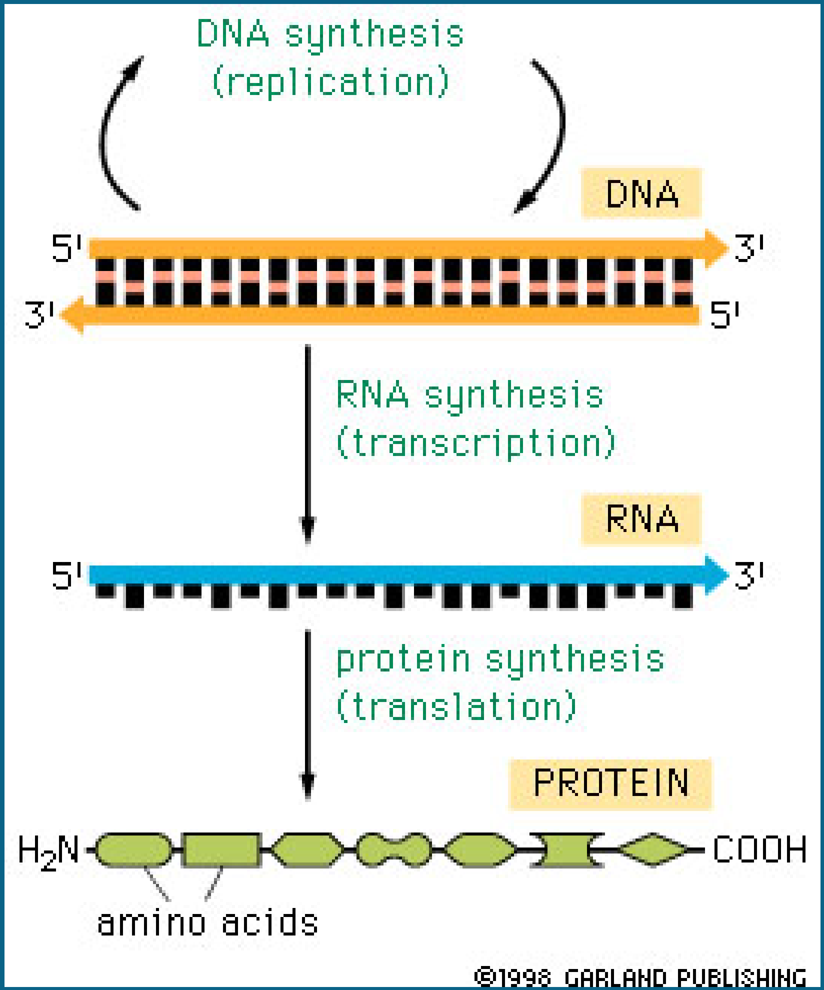
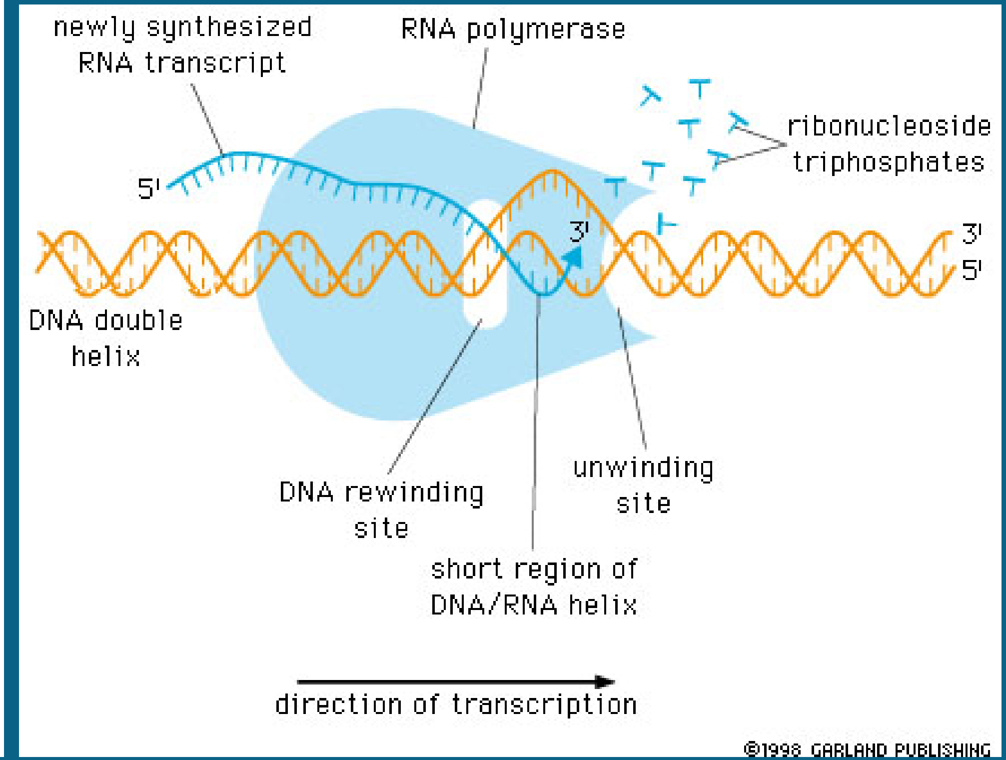
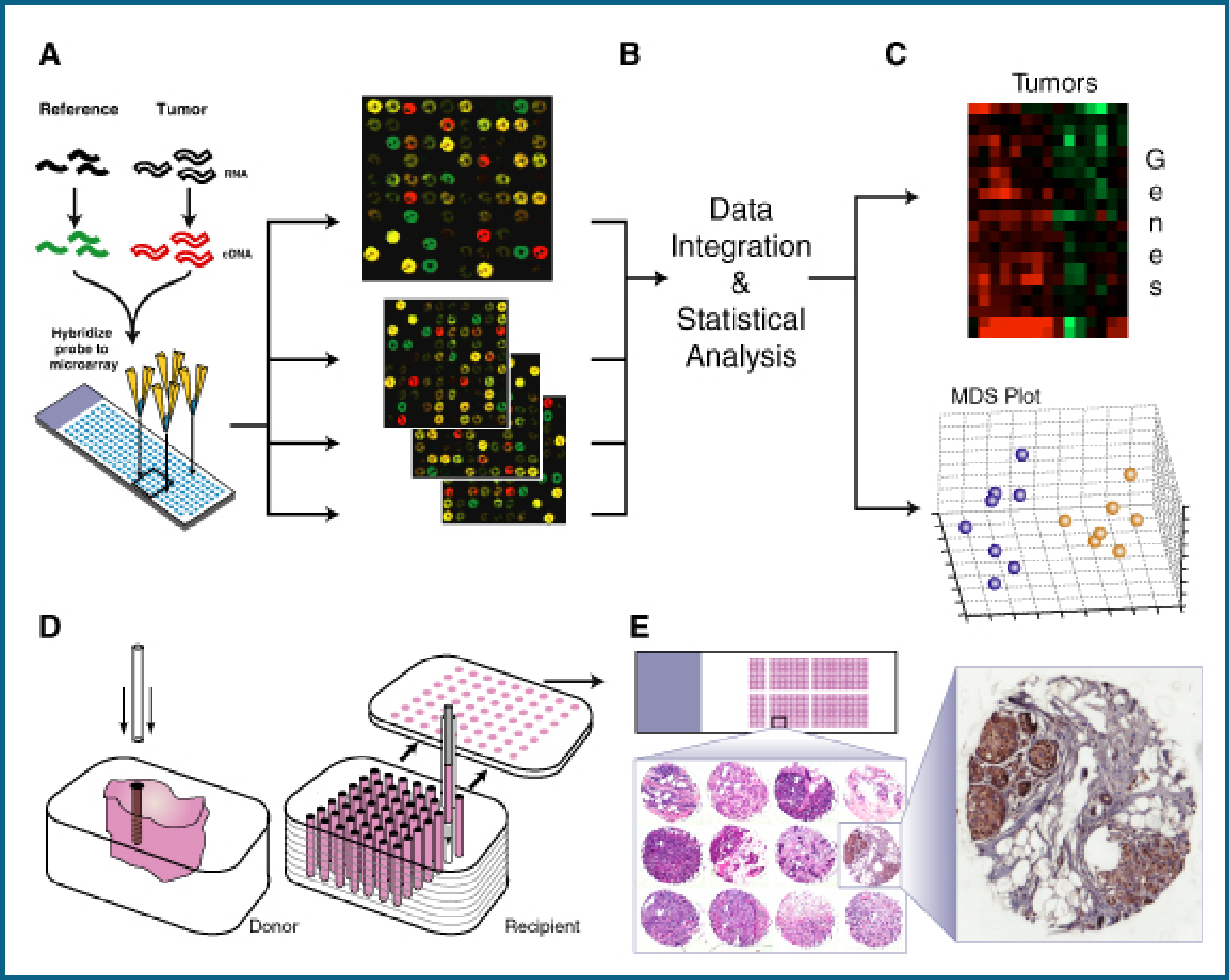
Inferring genetic networks from microarray gene expression data

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

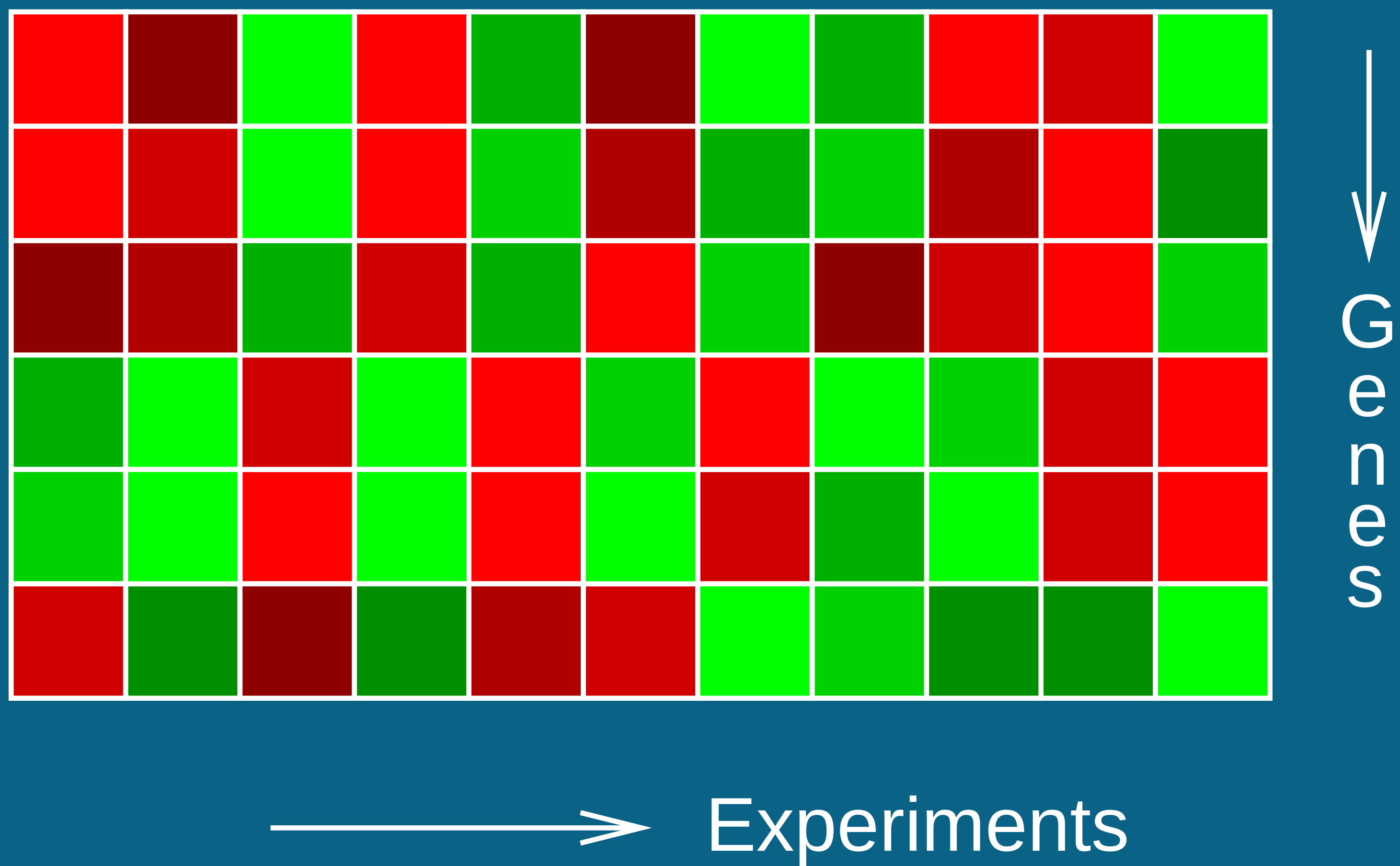


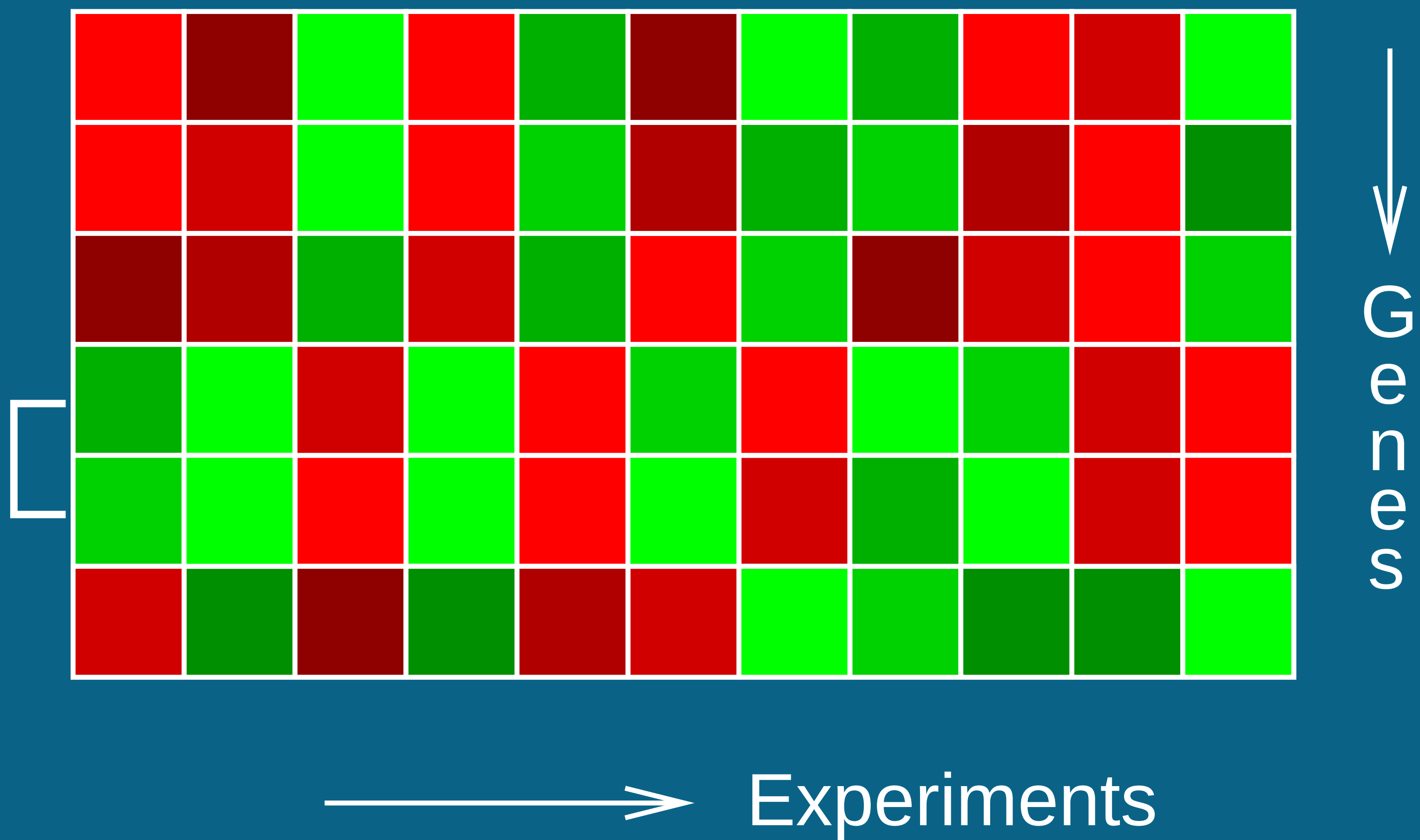


From http://www.csu.edu.au/faculty/health/biomed/subjects/molbol/images

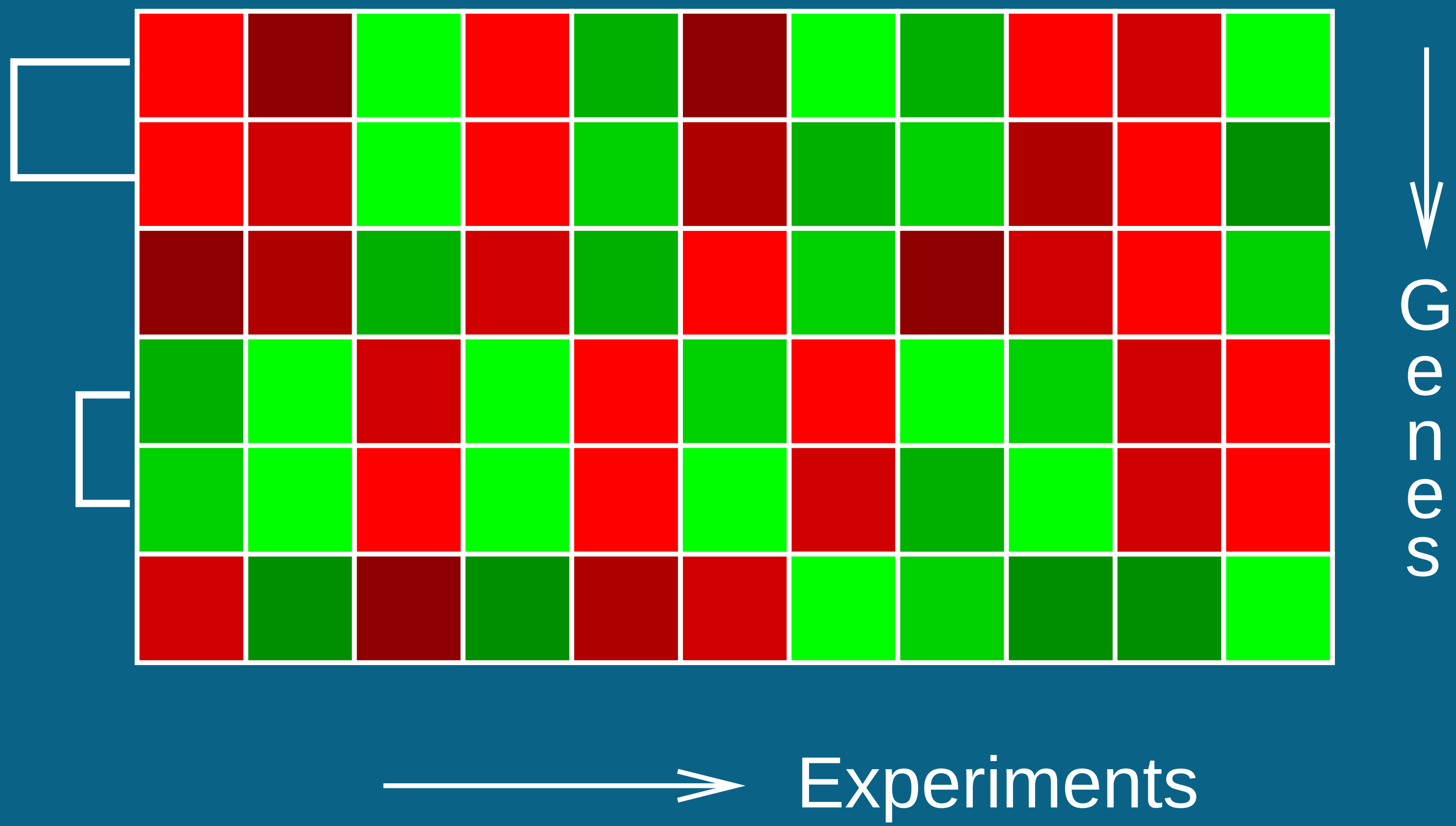


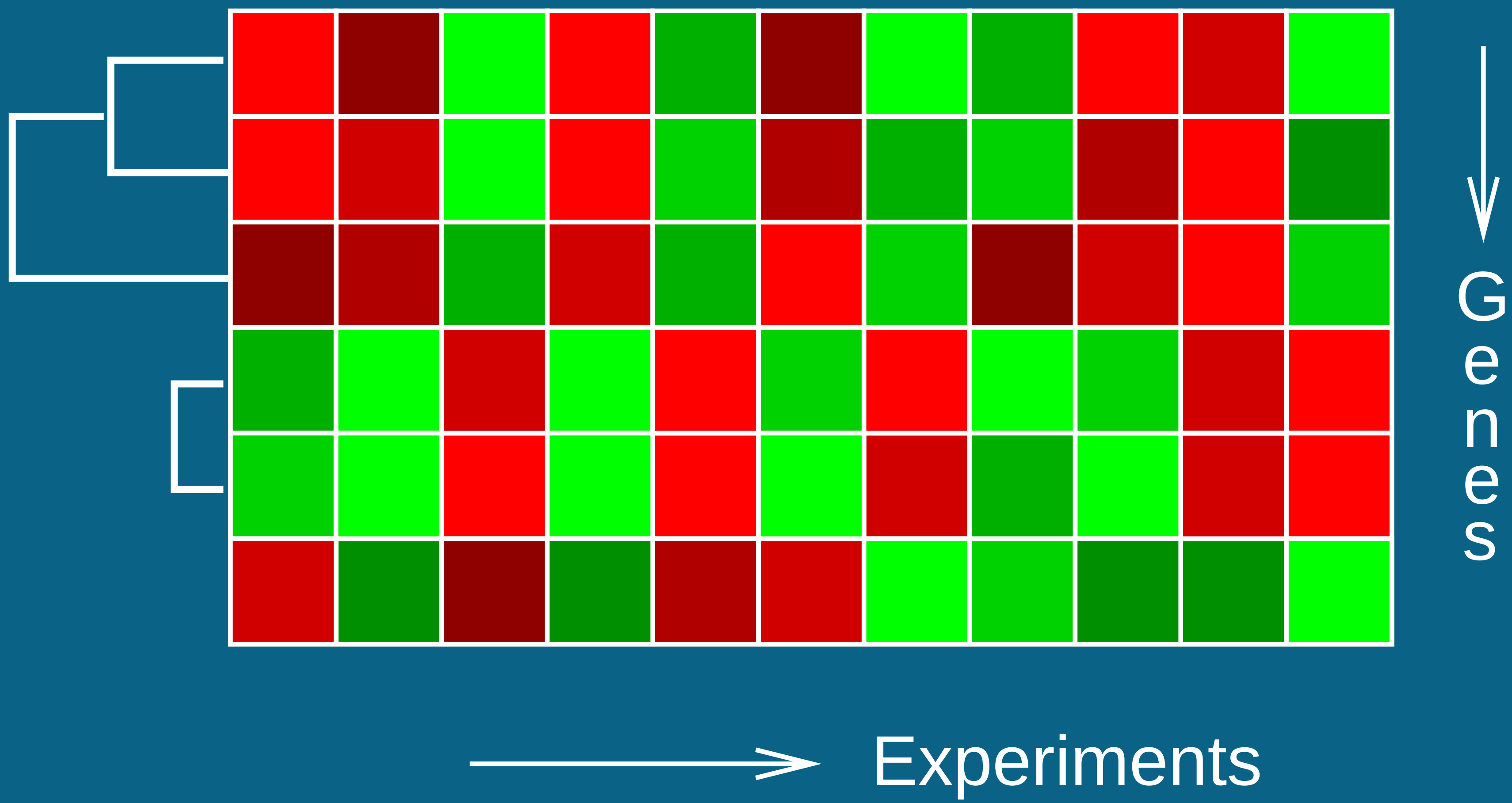
From http://www.nhgri.nih.gov/DIR/Microarray/NEJM_Supplement/

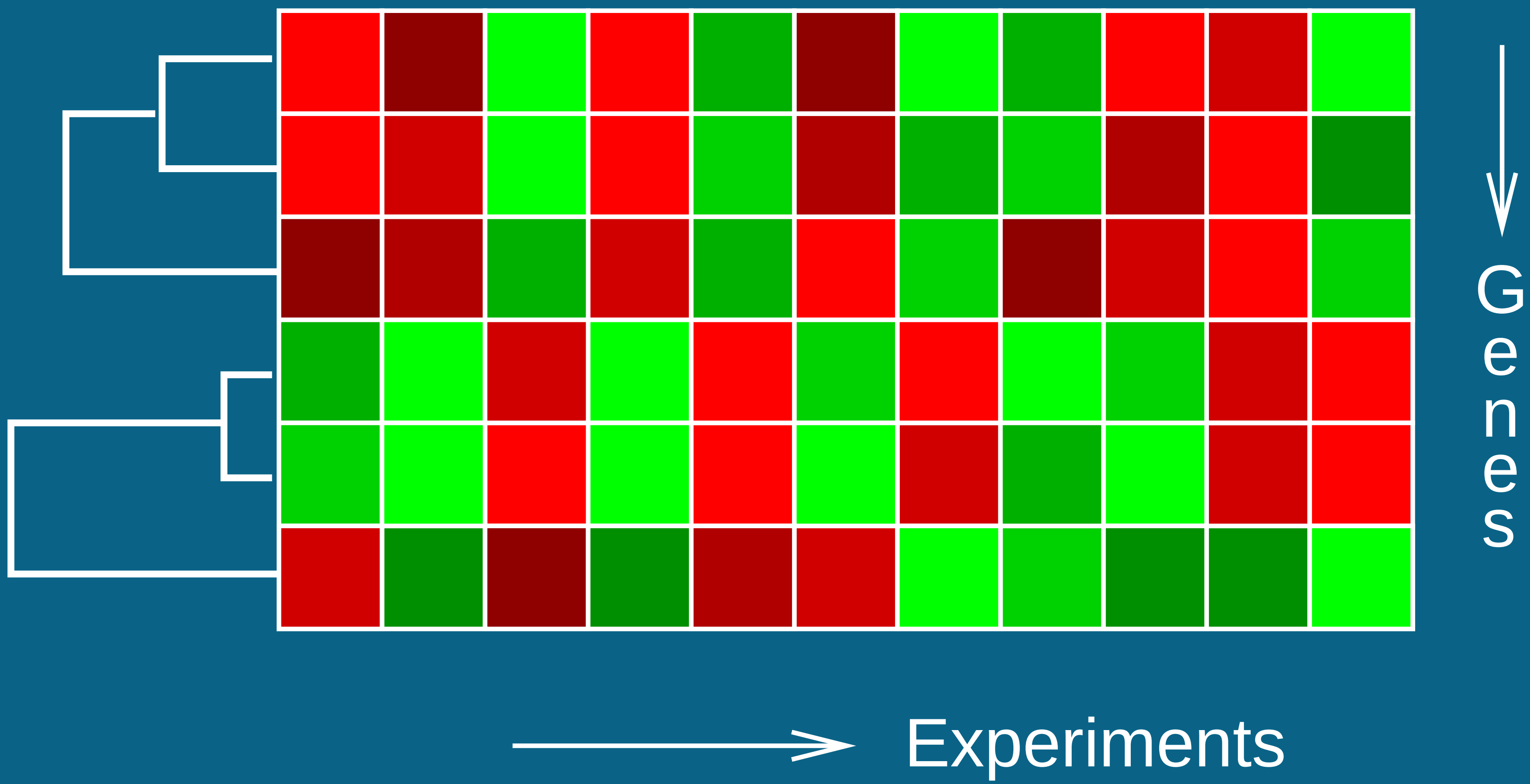

Genes
Experiments

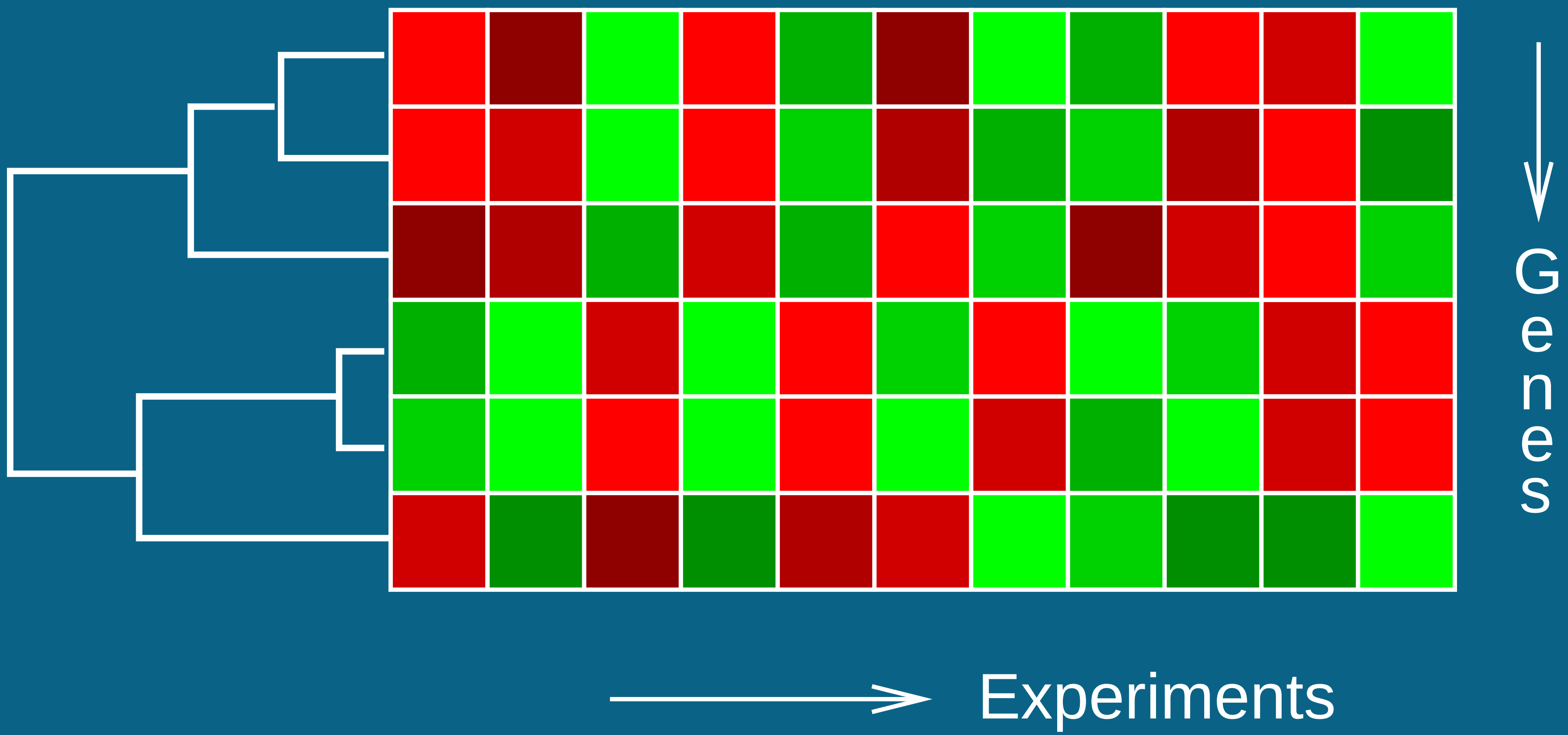


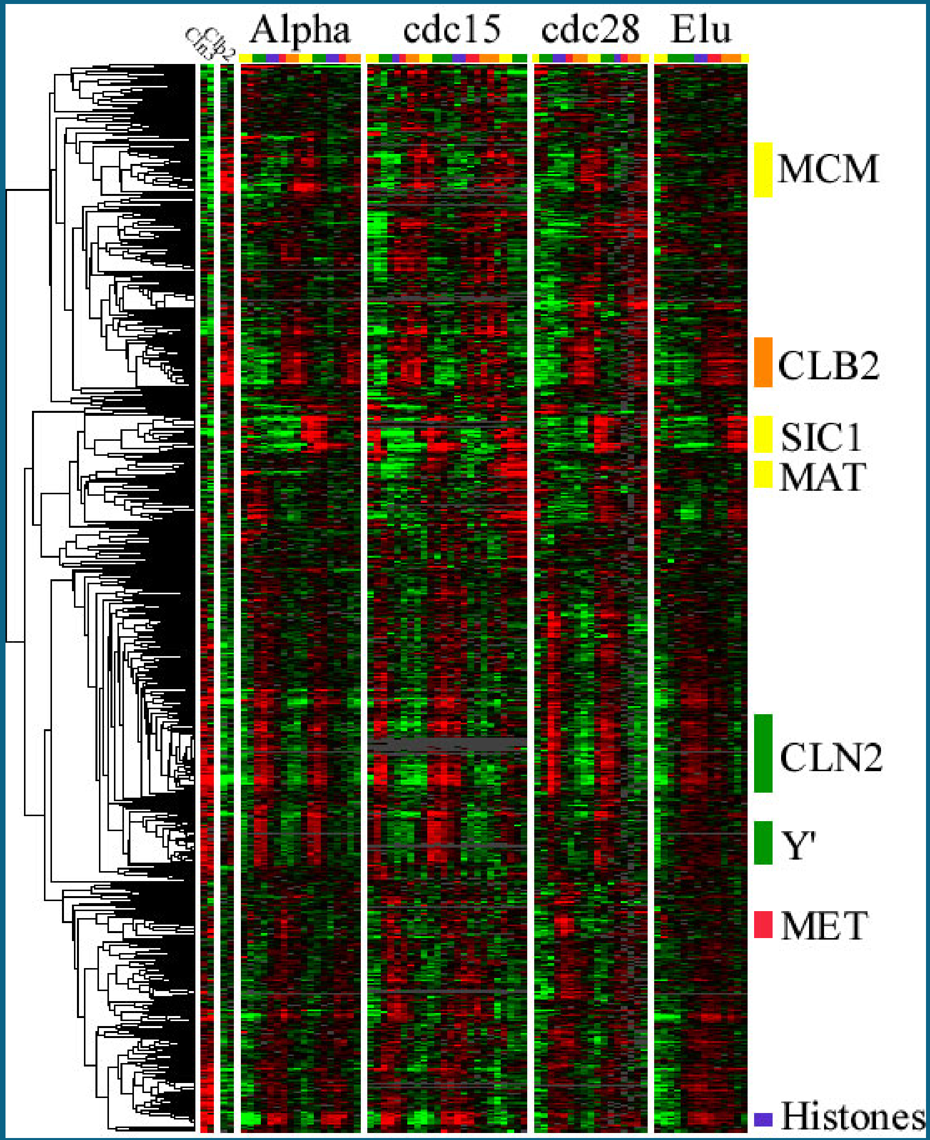
Genes
Experiments


Genes
Experiments

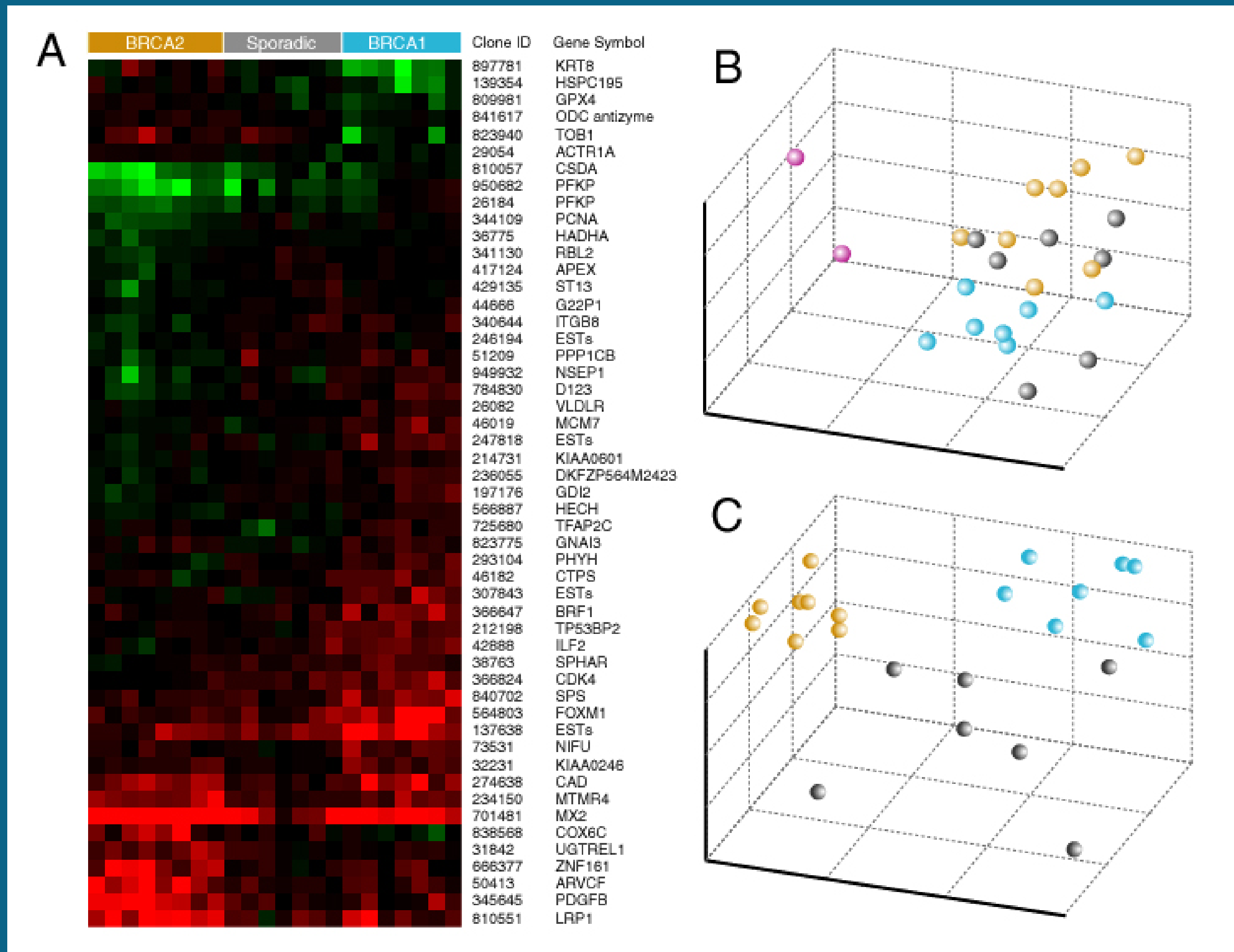

Genes
Experiments


Genes
Experiments

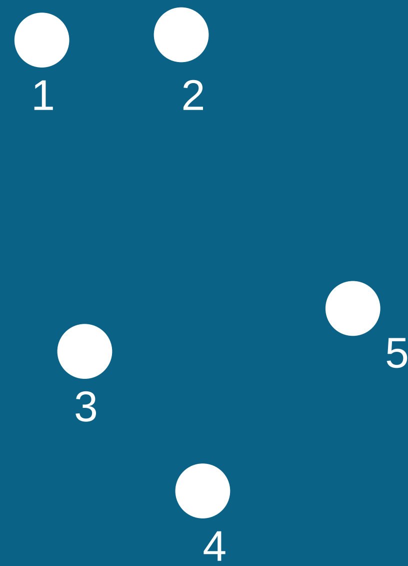

Genes
Experiments



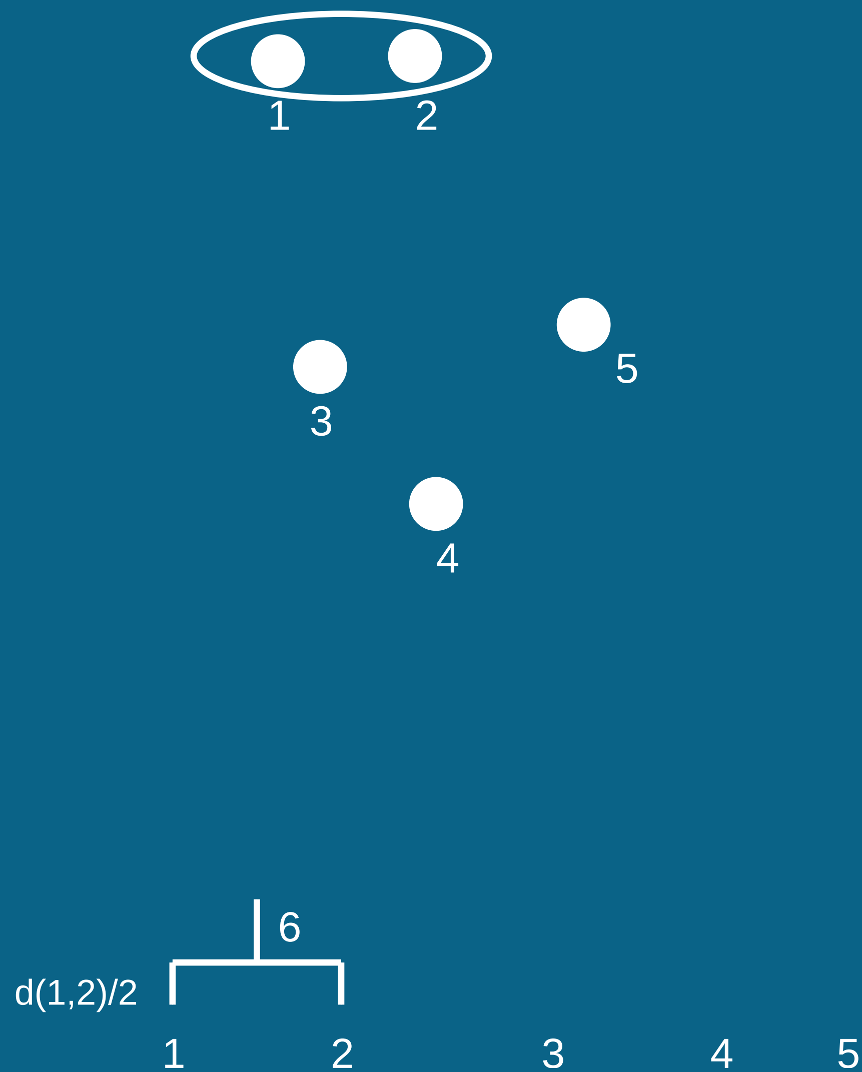
From Spellman et al., http://cellcycle-www.stanford.edu/

