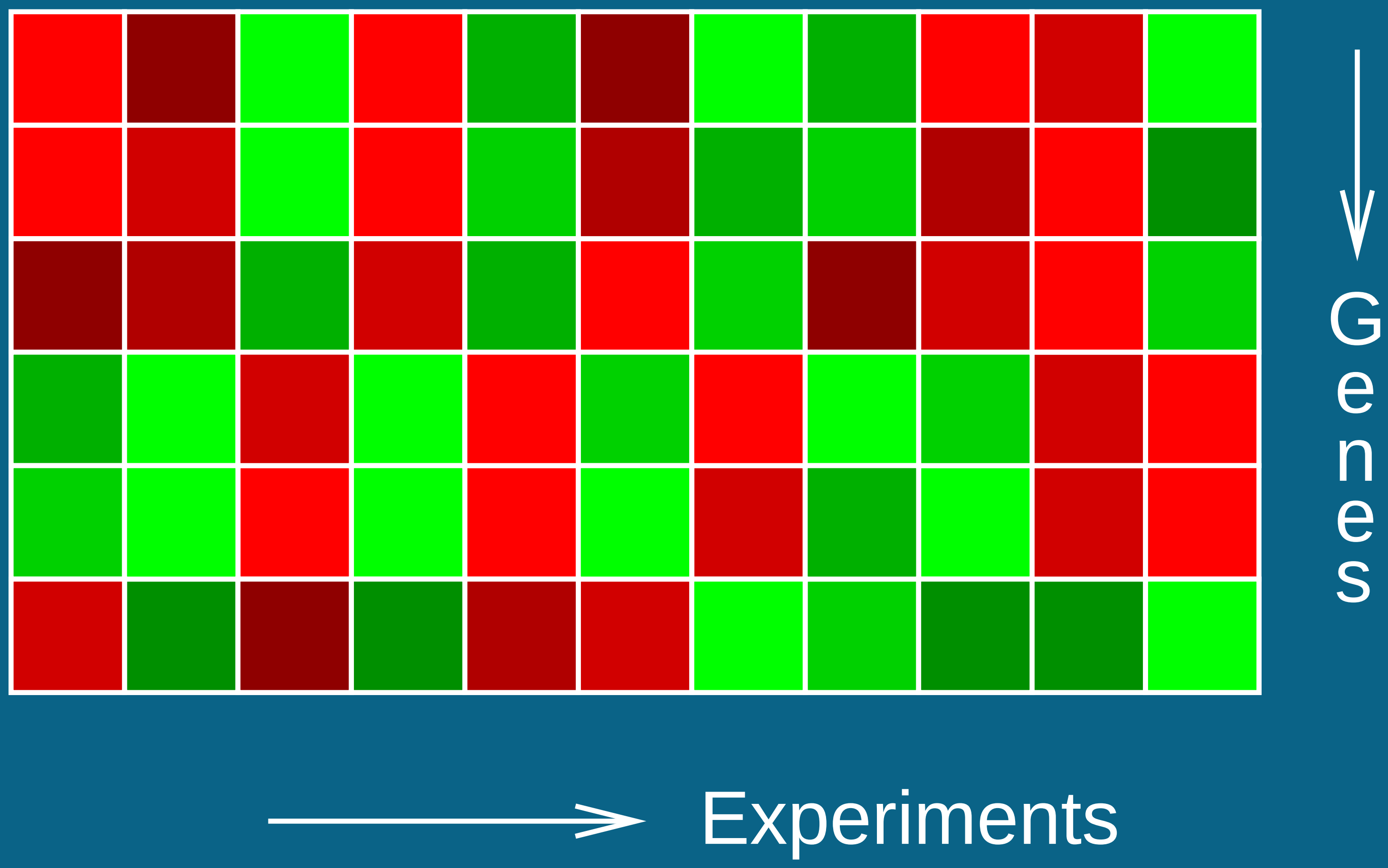
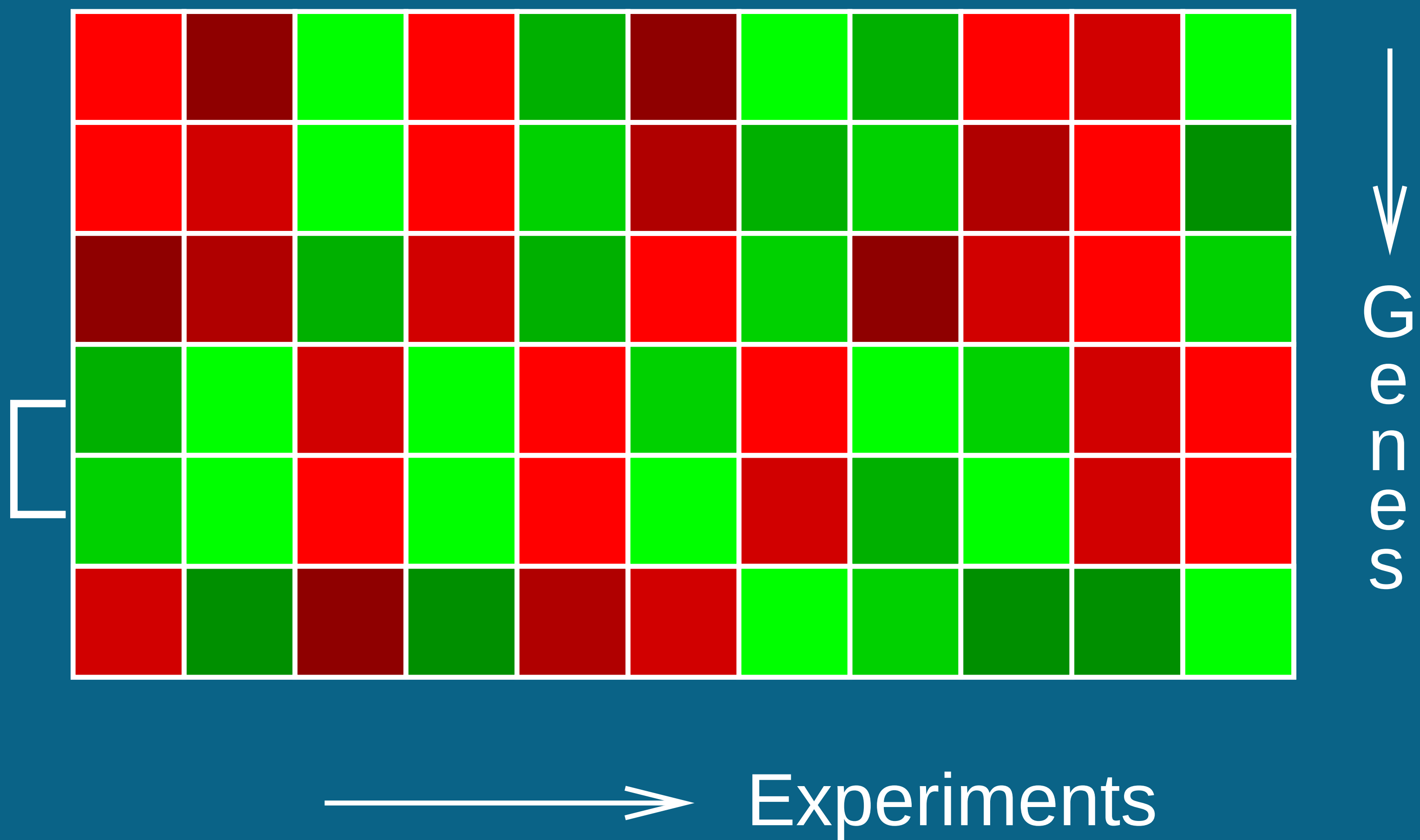
