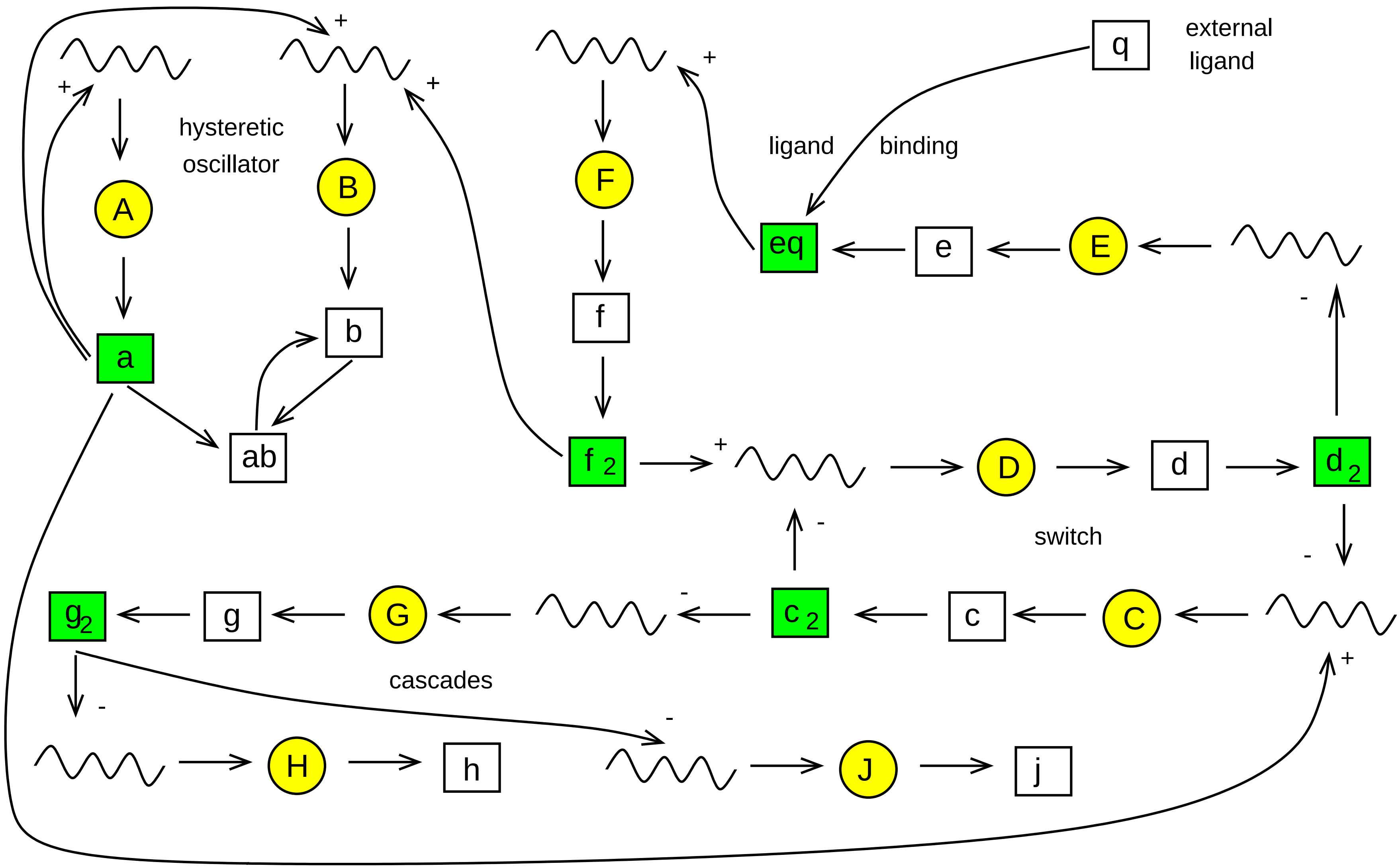
Binding / unbinding
Degradation
a2
Transcription factors
mRNA
Transcription
c
rC
Promoter
b2
Translation
Dimerization
c2
c
Undimerization
Protein

$$\frac{d}{dt}[a2.rC] = \lambda^{+}_{a_2.rC}[a_2][rC] - \lambda^{-}_{a_2.rC}[a_2.rC]$$

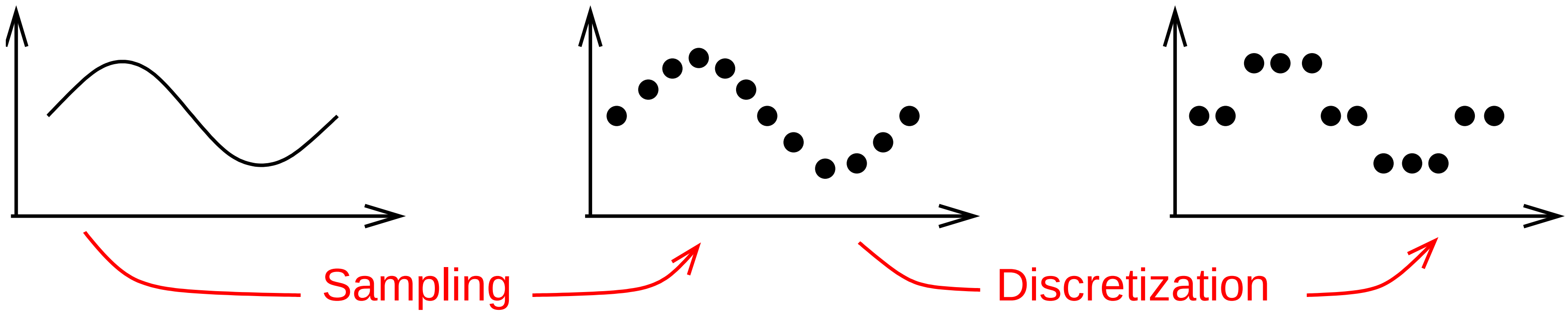
$$\frac{d}{dt}[C] = \lambda_{rC}[rC] + \lambda_{a_2.rC}[a_2.rC] + \lambda_{b_2.rC}[b_2.rC] - \lambda_{C}[C]$$