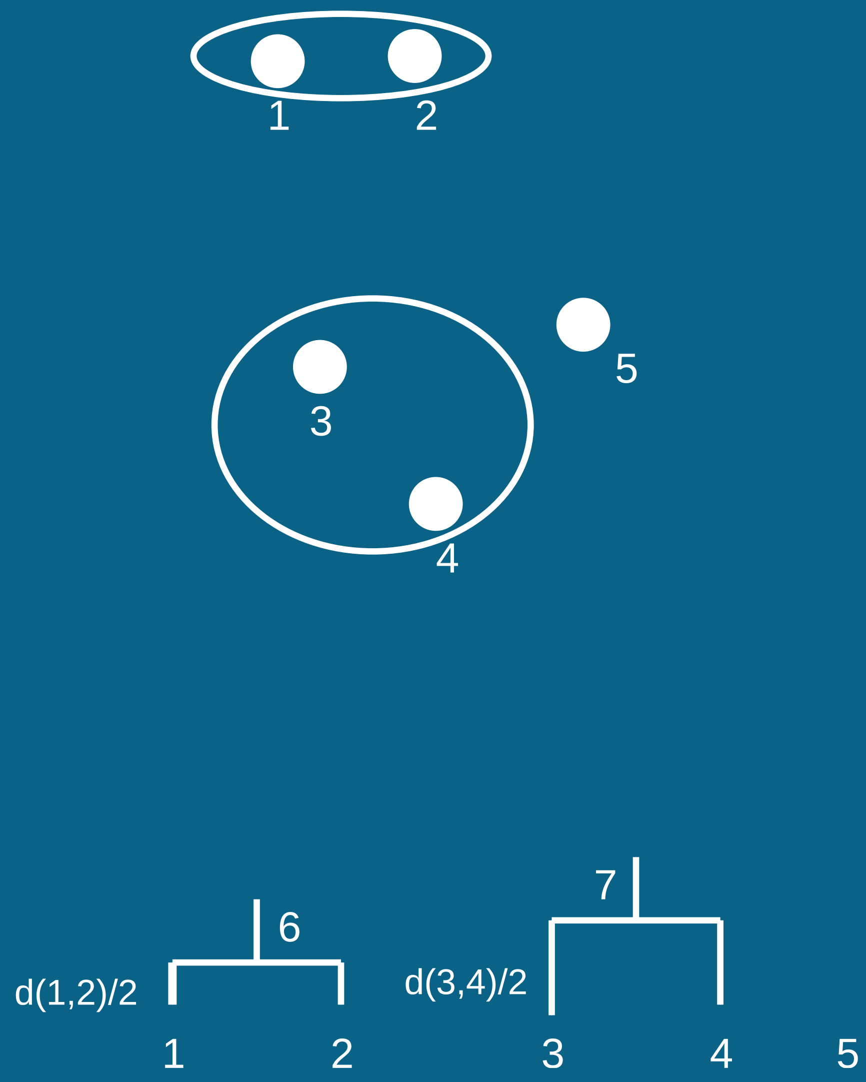


From http://www.nhgri.nih.gov/DIR/Microarray/NEJM_Supplement/

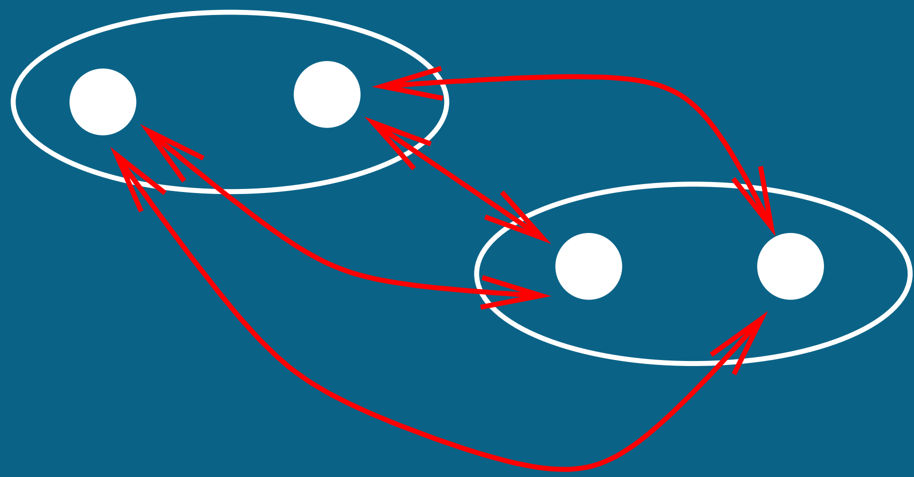


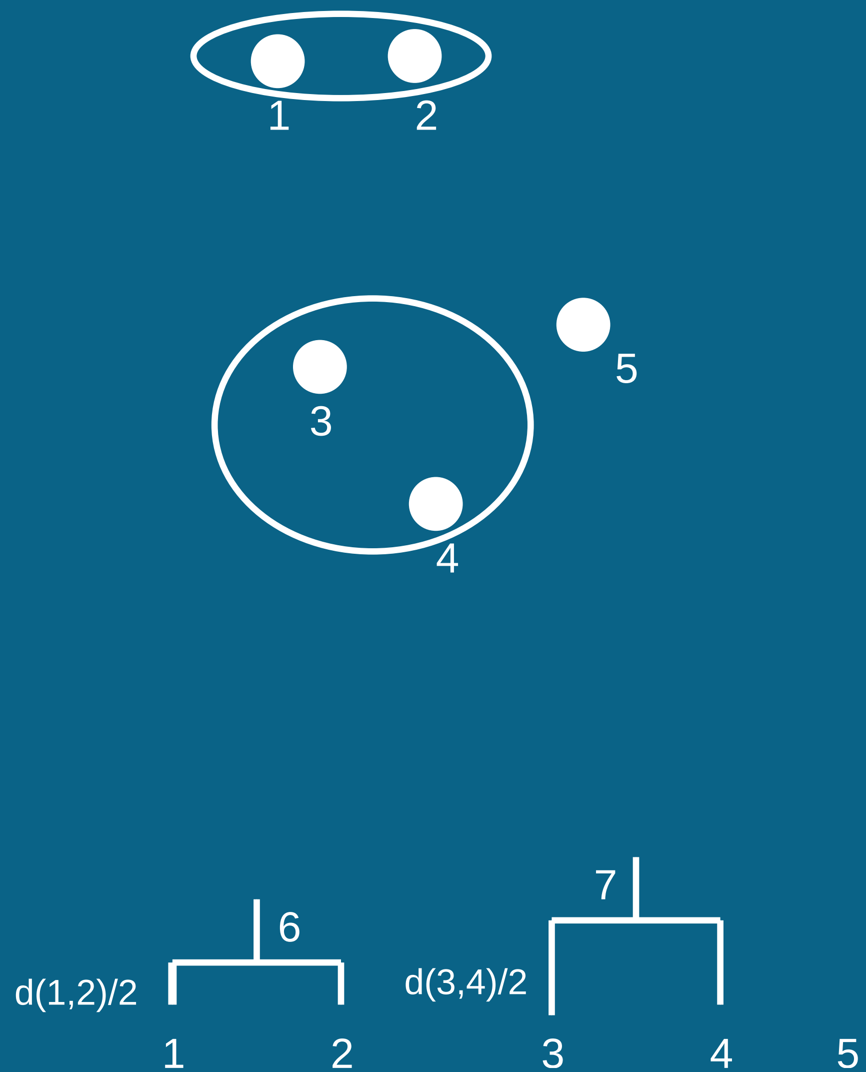
1 2 3 4 5

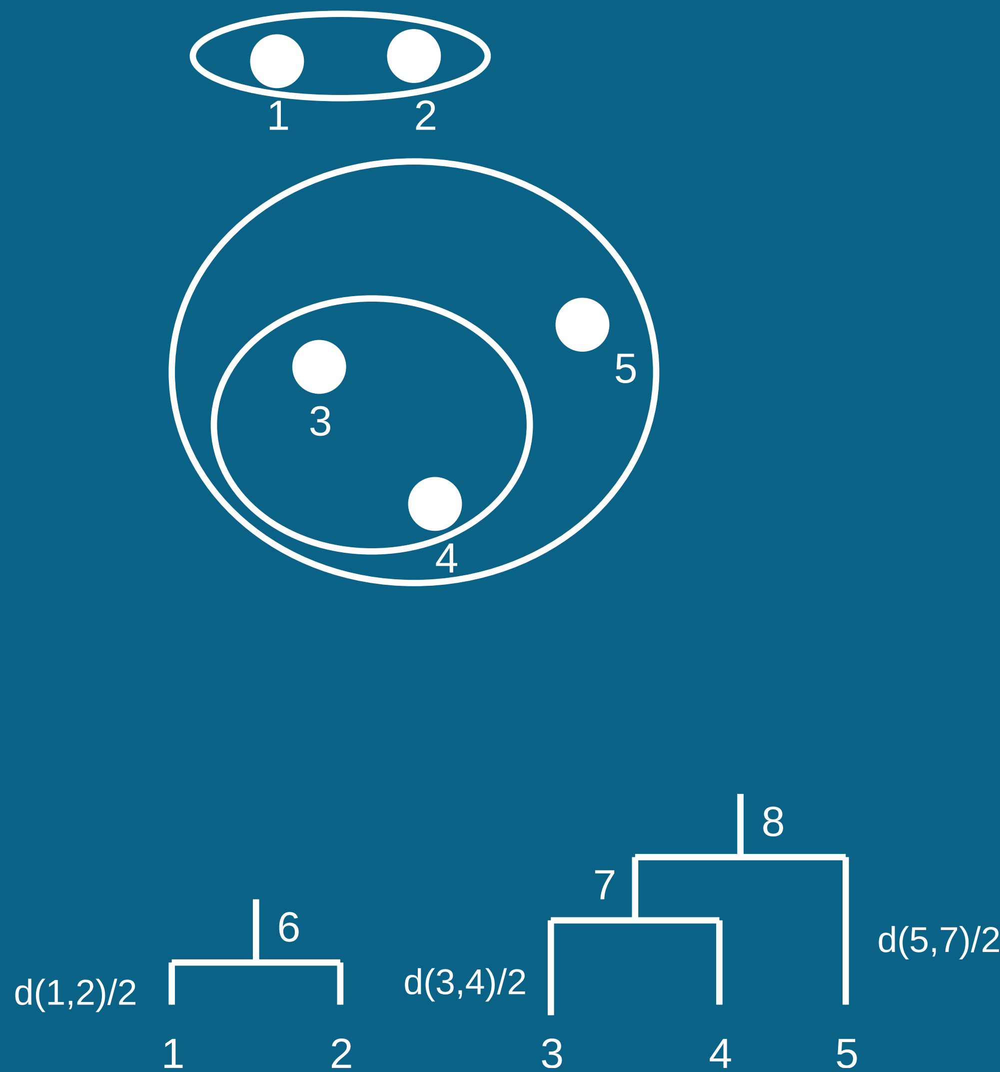

1
2
3
5
4
6
d(1,2)/2
1
2
3
4
5

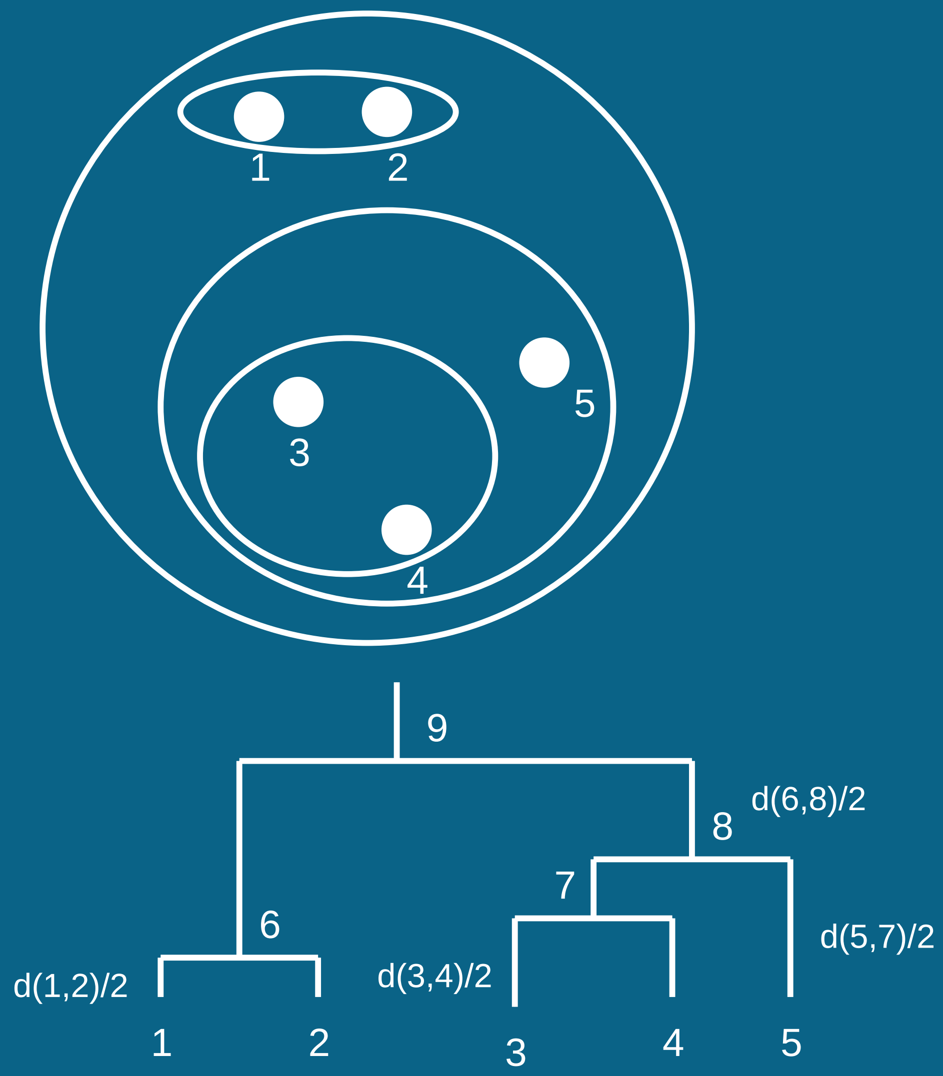

1
2
5
3
4
7
6
d(1,2)/2
d(3,4)/2
1
2
3
4
5

Distance between clusters

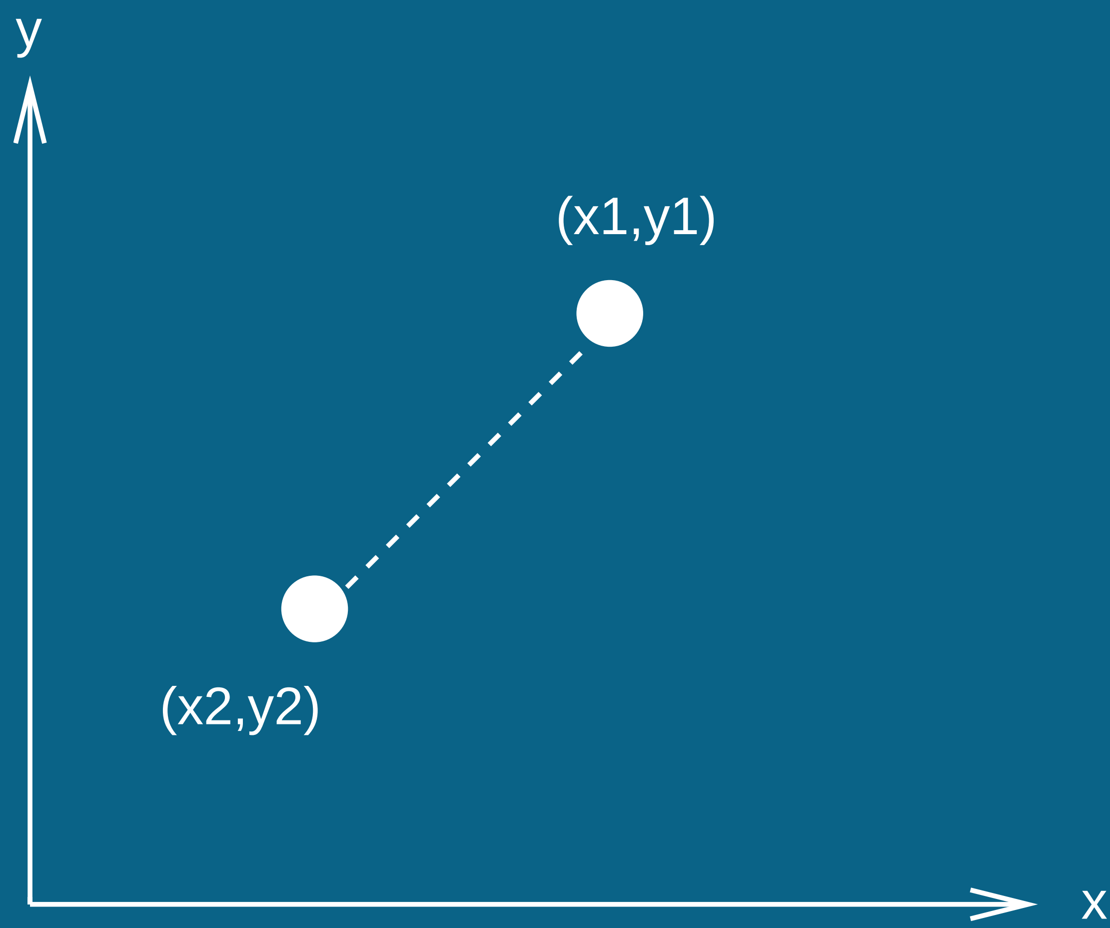



1
2
5
3
4
7
6
d(1,2)/2
d(3,4)/2
1
2
3
4
5

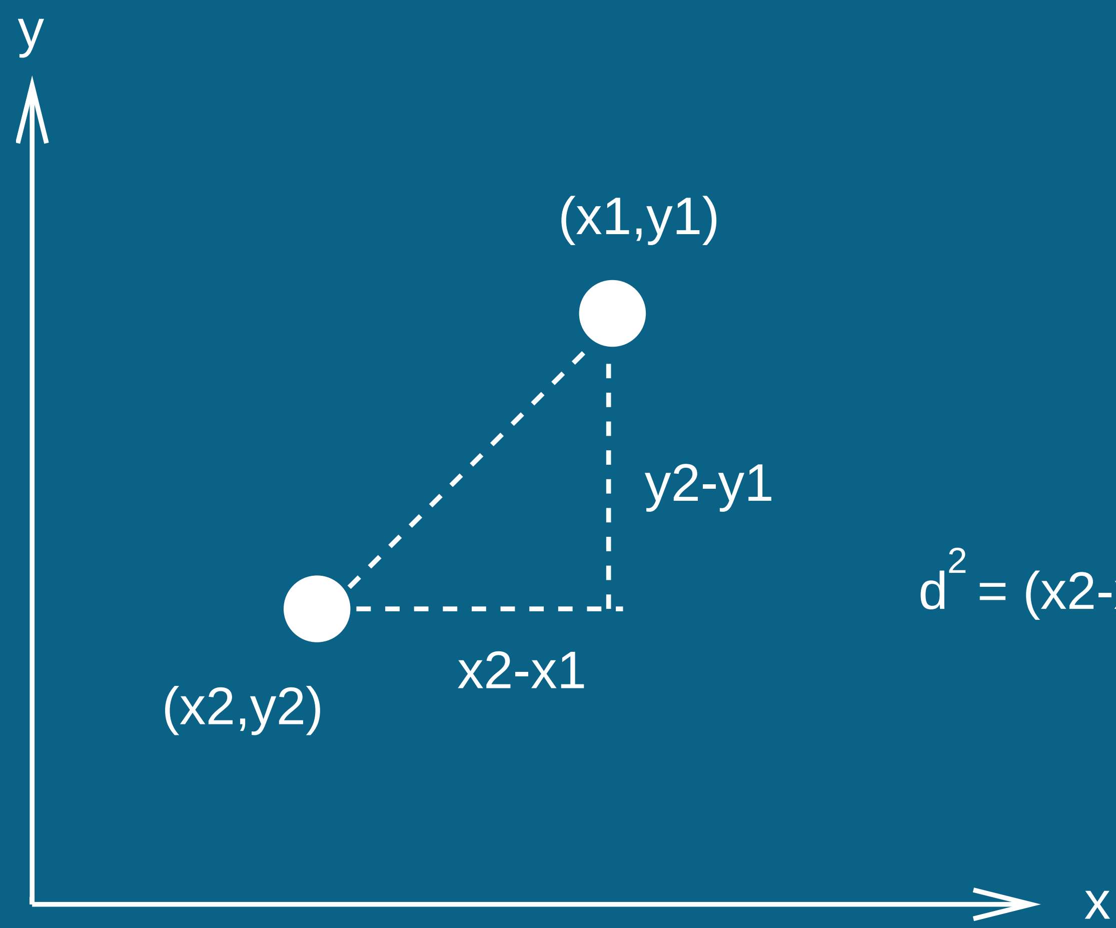

1
2
5
3
4
8
7
6
d(5,7)/2
d(3,4)/2
d(1,2)/2
1
2
3
4
5


1
2
5
3
4
9
d(6,8)/2
8
7
6
d(5,7)/2
d(1,2)/2
d(3,4)/2
1
2
3
4
5

Euclidean distance

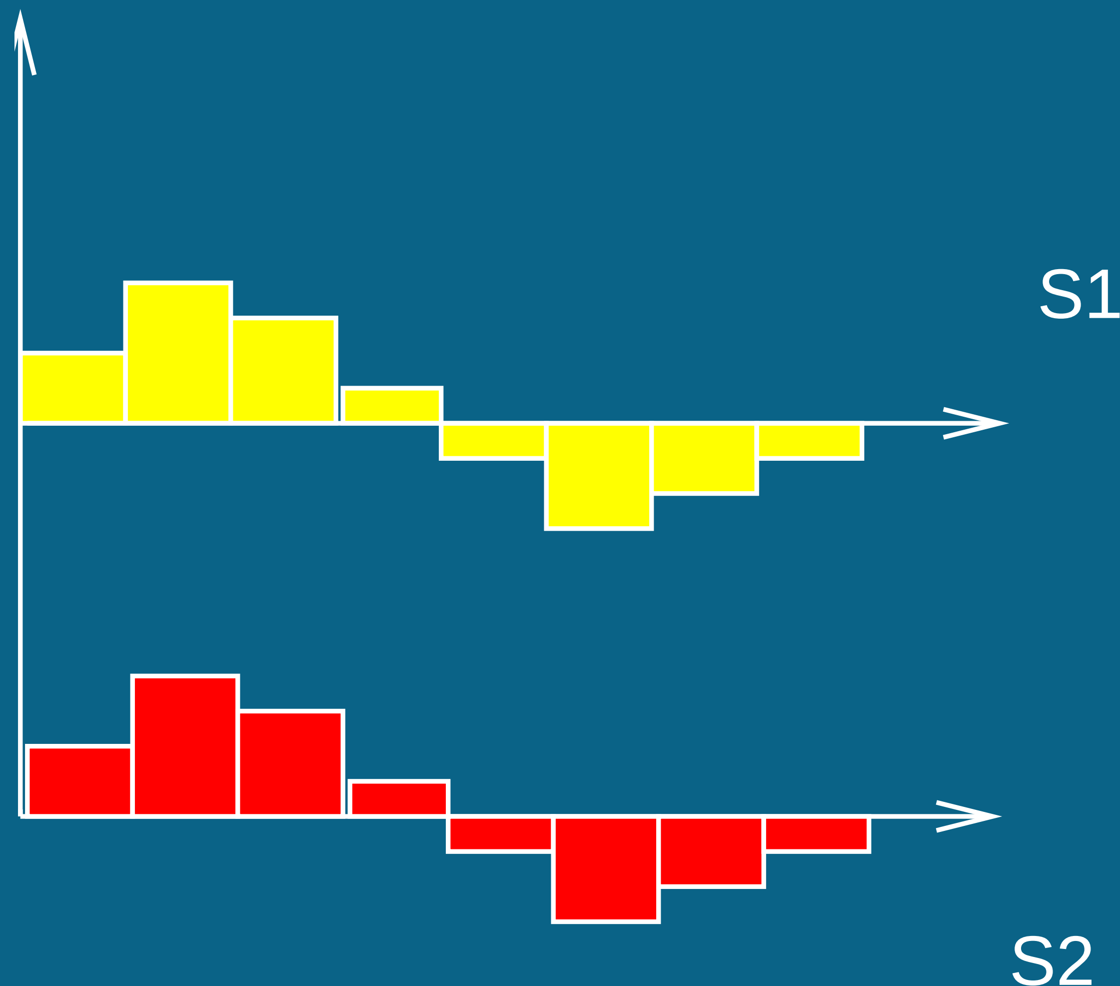


Euclidean distance



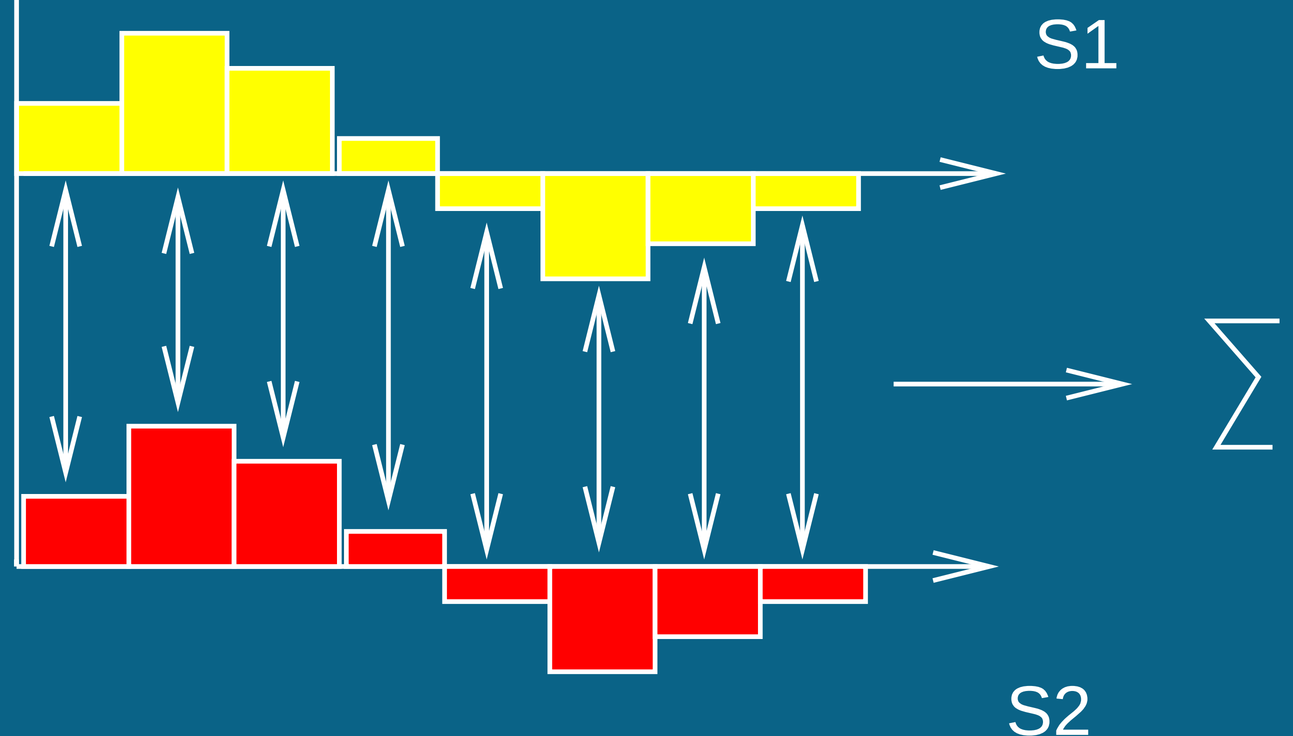
$$d^2 = (x2-x1)^2 + (y2-y1)^2$$

Correlation

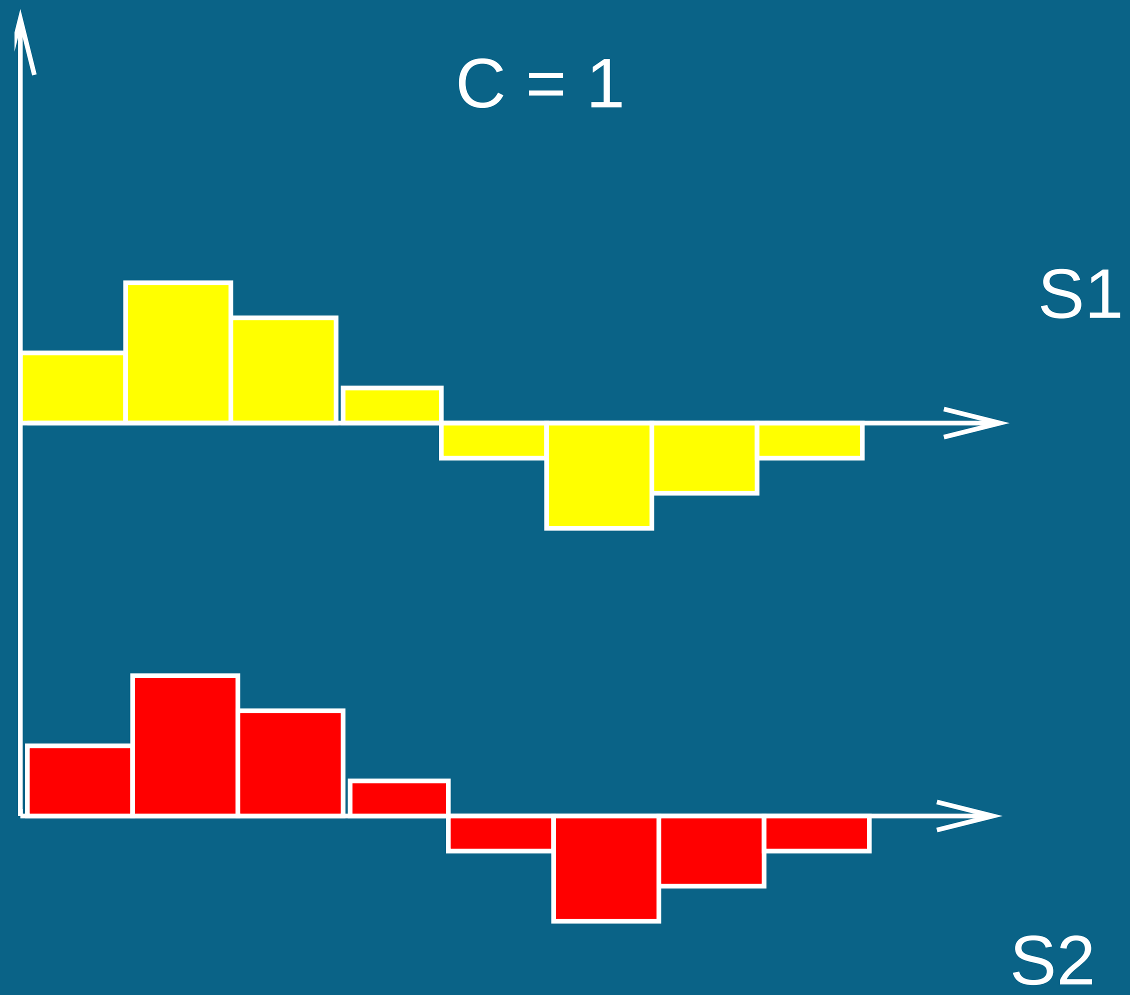


Correlation

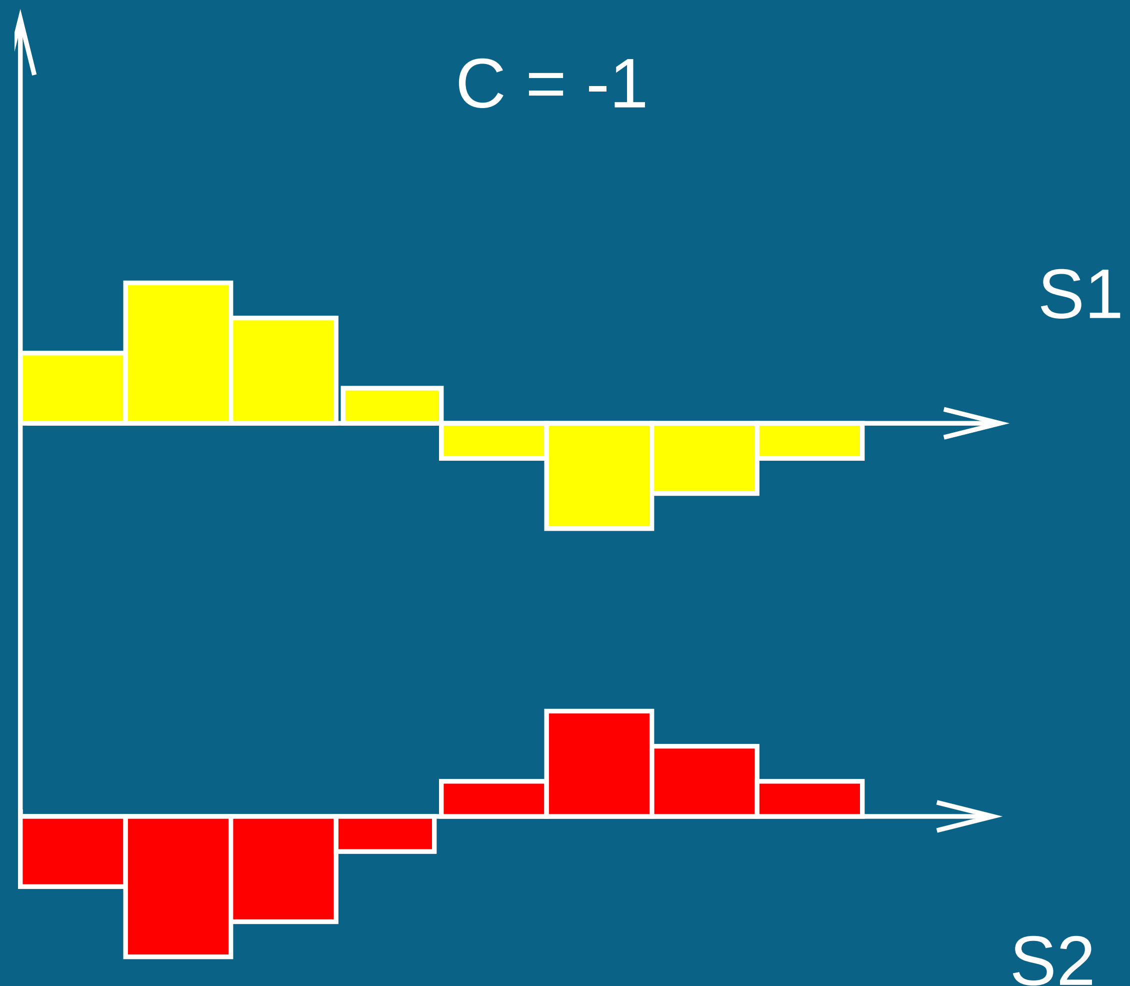
$$C = \frac{\langle S1-\mathrm{mean}(S1) | S2-\mathrm{mean}(S2) \rangle}{||(S1-\mathrm{mean}(S1)|| \; ||S2-\mathrm{mean}(S2)||}$$



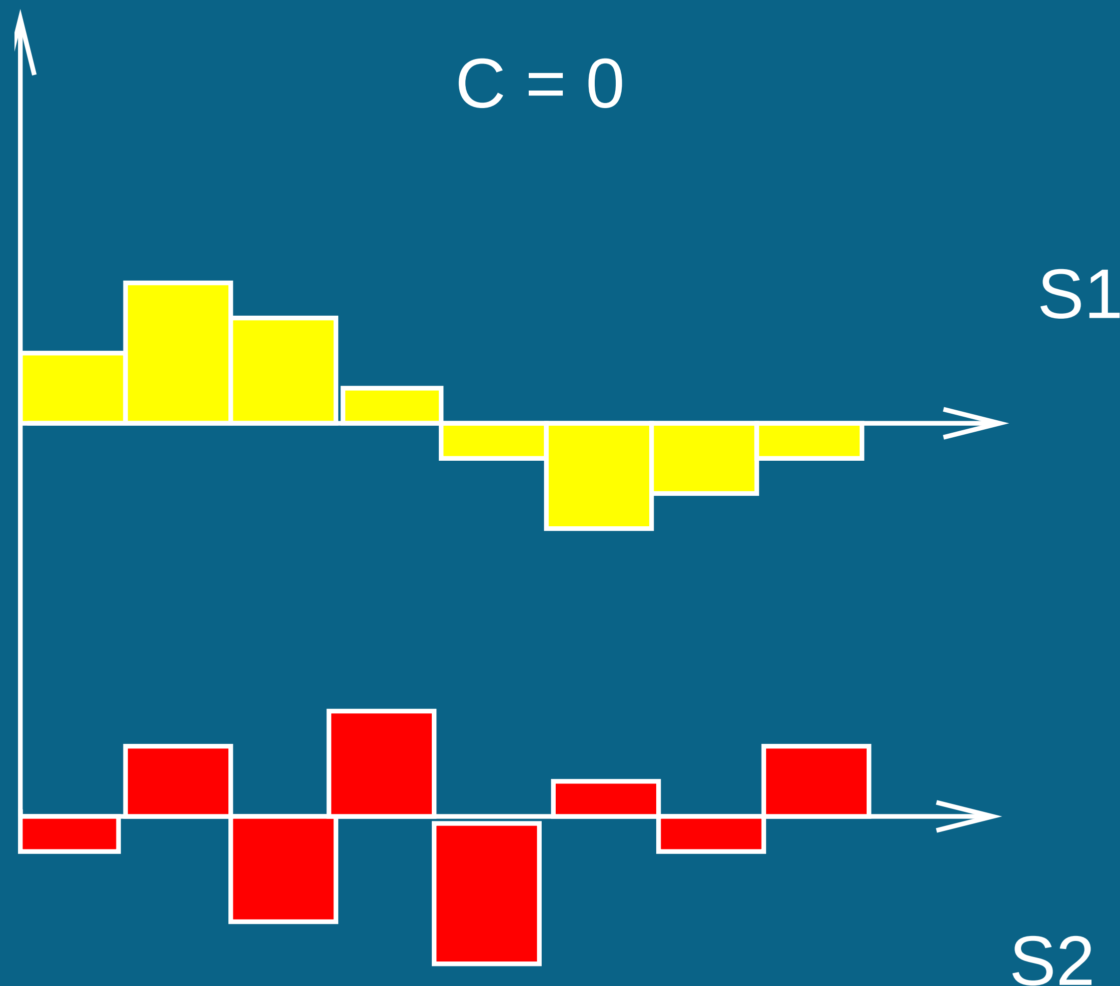
Correlation



Correlation

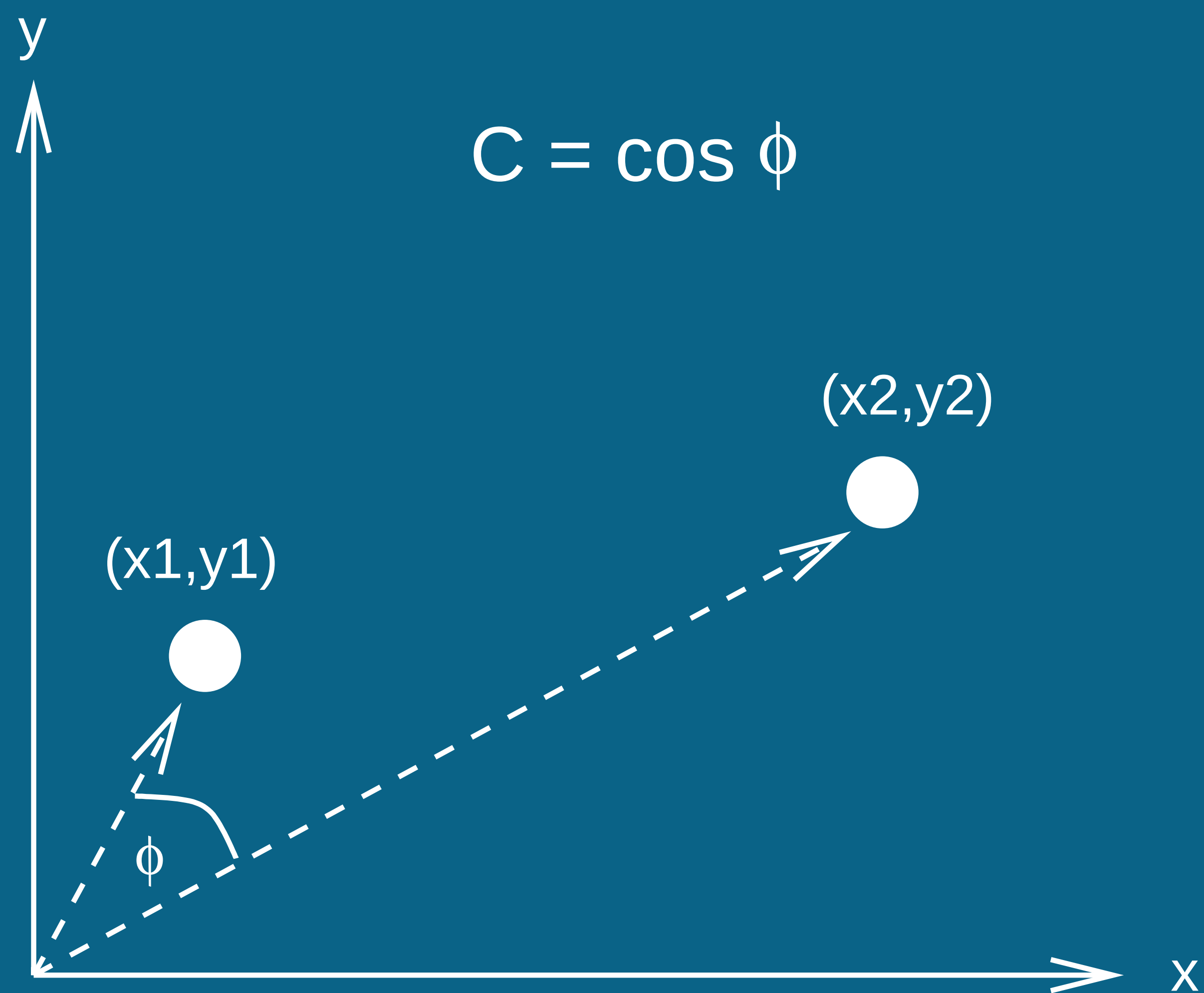


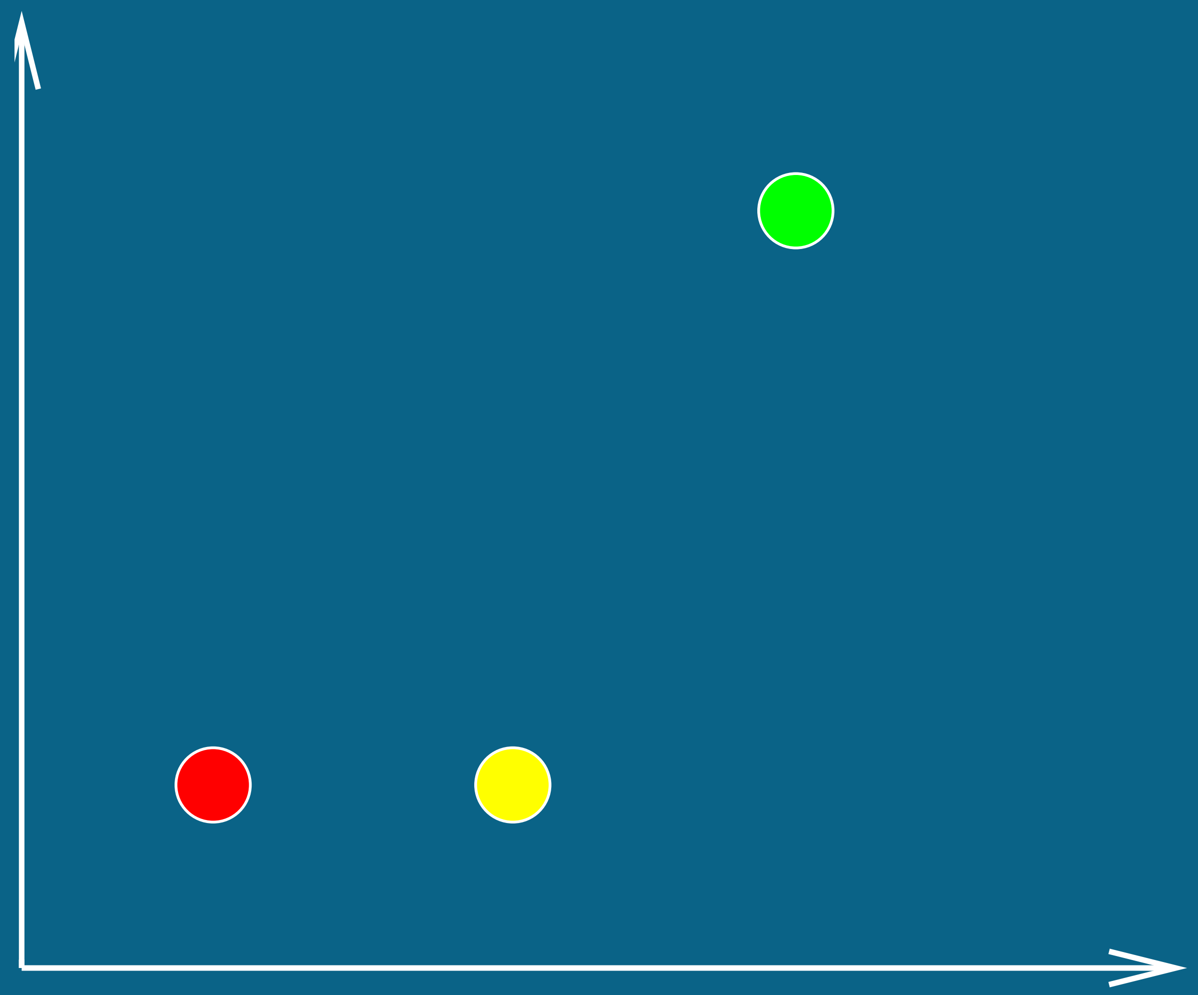
Correlation



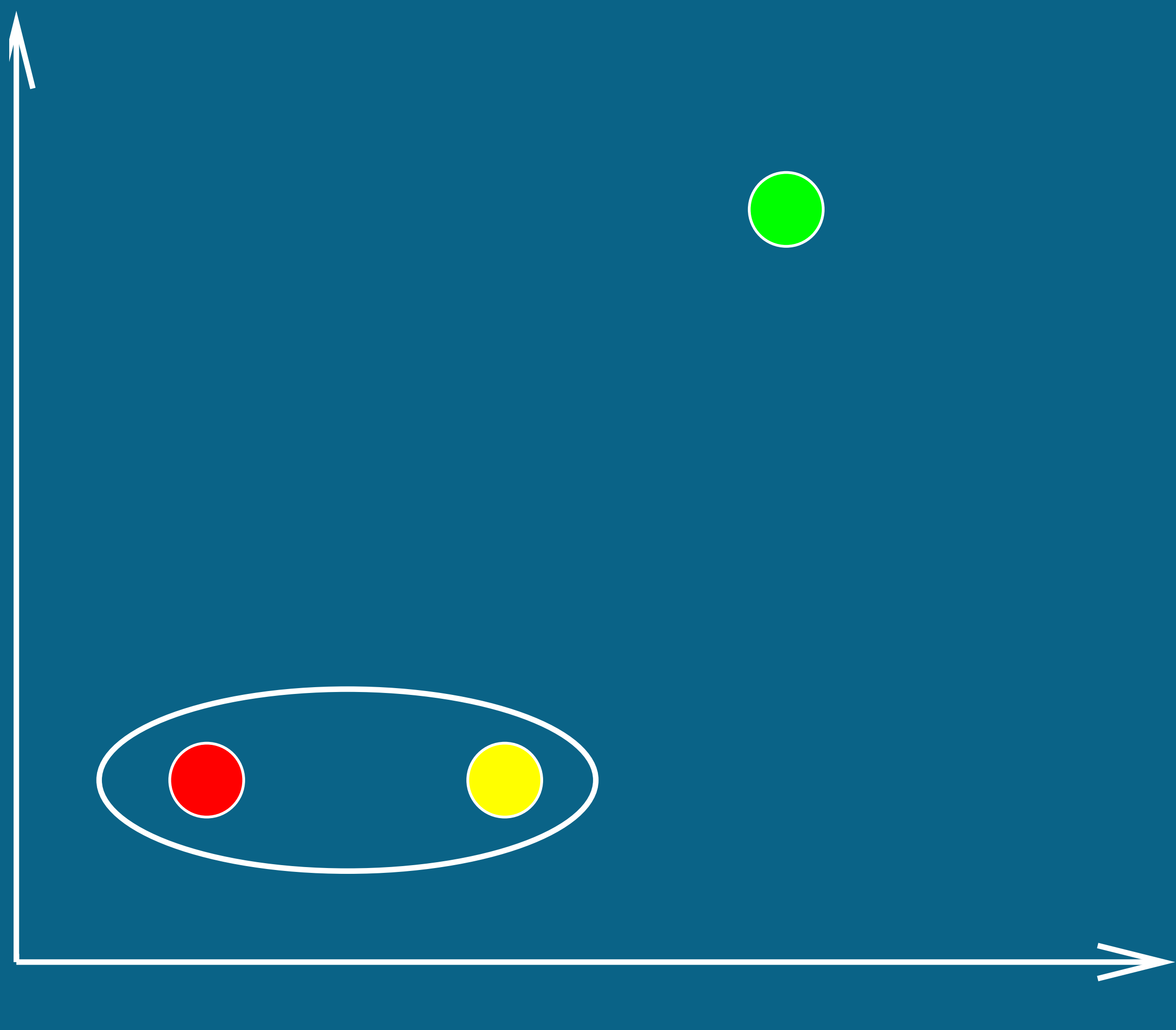
Correlation

$C = \cos \phi$

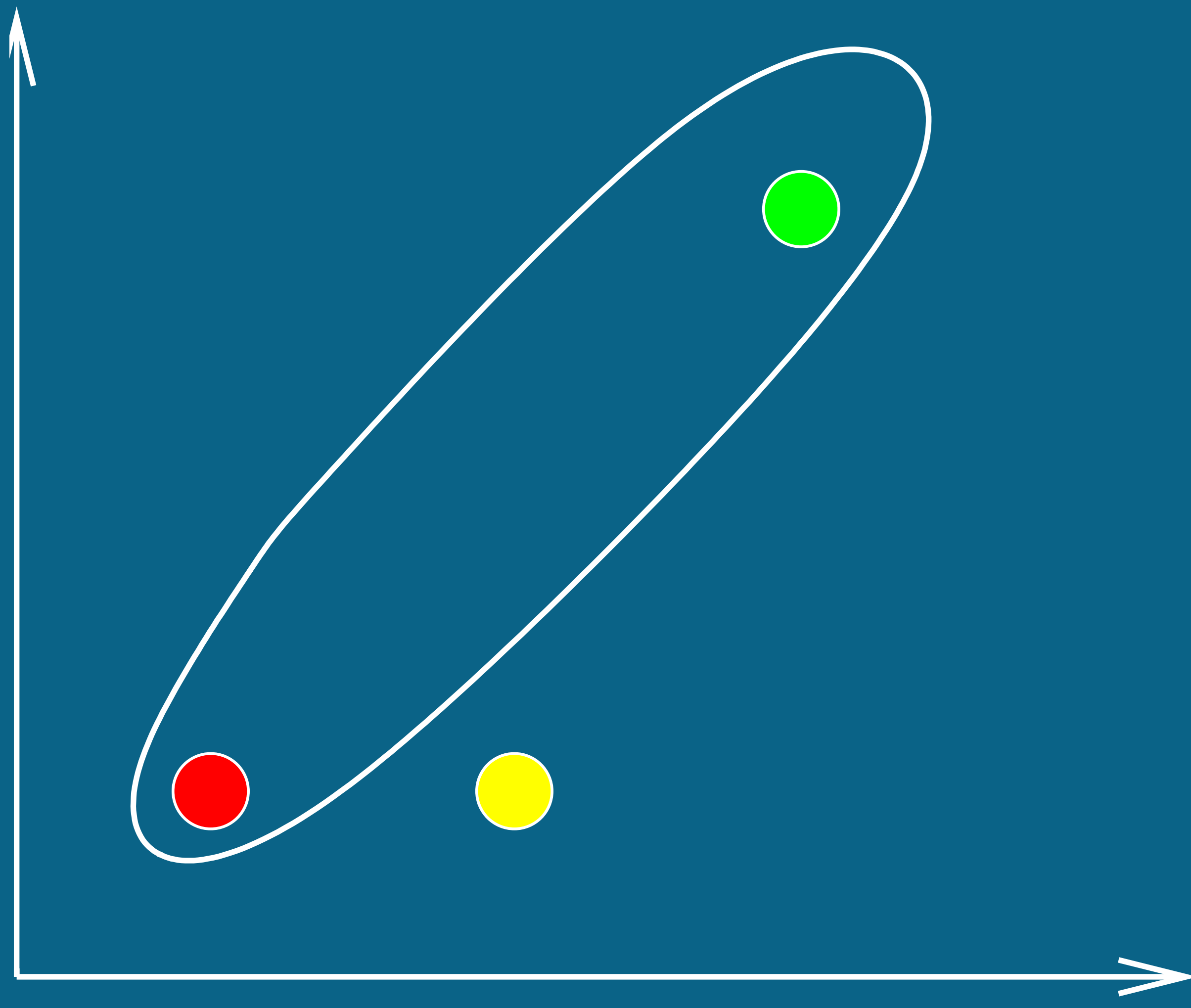




Which vectors are grouped together?