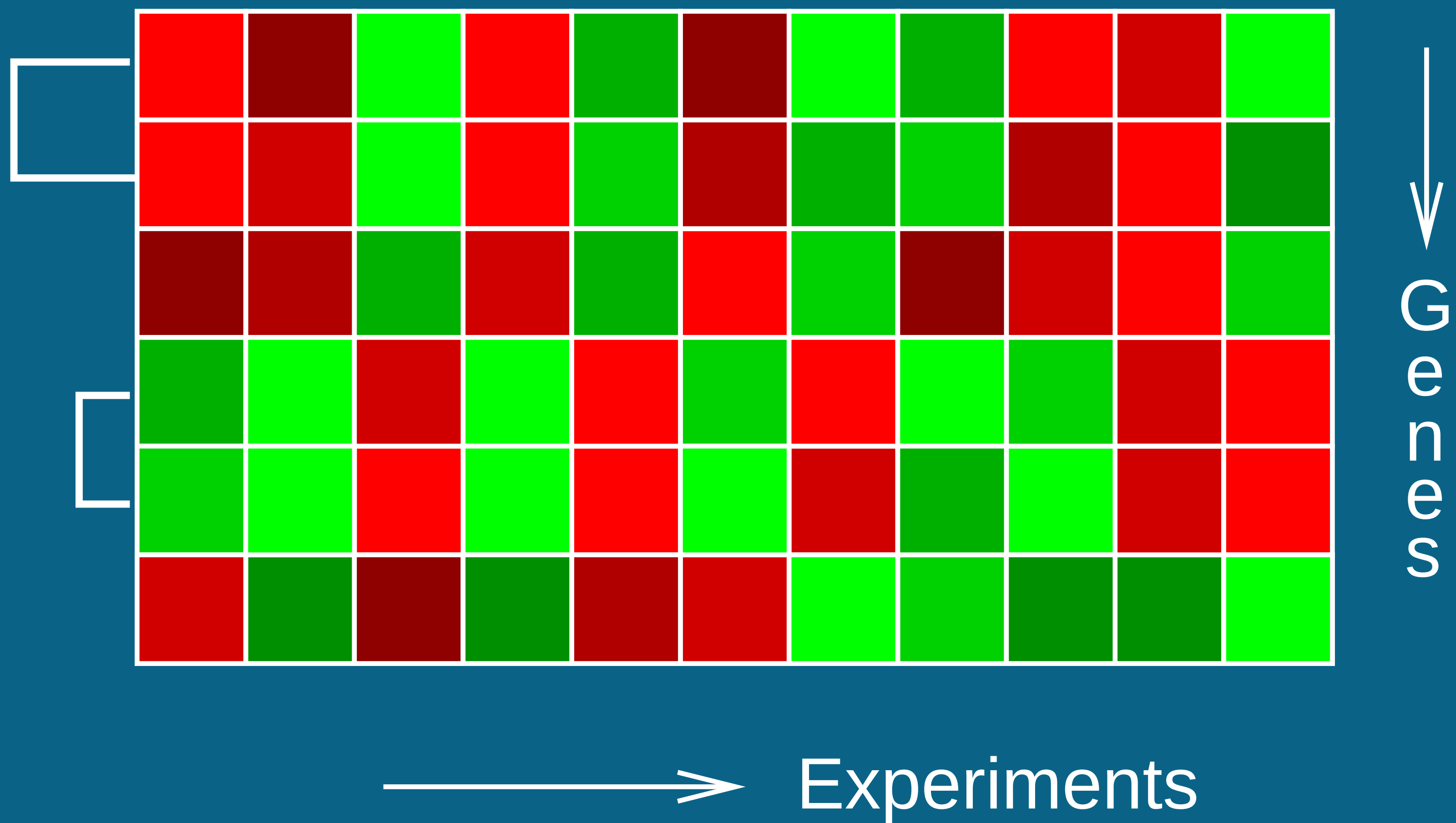
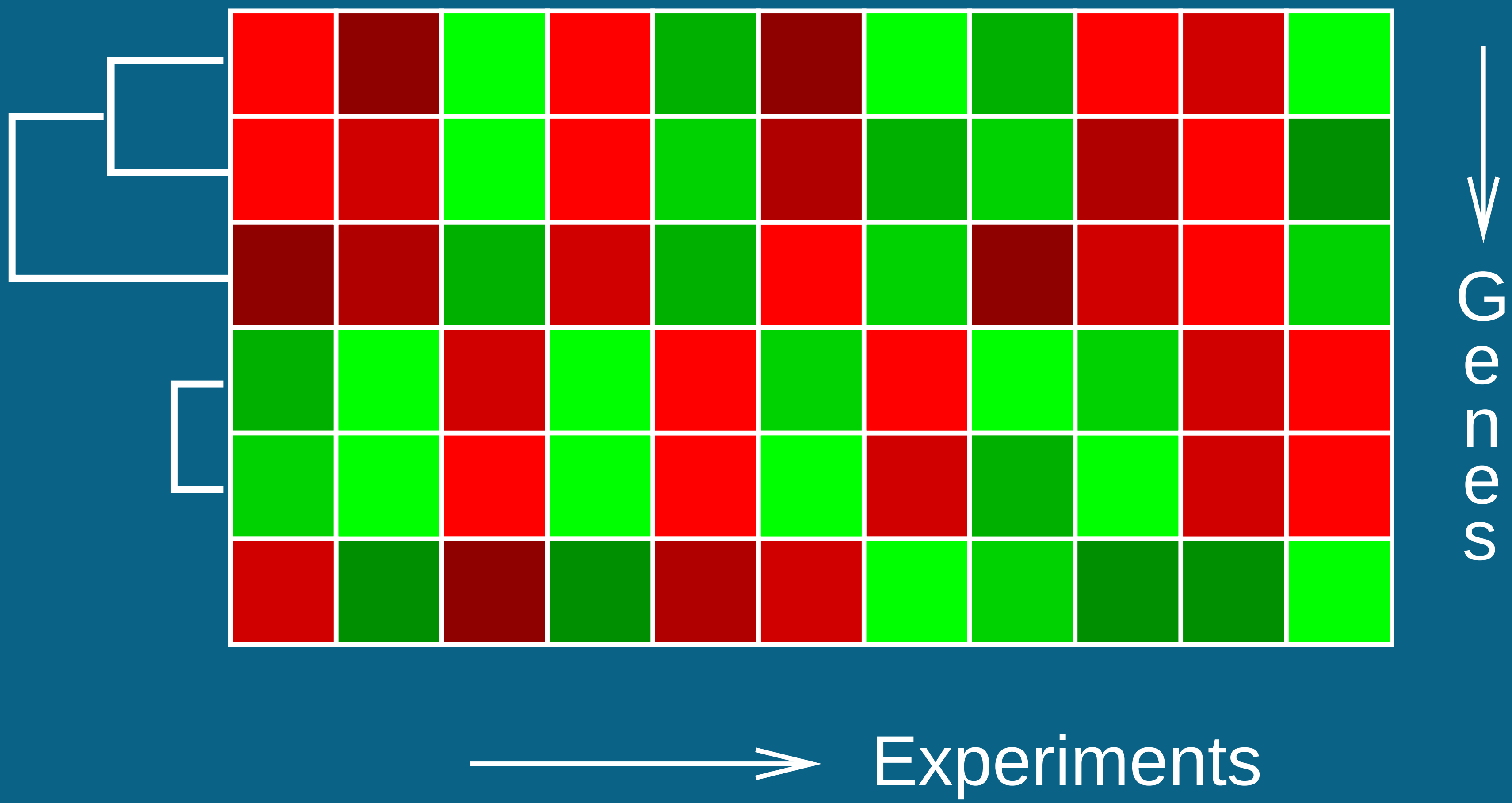
Inferring