$$\frac{d}{dt}[c] = \lambda_{Cc}[C] - \lambda_{c}[c]$$

$$\frac{d}{dt}[c_2] = \lambda^{+}_{cc}[c]^2 - \lambda^{-}_{cc}[c_2]$$

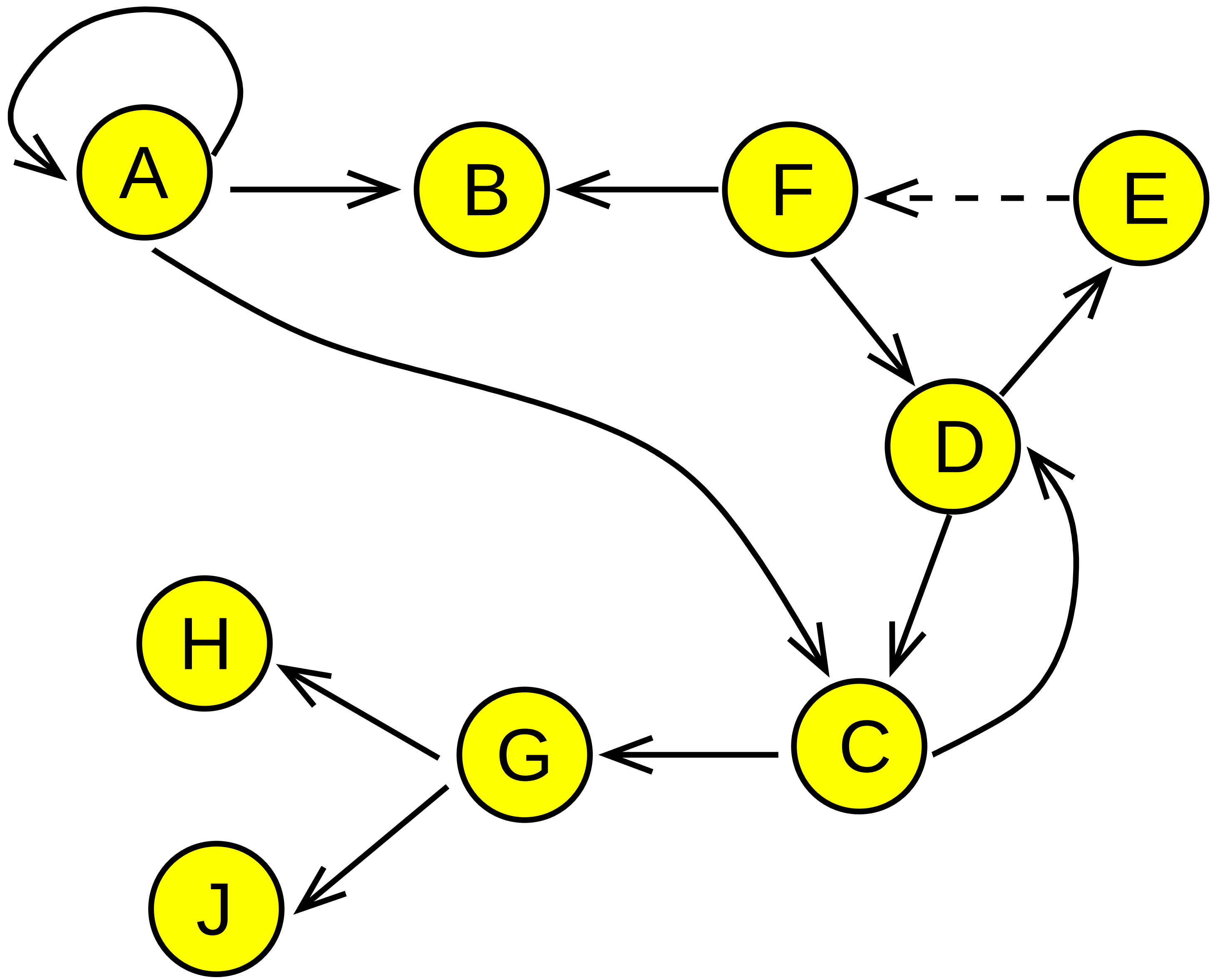


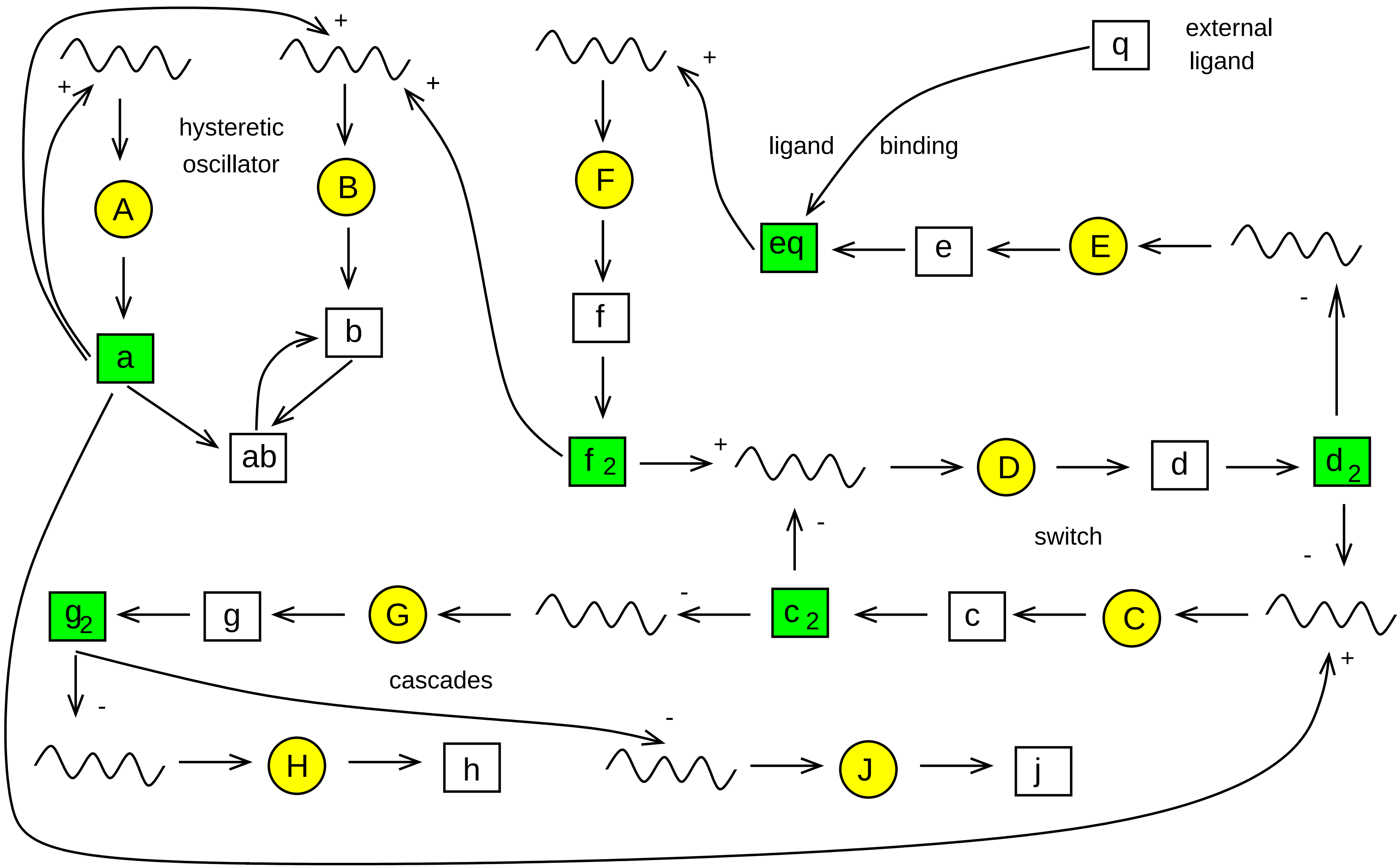
Zak et al., Int. Conf. Sys. Biol., 2001



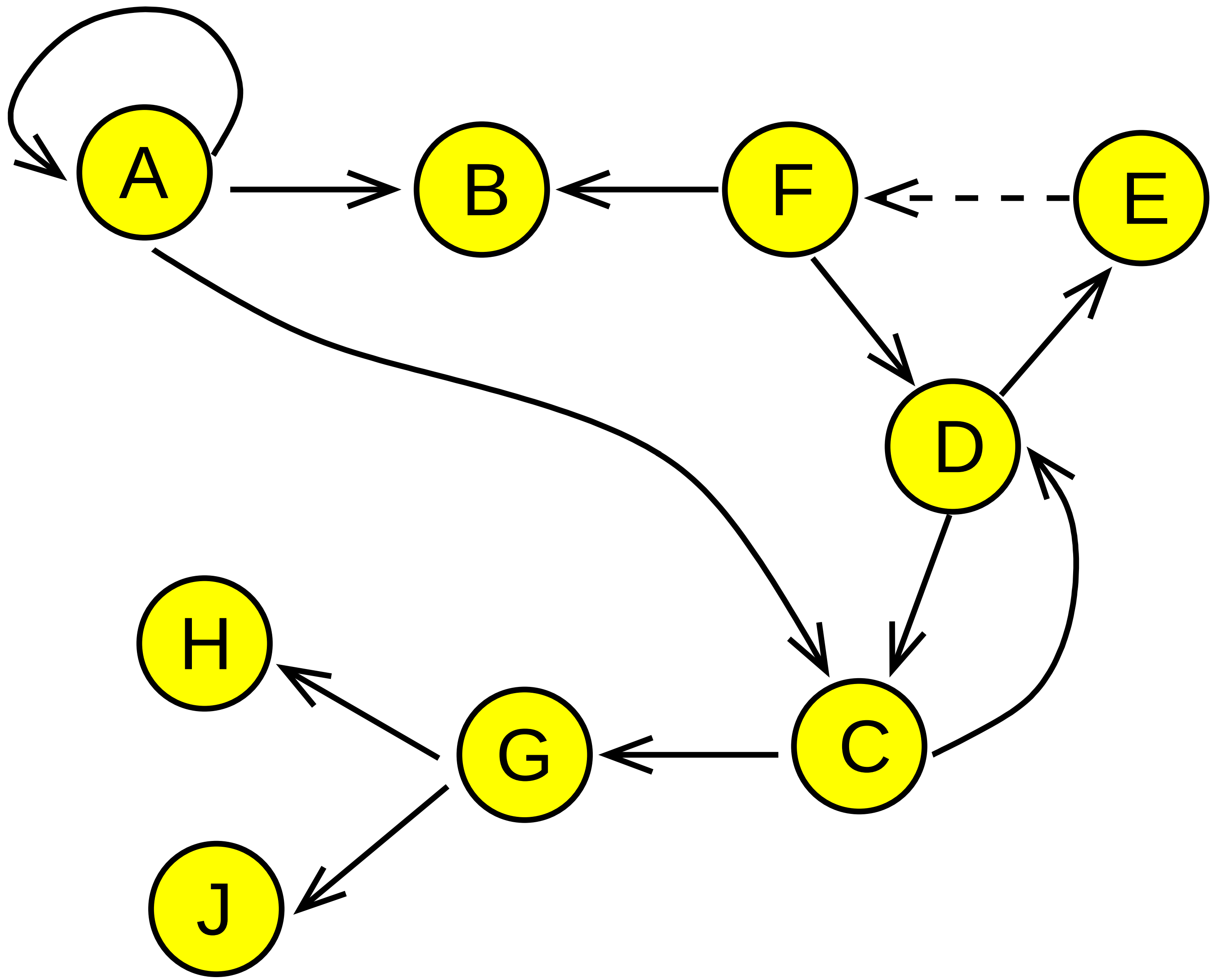