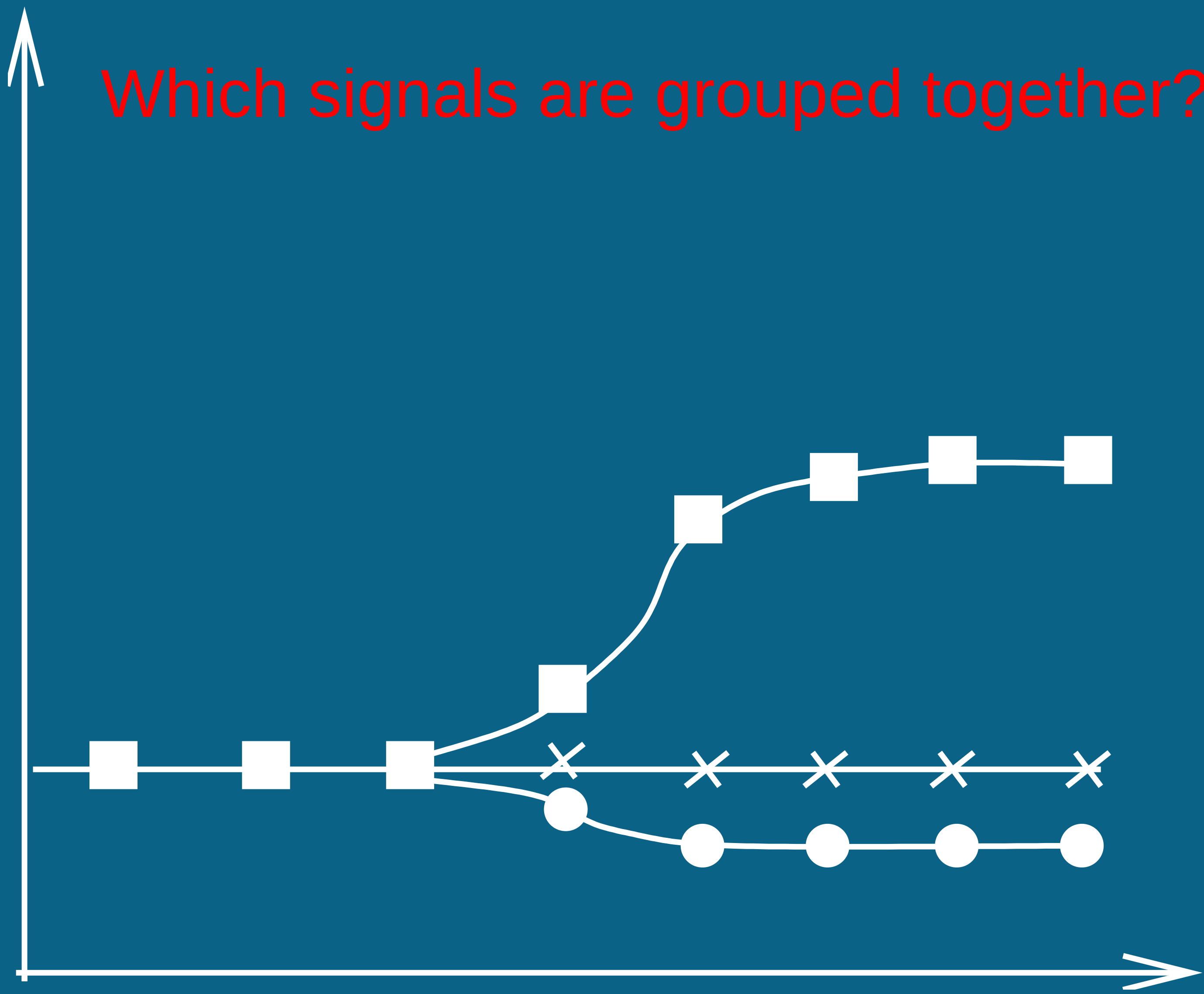


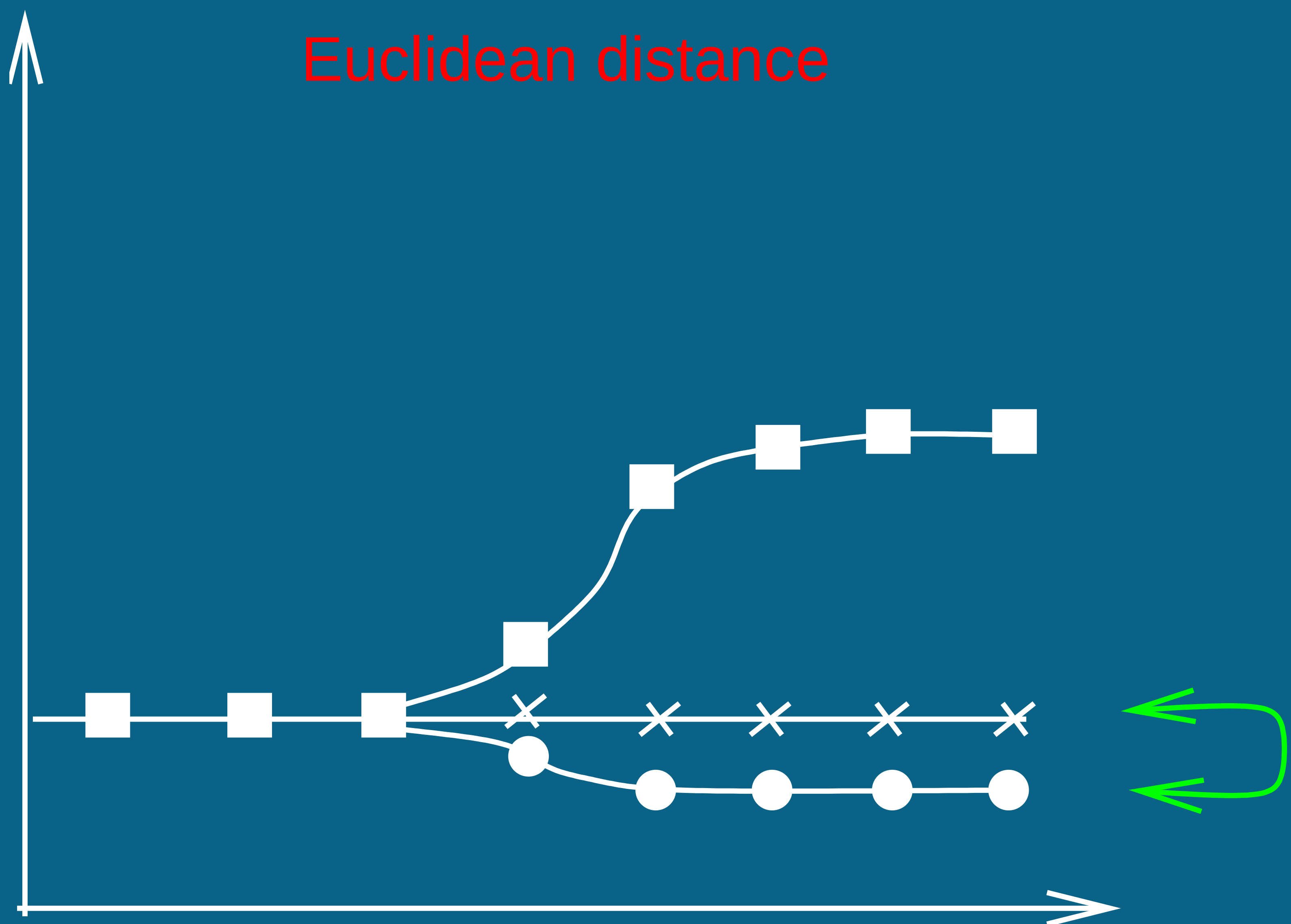
Euclidean distance



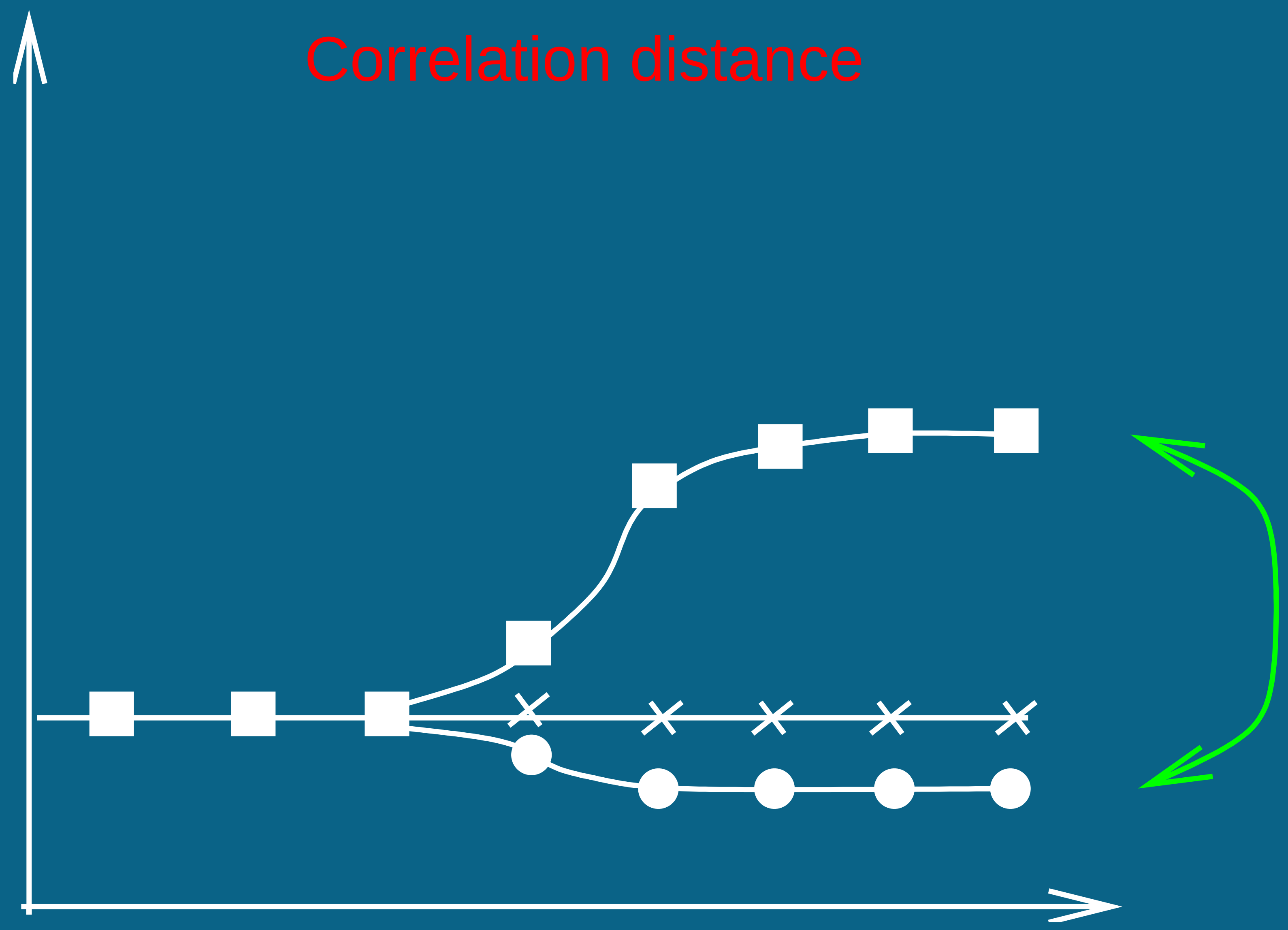
Correlation distance

Which signals are grouped together?

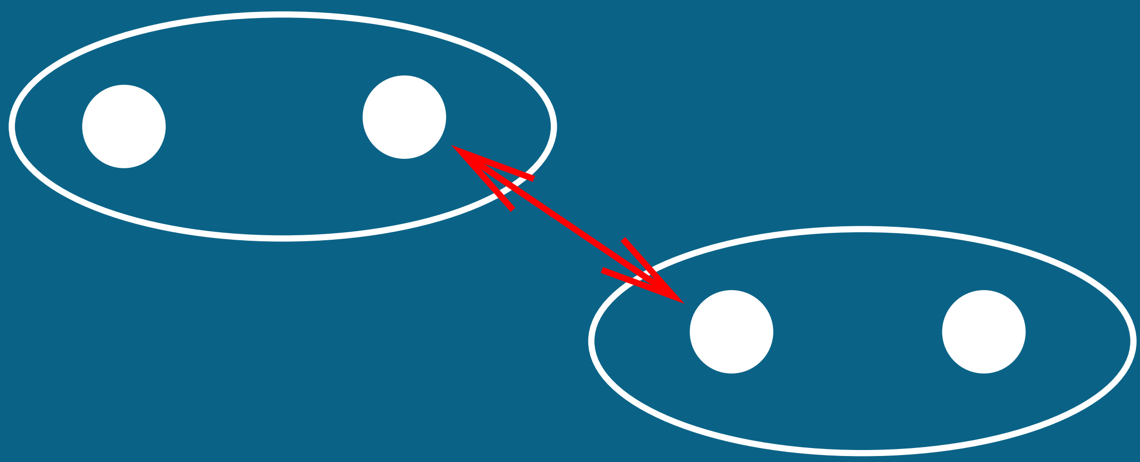



Euclidean distance

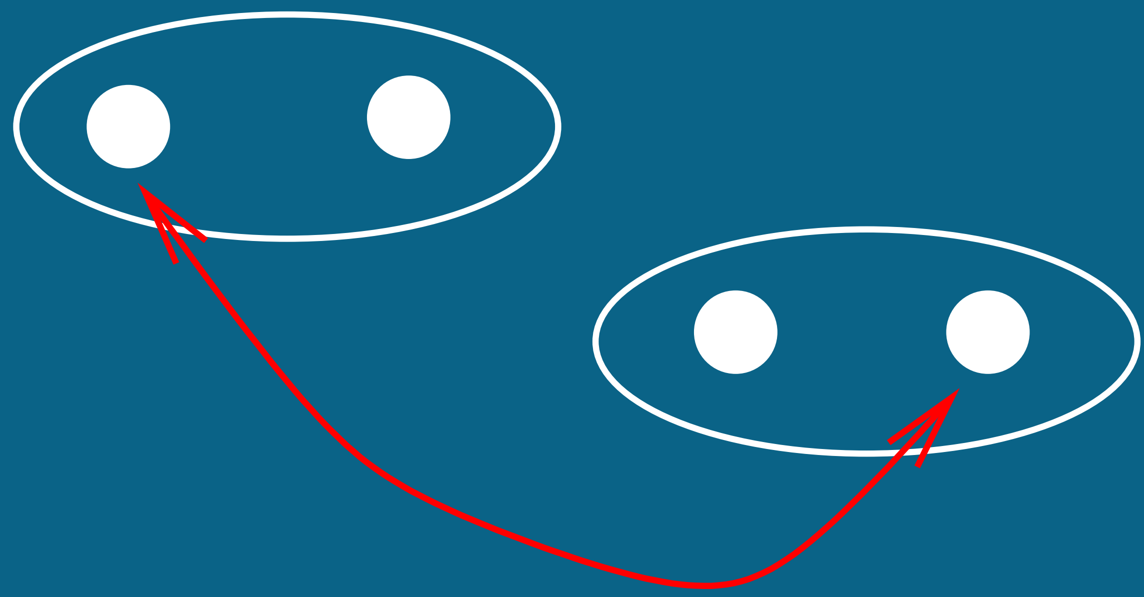
Correlation distance



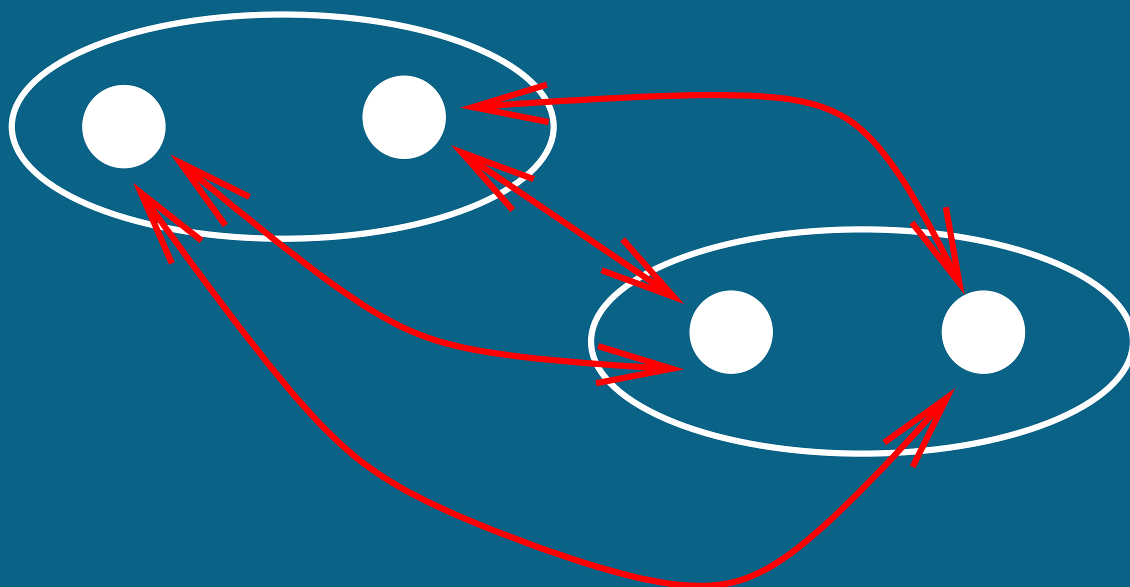
Single linkage



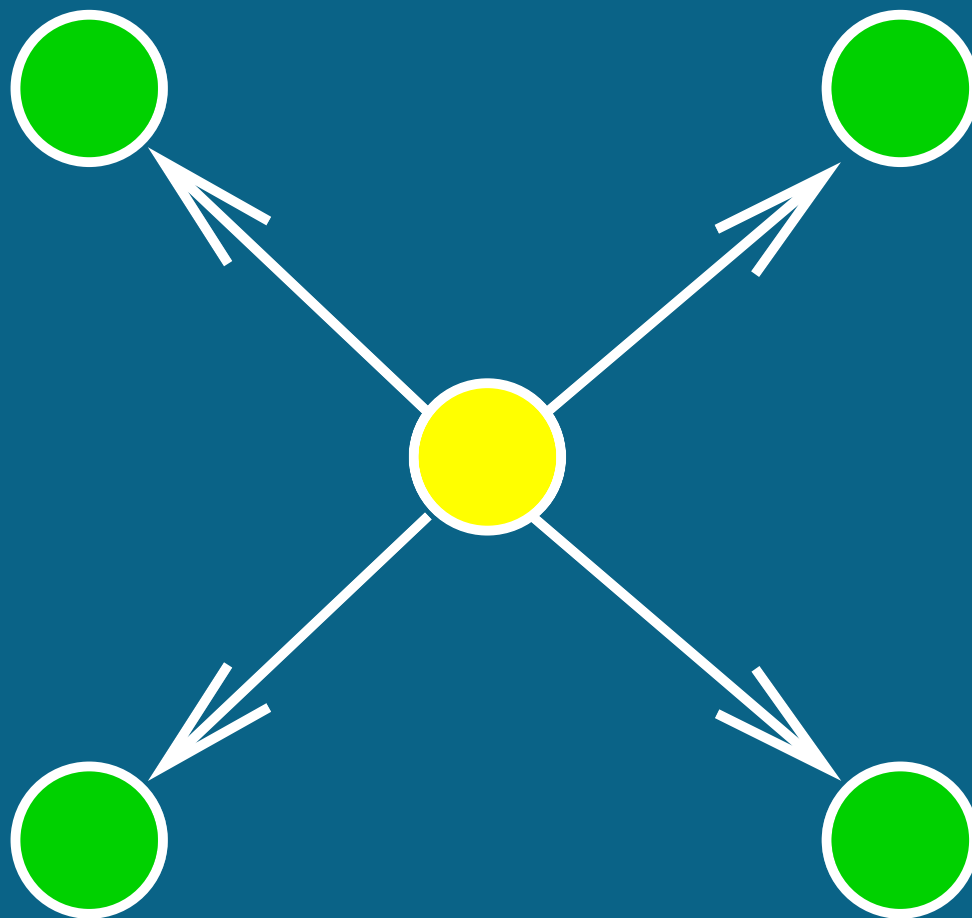
Complete linkage

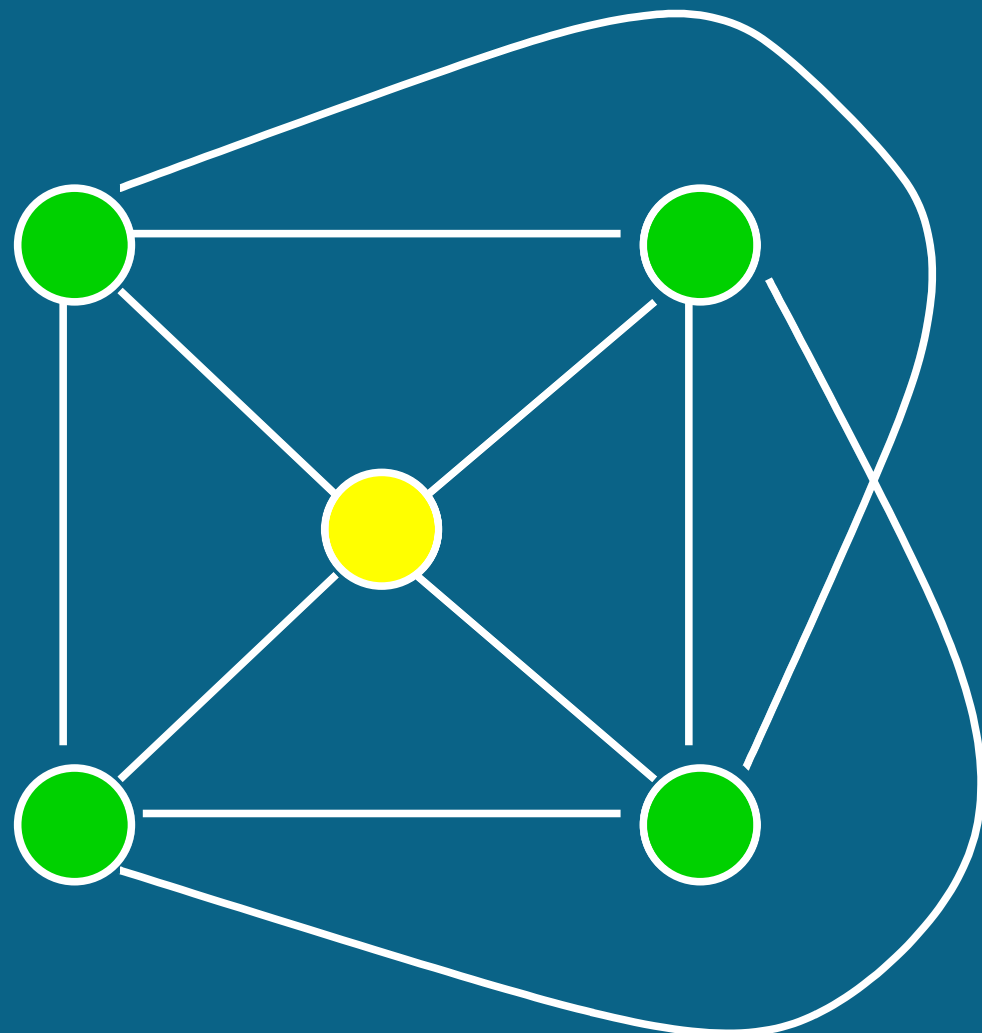


Average linkage



Disadvantage of clustering





Bayesian networks

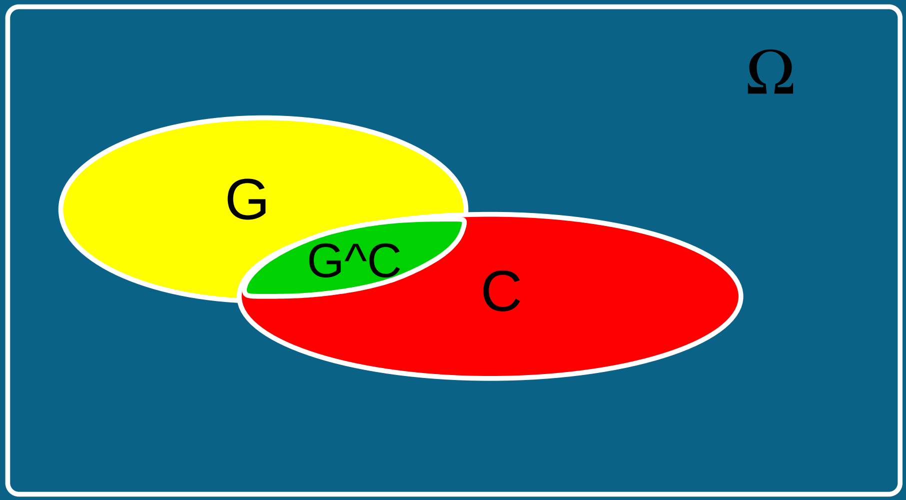
or

Directed acyclic graphs (DAGs)

- Probabilistic framework for inferring the dependence structure between multiple interacting quantities (here: expression levels of different genes).
- Probabilistic semantics: Fits the stochastic nature of both the biological processes and noisy experiments.
- Capable of handling noise and estimating the confidence in the different features of the network.

Revision: Bayes' Rule

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



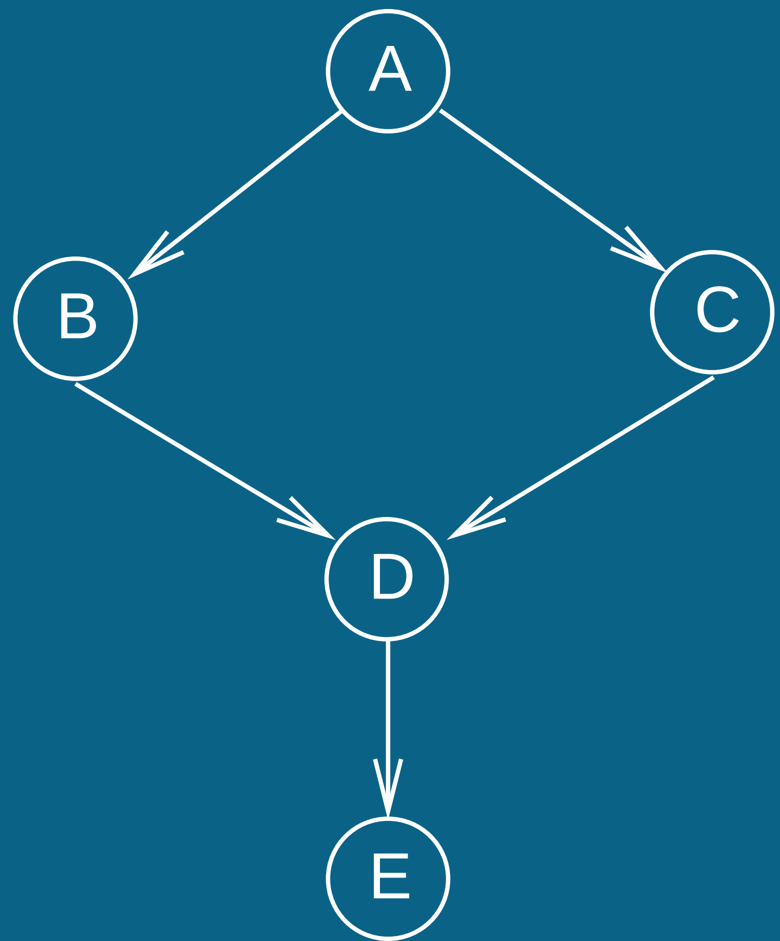
$$P(G|C) = \frac{P(G, C)}{P(C)}$$

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

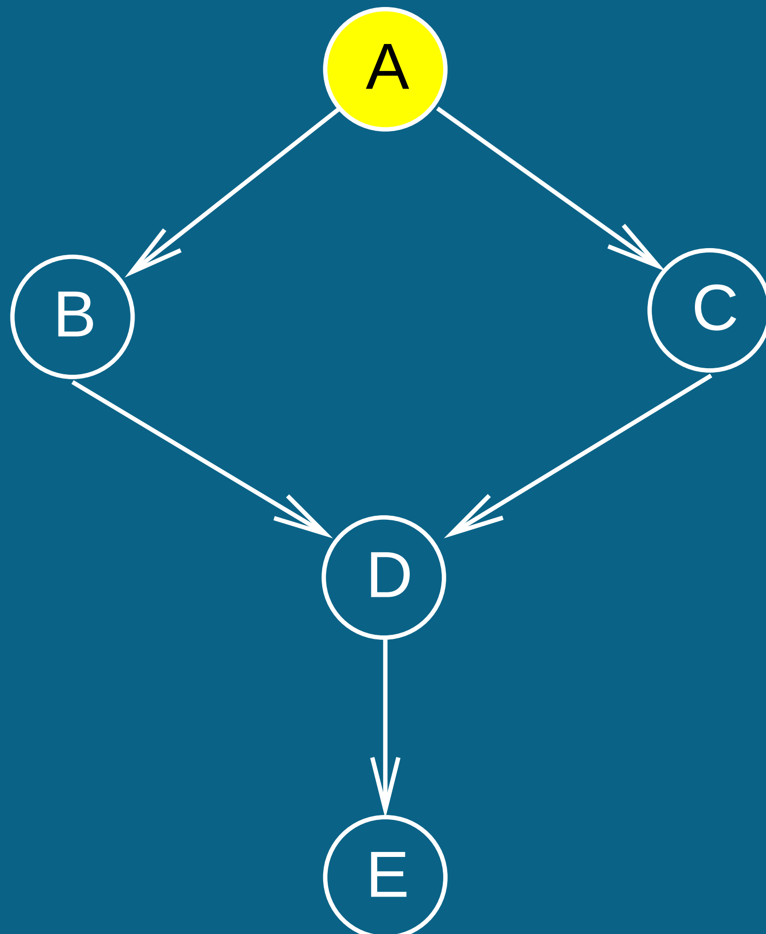
$$P(C|G) = \frac{P(G|C)P(C)}{P(G)}$$

Introduction to probabilistic graphical models

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

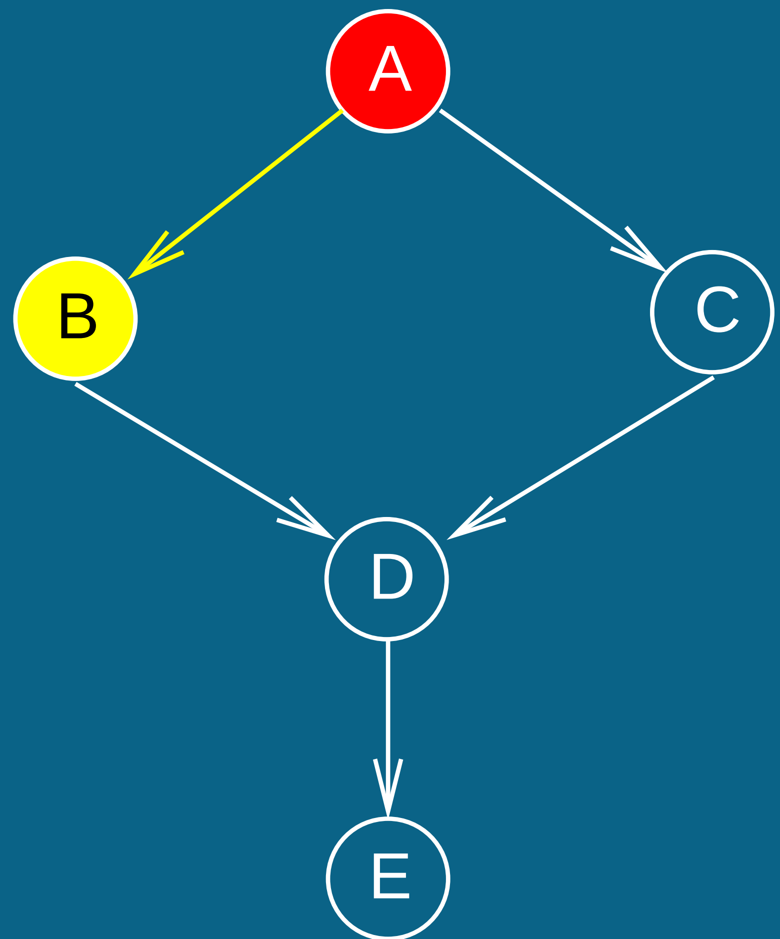


$$P(A, B, C, D, E) =$$



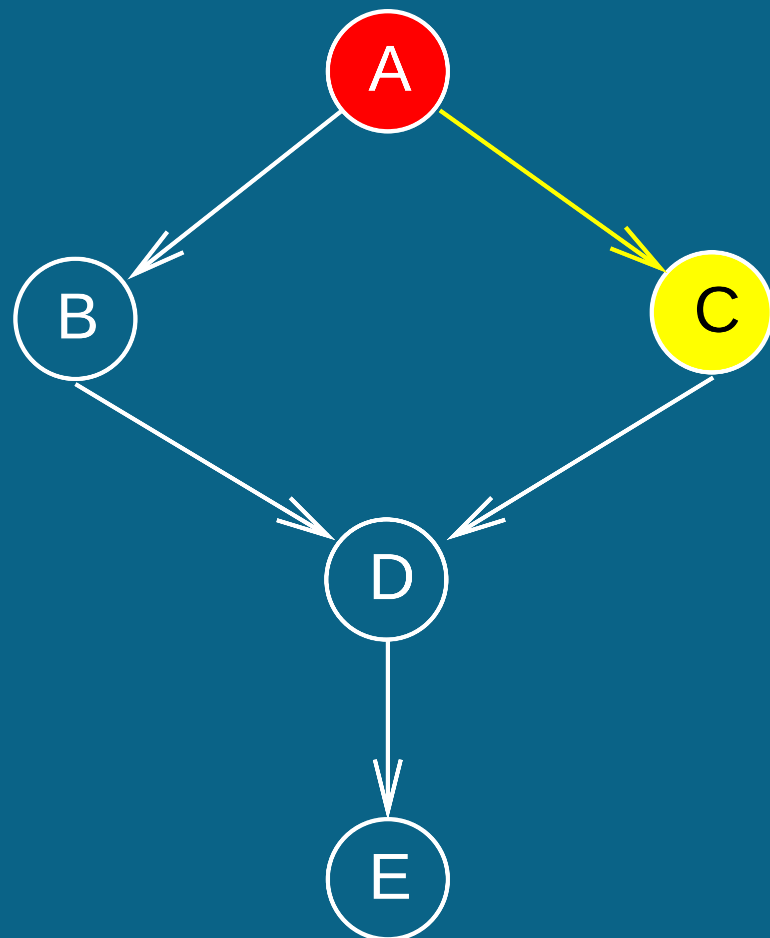
$$P(A, B, C, D, E) =$$

$$P(A)$$



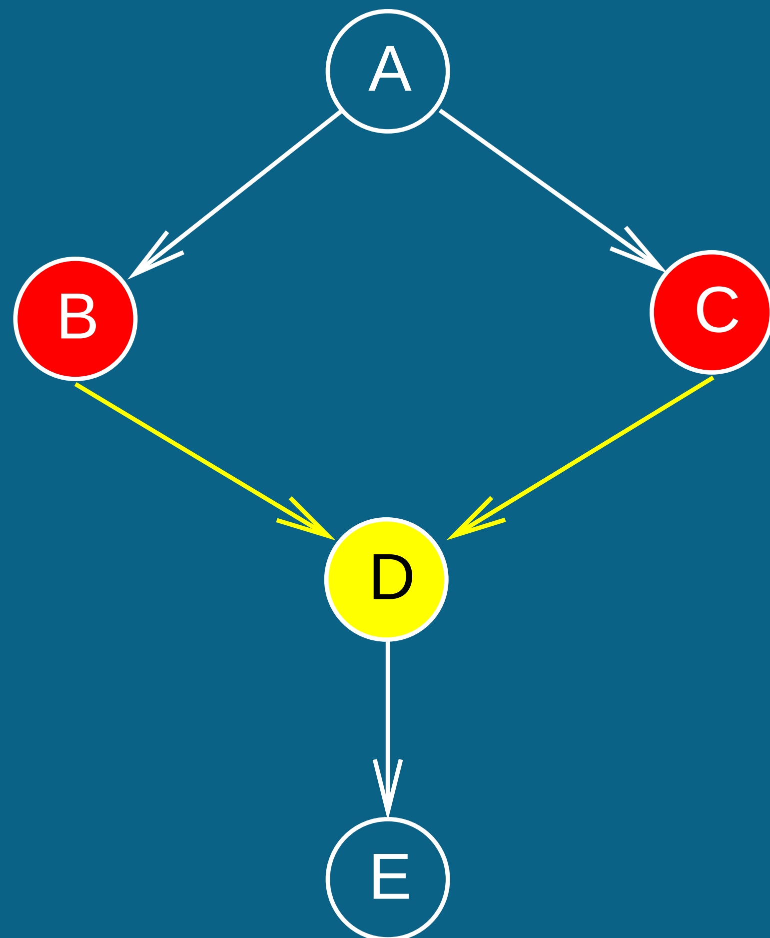
$$P(A, B, C, D, E) =$$

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



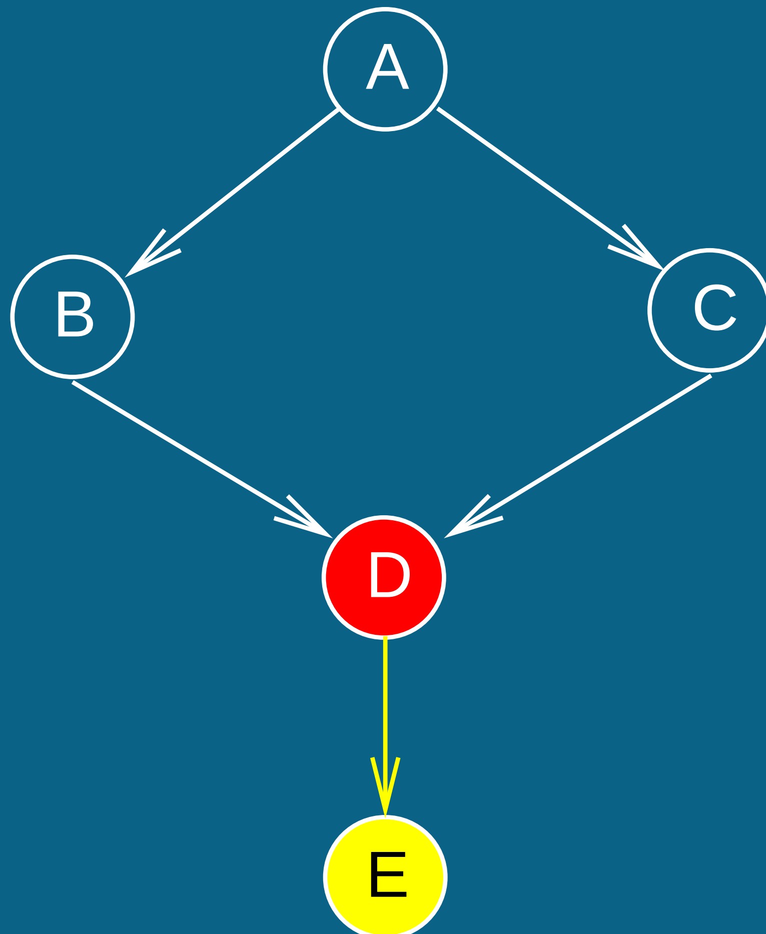
$$P(A, B, C, D, E) =$$

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



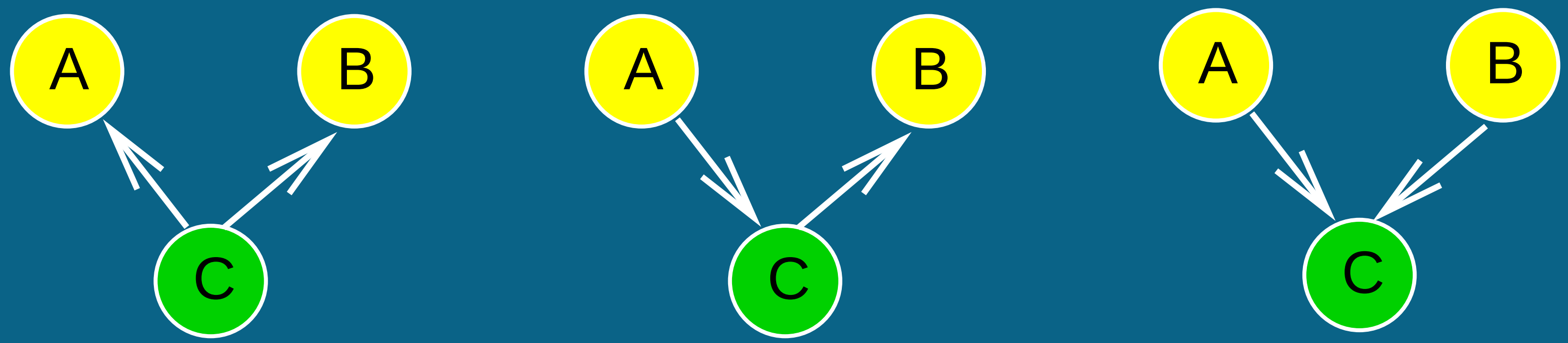
$$P(A, B, C, D, E) =$$

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



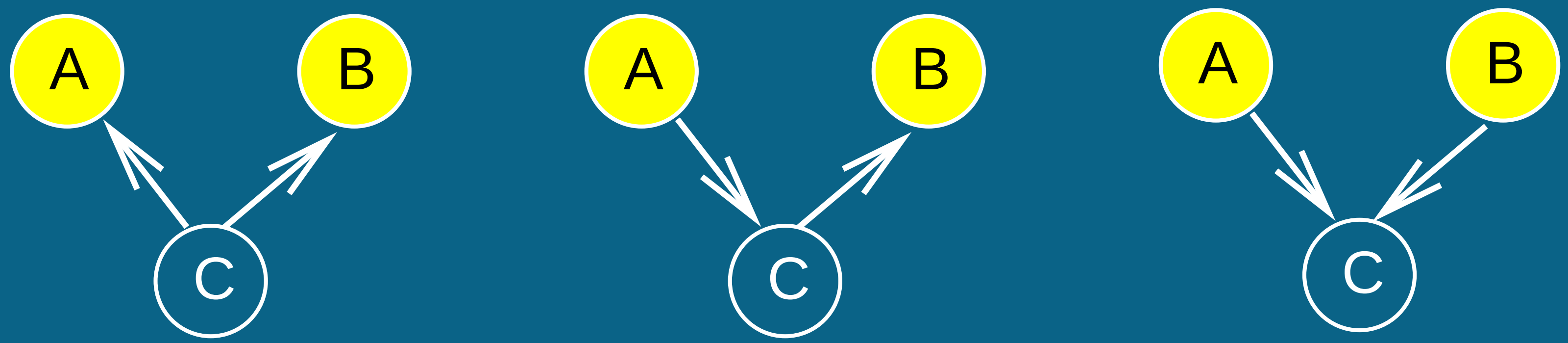
$$P(A, B, C, D, E) =$$

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



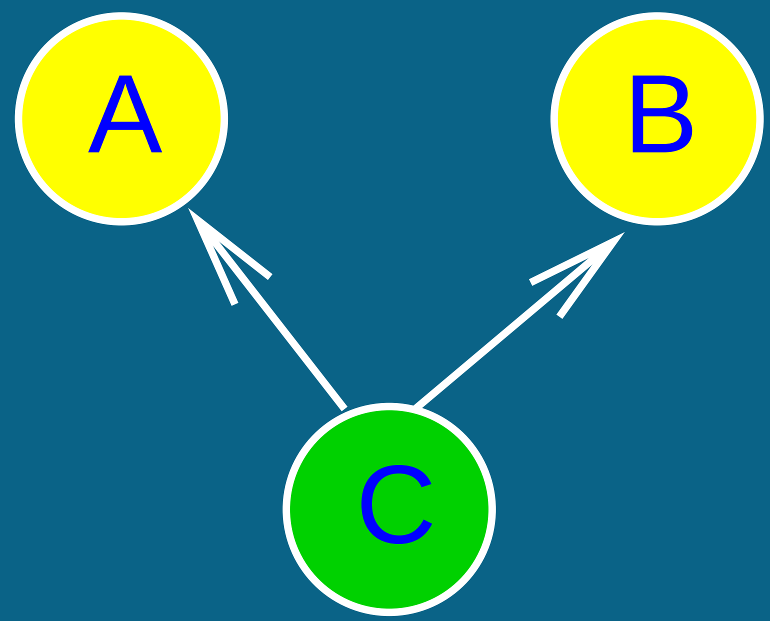
When are A and B conditionally independent given C?