genetic networks from microarray gene expression data

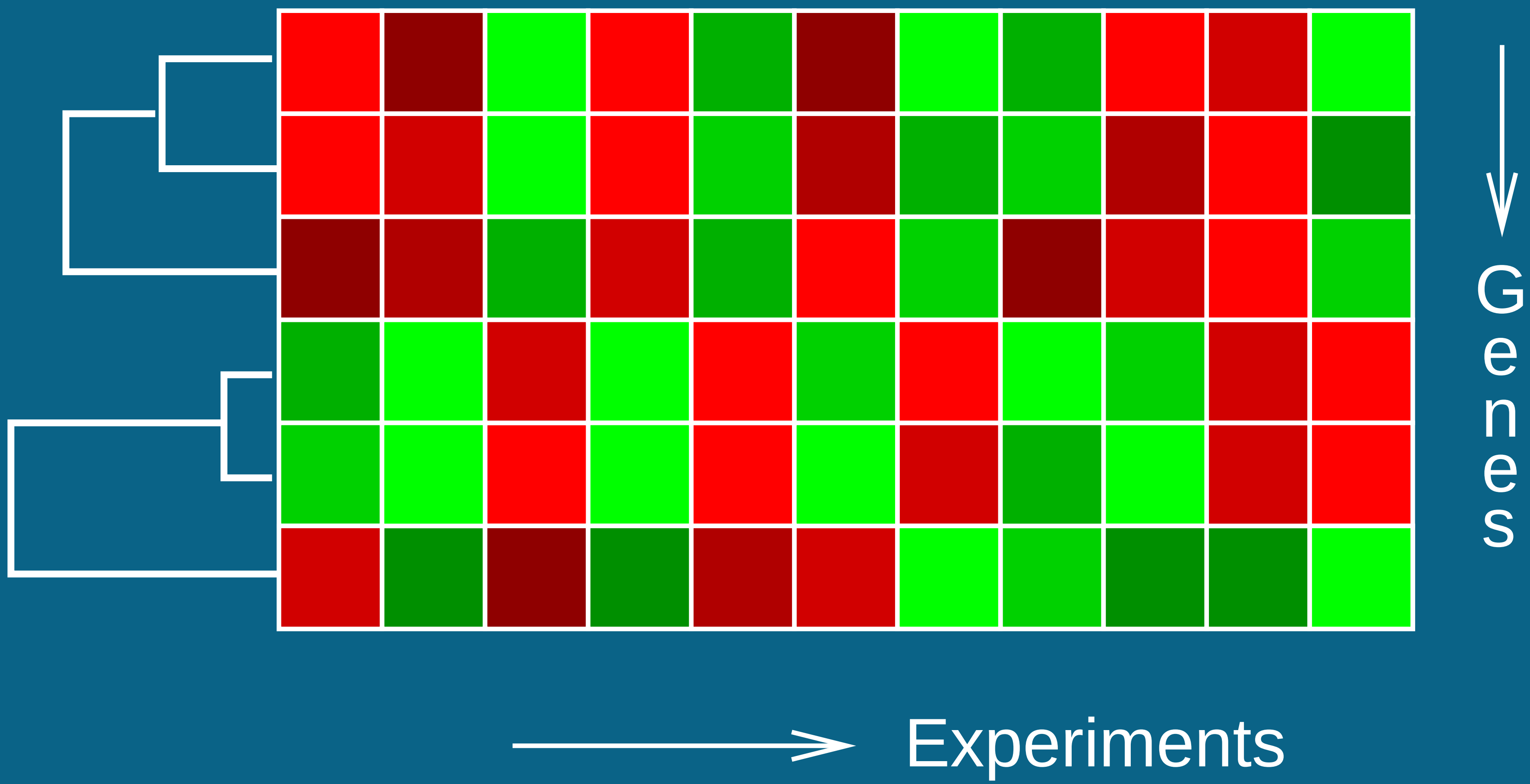