12 time points

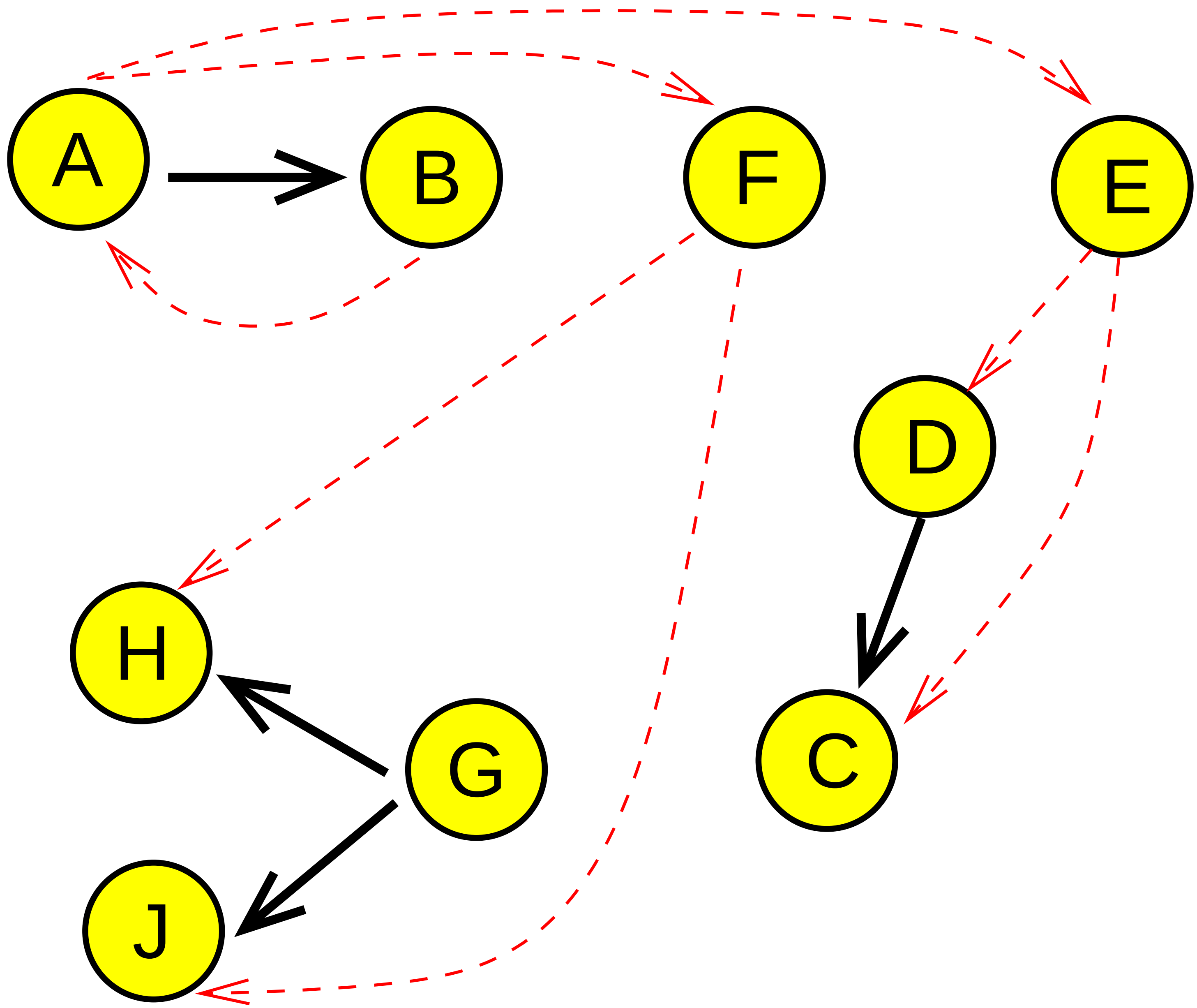
Recover the true genetic network
with
reverse engineering.


A
B
F
E
D
H
G
C
J



+
+
+
A
a
hysteretic
oscillator
B
b
ab
F
f
f_2
+
q
external
ligand
ligand
binding
eq
e
E
-
+
D
d
d_2
switch
-
-
g_2
g
G
-
c_2
c
C
+
-
cascades
-
H
h
J
j


A
B