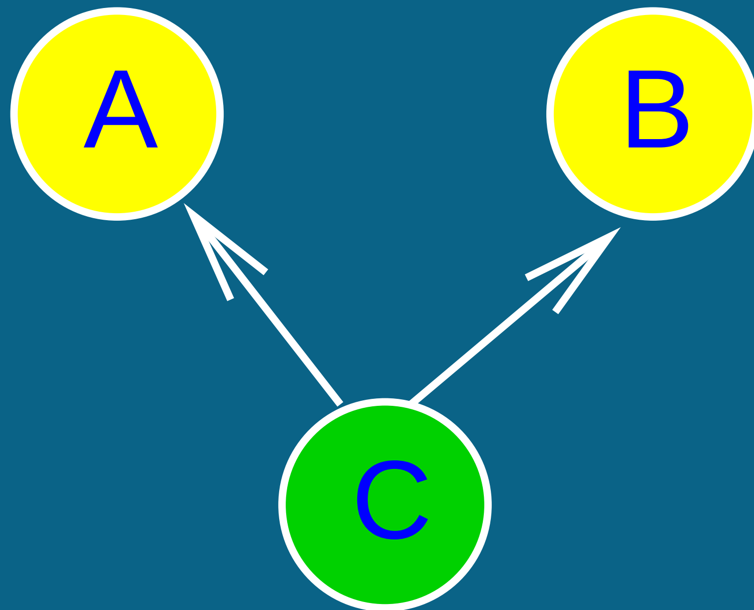
$$P(A, B|C) = P(A|C)P(B|C)$$



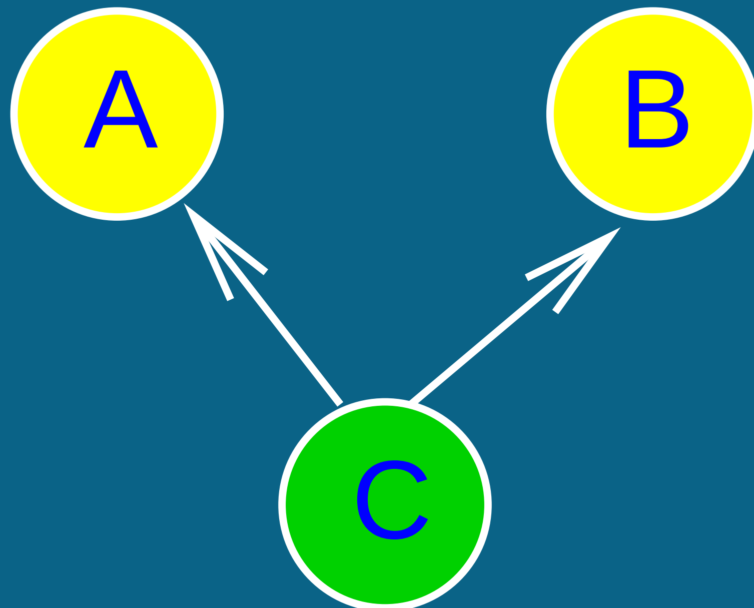
When are A and B marginally independent?

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


A
B
C

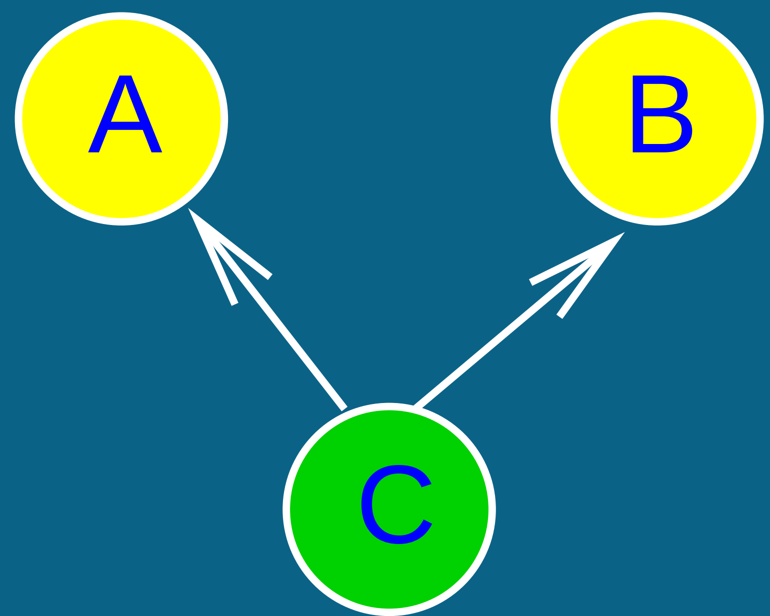


$$P(A, B, C) = P(A|C)P(B|C)P(C)$$



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

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

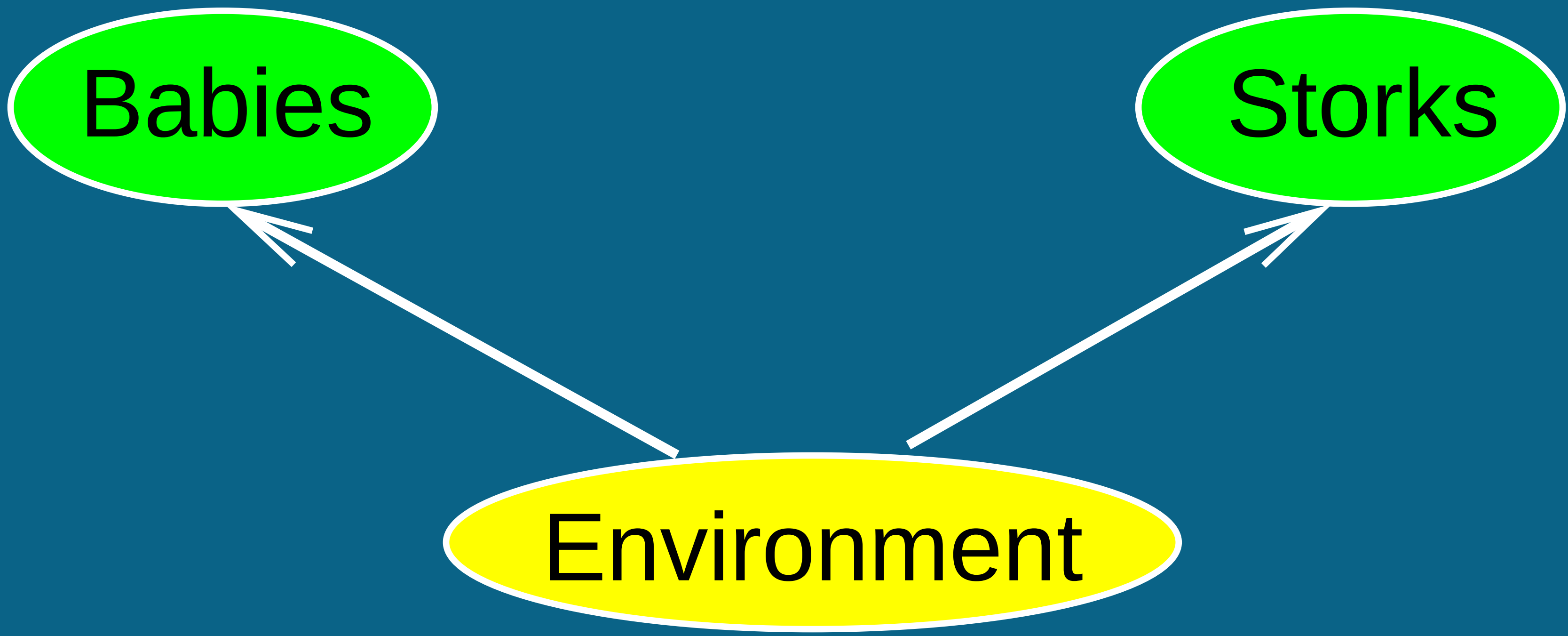


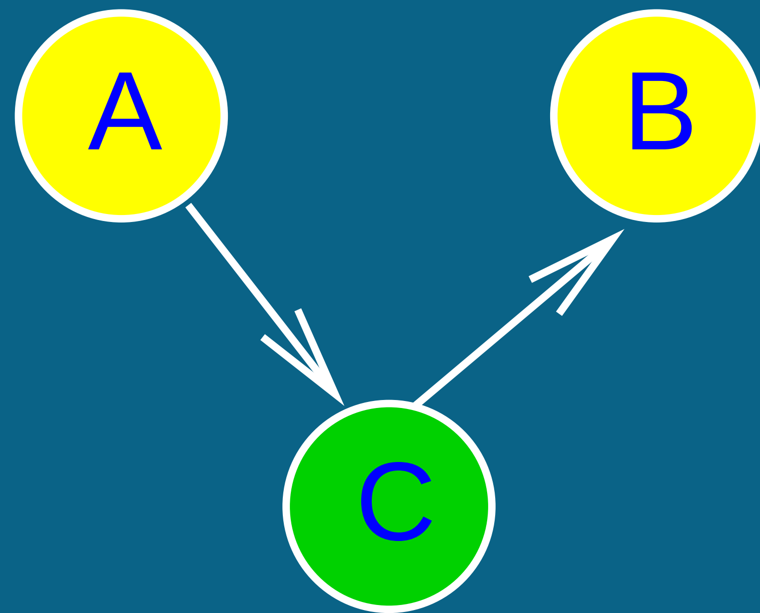
$$P(A,B,C) = P(A|C)P(B|C)P(C)$$

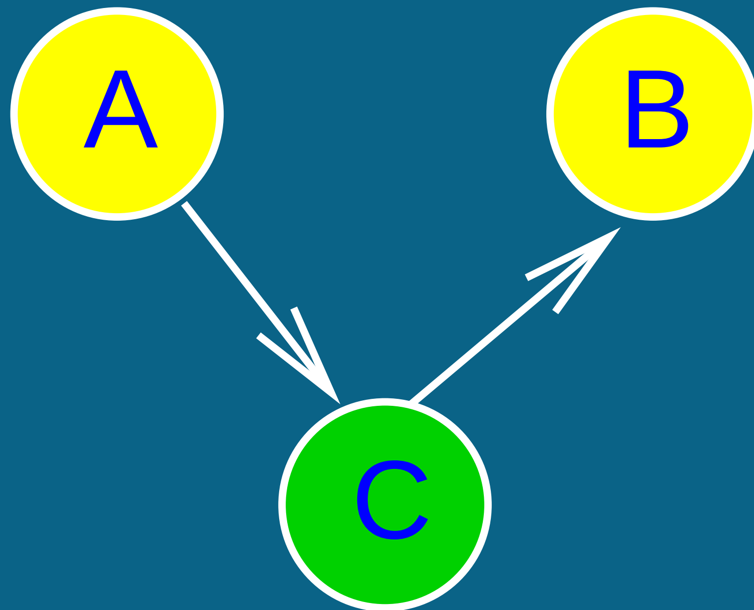
$$P(A,B|C) = \frac{P(A,B,C)}{P(C)} = P(A|C)P(B|C)$$

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

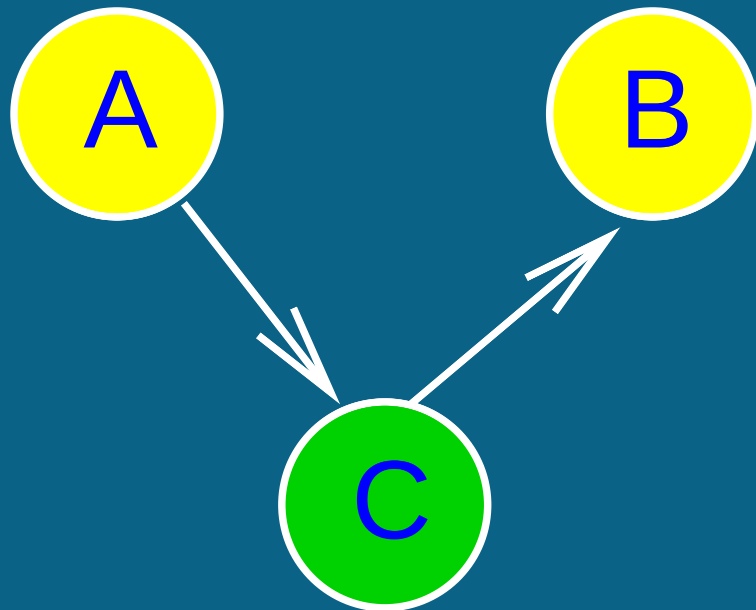

Babies
Storks


Babies
Storks
Environment


A
B
C

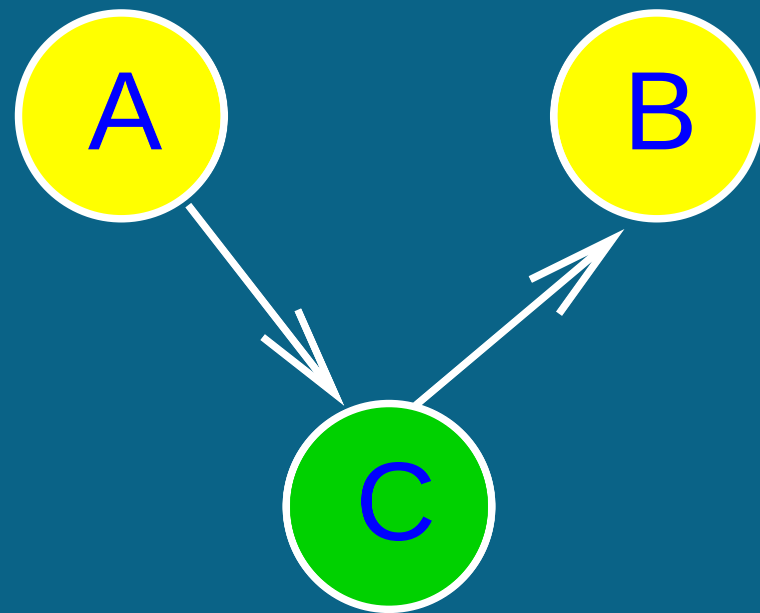


$$P(A,B,C) = P(B|C)P(C|A)P(A)$$



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

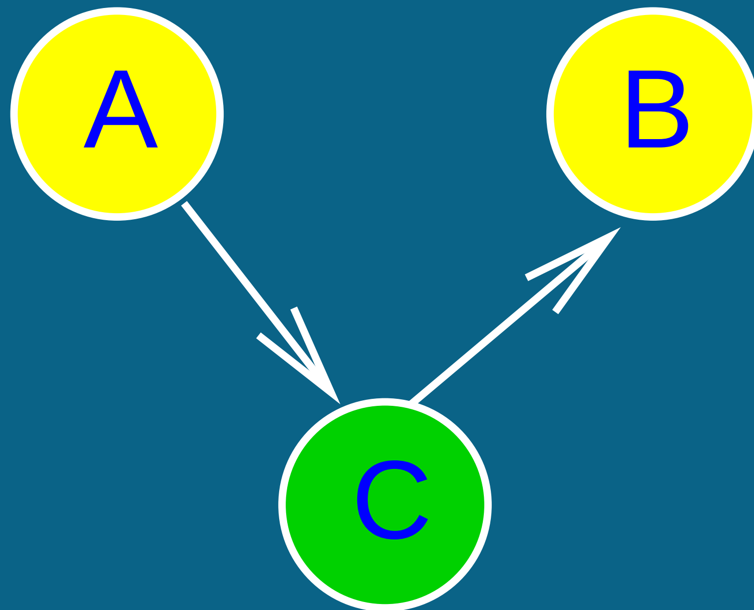
$$P(A,B|C) = \frac{P(A,B,C)}{P(C)} = P(B|C)\frac{P(C|A)P(A)}{P(C)}$$



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

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

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



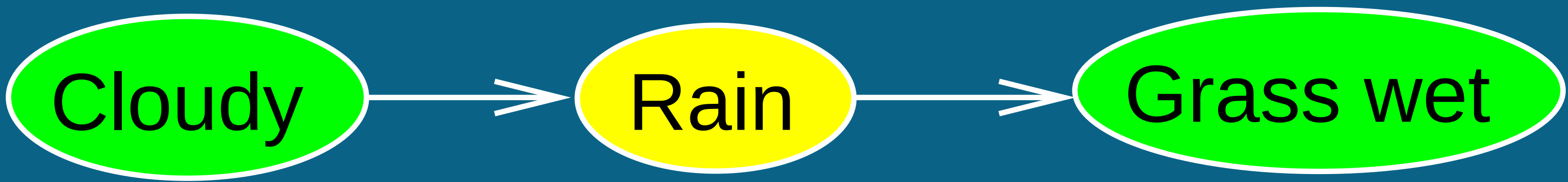
$$P(A,B,C) = P(B|C)P(C|A)P(A)$$

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

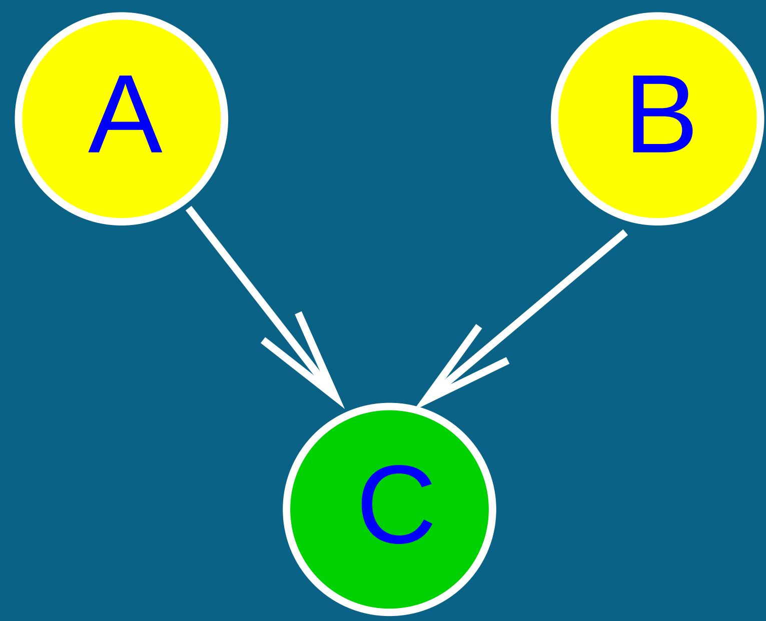
$$P(A,B|C) = P(B|C)P(A|C)$$

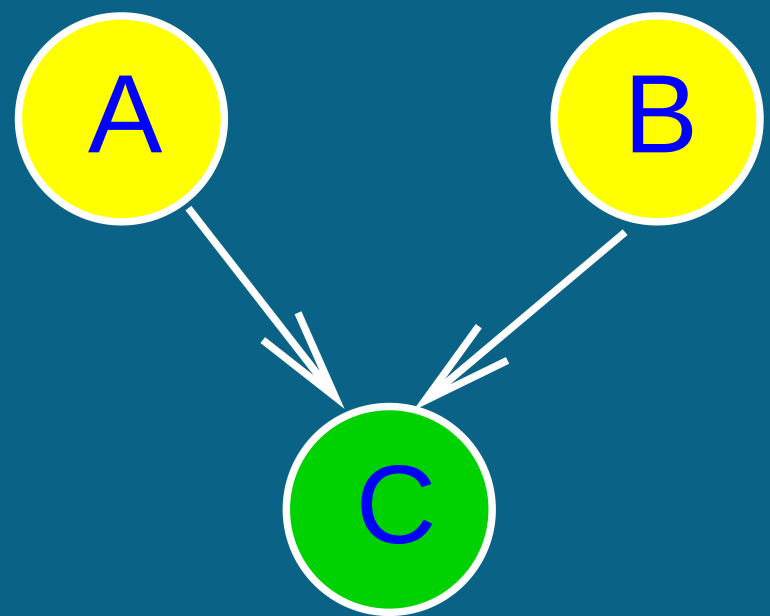
But: $P(A,B) \neq P(A)P(B)$


Cloudy
Rain
Grass wet

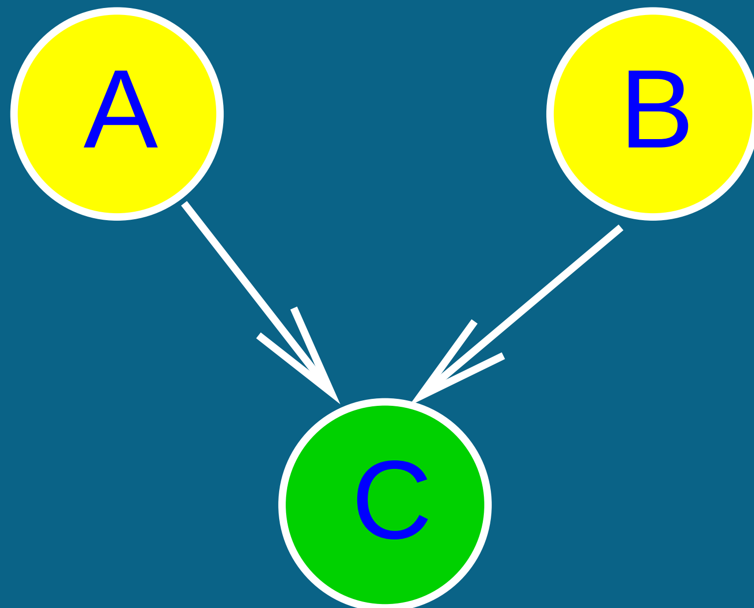


Cloudy
Rain
Grass wet


A
B
C

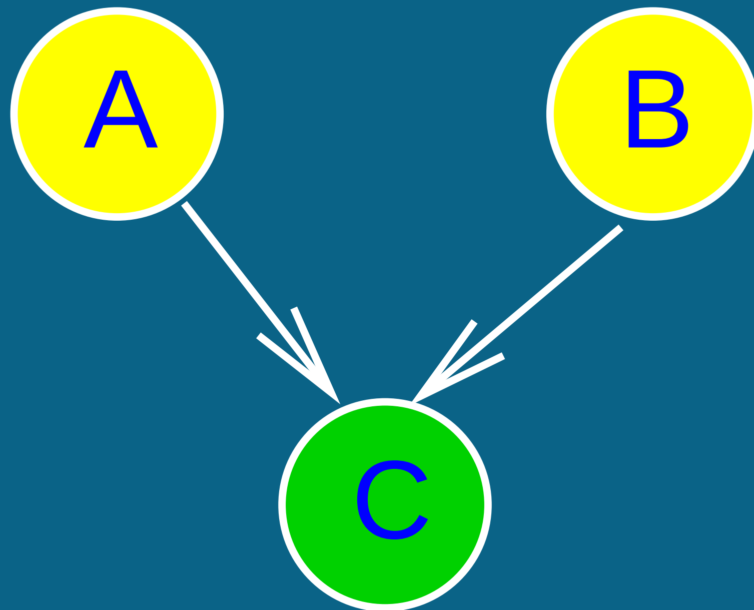


$$P(A, B, C) = P(C|A, B)P(A)P(B)$$



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

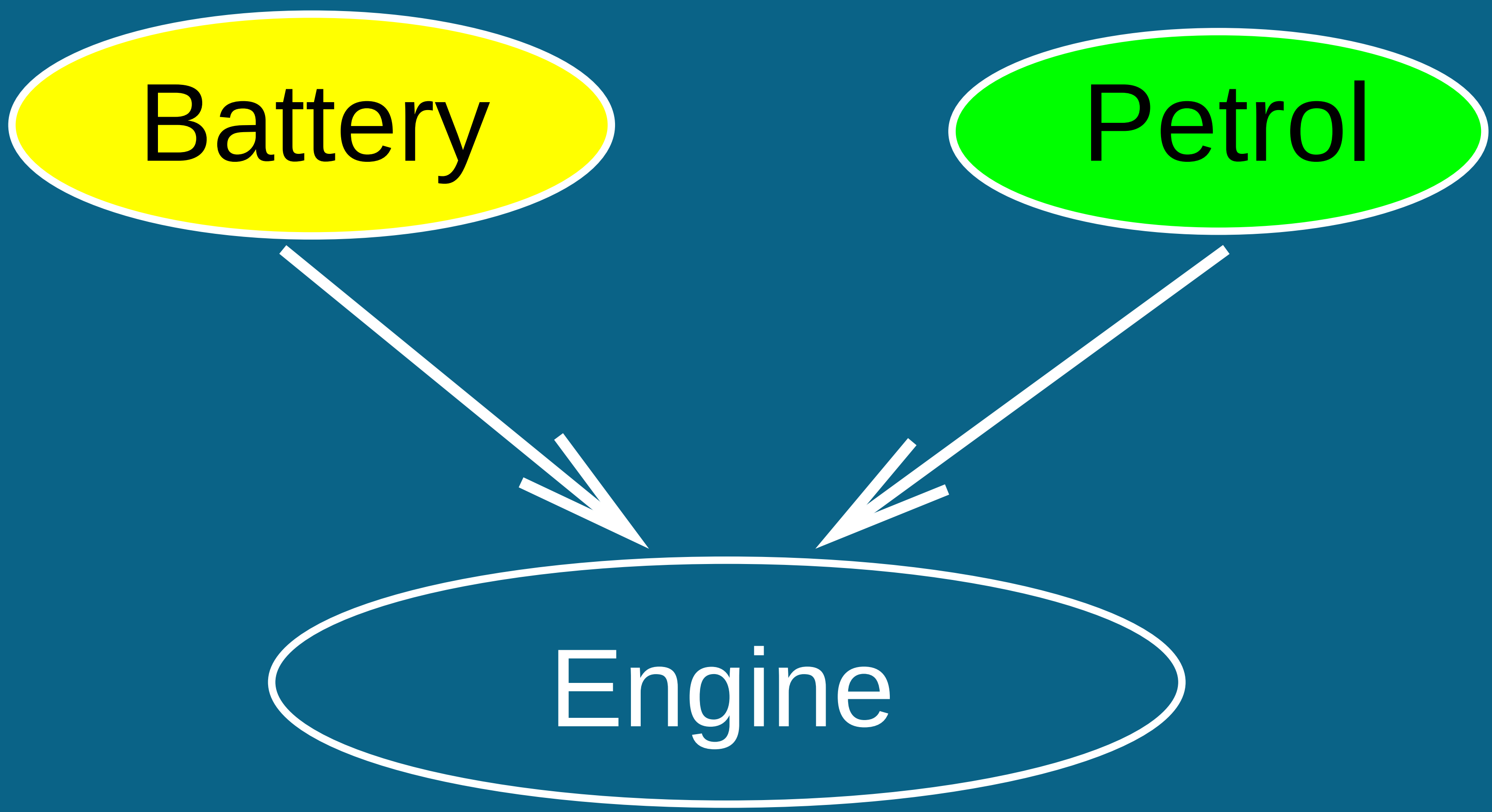
$$P(A,B) = \sum_{C} P(A,B,C) = P(A)P(B)$$

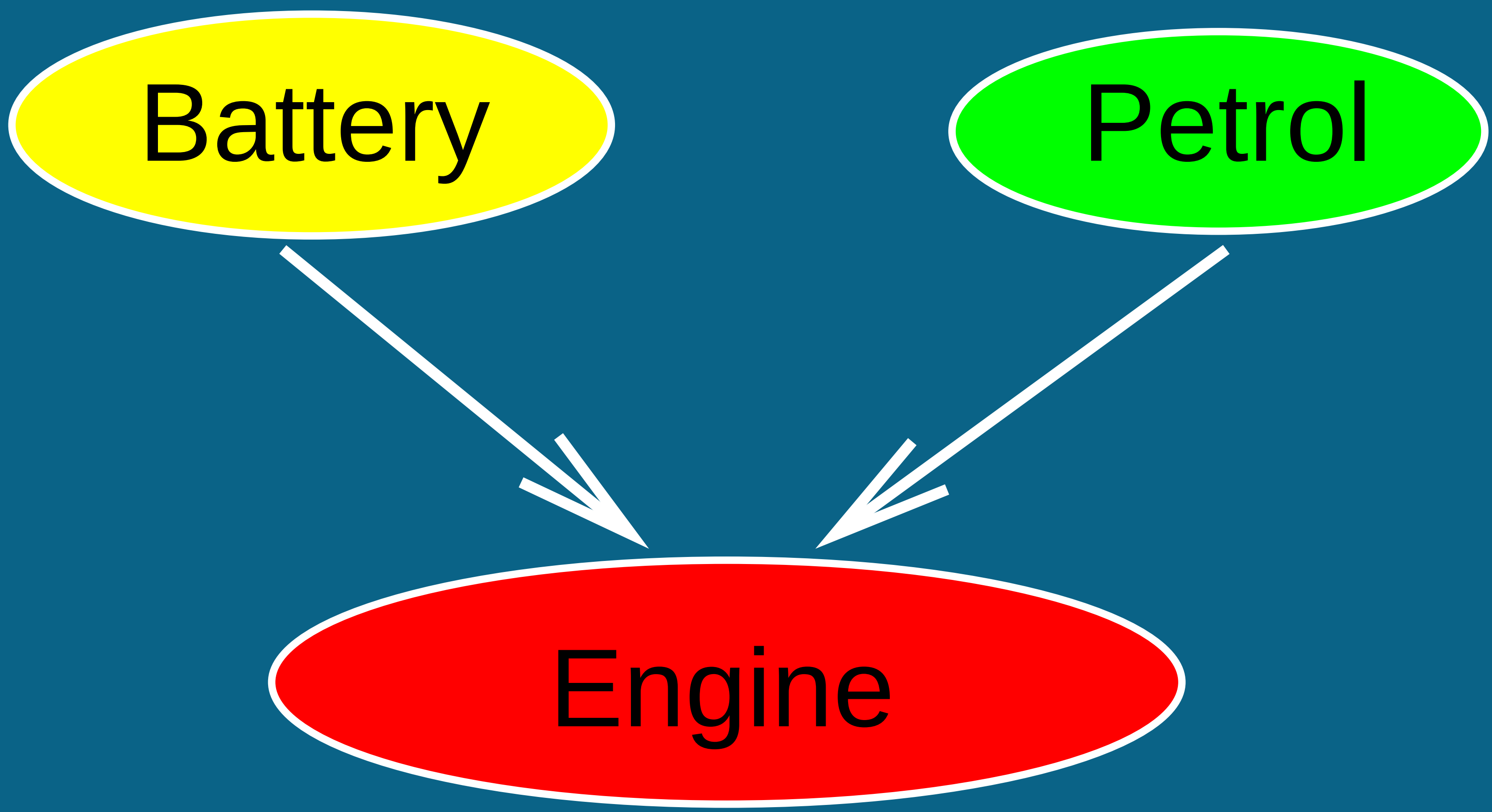


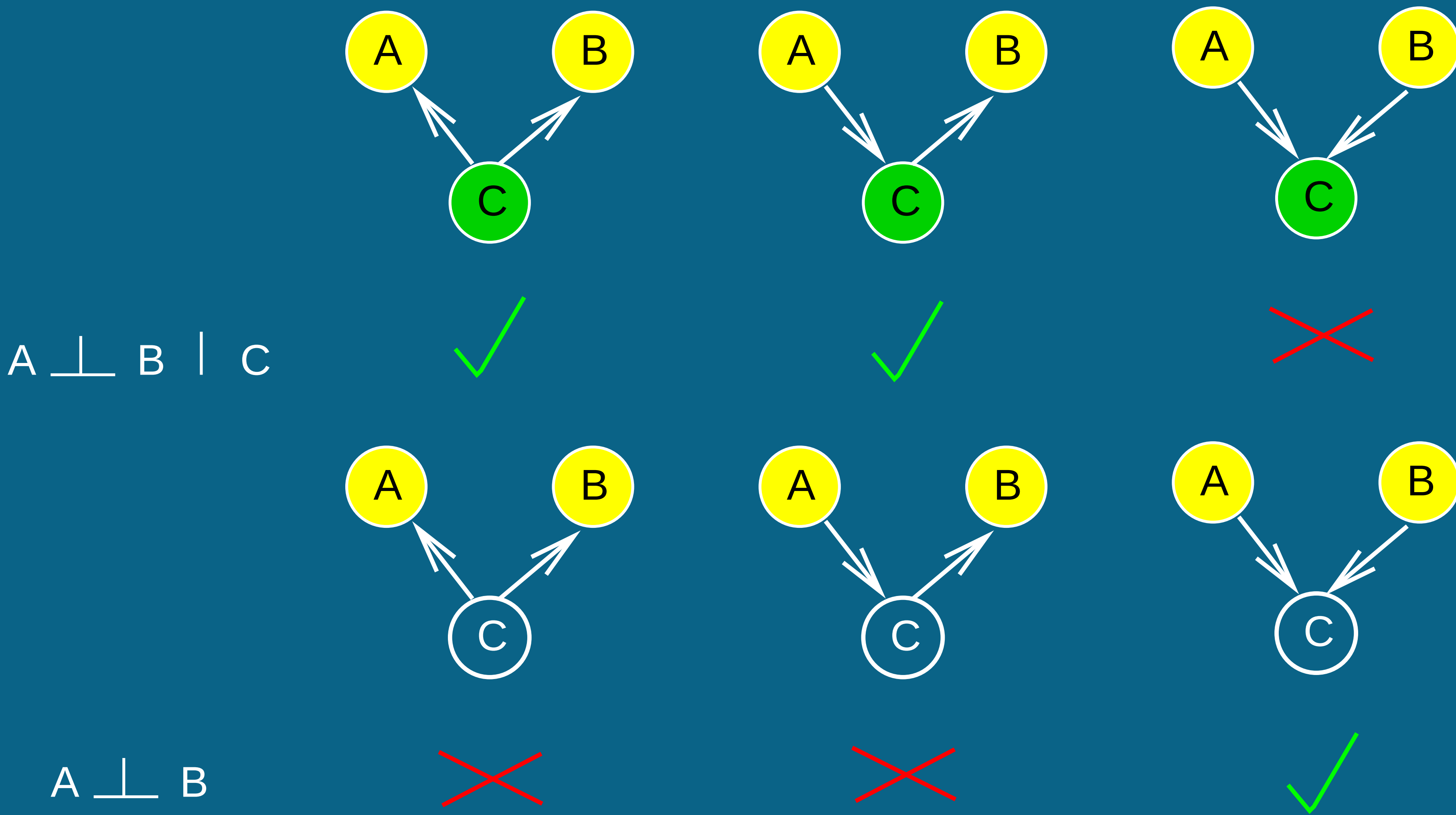
$$P(A,B,C) = P(C|A,B)P(A)P(B)$$

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

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


Battery
Petrol
Engine


Battery
Petrol
Engine


$A \perp B \mid C$
$A \perp B$

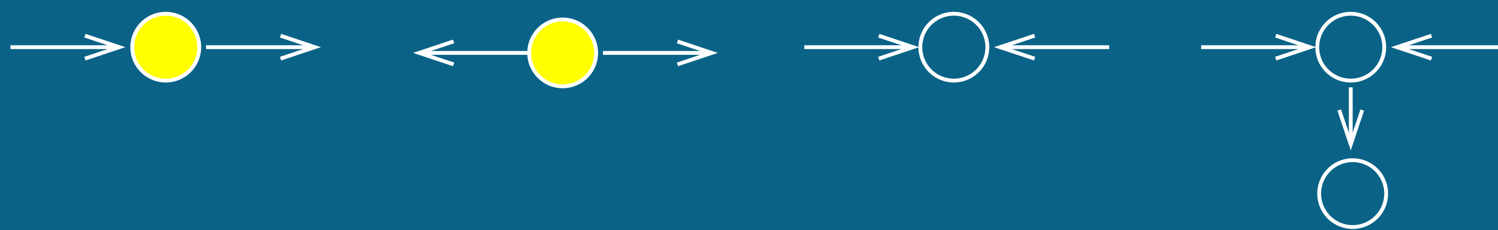
d-separation

- Let A, B be two nodes, and let Z be a group of nodes.
- A path from A to B is blocked
  - if there is a node $C \in Z$ which is head-to-tail or tail-to-tail with respect to the path, or
  - if the node C is head-to-head and neither C nor any of its decendents are in Z.
- $A \perp B | Z$ iff all possible paths are blocked.

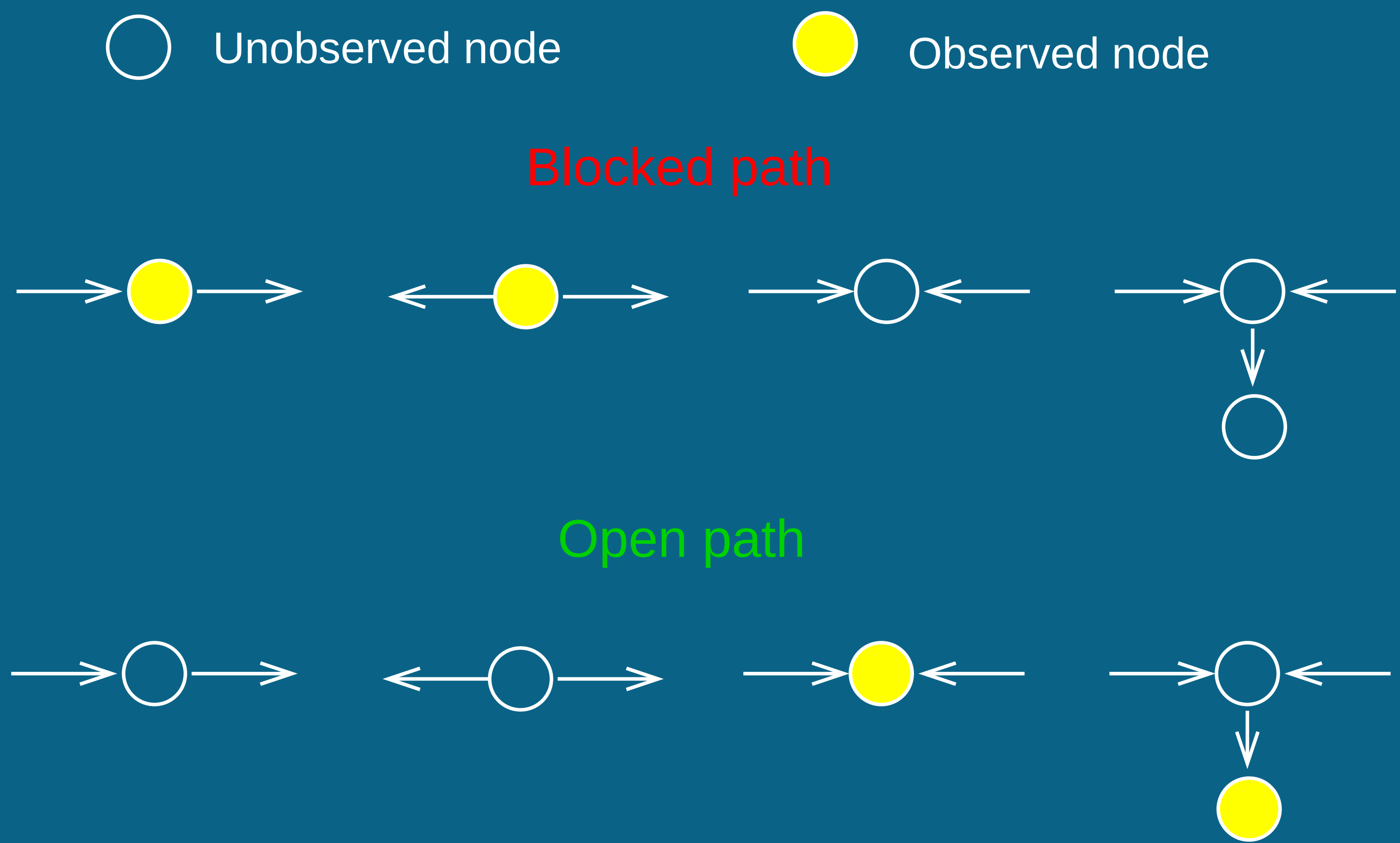
Unobserved node

Observed node

Blocked path

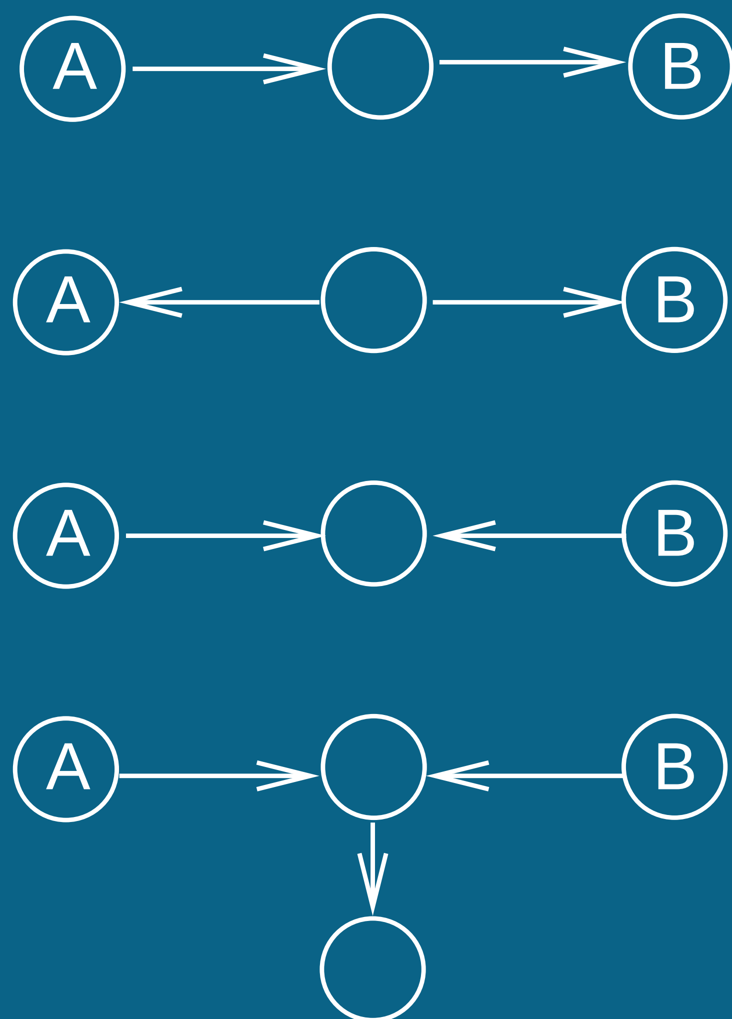


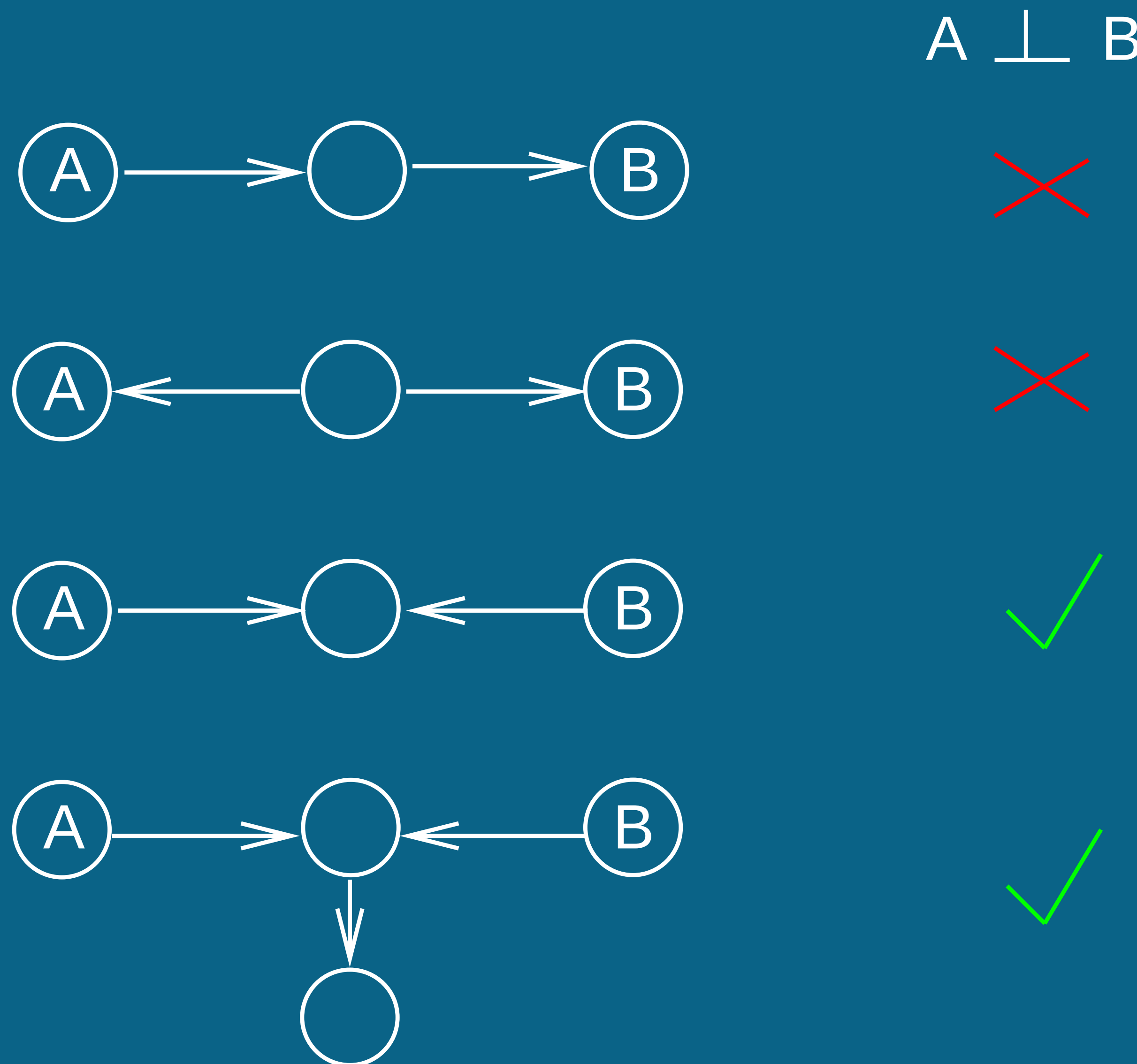
Open path



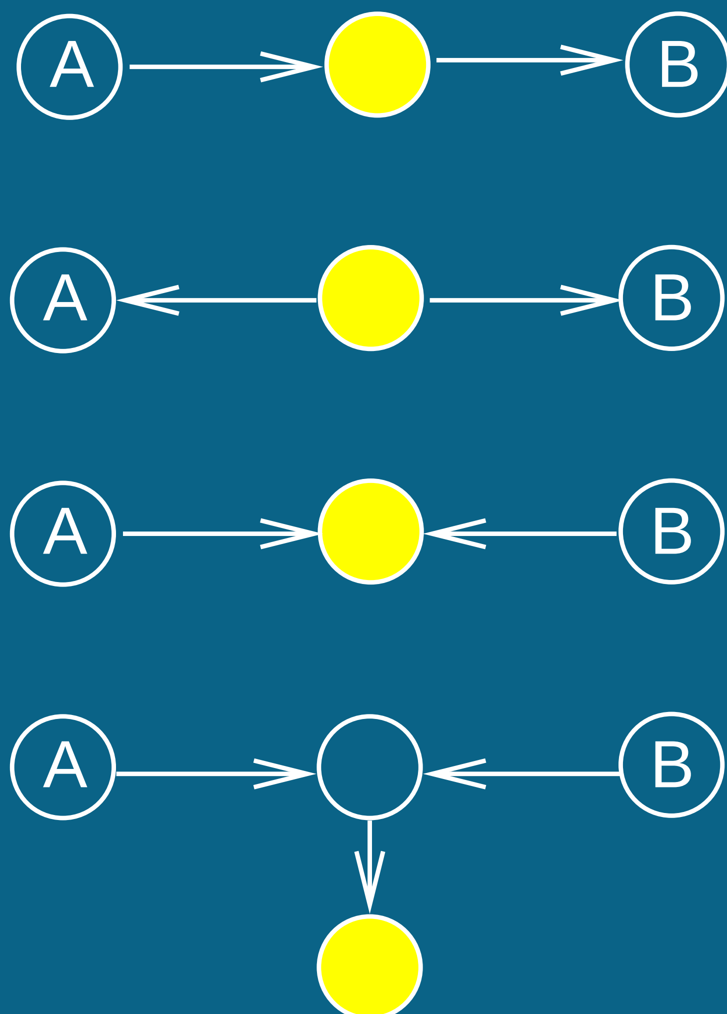
- Two variables A, B are d-separated if every path from A to B is blocked.
- $A \perp B$ iff A and B are d-separated.