Dirk Husmeier
Biomathematics & Statistics Scotland (BioSS)
JCMB, The King's Buildings, Edinburgh EH9 3JZ
United Kingdom
http://www.bioss.ac.uk/∼dirk

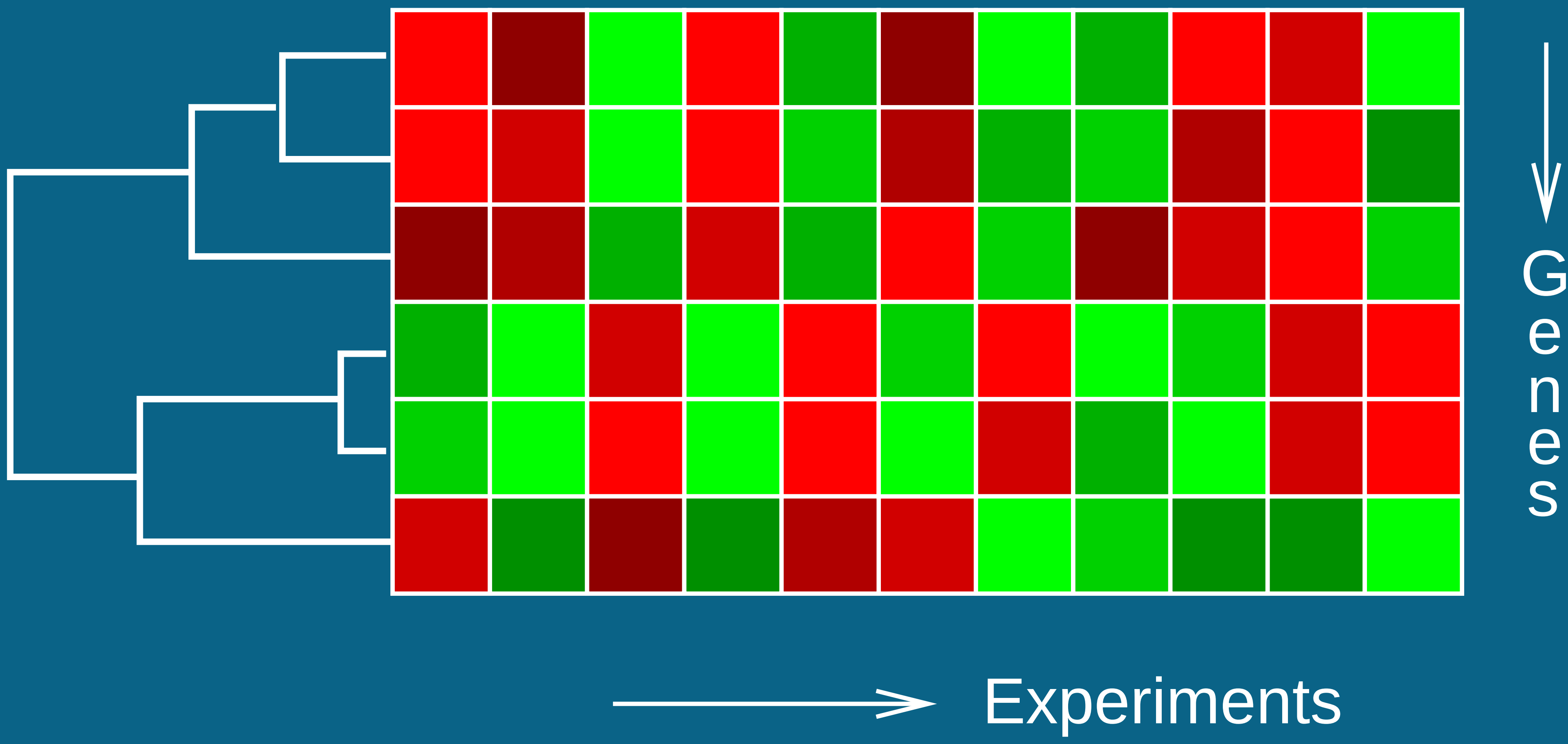