F
E
D
H
G
C
J


A
B
F
E
D
H
G
C
J

Simulation Experiments

Ligand injection for 10 minutes.

Equilibrium

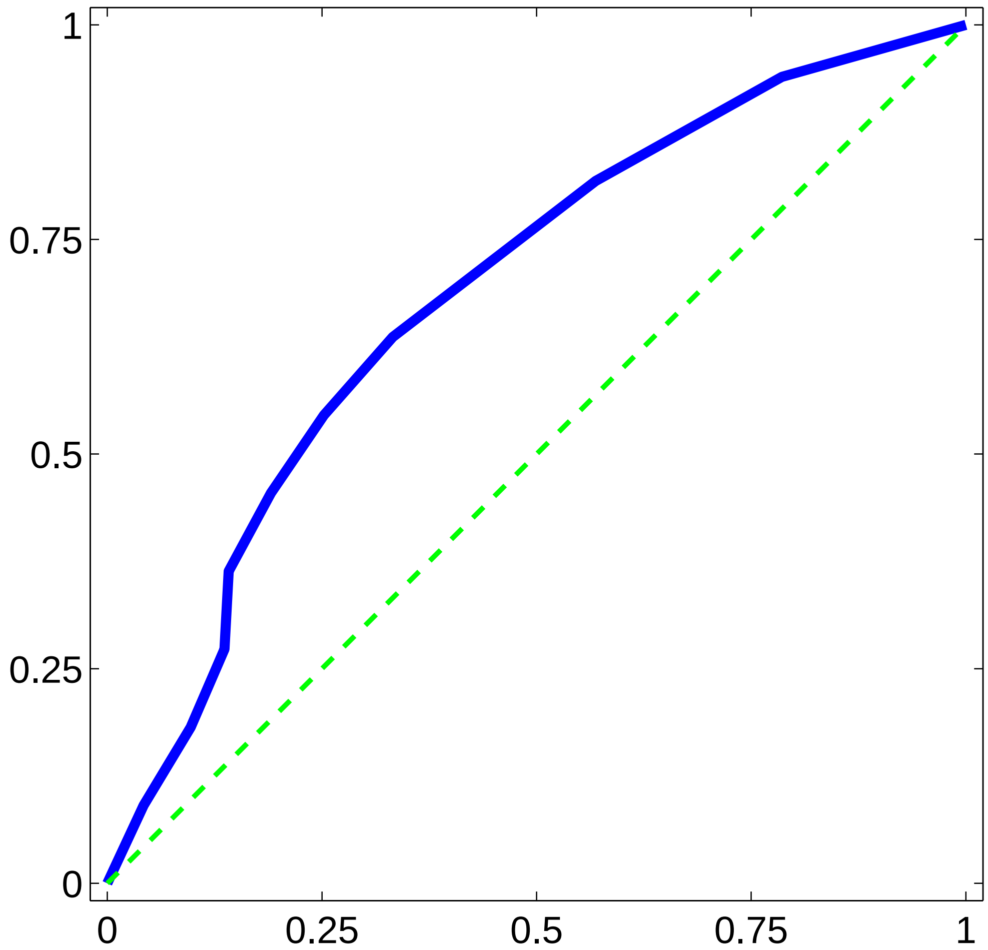
12 data points collected over 4000 min
in equi-distant intervals.

Disequilibrium

12 data points collected over 500 min
in equi-distant intervals.