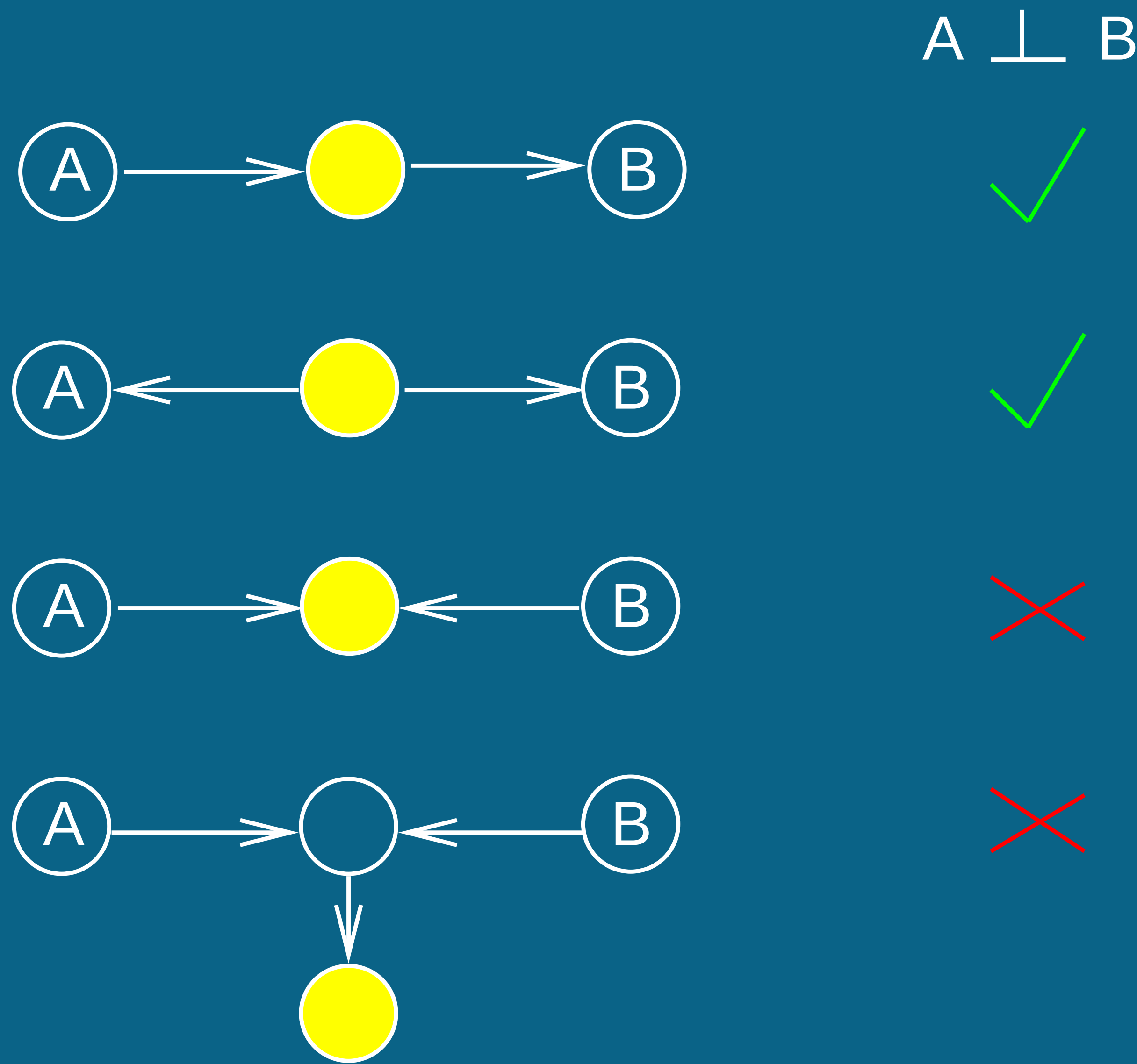
$A \perp B$

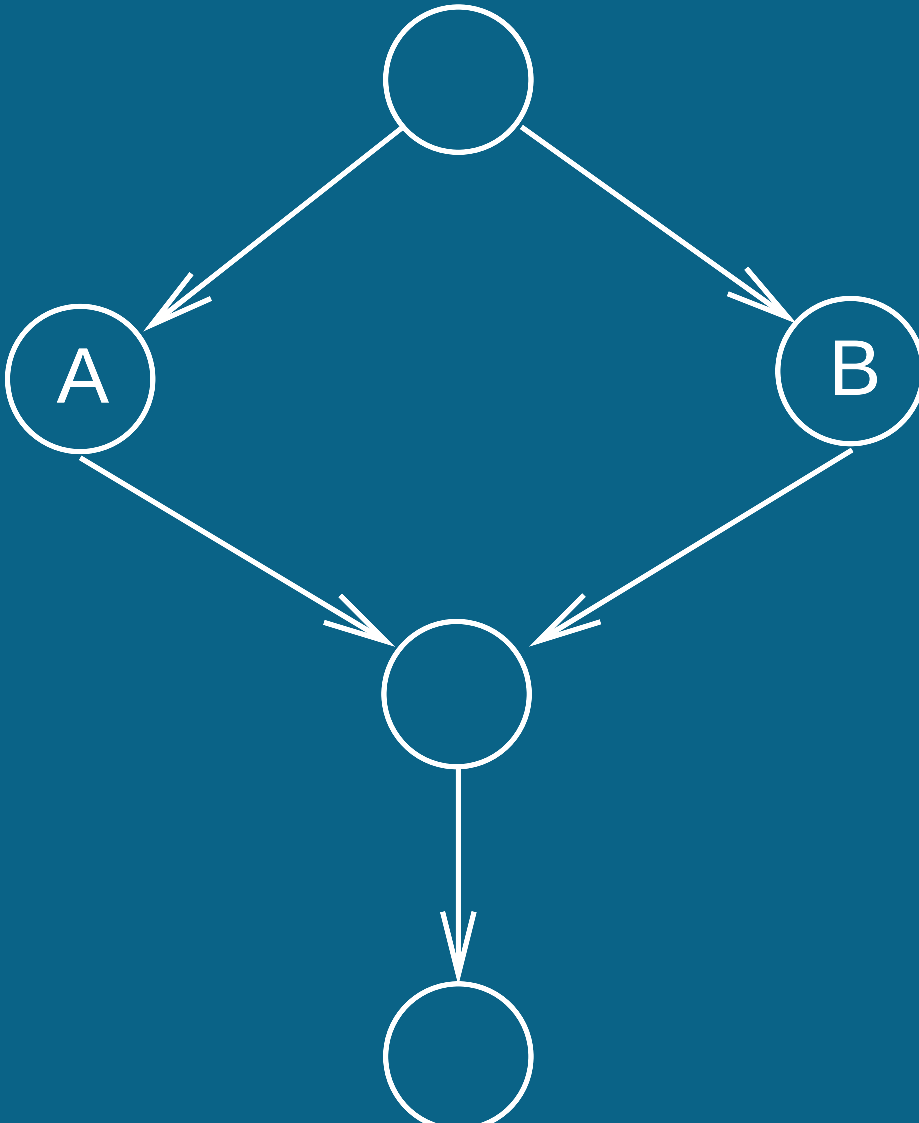



A ⊥ B
A
B
A
B
A
B
A
B

$A \perp B$

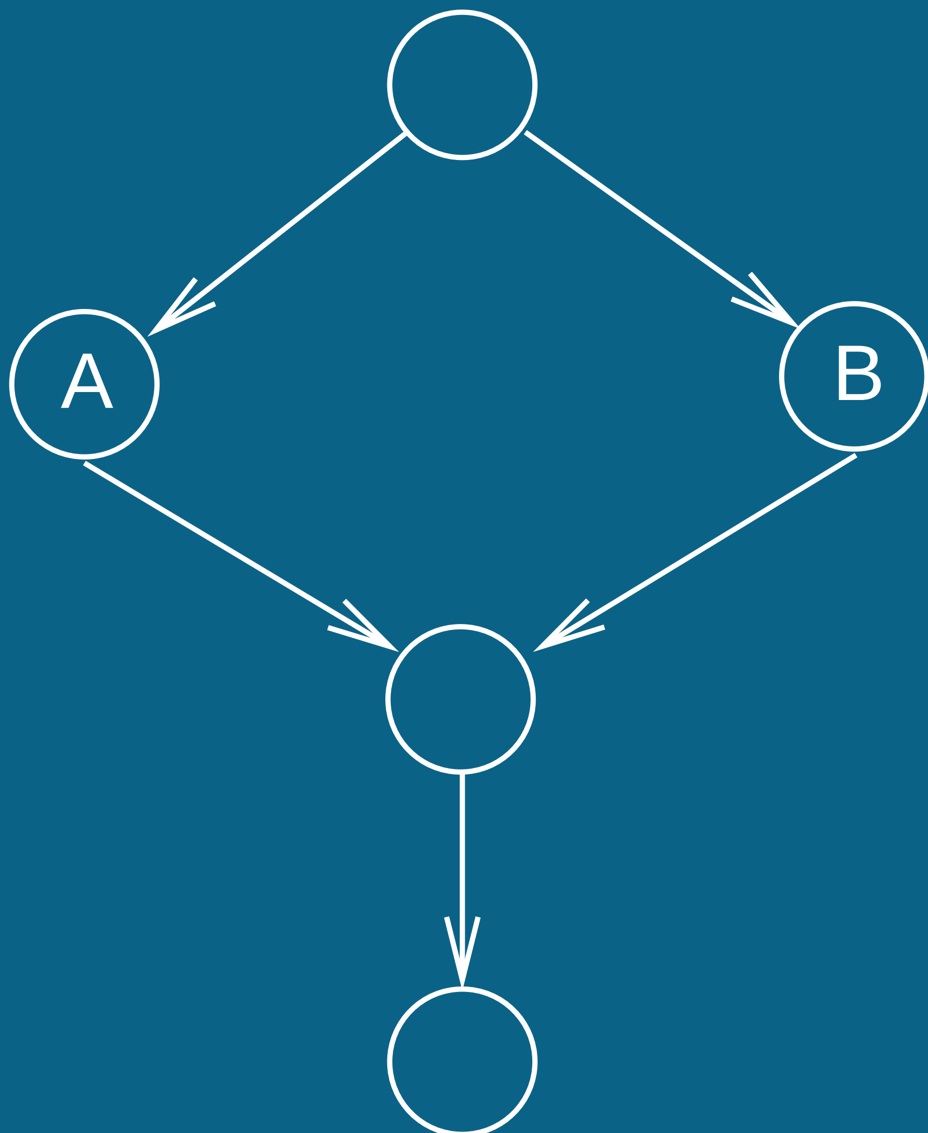



A ⊥ B
A
B
A
B
A
B
A
B



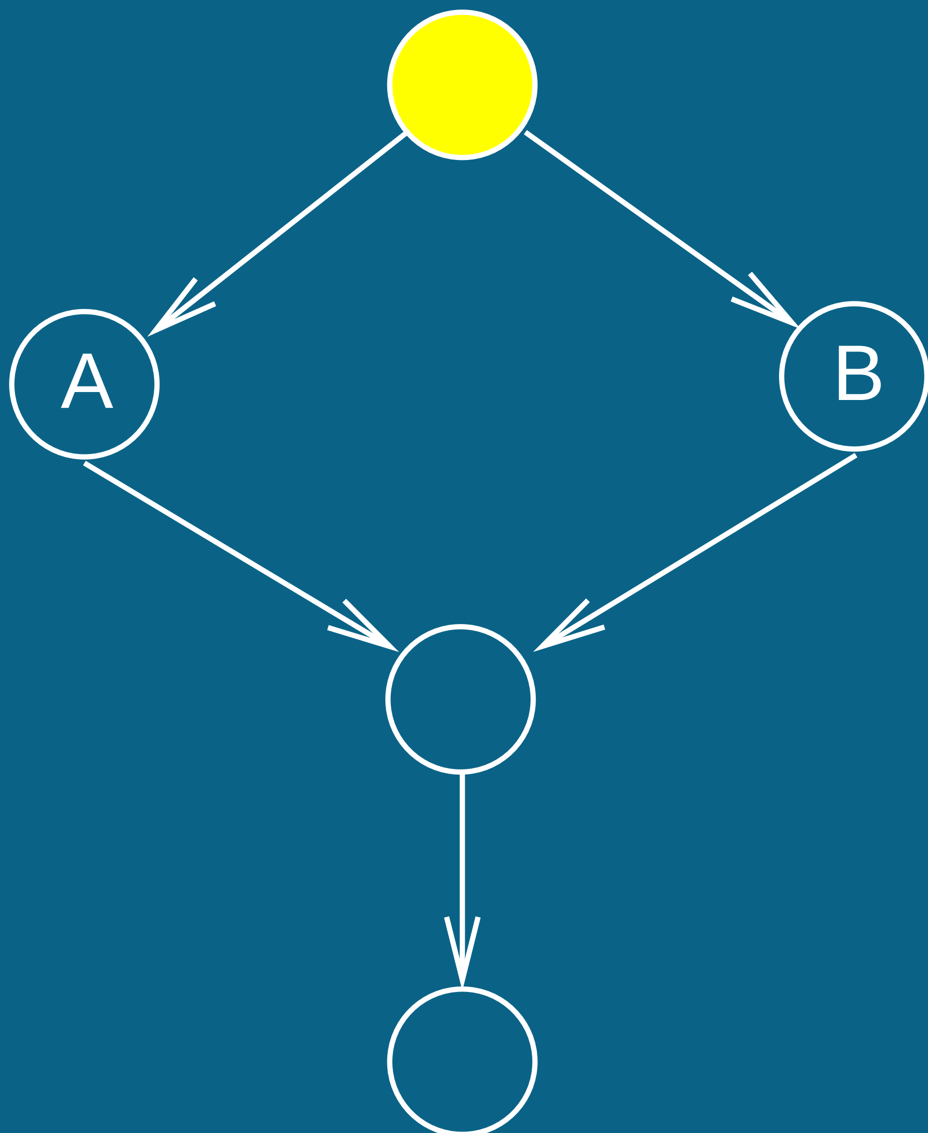
$A \perp B$

?



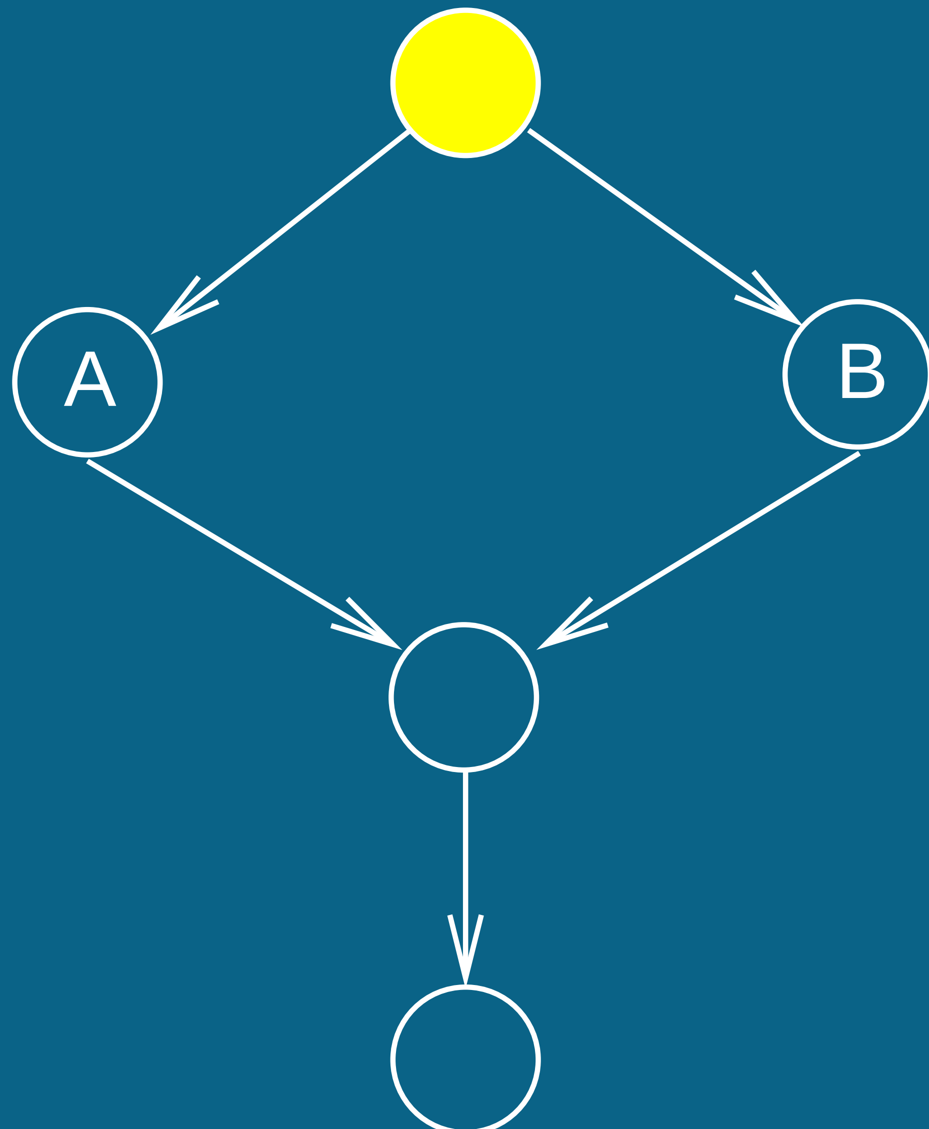
$A \perp B$

No



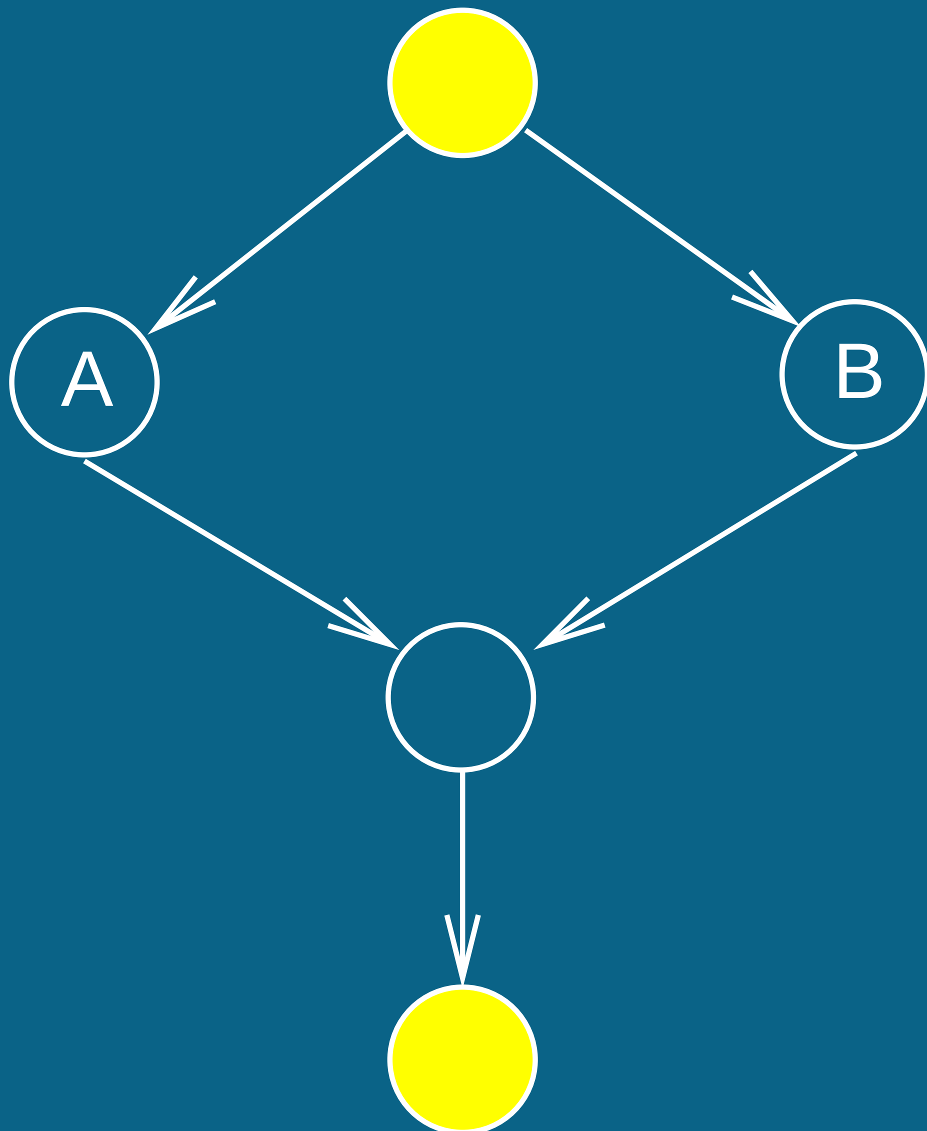
$A \perp B$

?



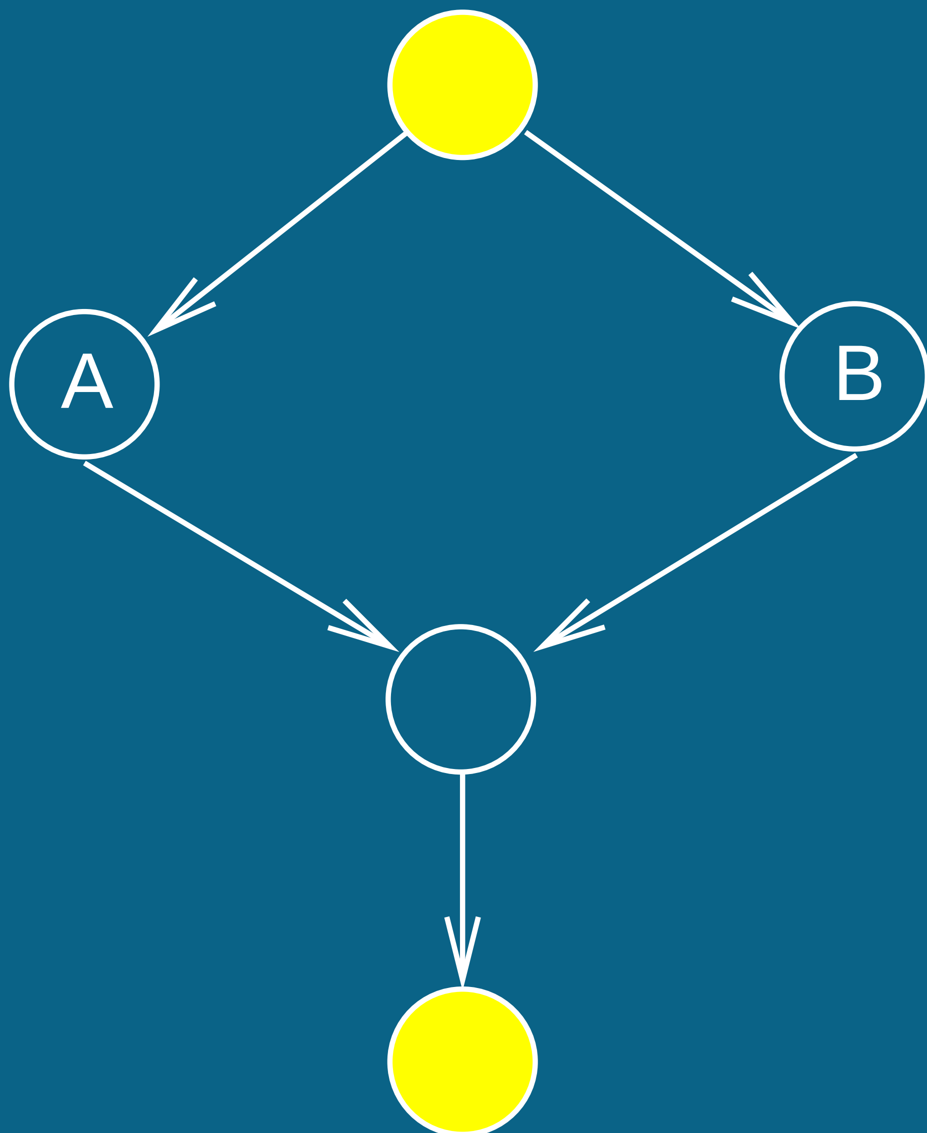
$A \perp B$

Yes



$A \perp B$

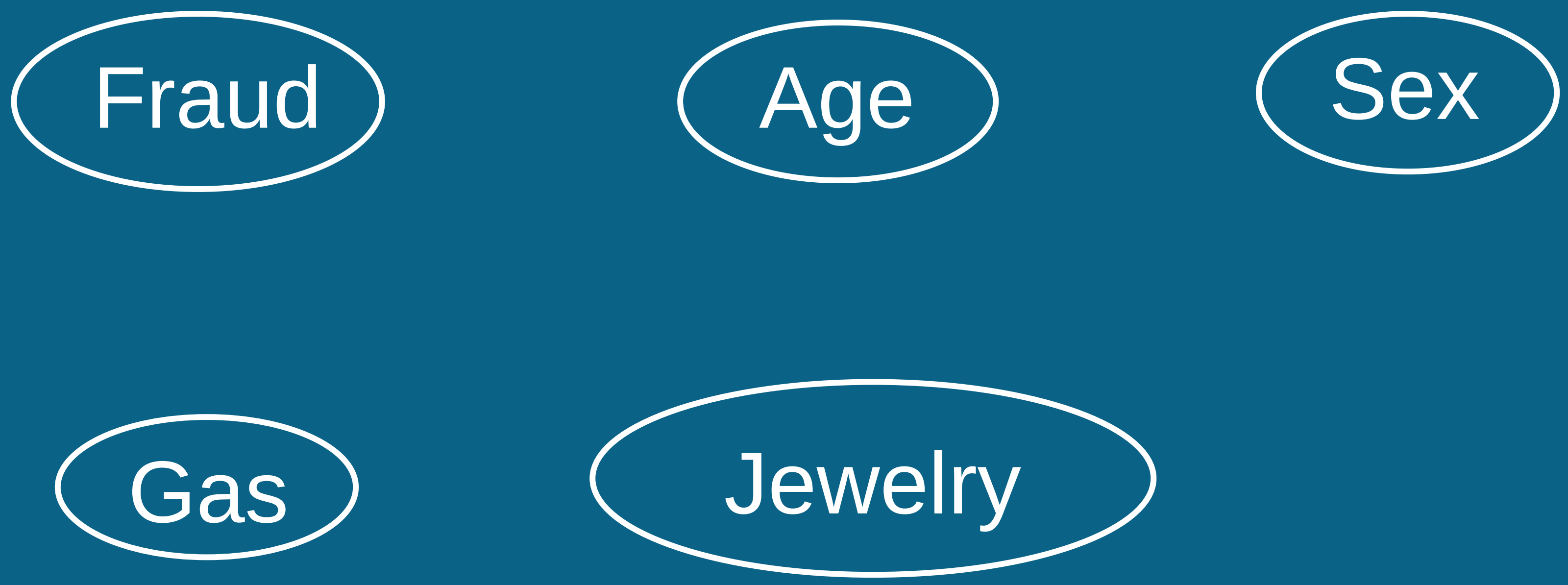
?

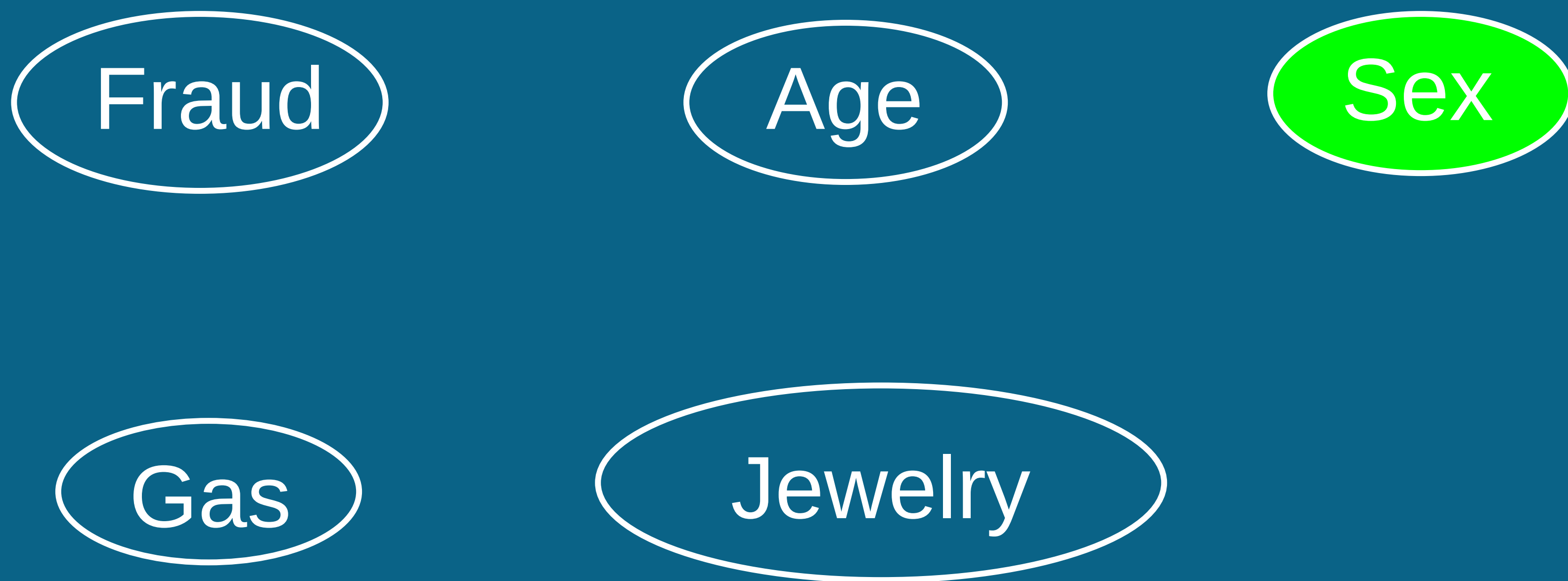


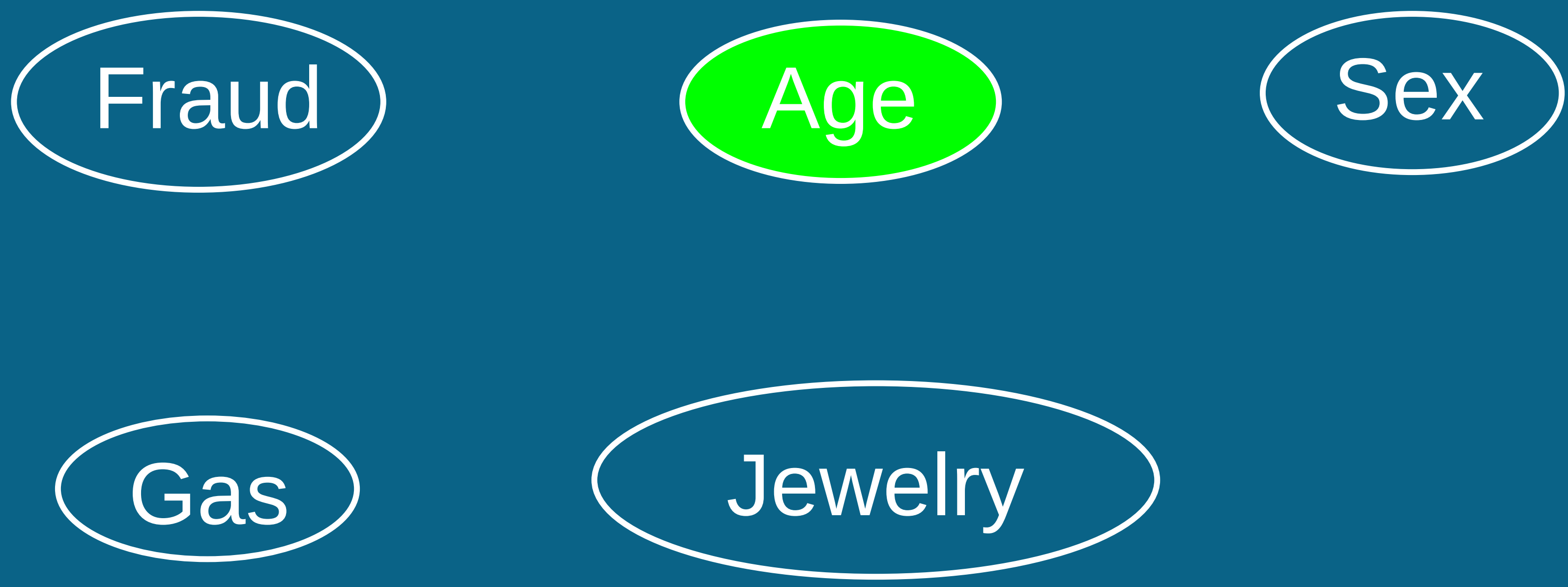
$A \perp B$

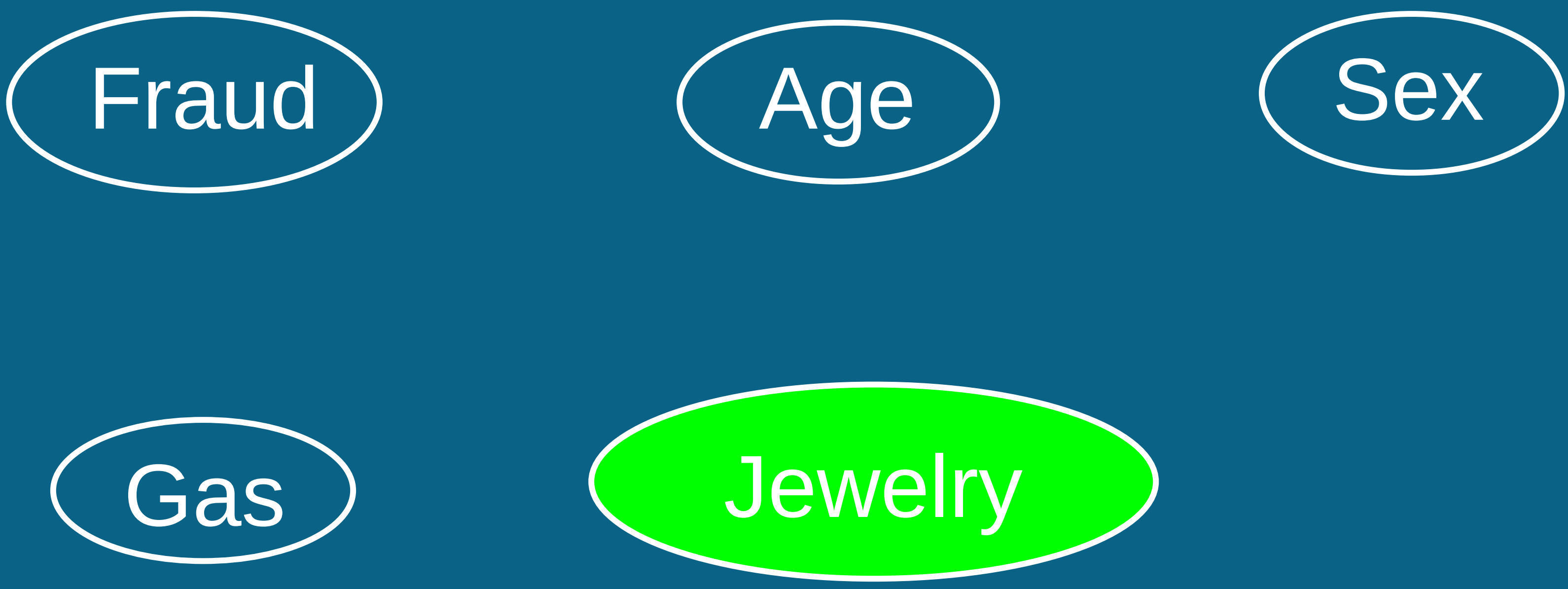
No

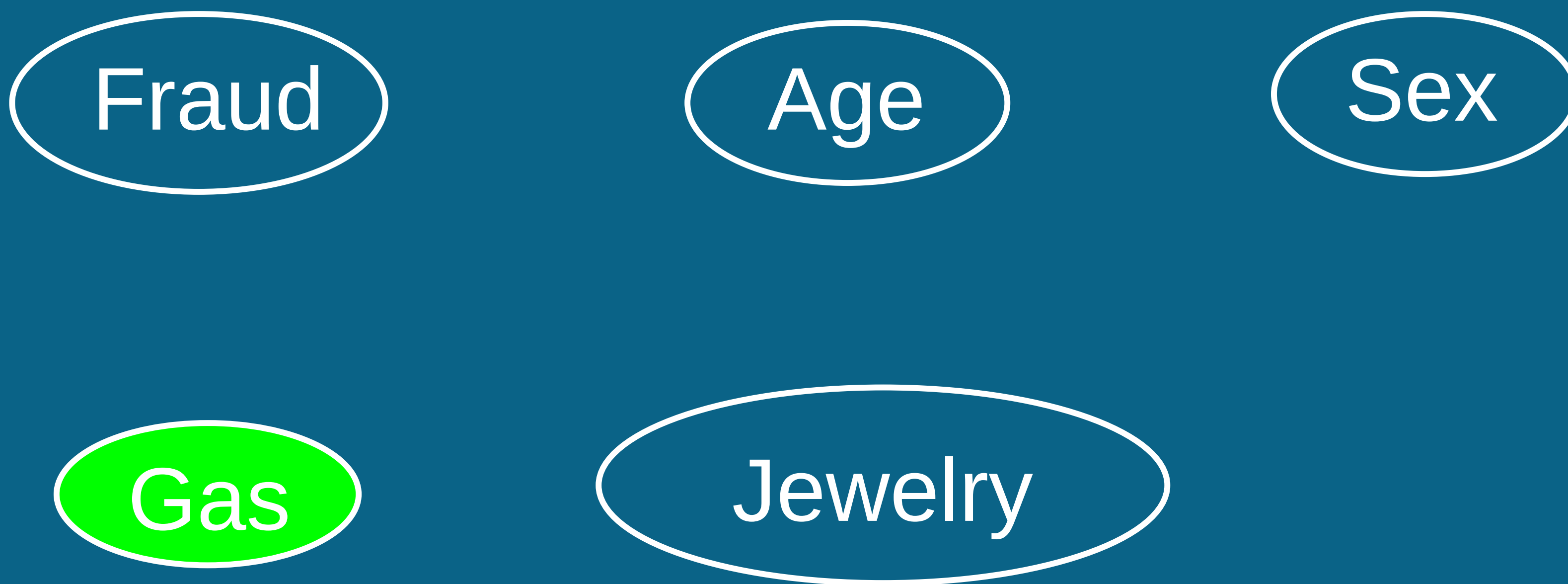
Can we learn network structures from data?