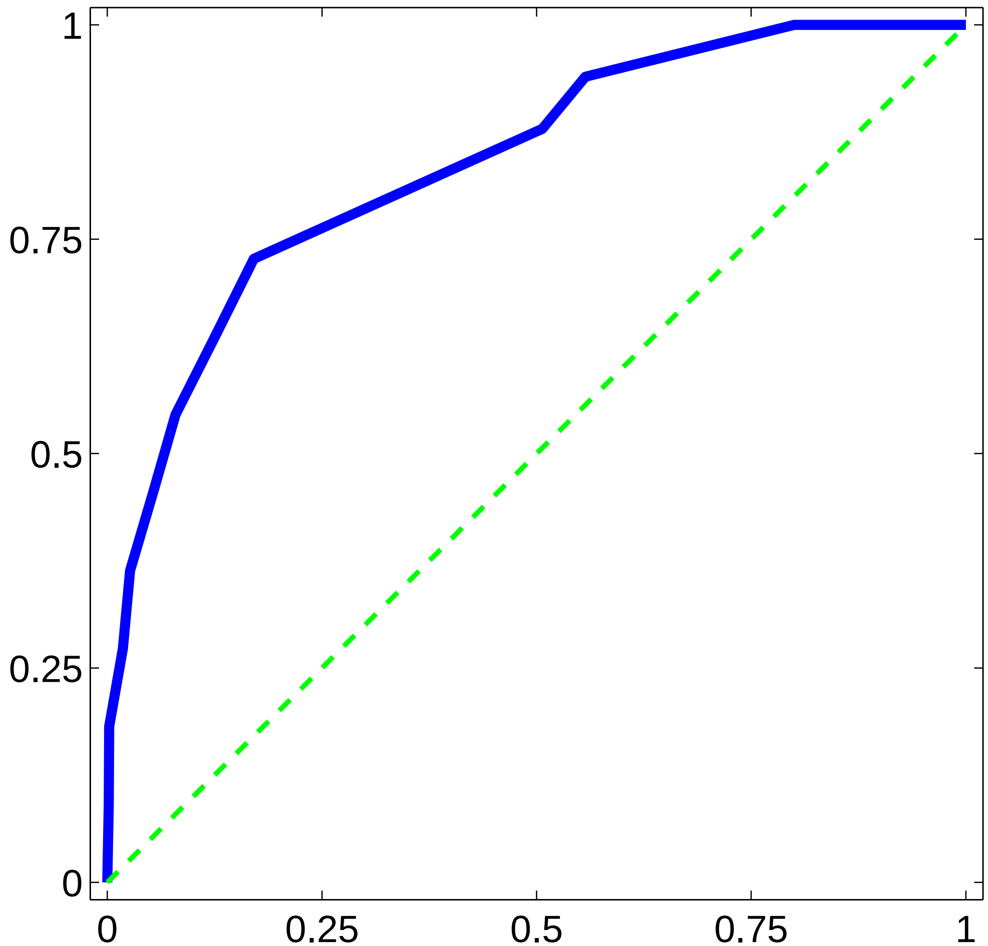
ROC curve: Equilibrium

True positives (vertical axis) $\longleftrightarrow$ False positives (horizontal axis)