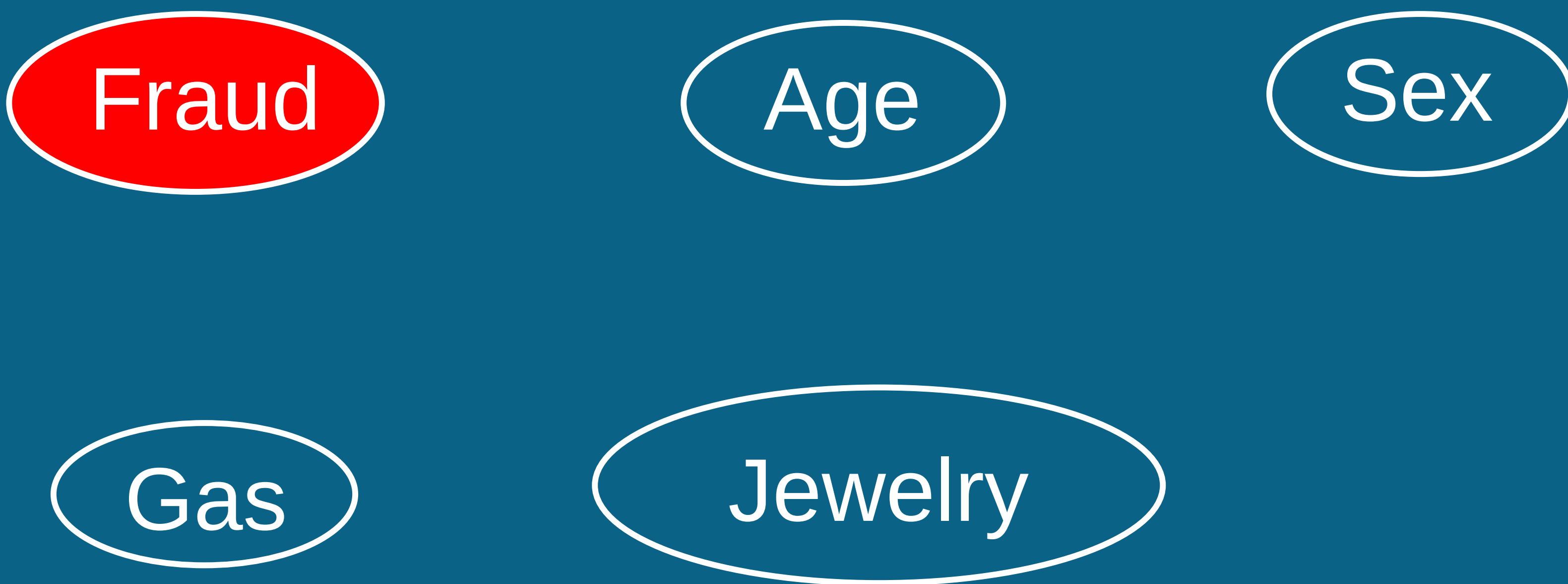

Fraud
Age
Sex
Gas
Jewelry

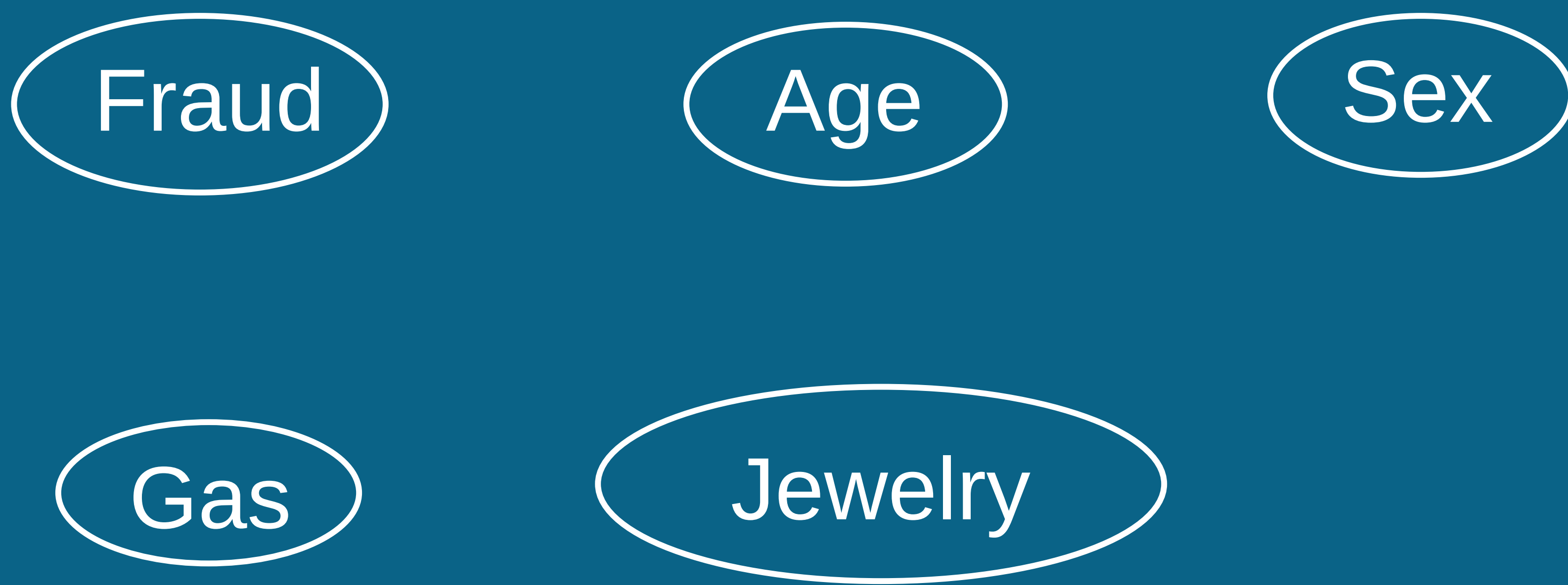

Fraud
Age
Sex
Gas
Jewelry

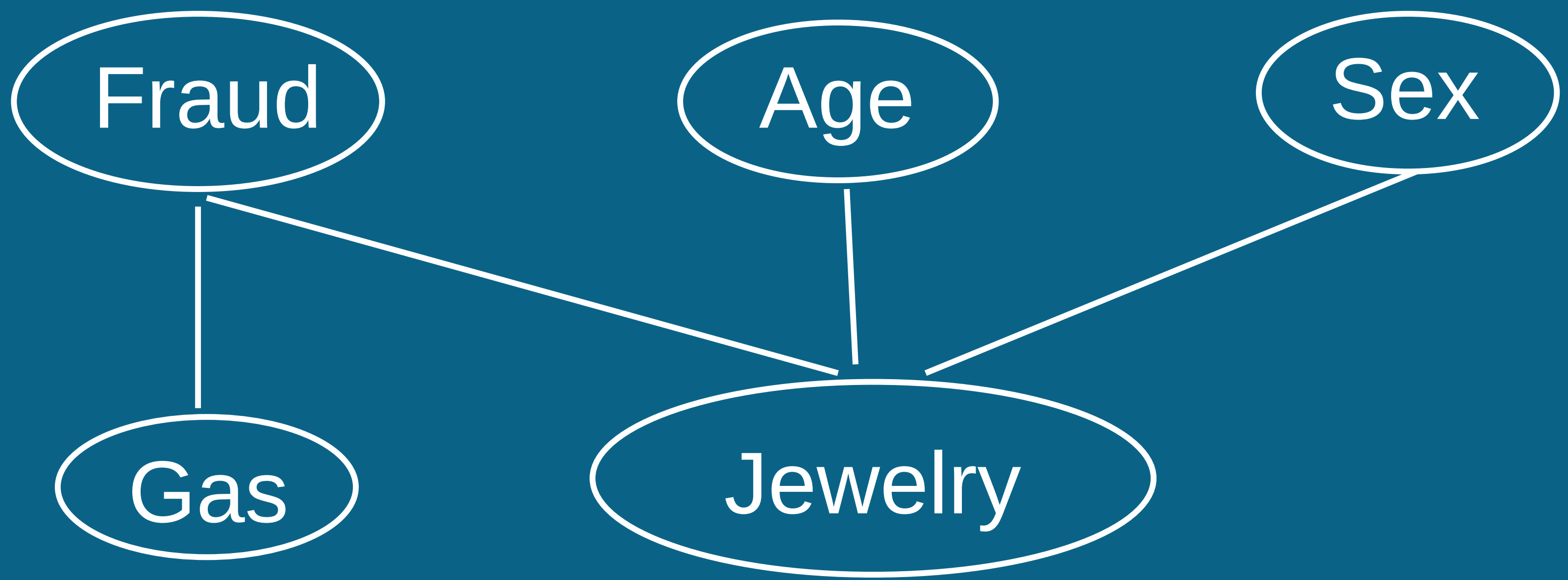

Fraud
Age
Sex
Gas
Jewelry

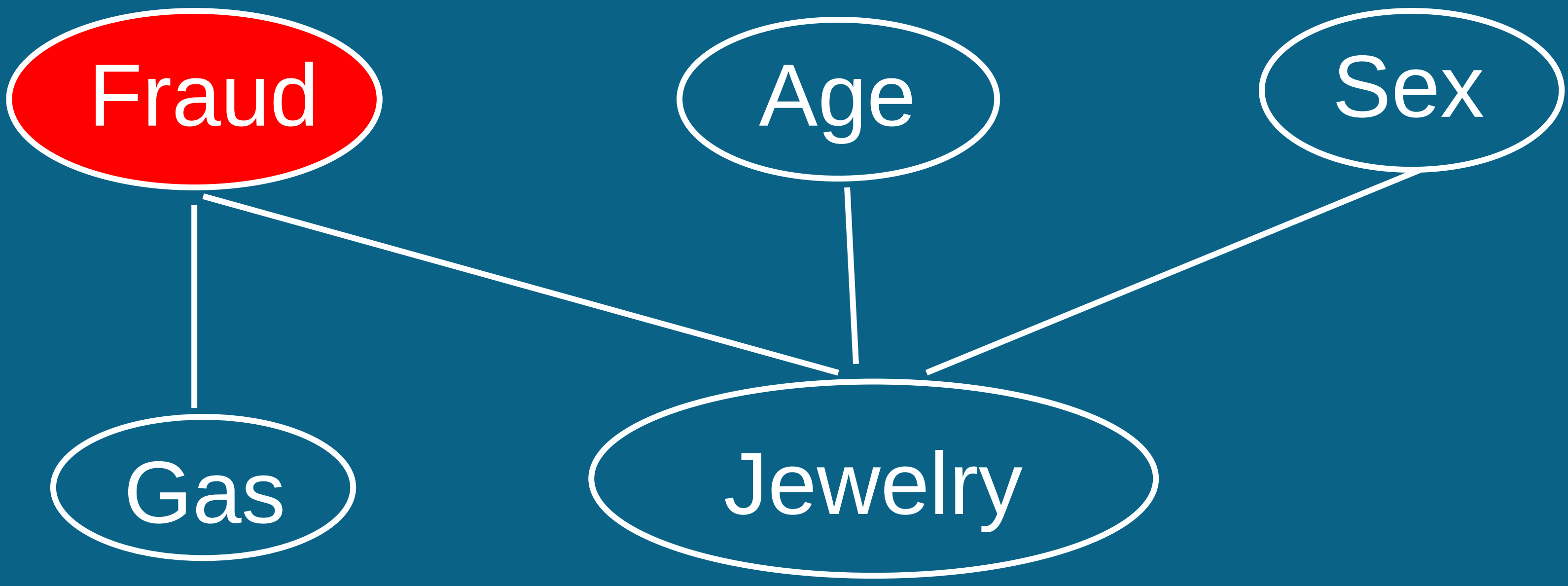

Fraud
Age
Sex
Gas
Jewelry

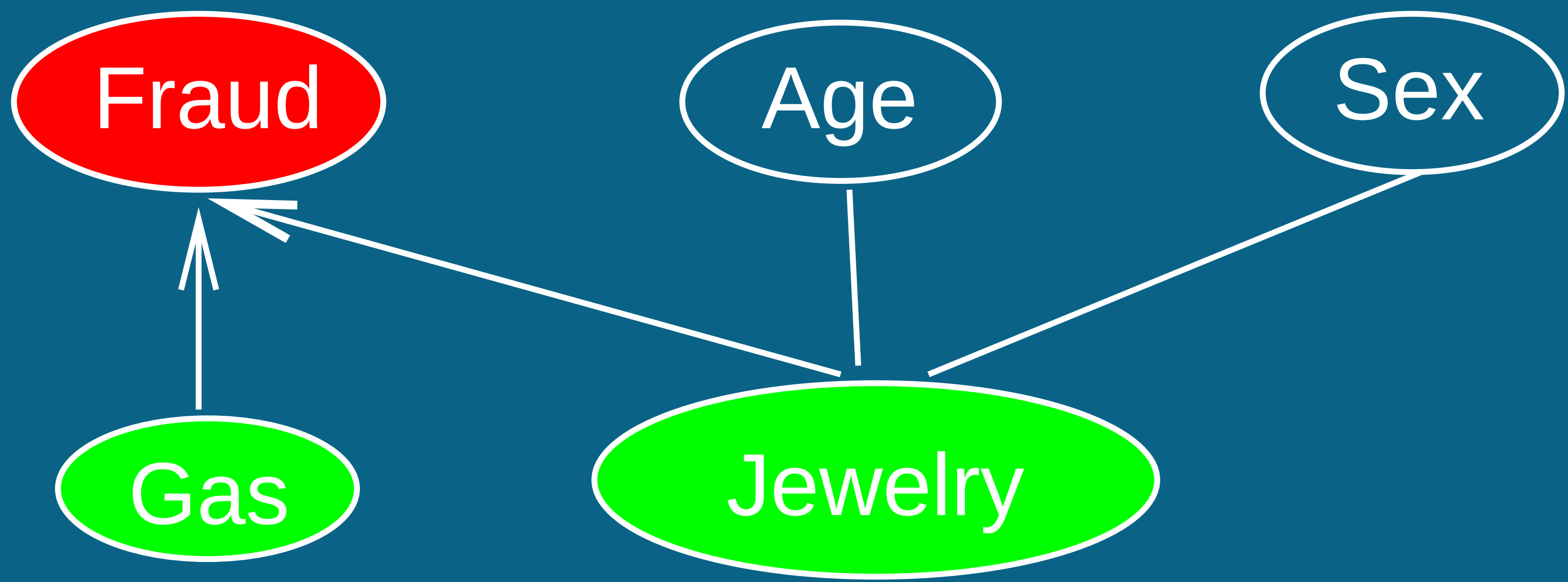

Fraud
Age
Sex
Gas
Jewelry


Fraud
Age
Sex
Gas
Jewelry


Fraud
Age
Sex
Gas
Jewelry

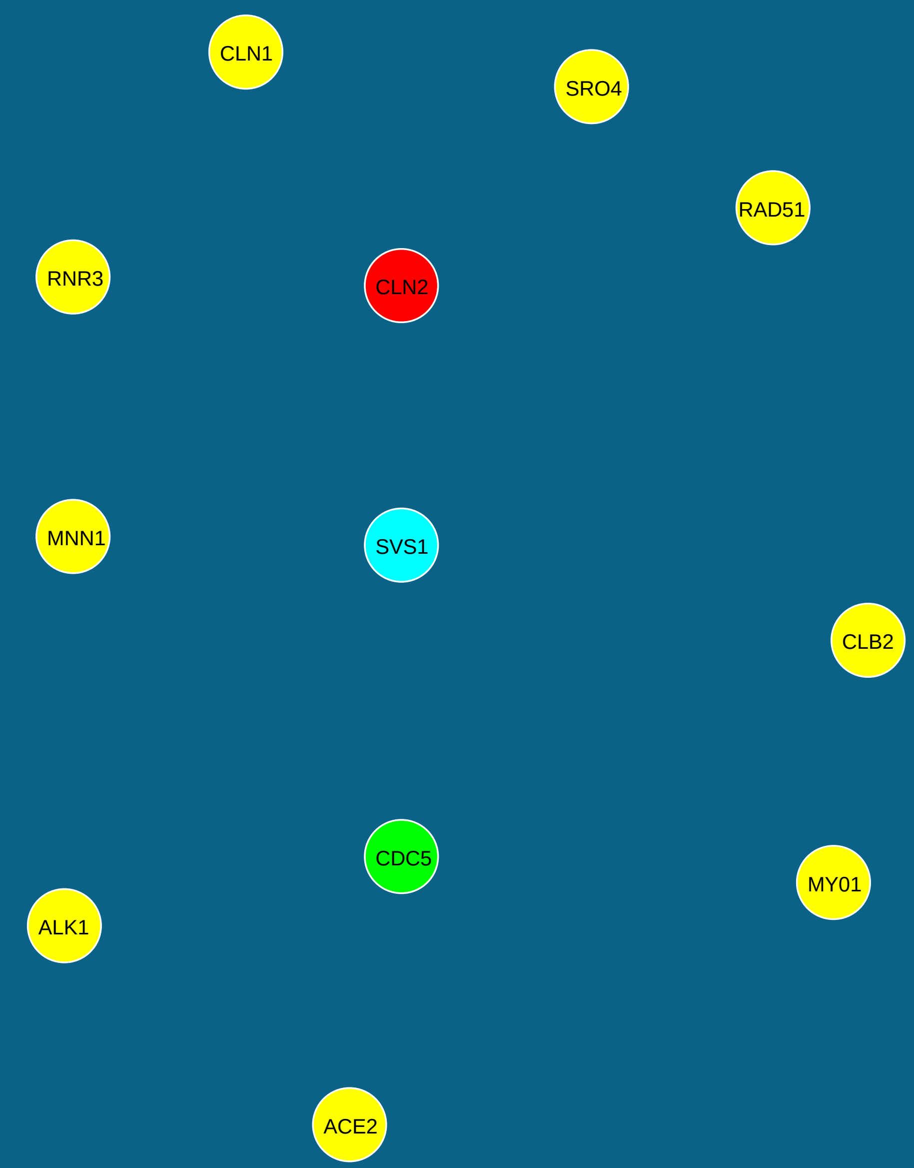

Fraud
Age
Sex
Gas
Jewelry

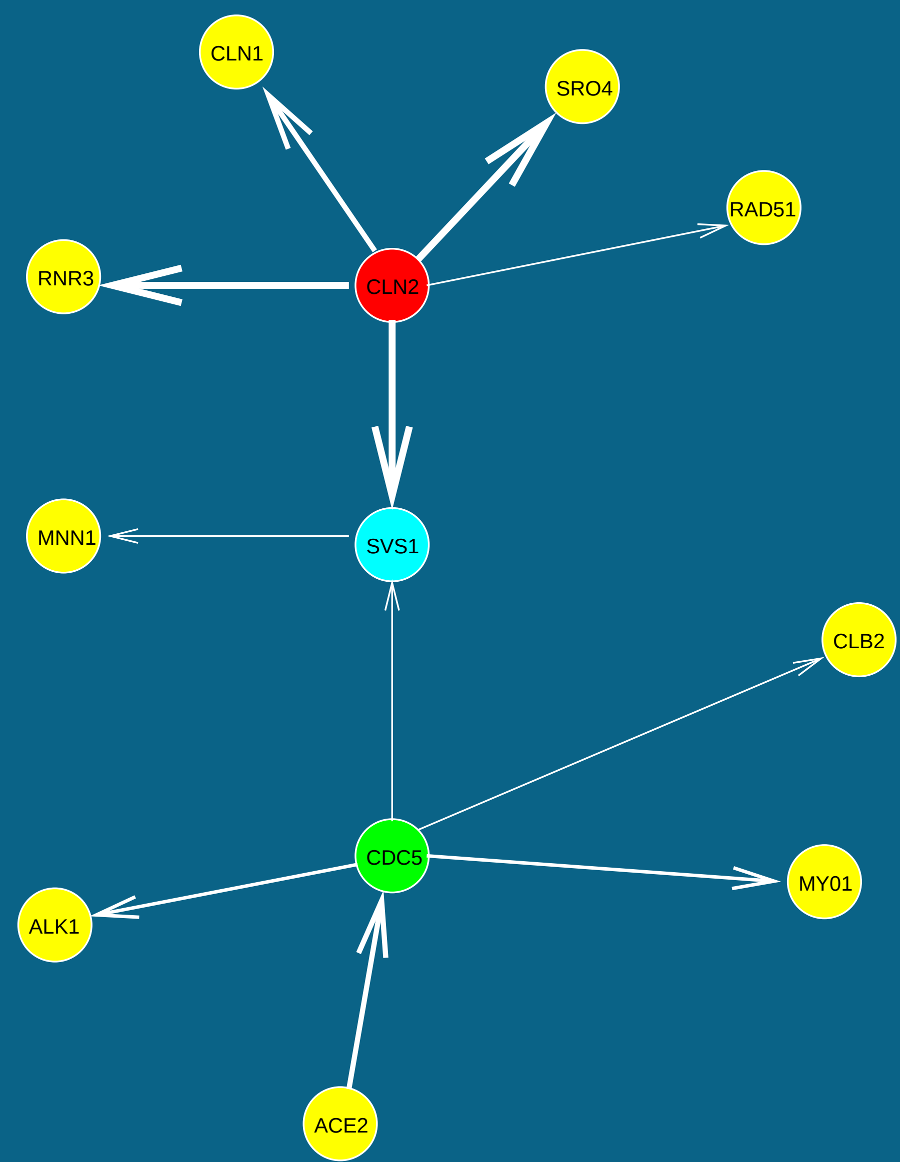

Fraud
Age
Sex
Gas
Jewelry



Fraud
Age
Sex
Gas
Jewelry

Application: Genetic Networks


CLN1
SRO4
RAD51
RNR3
CLN2
MNN1
SVS1
CLB2
CDC5
MY01
ALK1
ACE2


CLN1
SRO4
RAD51
RNR3
CLN2
MNN1
SVS1
CLB2
CDC5
ALK1
MY01
ACE2

Bayesian inference

Bayesian inference

Heads Tails

Probability θ $1-\theta$

Data

N tosses, k observations of "heads"

Bayesian inference

Heads Tails

Probability θ $1-\theta$

Data

N tosses, k observations of "heads"

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

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

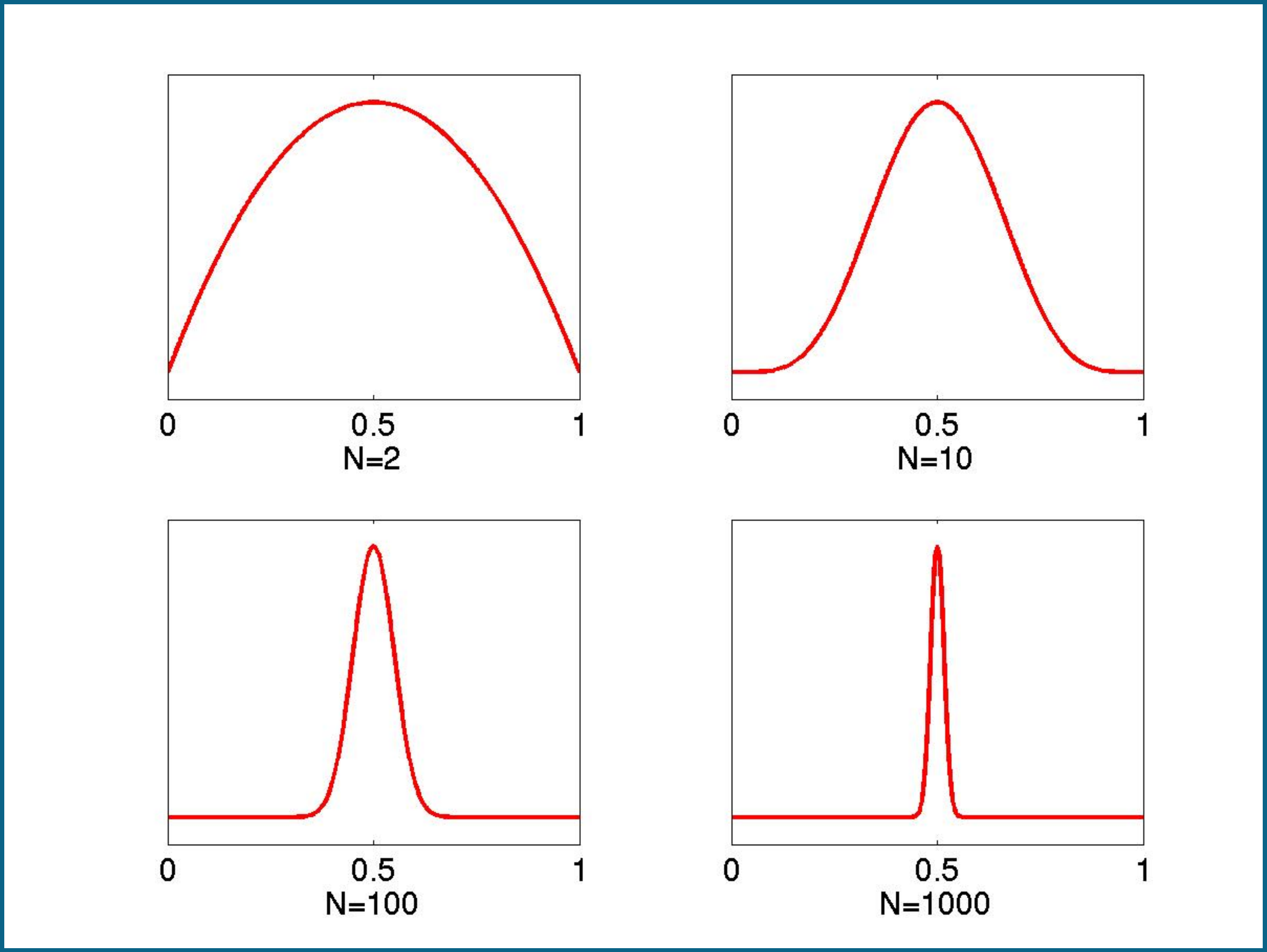
$$P(D|\theta) \propto \theta^k(1-\theta)^{N-k}$$

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

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

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

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



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

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

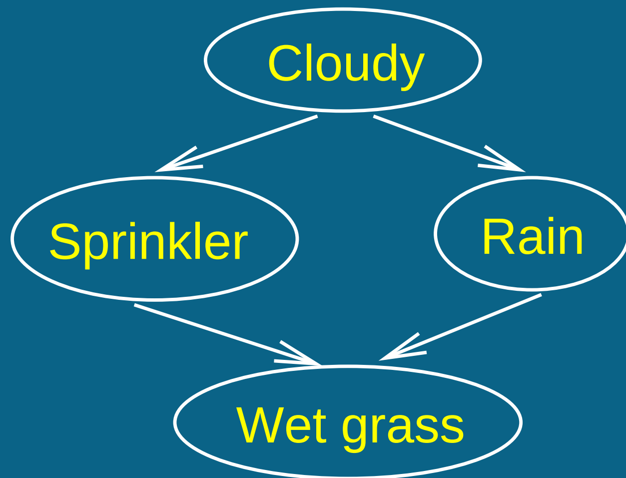
$$P(D|M) = \int P(D|\theta, M)P(\theta|M)d\theta$$

When is the integral analytically tractable?

- 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.

Example: multinomial CPD

P(true)	P(false)
0.5	0.5



Cloudy	P(true)	P(false)
true	0.3	0.7
false	0.8	0.2

Cloudy	P(true)	P(false)
true	0.8	0.2
false	0.0	1.0

Sprinkler	Rain	P(true)	P(false)
true	true	1.0	0.0
true	false	0.8	0.2
false	true	0.9	0.1
false	false	0.0	1.0

Multinomial CPD

Disadvantage

- Requires discretization $\longrightarrow$ Loss of information.
- Number of free parameters exponential in the number of parents.

Advantage

- Can model nonlinear interactions.

Linear Gaussian CPD

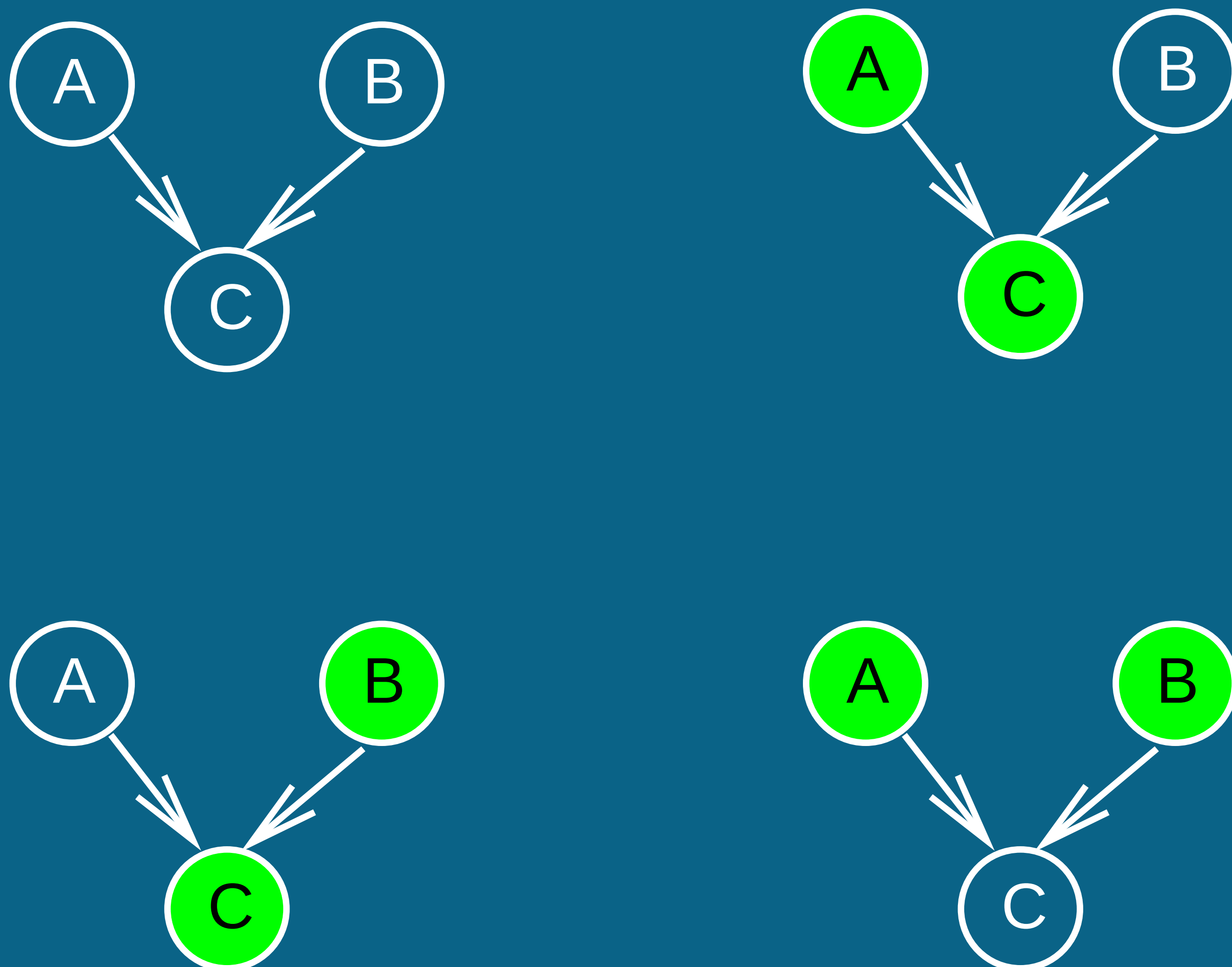
$$P(A|P_1,\ldots,P_n) = N\left(w_0 + \sum_{i=1}^{n} w_i P_i, \sigma^2\right)$$

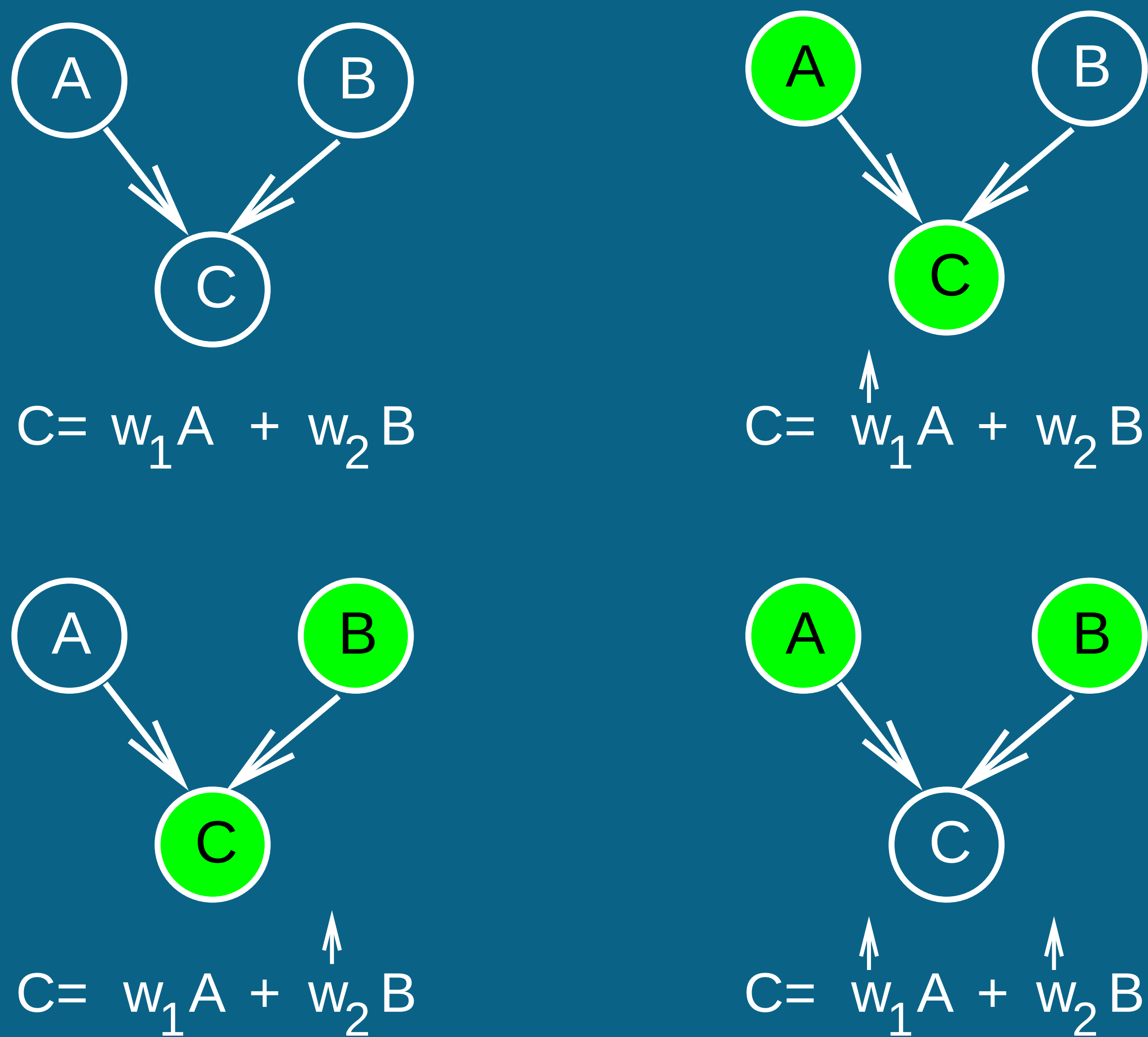
Advantage

- Number of free parameters linear in the number of parents.
- No discretization $\longrightarrow$ No loss of information.

Disadvantage

- Can only model linear interactions.


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



A
B
C
C= w_1A + w_2B
A
B
C
C= w_1A + w_2B
A
B
C
C= w_1A + w_2B
A
B
C
C= w_1A + w_2B

Structure learning

Objective

Find the best network structure M, that is, the one that maximizes $P(M|D)$, or find $P(M|D)$.

Naive approach

Compute $P(M|D)$ for all possible network structures.

Objective

Find the best network structure M, that is, the one that maximizes $P(M|D)$, or find $P(M|D)$.

Naive approach

Compute $P(M|D)$ for all possible network structures.

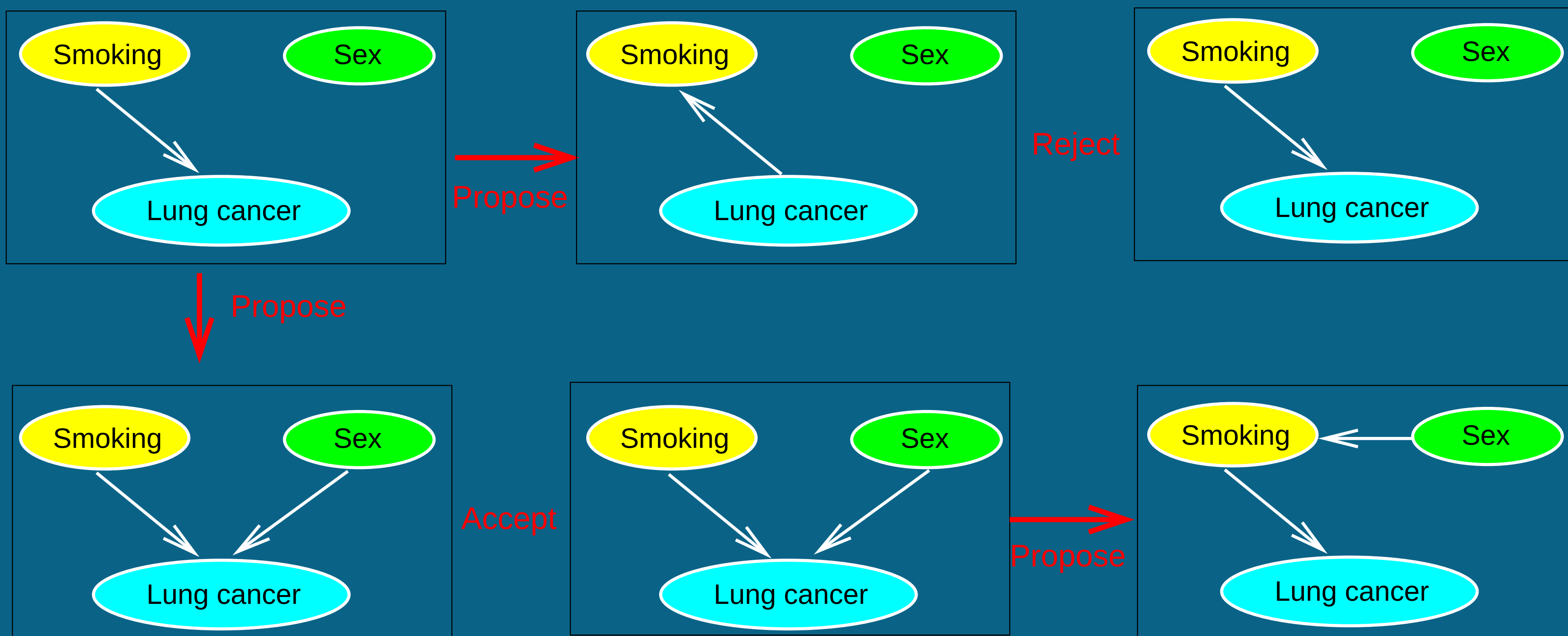
Problem

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

Example

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}

Markov chain Monte Carlo (MCMC)



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

Objective: Sample from the posterior distribution

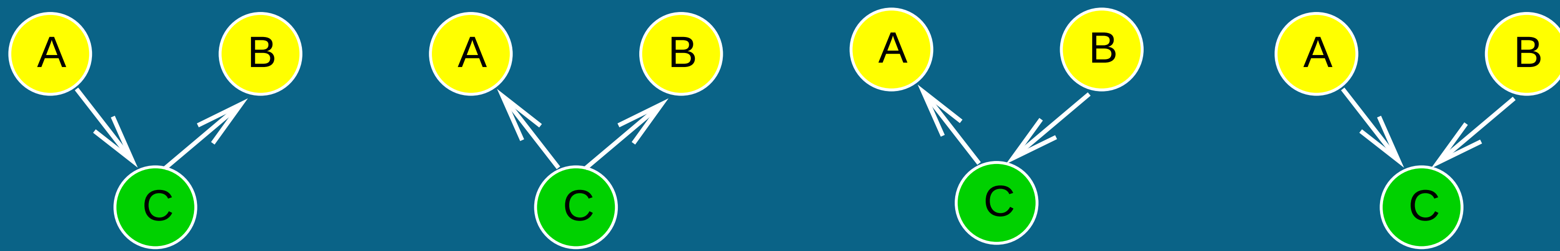
$$P(M_k|D) = \frac{P(D|M_k)P(M_k)}{\sum_i P(D|M_i)P(M_i)}$$

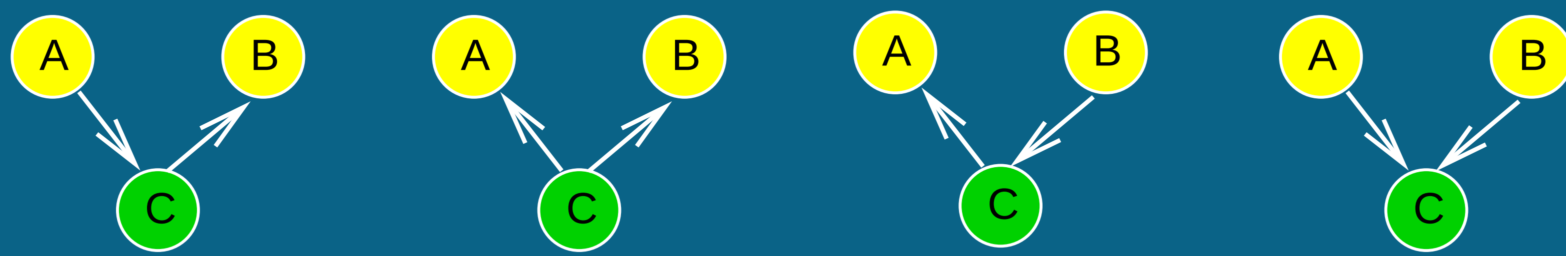
Direct approach intractable due to $\sum_i P(D|M_i)P(M_i)$

Markov chain Monte Carlo:

- 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 $\min\left\{1, \frac{P(D|M_{new})P(M_{new})}{P(D|M_{old})P(M_{old})} \times \frac{Q(M_{old}|M_{new})}{Q(M_{new}|M_{old})}\right\}$

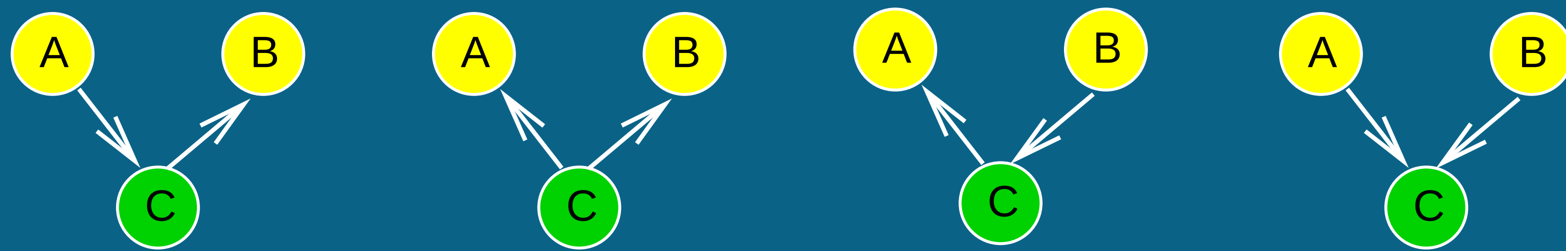
Equivalent classes and PDAGs


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



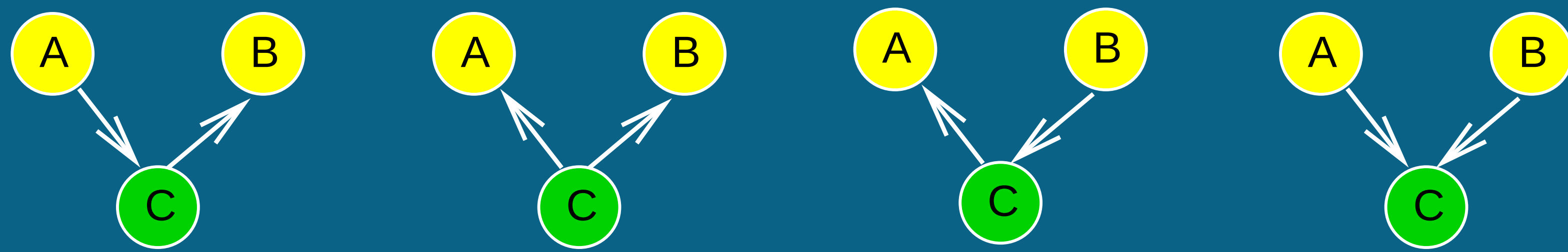
$$P(A,B,C) =$$

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



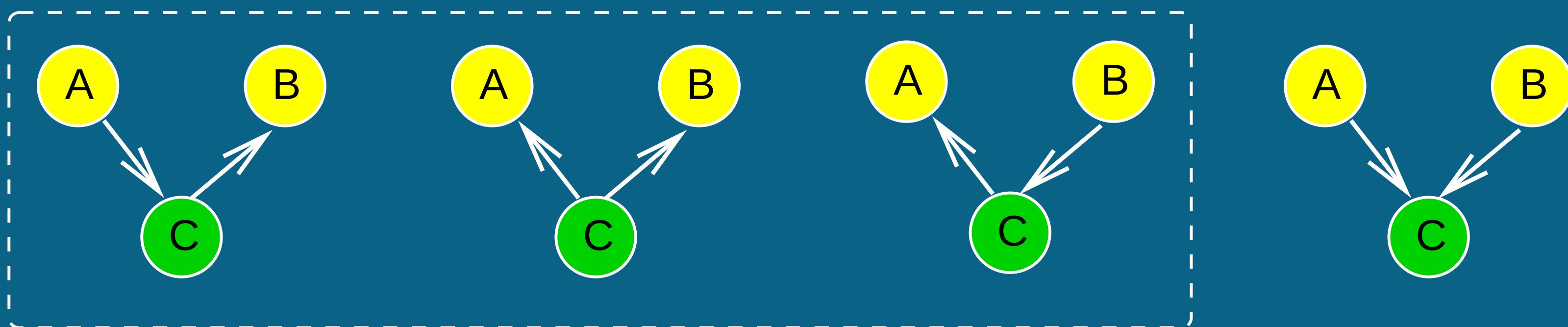
$$P(A,B,C) =$$

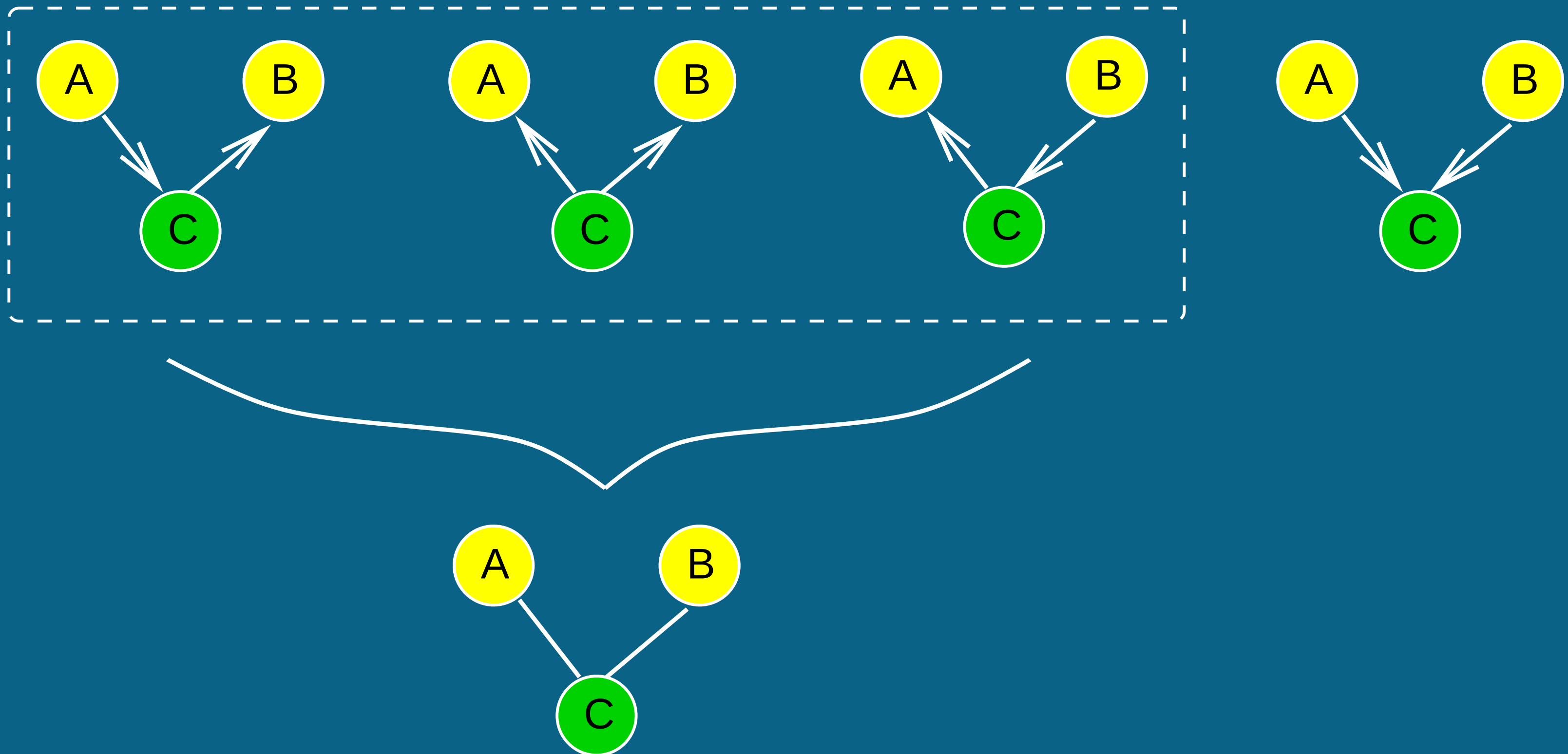
$$P(B|C)\,\underbrace{P(C|A)\,P(A)}_{P(A|C)\,P(C)} \qquad P(A|C)\,P(B|C)\,P(C) \qquad P(A|C)\,\underbrace{P(C|B)\,P(B)}_{P(B|C)\,P(C)} \qquad P(C|A,B)\,P(A)\,P(B)$$

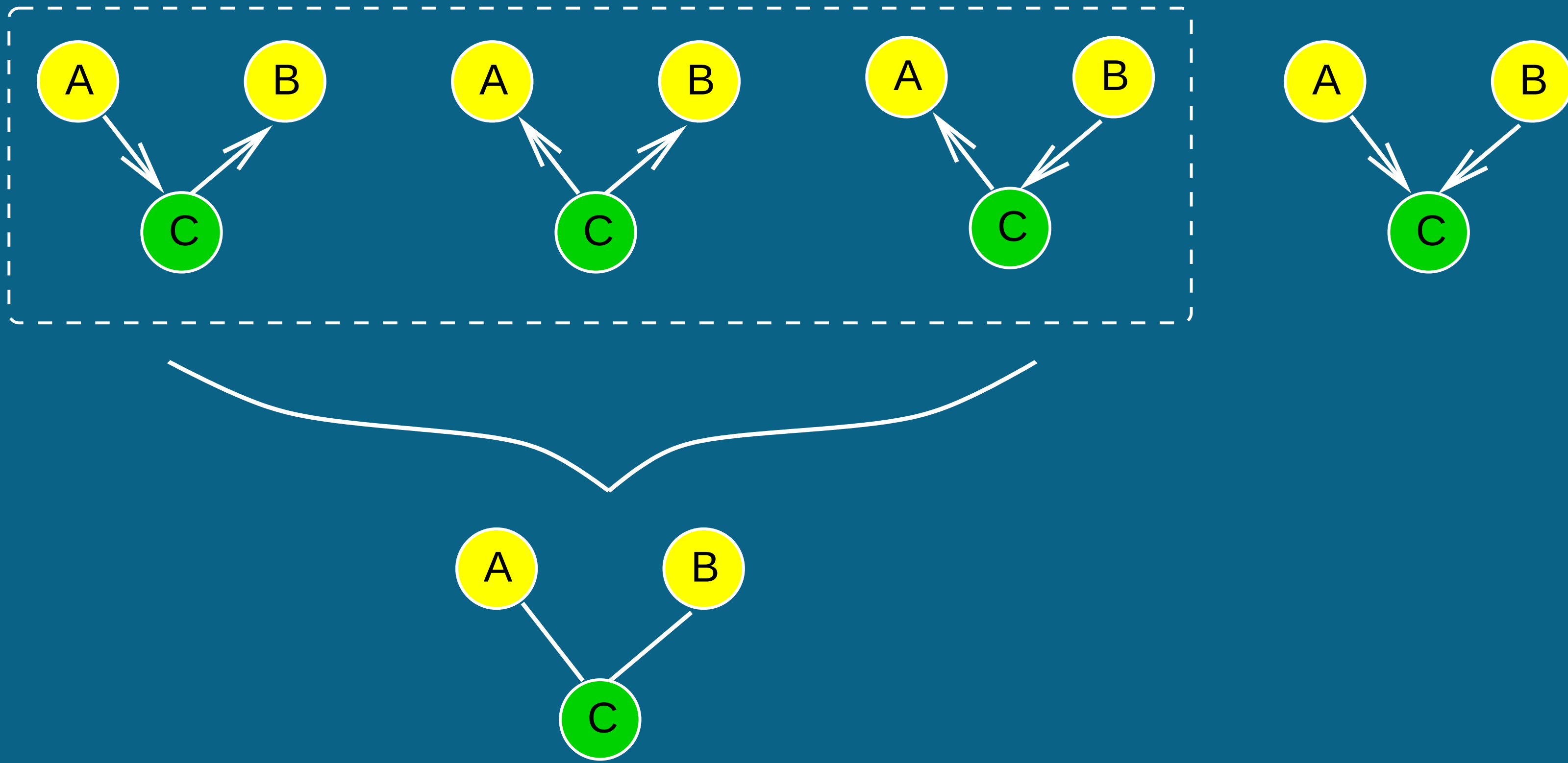


$$P(A,B,C) =$$

$$P(B|C)\,\underbrace{P(C|A)\,P(A)}_{P(A|C)\,P(C)} \qquad P(A|C)\,P(B|C)\,P(C) \qquad P(A|C)\,\underbrace{P(C|B)\,P(B)}_{P(B|C)\,P(C)} \qquad P(C|A,B)\,P(A)\,P(B)$$



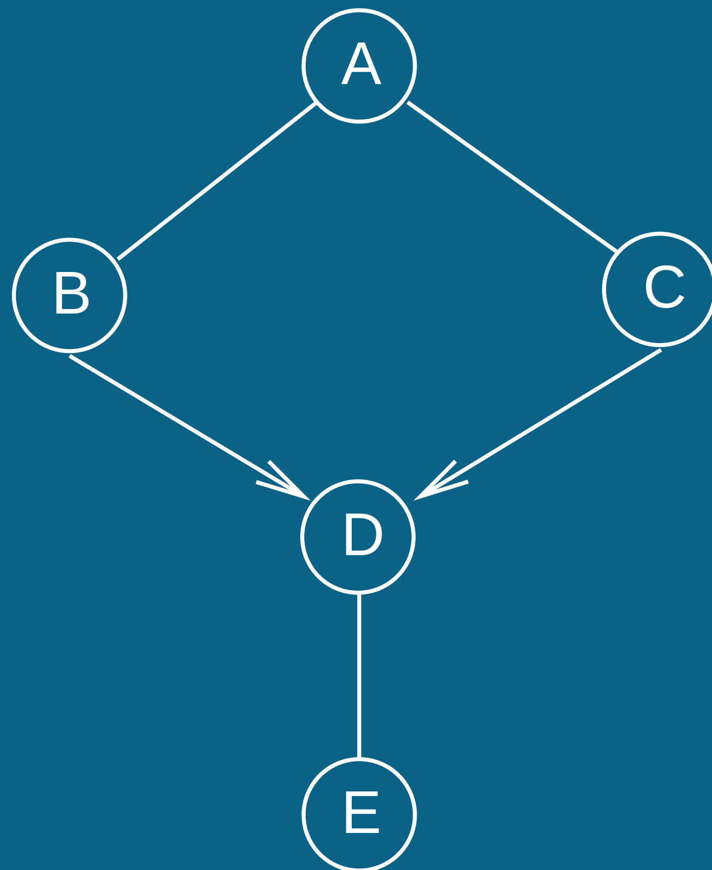

A
B
C
A
B
C
A
B
C
A
B
C
A
B
C



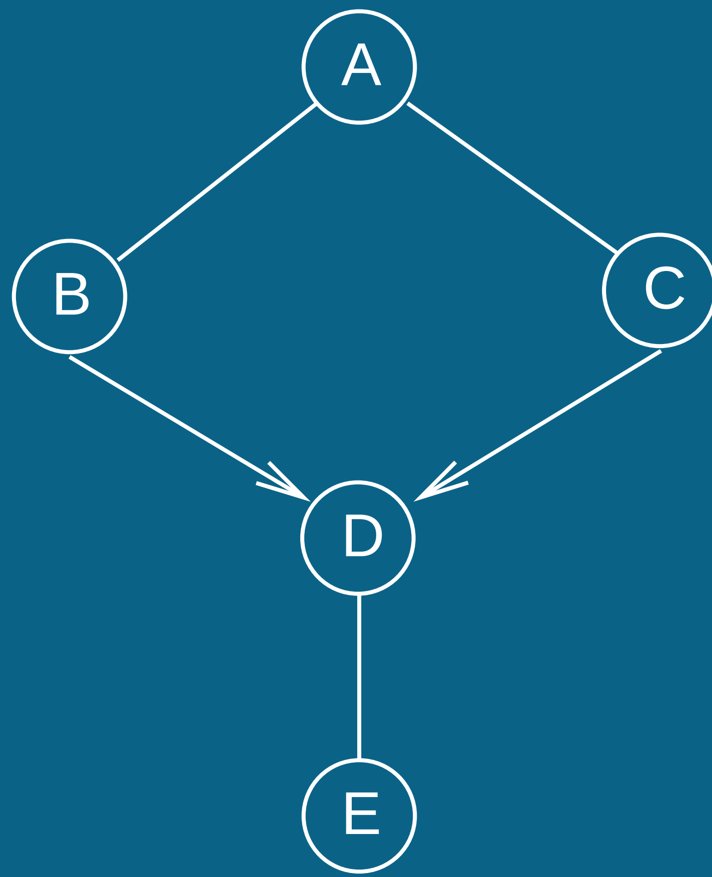
- Two DAGs are equivalent iff they have the same skeleton (= the underlying undirected graph) and the same v-structure.

- 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).



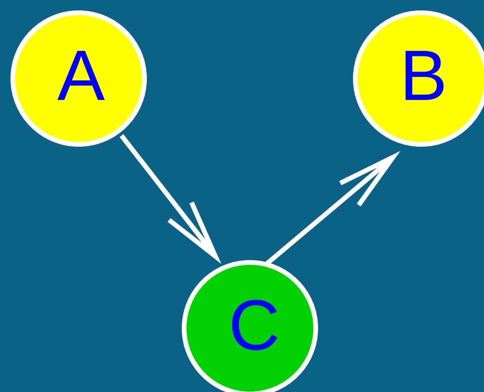
An equivalence class of DAGs can be represented by a PDAG (partially directed acyclic graph).



We can only learn PDAGs from the data!

Can we distinguish between
correlation and causality ?

Can we distinguish between
correlation and causality ?



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

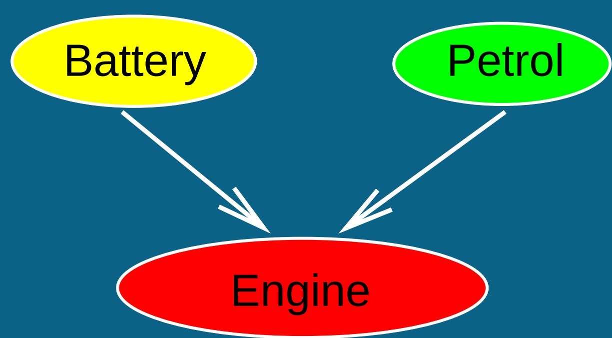
Causal network: A node is independent of its earlier causes, given its immediate causes.

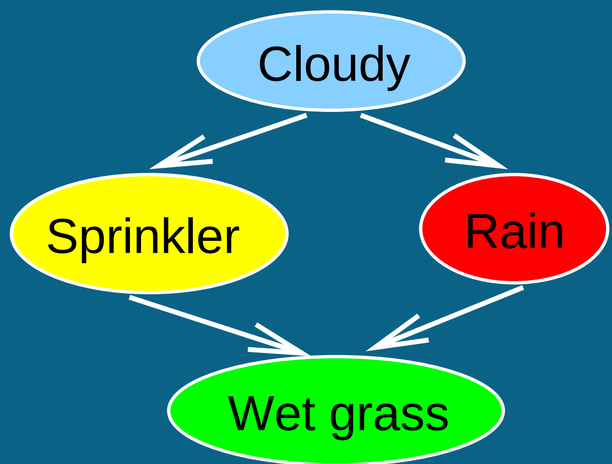
Intuitive notion

Causal explanation of a dependency model

=

Conditional independence facts entailed by the corresponding DAG



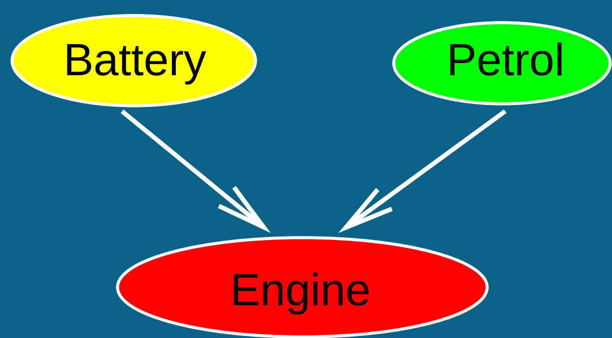


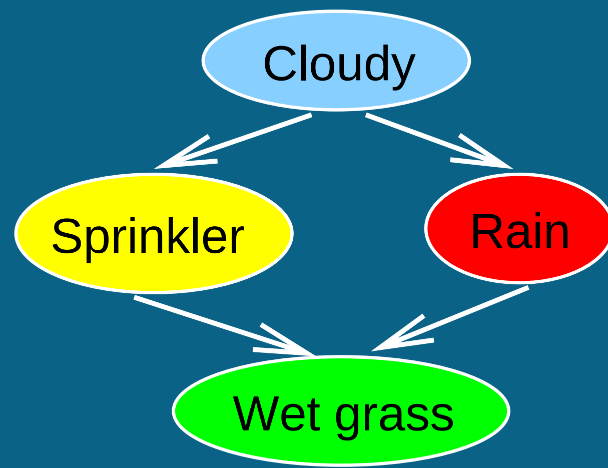
Intuitive notion

Causal explanation of a dependency model

=

Conditional independence facts entailed by the corresponding DAG



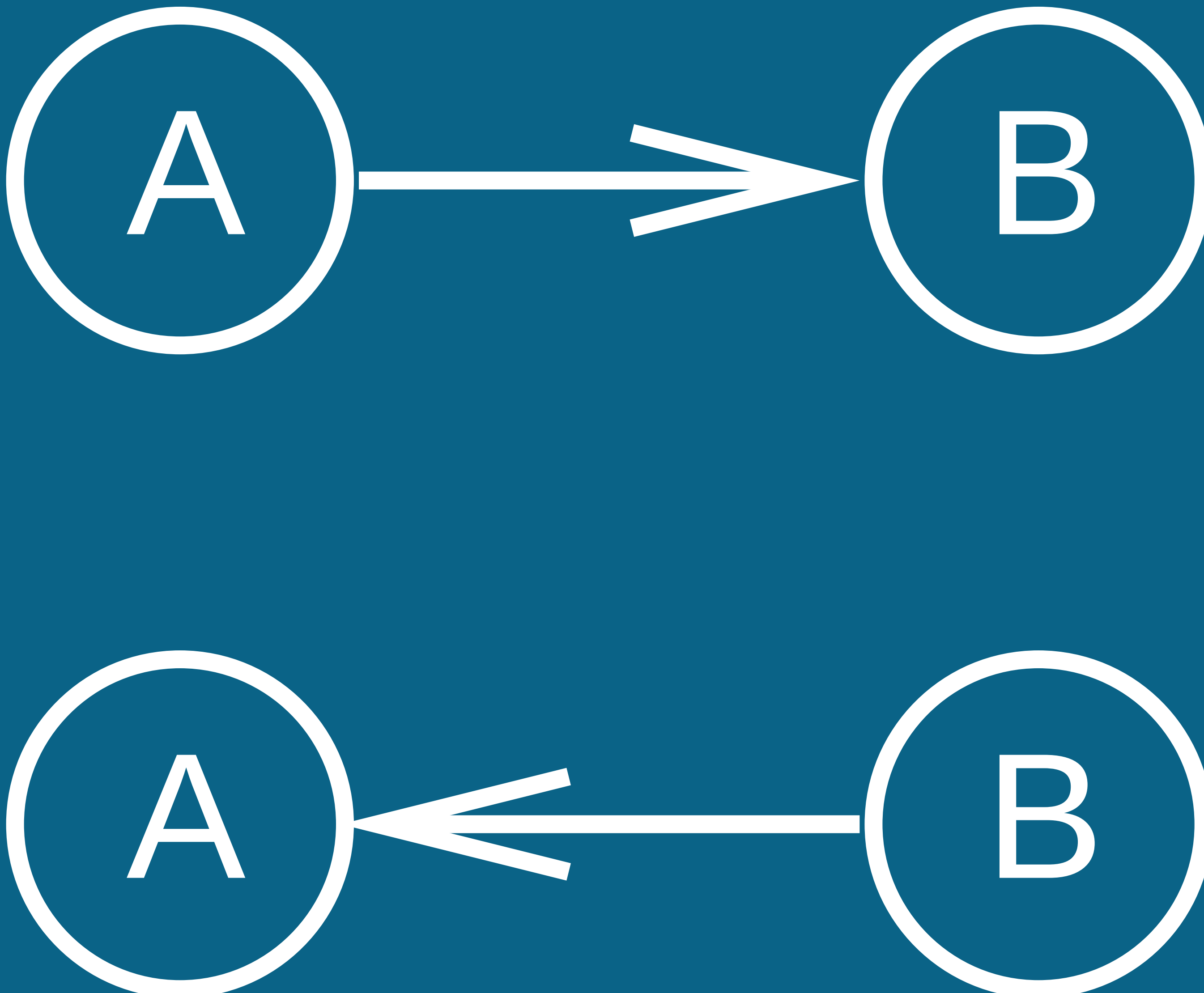


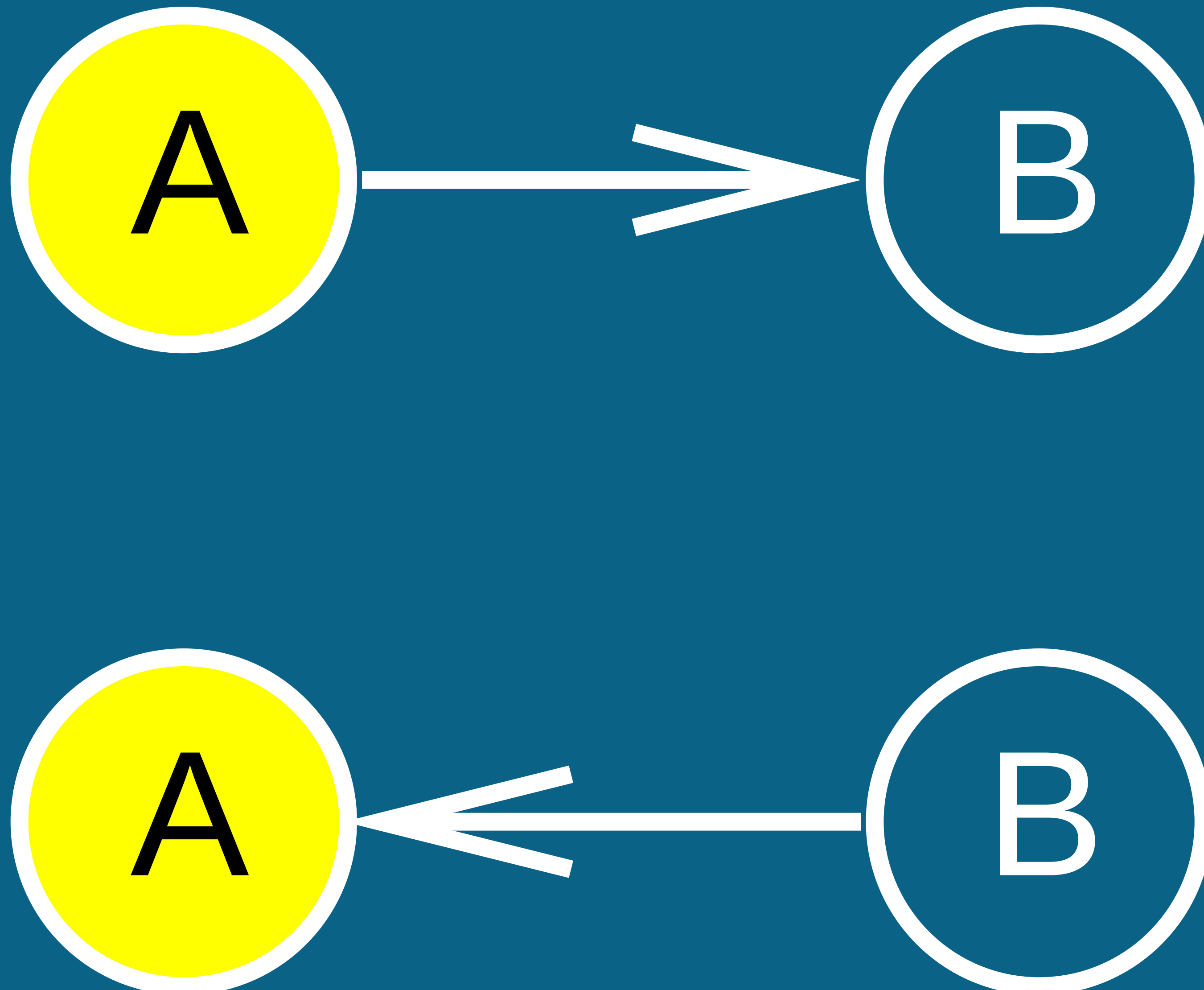
Causal Markov assumption

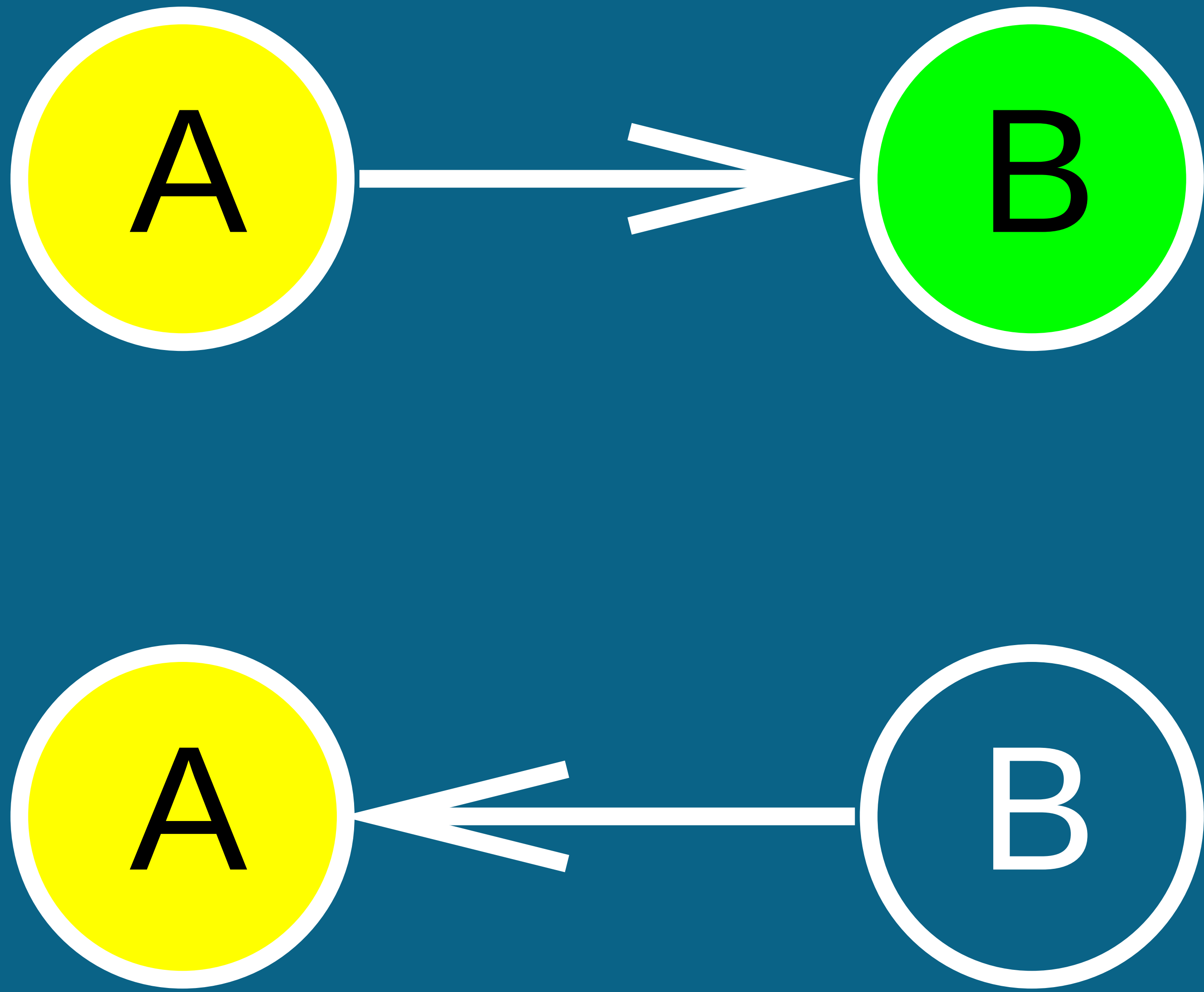
A domain with causal relationships given by a graph $\mathcal{G}$ satisfies the conditional independencies of the corresponding Bayesian network.

- 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

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

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

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

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

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

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.

Can we distinguish between
correlation and causality
without interventions?

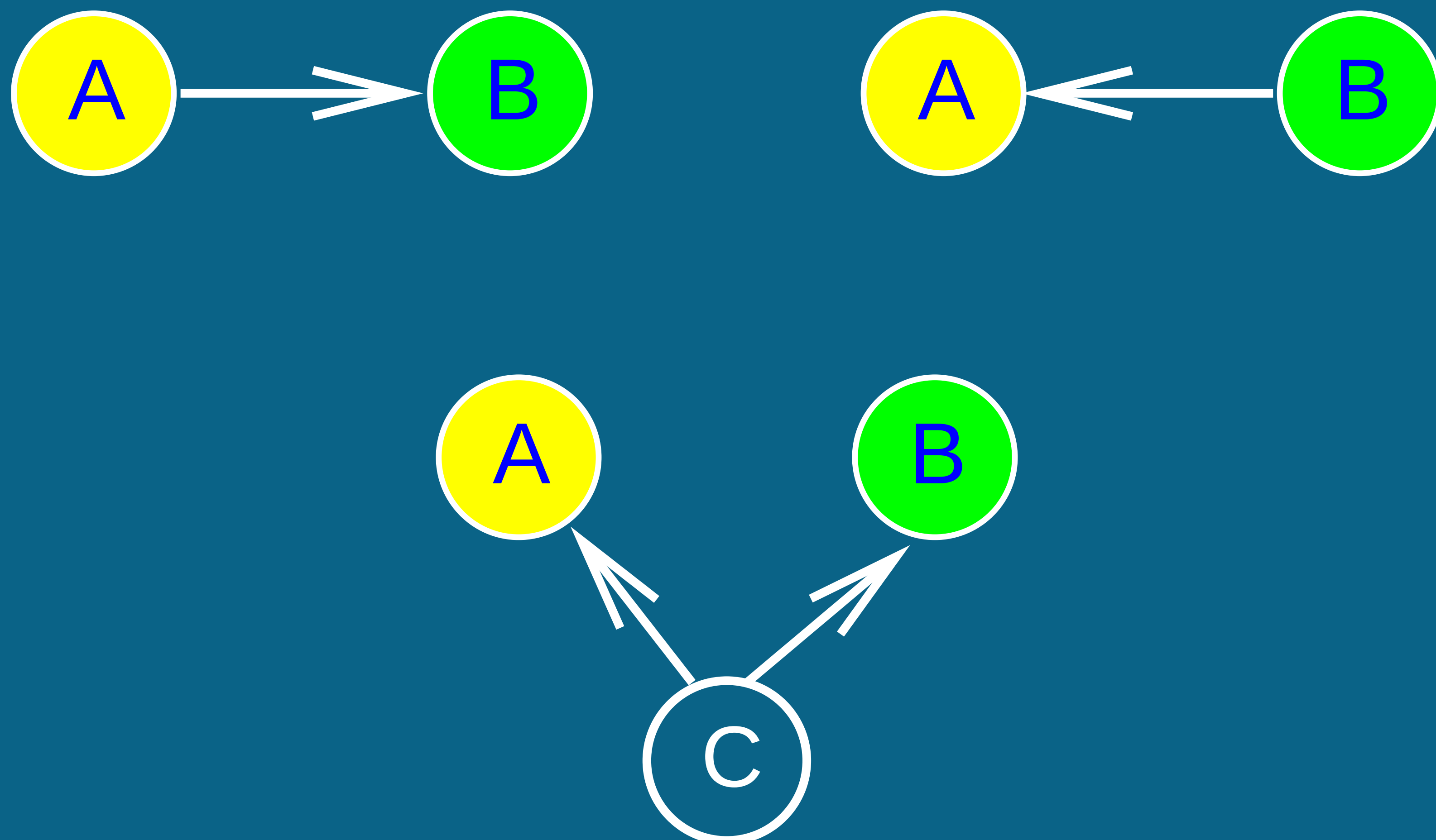
Can we distinguish between correlation and causality without interventions?

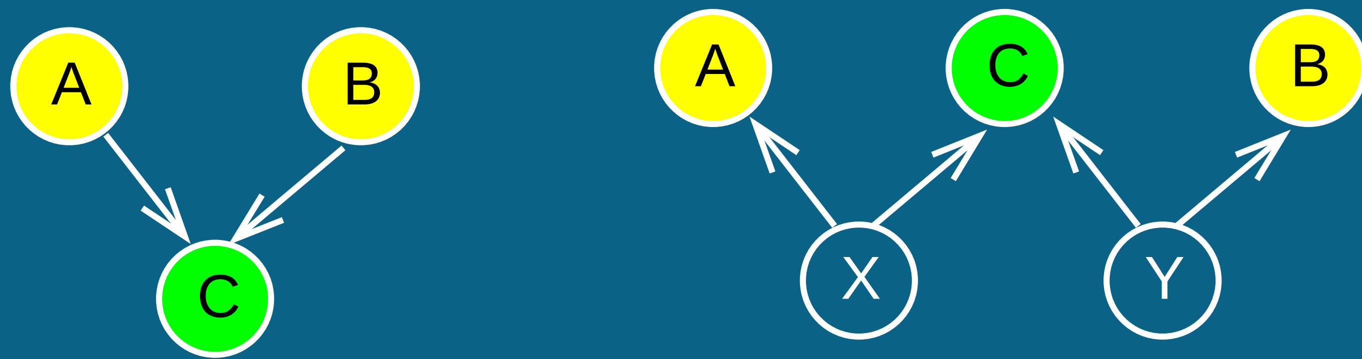
- Assume you have a complete data set.
- Assume you have learned a PDAG from this data.
- By the causal Markov assumption, the graphs in the corresponding equivalence class must include the true causal graph.
- If there is a directed path from X to Y in the PDAG, then X causes Y.

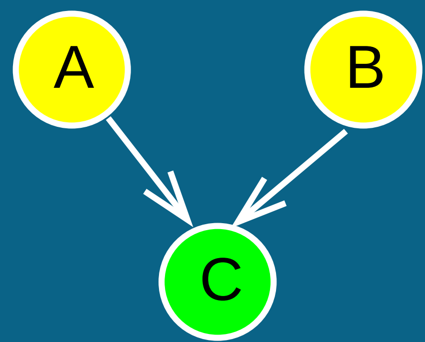
Can we distinguish between
correlation and causality
without interventions?

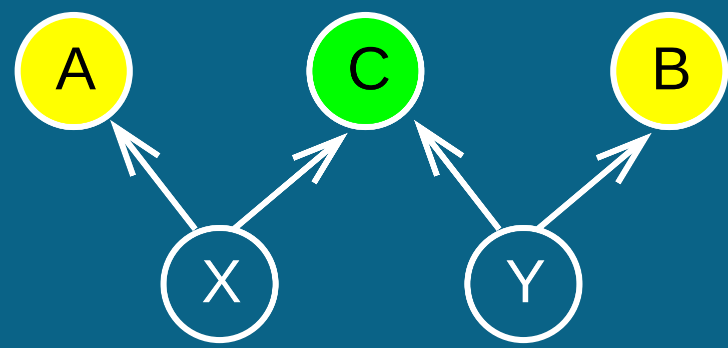
- Assume you have a complete data set.
- Assume you have learned a PDAG from this data.
- By the causal Markov assumption, the graphs in the corresponding equivalence class must include the true causal graph.
- If there is a directed path from X to Y in the PDAG, then X causes Y.

Problem: unobserved latent variables

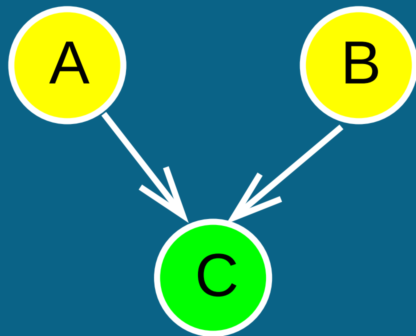


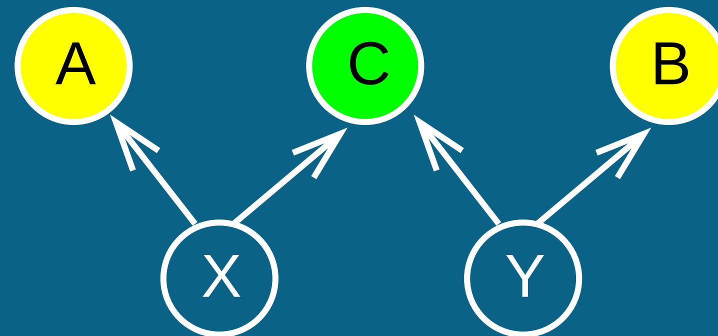

A
B
C
A
C
B
X
Y





$$
\begin{aligned}
P(A,B,C,X,Y) &= P(A|X)P(X)P(C|X,Y)P(B|Y)P(Y) \\
&= P(X|A)P(A)P(C|X,Y)P(Y|B)P(B)
\end{aligned}
$$





$$
\begin{aligned}
P(A,B,C,X,Y) &= P(A|X)P(X)P(C|X,Y)P(B|Y)P(Y) \\
&= P(X|A)P(A)P(C|X,Y)P(Y|B)P(B)
\end{aligned}
$$

$$
\begin{aligned}
P(A,B,C) &= \sum_X \sum_Y P(A,B,C,X,Y) \\
&= P(A)P(B) \sum_X \sum_Y P(C|X,Y)P(X|A)P(Y|B) \\
&= P(A)P(B)P(C|A,B)
\end{aligned}
$$

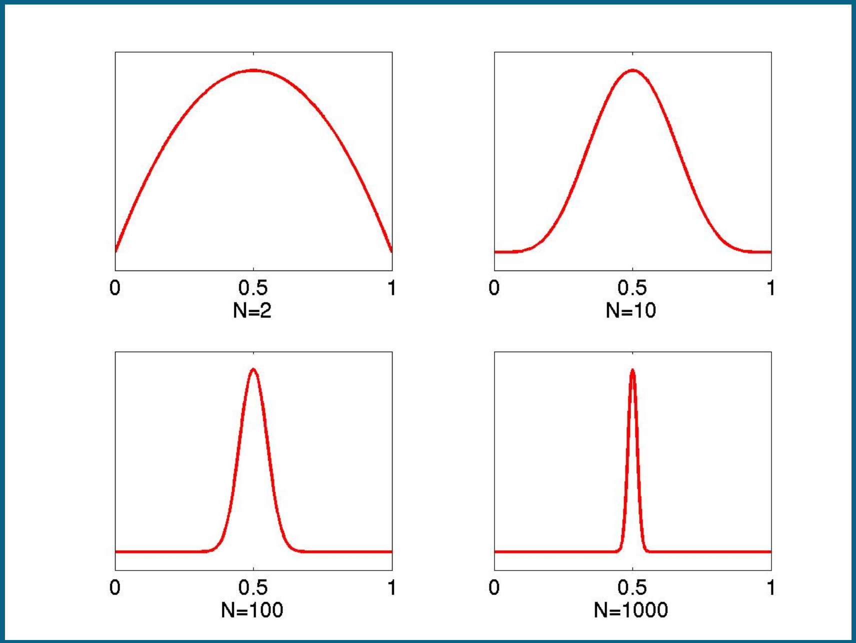
Take a directed path in a PDAG
as an indication of a causal relation,
to be tested experimentally.

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.



- Posterior probability $P(M|D)$ diffused: We cannot list all the networks that are plausible given the data.

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: $P(f|D) = \frac{1}{T} \sum_{i=1}^{T} f(M_i)$

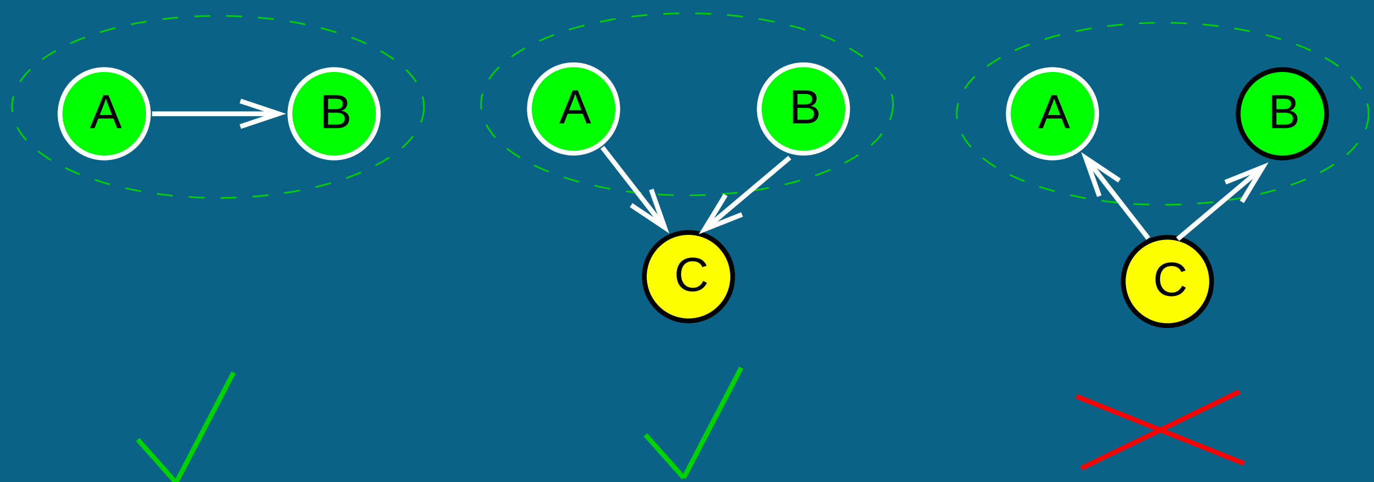
where $\{M_i\}$ is a sample from the posterior obtained with MCMC.

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.

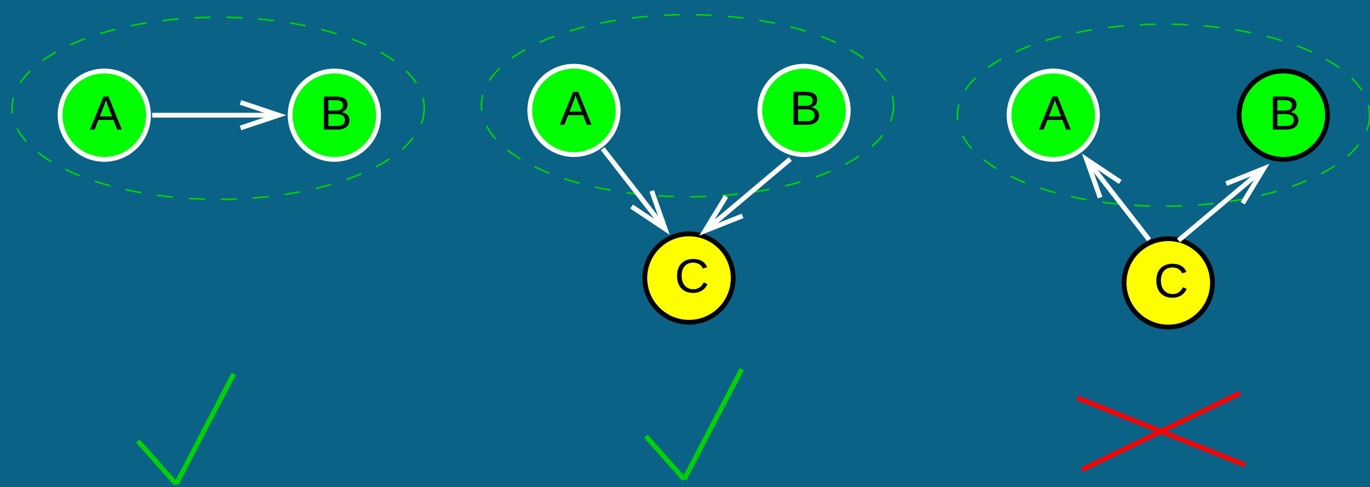


- Parent-child: One gene regulating another.

- Spouse relations: Two genes co-regulating another.

Markov neighbours

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



- Parent-child: One gene regulating another.

- Spouse relations: Two genes co-regulating another.

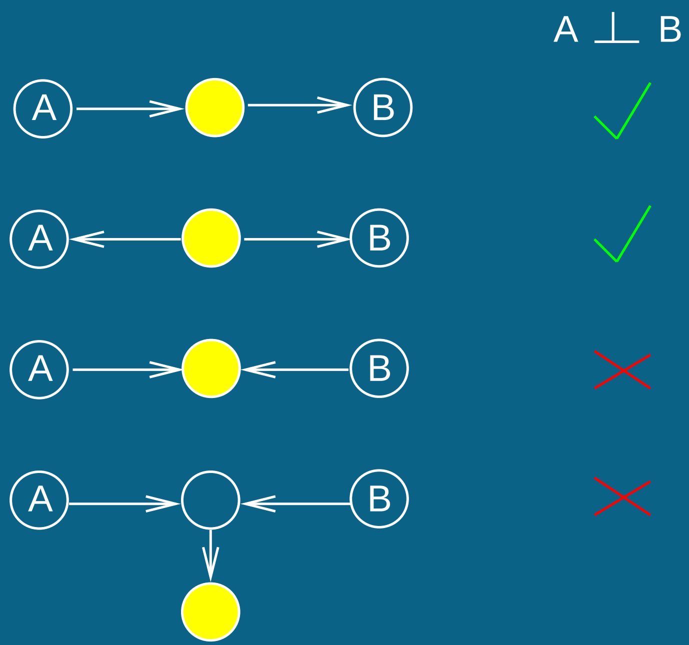
- Indication that two genes are related in some joint biological interaction or process.

Separators

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

Separators

- Given that A and B are indirectly dependent, what 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 PDAG 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 PDAG contain a directed path from A to B?
- Indication that A might be a causal ancestor of B.