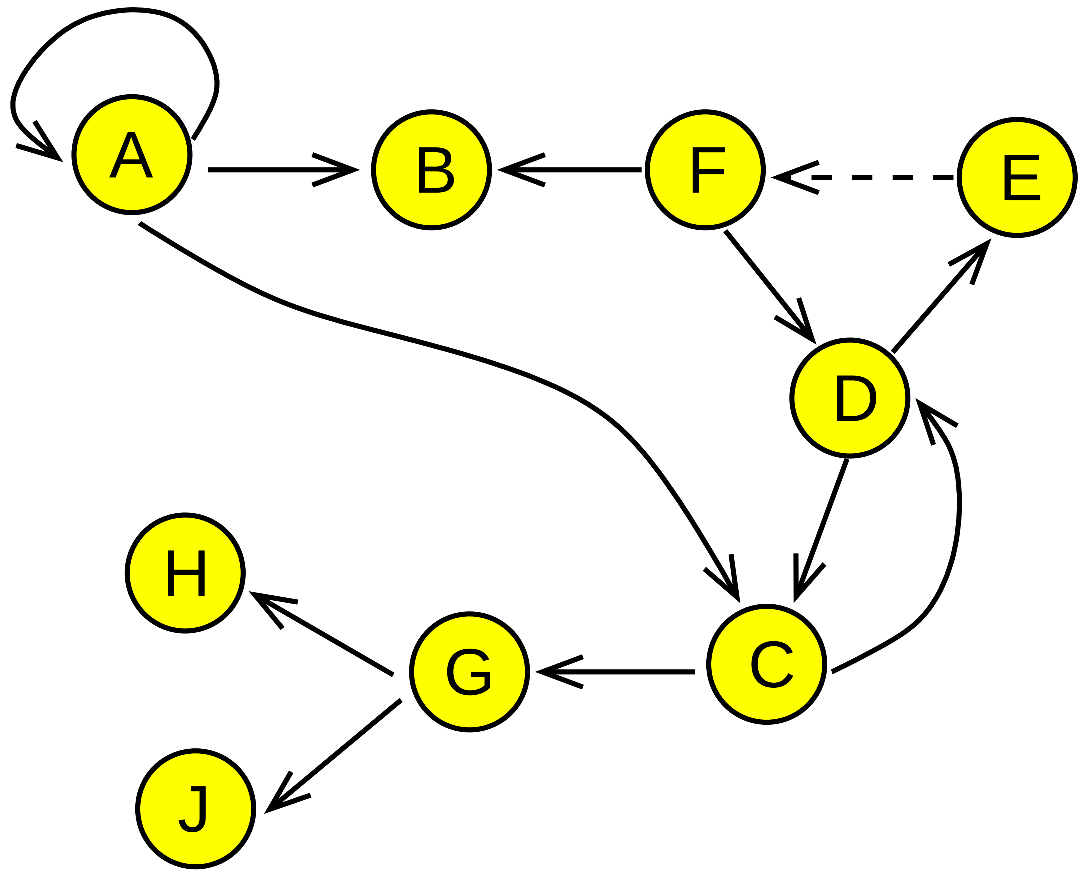


ROC curve: Disequilibrium

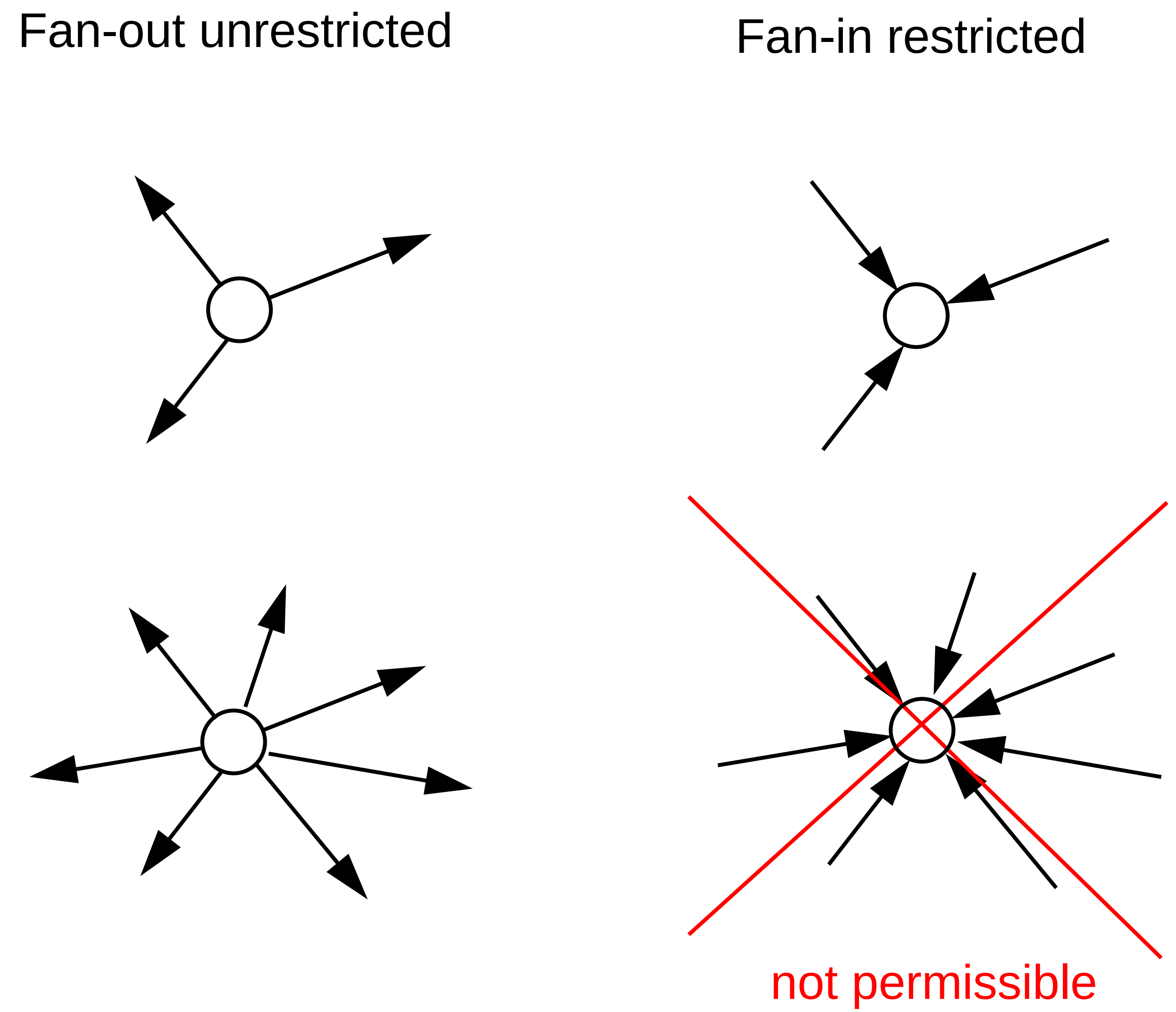
True positives (vertical axis) $\longleftrightarrow$ False positives (horizontal axis)



Structure Prior



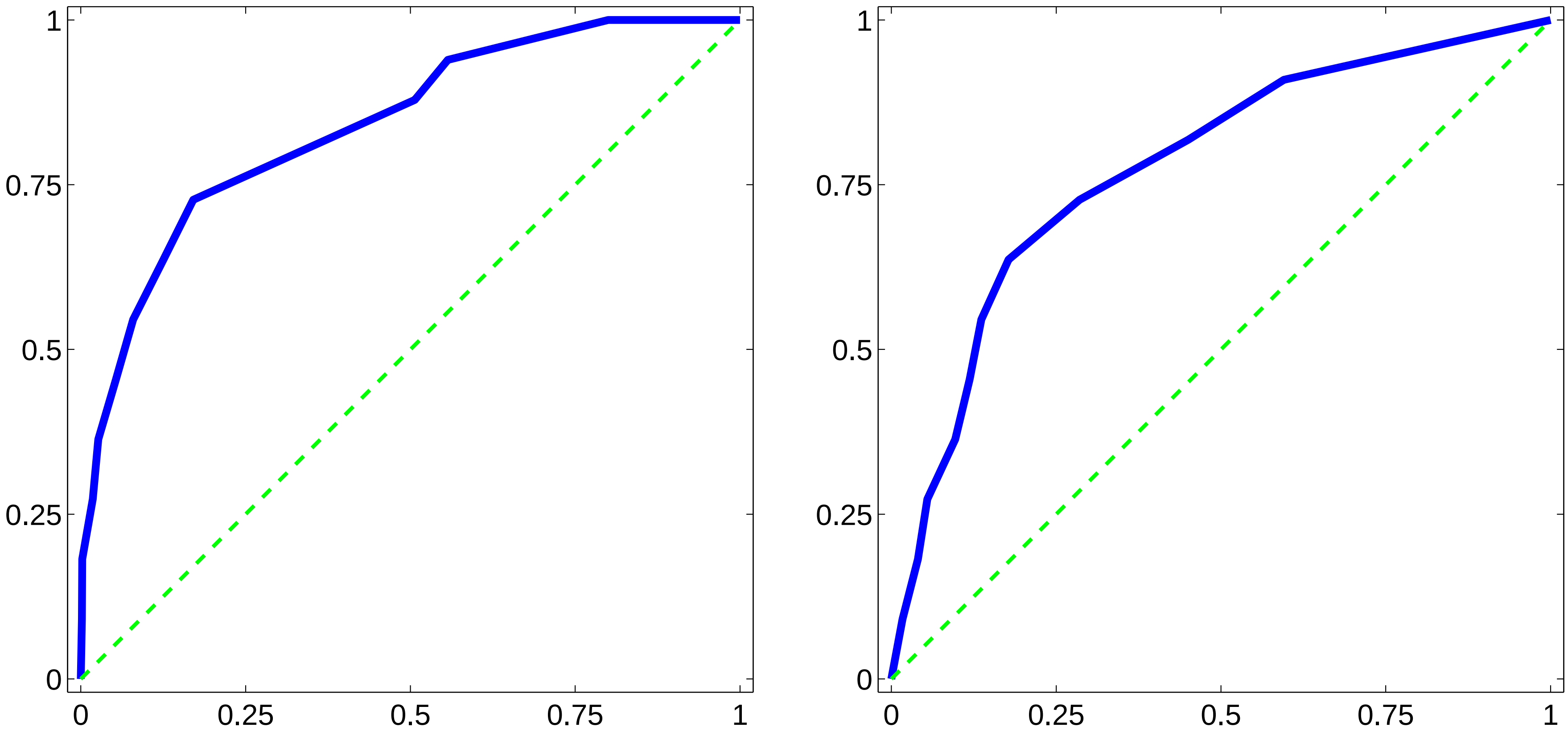
Max fan-in = 2, 3, 4


Fan-out unrestricted
Fan-in restricted
not permissible

ROC curves

True positives (vertical axis) $\longleftrightarrow$ False positives (horizontal axis)

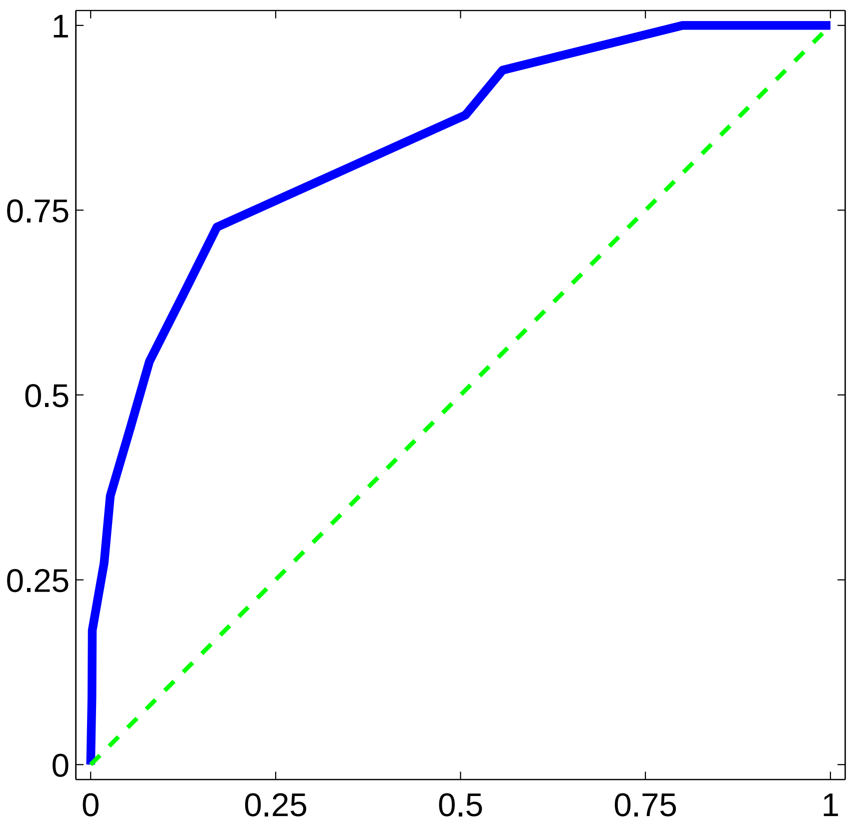
Left: max fan-in $= 2$ Right: max fan-in $= 3$

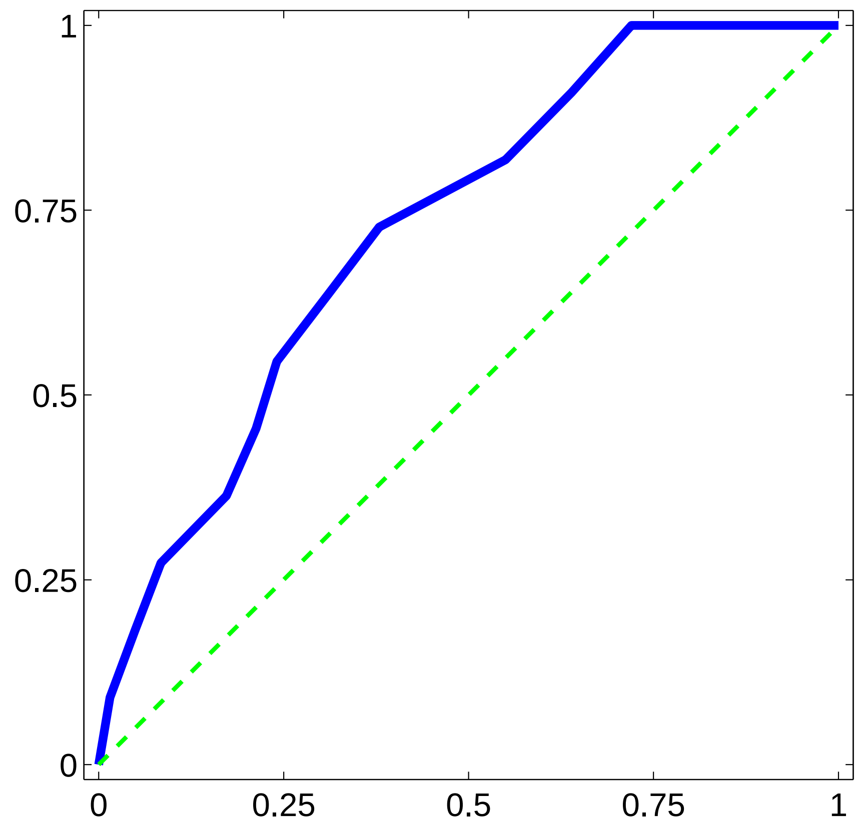


ROC curves

True positives (vertical axis) $\longleftrightarrow$ False positives (horizontal axis)

Left: max fan-in $= 2$ Right: max fan-in $= 4$





Sequence information

$$\frac{P(y \rightarrow rX | r \in B[y])}{P(y \rightarrow rX | r \notin B[y])} = 2$$