Subnetworks

- $e_{A,B}$: Markov edge between A and B
- Confidence $P(e_{A,B}|D)$
- Identify high-scoring edges $P(e_{A,B}|D) \geq \theta$

Subnetworks

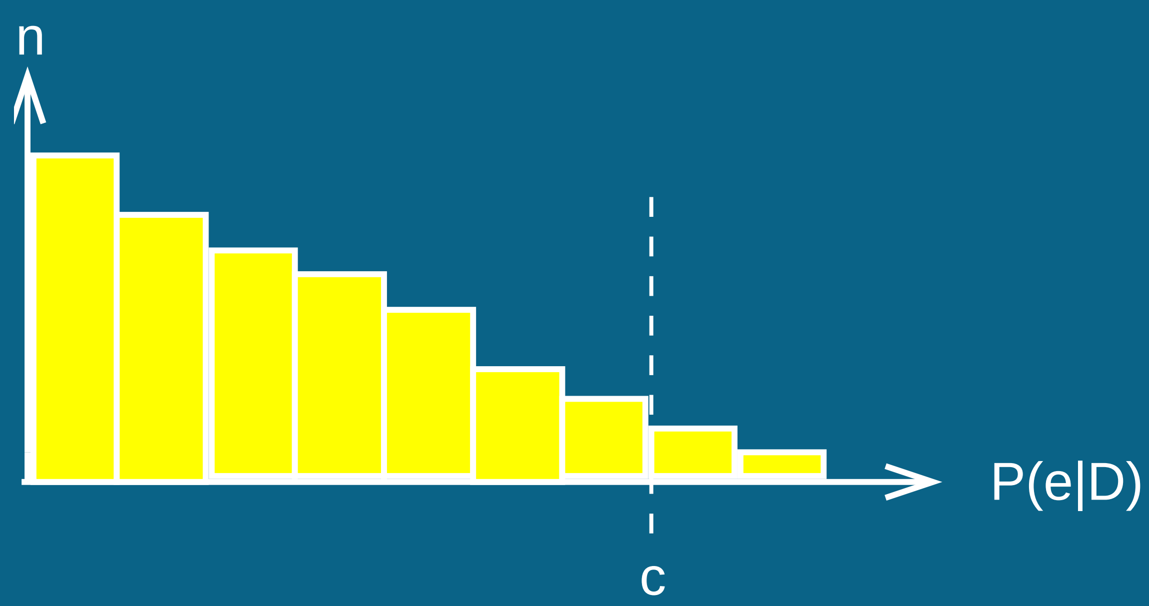
- $e_{A,B}$: Markov edge between A and B
- Confidence $P(e_{A,B}|D)$
- Identify high-scoring edges $P(e_{A,B}|D) \geq \theta$
- Identify subnetworks composed of high-scoring edges.
- Which subnetworks are significant?

Subnetworks

- $e_{A,B}$: Markov edge between A and B
- Confidence $P(e_{A,B}|D)$
- Identify high-scoring edges $P(e_{A,B}|D) \geq \theta$
- Identify subnetworks composed of high-scoring edges.
- Which subnetworks are significant?
- $\mathcal{N}$: Complete set of nodes, $N = |\mathcal{N}|$ total number of nodes.
- $\mathcal{U}$: Subnetwork-containing subset of $\mathcal{N}$, with $U = |\mathcal{U}|$ nodes.

- Total number of Markov edges in $\mathcal{U}$: $K = \binom{U}{2}$
- Number of high-scoring Markov edges: $h = \left|\left\{e_{X,Y} | P(e_{X,Y}|D) > \theta\right\}\right|$

- Total number of Markov edges in $\mathcal{U}$: $K = \binom{U}{2}$
- Number of high-scoring Markov edges: $h = \left| \left\{ e_{X,Y} | P(e_{X,Y}|D) > \theta \right\} \right|$
- $F(c) = P\Big(P(e_{X,Y}|D) \geq c \Big)$



- Null hypothesis:
  Confidence of each edge is sampled independently from F.

- Null hypothesis:
  Confidence of each edge is sampled independently from F.

- Probability of finding the high-scoring subnetwork in $\mathcal{U}$ under the null-hypothesis: $P = \binom{K}{h} F(\theta)^h \Big[1 - F(\theta)\Big]^{K-h}$

- Null hypothesis:
  Confidence of each edge is sampled independently from F.

- Probability of finding the high-scoring subnetwork in $\mathcal{U}$ under the null-hypothesis: $P = \binom{K}{h} F(\theta)^h \Big[1 - F(\theta)\Big]^{K-h}$

- Expected number of similar high-scoring subnetwork under the null-hypothesis: $E = \binom{N}{U} P = \binom{N}{U} \binom{K}{h} F(\theta)^h \Big[1 - F(\theta)\Big]^{K-h}$

- Null hypothesis:
  Confidence of each edge is sampled independently from F.

- Probability of finding the high-scoring subnetwork in $\mathcal{U}$ under the null-hypothesis: $P = \binom{K}{h} F(\theta)^h \Big[1 - F(\theta)\Big]^{K-h}$

- Expected number of similar high-scoring subnetwork under the null-hypothesis: $E = \binom{N}{U} P = \binom{N}{U}\binom{K}{h} F(\theta)^h \Big[1 - F(\theta)\Big]^{K-h}$

- Subnetwork is significant if $E \ll 1$, e.g., if $E \leq e^{-5}$

Experimental Results

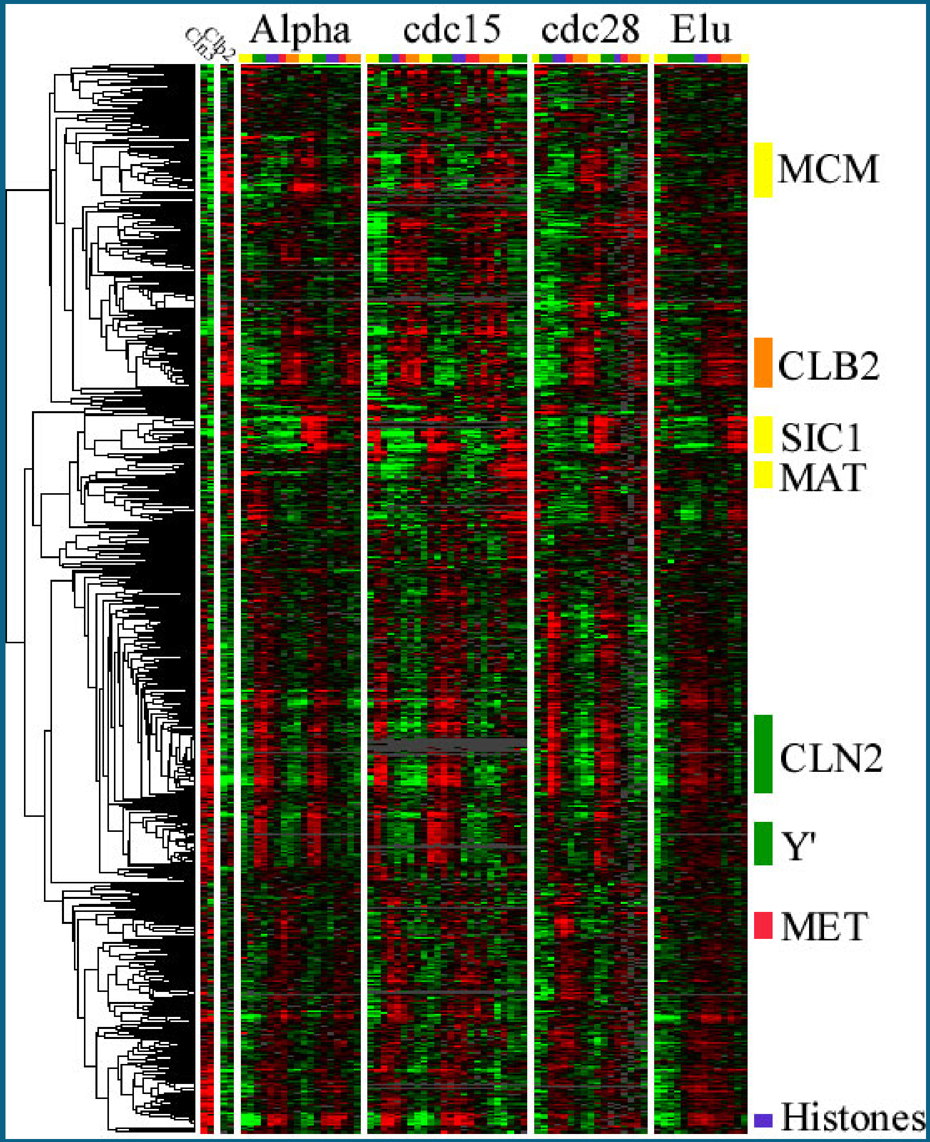
- Friedman, Linial, Nachman, Pe'er (2000)
  Journal of Computational Biology 7 (3/4): 601-620
  http://www.cs.huji.ac.il/labs/compbio/expression/#papers

- Pe'er, Regev, Elidan, Friedman (2001)
  Bioinformatics S1: 215-224
  http://www.cs.huji.ac.il/labs/compbio/ismb01/

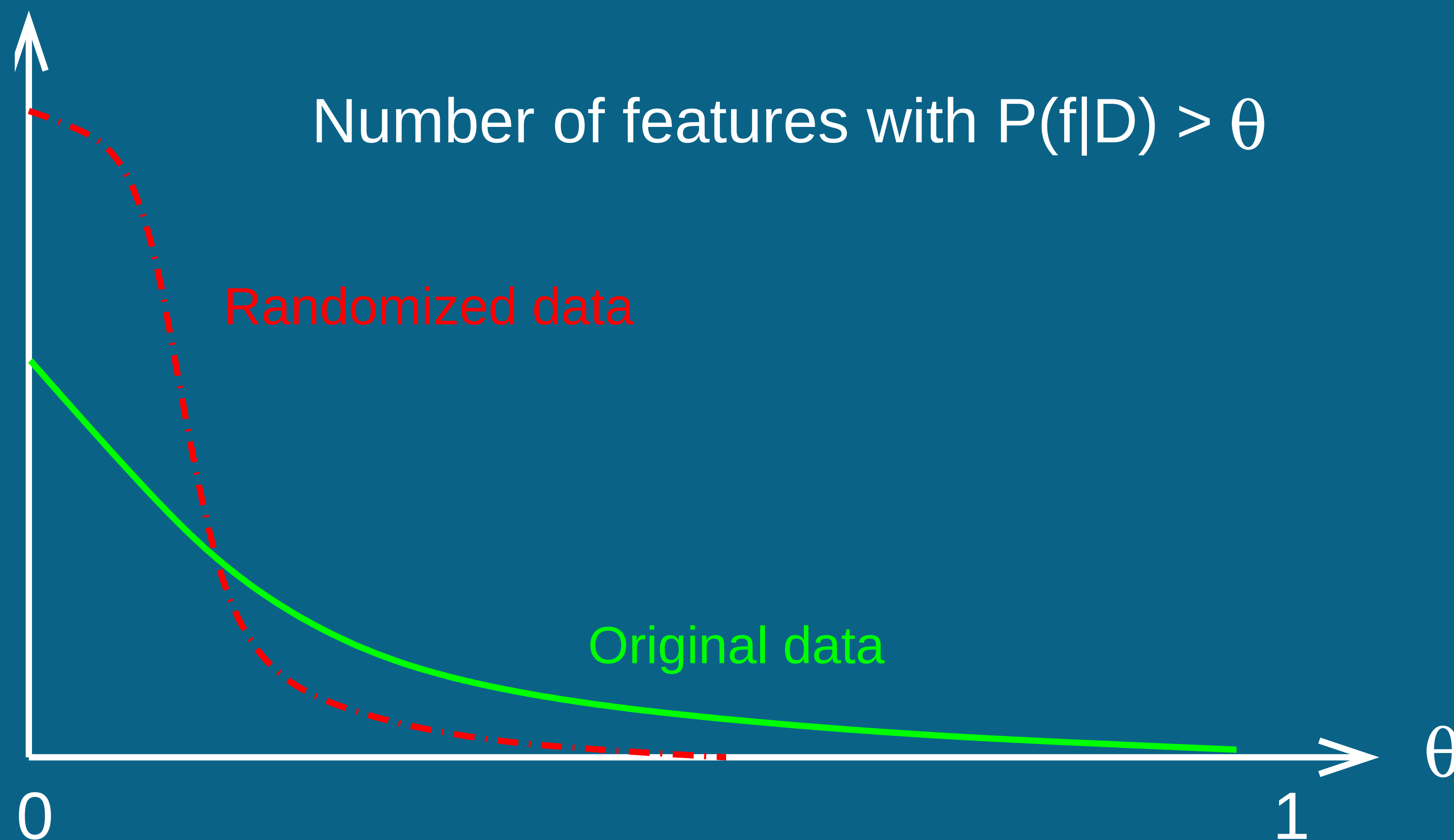
- Spellman, Sherlock, Zhang, Iyer, Anders, Eisen, Brown, Botstein, Futcher (1998)
  Molecular Biology of the Cell 9 (12) :3273-97
  http://cellcycle-www.stanford.edu/

- 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) > θ
Randomized data
Original data
θ
0
1

Order relations

Confidence in X being an ancestor of Y:
$P(X \longrightarrow Y|D)$

Dominance score of X: $\sum_{Y} P(X \longrightarrow Y|D)$

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:
$P(X \longrightarrow Y|D)$

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

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

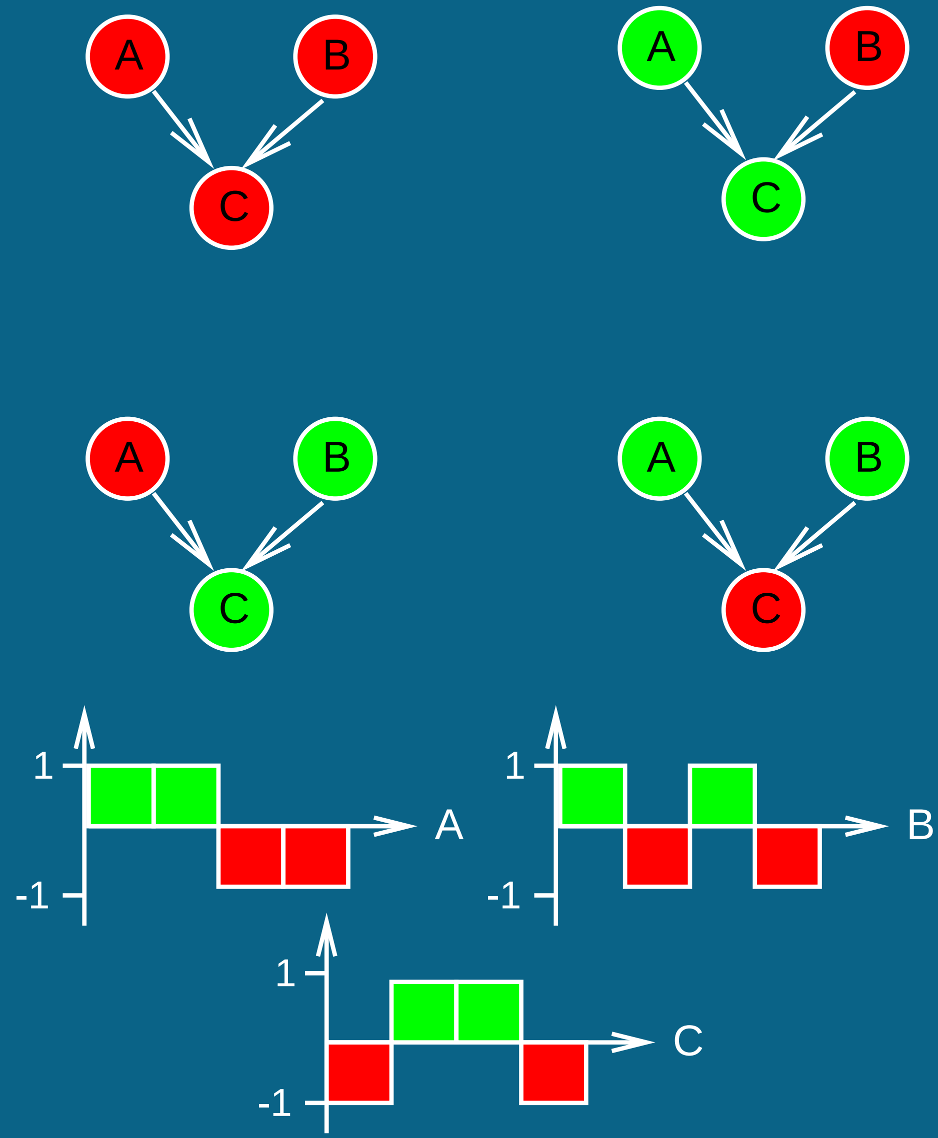
Markov relations

$P(X \leftrightarrow Y|D)$: 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.



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

Separator relations

Explain away dependencies.

Separator relations

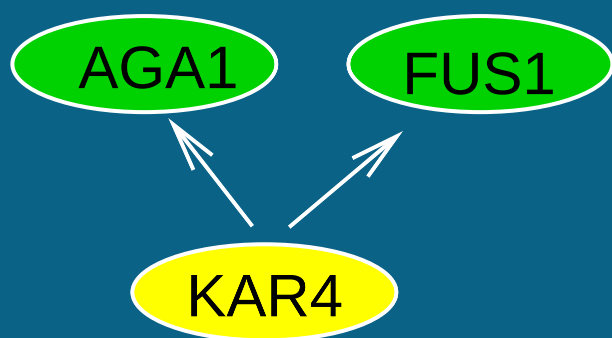
Explain away dependencies.

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

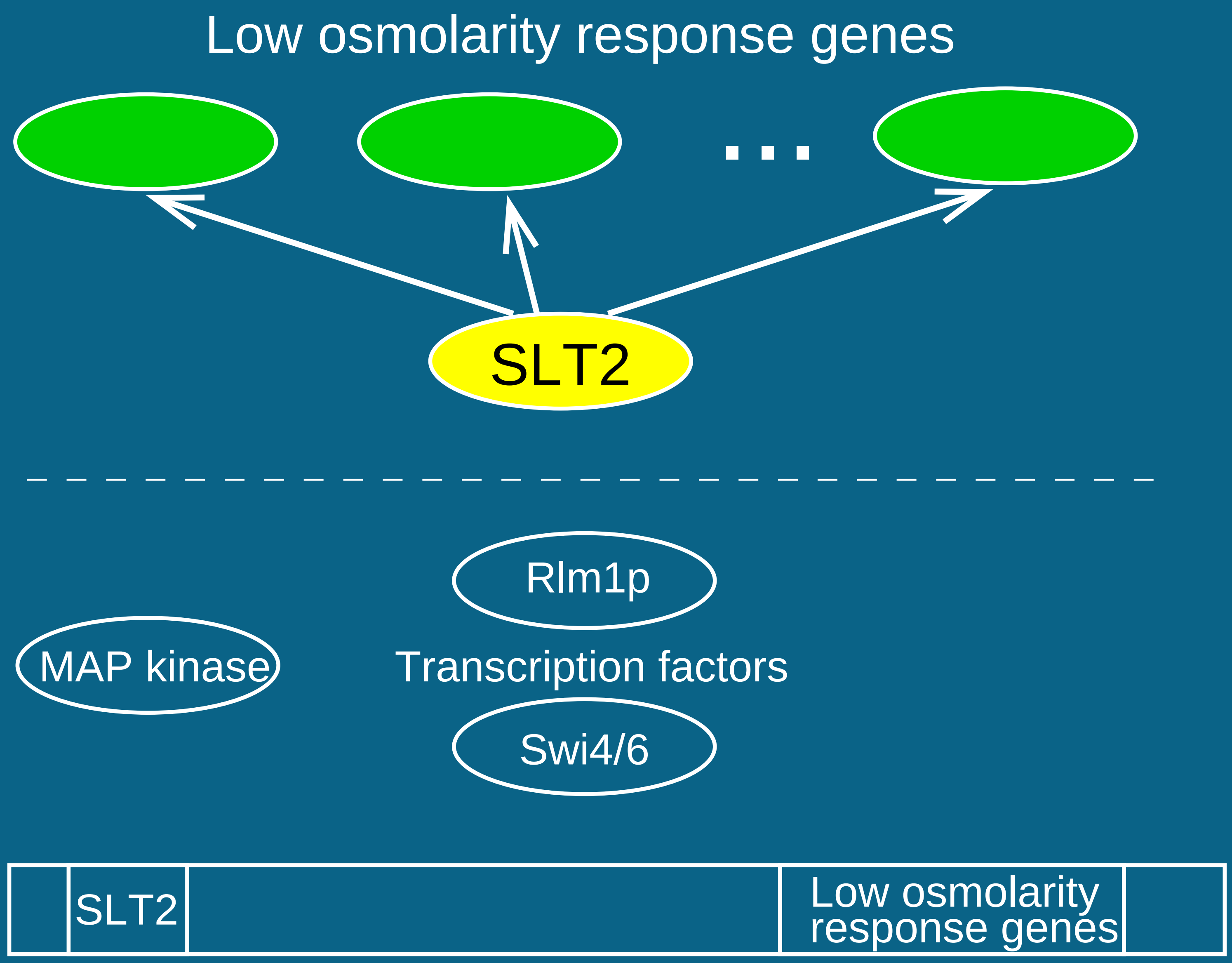
Separator relations

Explain away dependencies.

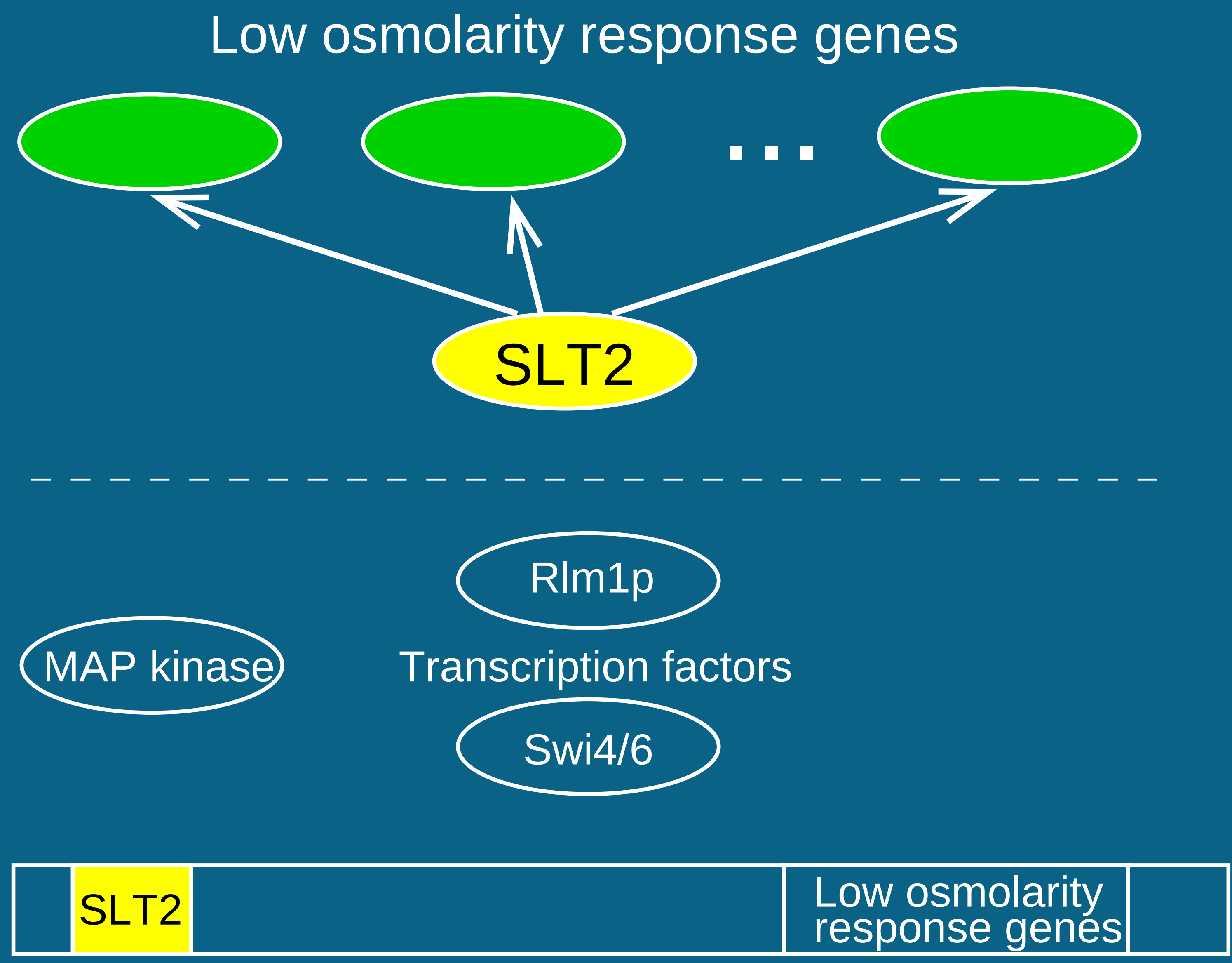
- Separators: Known transcriptional regulators
- Separated genes: Share functional roles

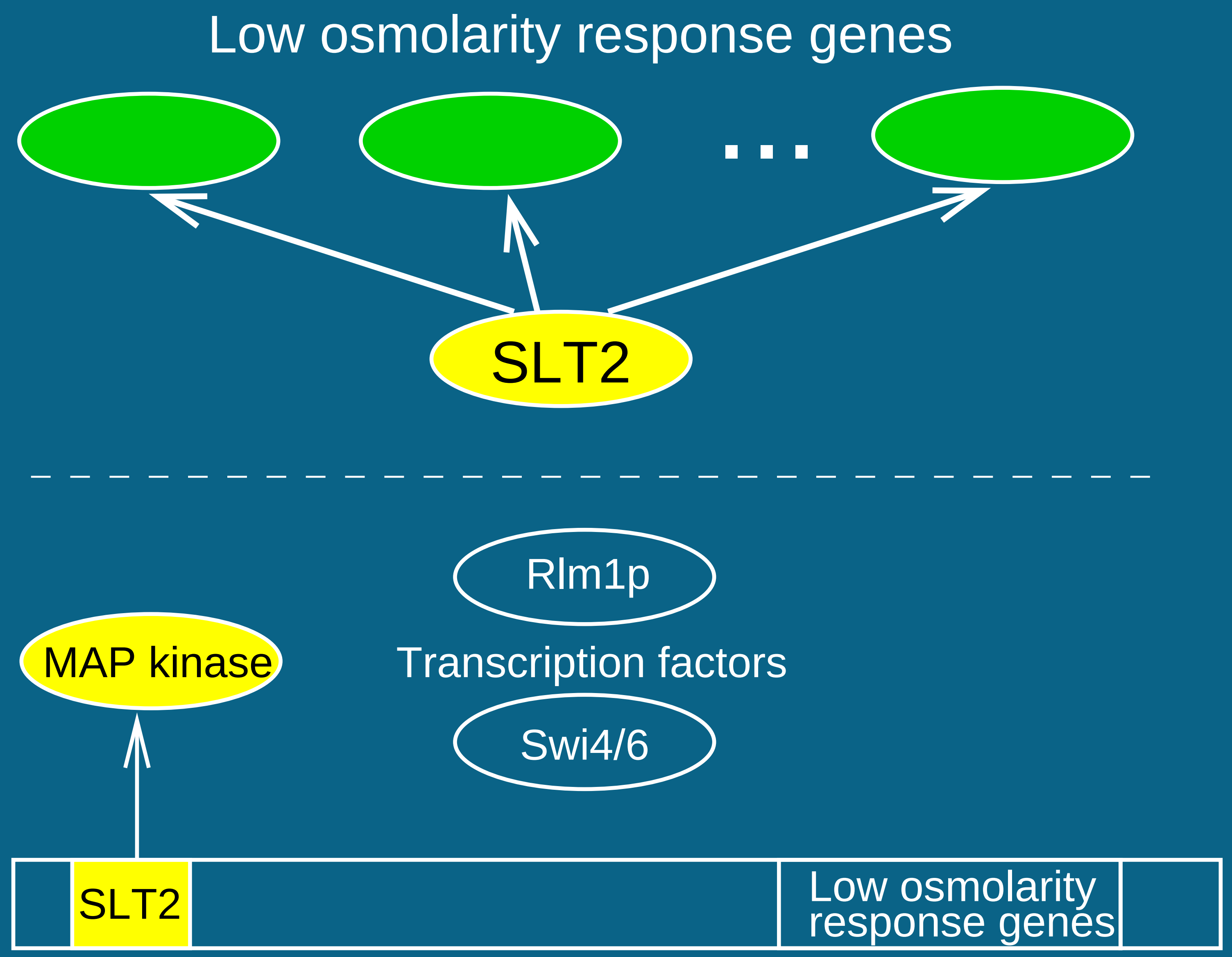


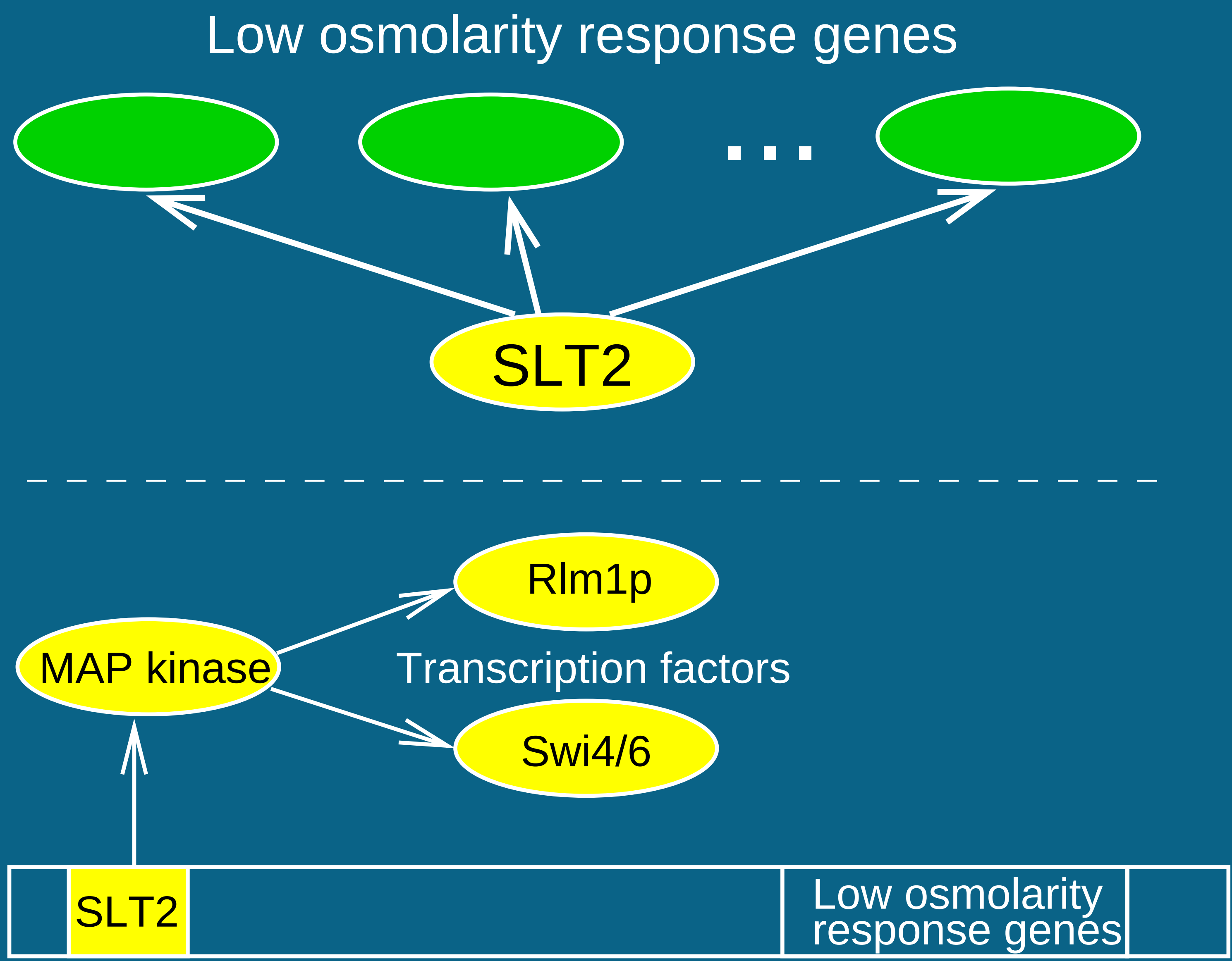
KAR4	Transcriptional regulator of nuclear fusion genes.
AGA1,FUS1	Cell fusion genes

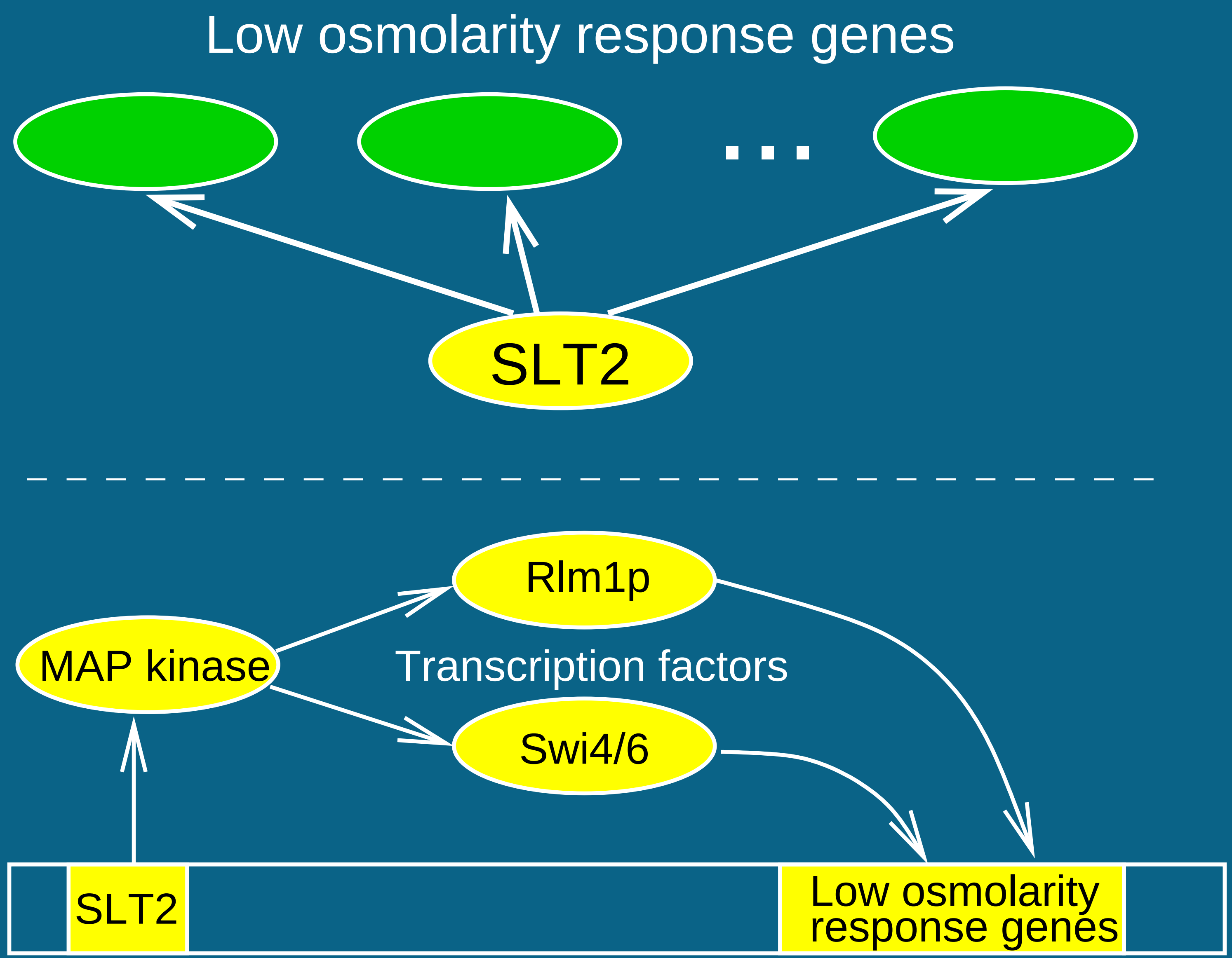


Low osmolarity response genes
...
SLT2
Rlm1p
MAP kinase
Transcription factors
Swi4/6
SLT2
Low osmolarity
response genes


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

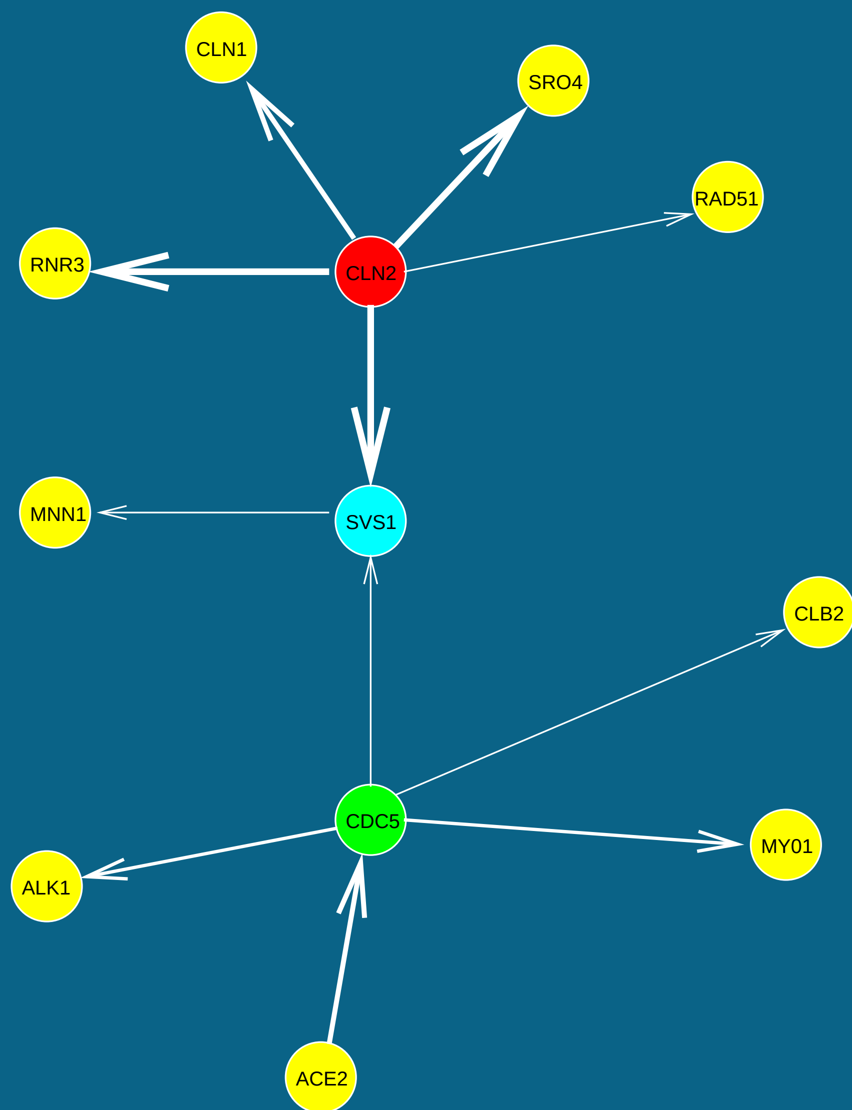

Low osmolarity response genes
...
SLT2
Rlm1p
MAP kinase
Transcription factors
Swi4/6
SLT2
Low osmolarity
response genes


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


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

Subnetworks

Subnetworks



Discussion

Discussion

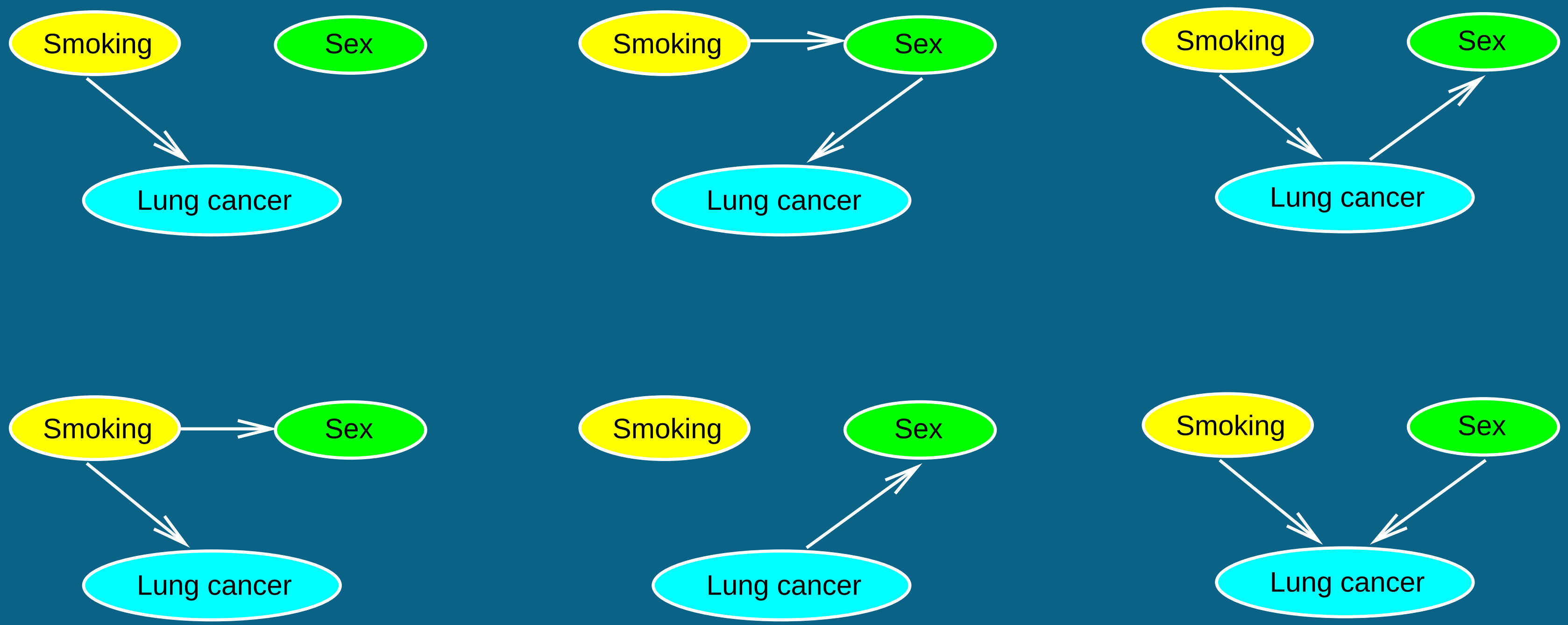
- Modelling the dependence on time explicitly $\longrightarrow$ dynamical Bayesian networks (David Wilde).

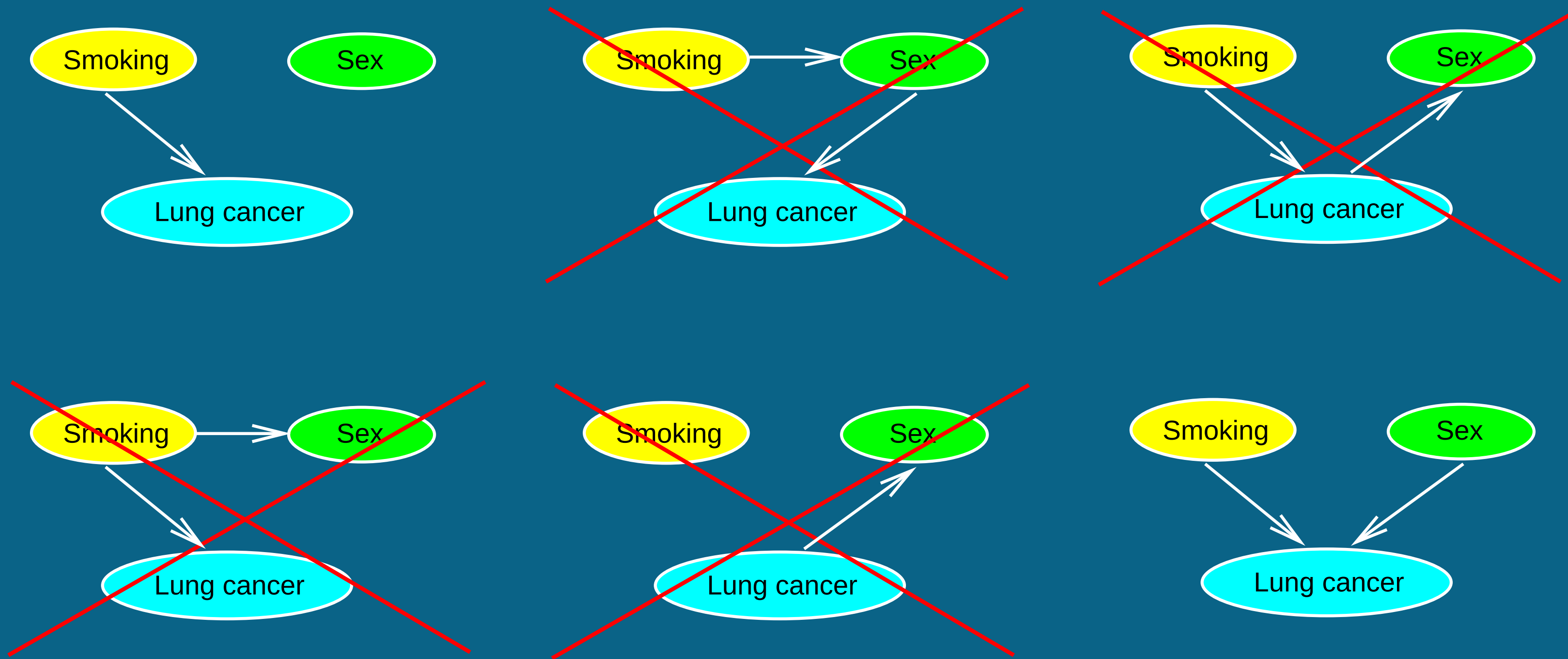
Discussion

- Modelling the dependence on time explicitly $\longrightarrow$ dynamical Bayesian networks (David Wilde).
- Integrating different types of data: Sequence data $\rightarrow$ transcription factor binding sites (Eran Segal).

Discussion

- Modelling the dependence on time explicitly $\longrightarrow$ dynamical Bayesian networks (David Wilde).
- Integrating different types of data: Sequence data $\rightarrow$ transcription factor binding sites (Eran Segal).
- Use more informative priors.


Smoking
Sex
Lung cancer
Smoking
Sex
Lung cancer
Smoking
Sex
Lung cancer
Smoking
Sex
Lung cancer
Smoking
Sex
Lung cancer
Smoking
Sex
Lung cancer



Smoking
Sex
Lung cancer
Smoking
Sex
Lung cancer
Smoking
Sex
Lung cancer
Smoking
Sex
Lung cancer
Smoking
Sex
Lung cancer
Smoking
Sex
Lung cancer

- 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