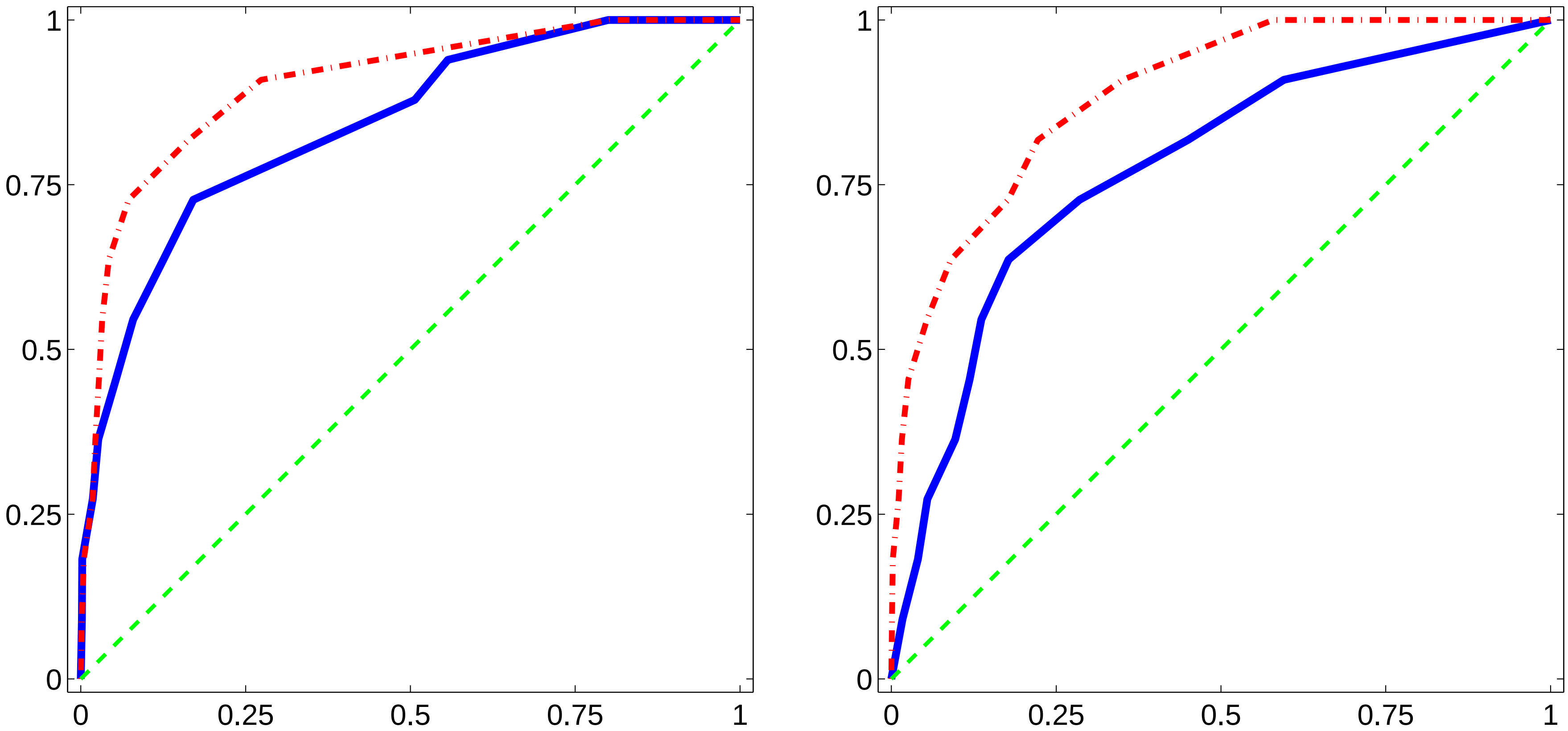
$y \rightarrow rX$ denotes the event that transcription factor y binds to the promoter r upstream of gene X, and $B[y]$ is the set of (known) binding motifs for y.

In words: The equation expresses that on identifying a binding motif for transcription factor y in the upstream region of gene X, this transcription factor is twice as likely to bind to X than in the absence of such a motif.

ROC curves

True positives (vertical axis) $\longleftrightarrow$ False positives (horizontal axis)

Left: max fan-in $= 2$ Right: max fan-in $= 3$



Conclusions

Conclusions

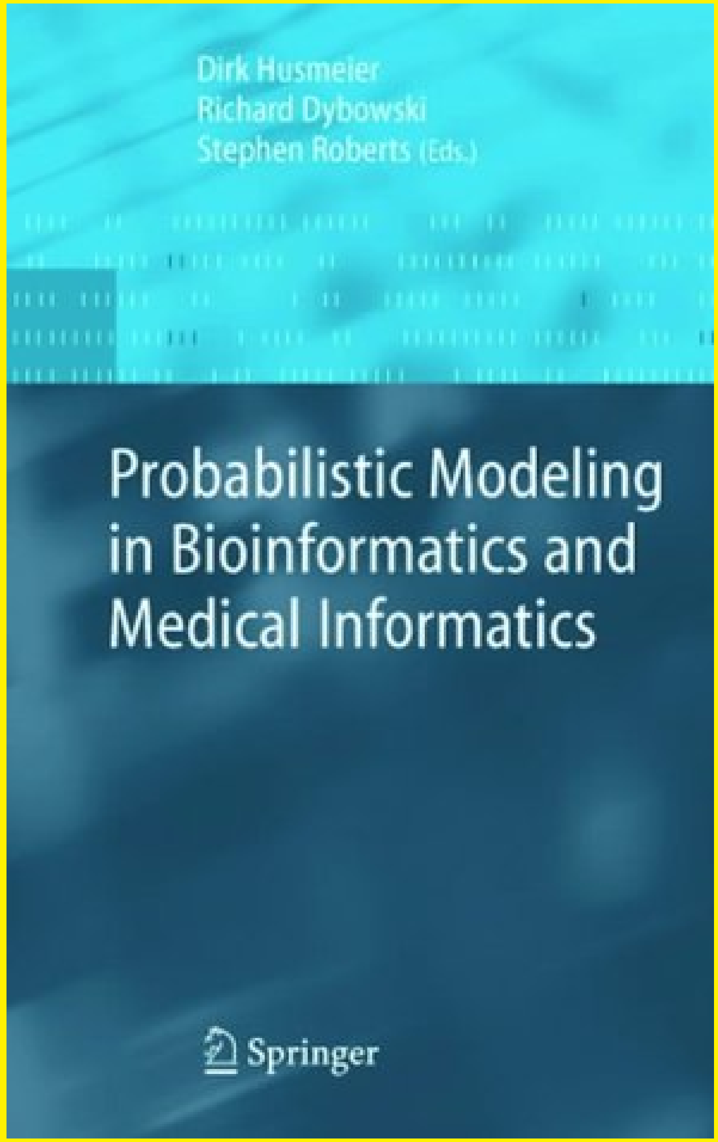
- Learning the global network $\longrightarrow$ impossible
- Intrinsic uncertainty due to lack of data

Conclusions

- Learning the global network $\longrightarrow$ impossible
- Intrinsic uncertainty due to lack of data
- Inference local substructures possible, but obscured by noise.

Conclusions

- Learning the global network $\longrightarrow$ impossible
- Intrinsic uncertainty due to lack of data
- Inference local substructures possible, but obscured by noise.
- Biologically realistic priors important
- Integrating post-genomic data.


Dirk Husmeier
Richard Dybowski
Stephen Roberts (Eds.)
Probabilistic Modeling in Bioinformatics and Medical Informatics
Springer