Genes
Experiments

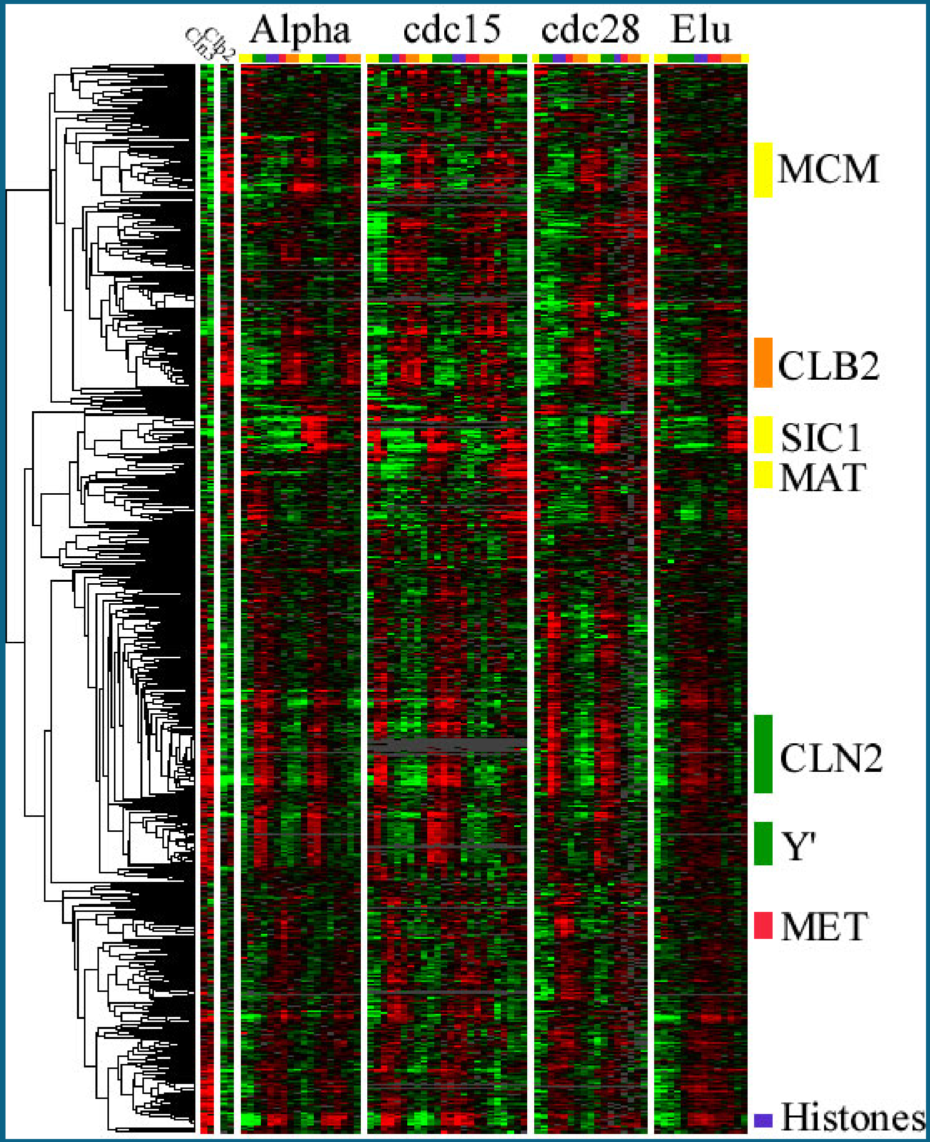

Genes
Experiments


Genes
Experiments


Genes
Experiments


Genes
Experiments


Genes
Experiments

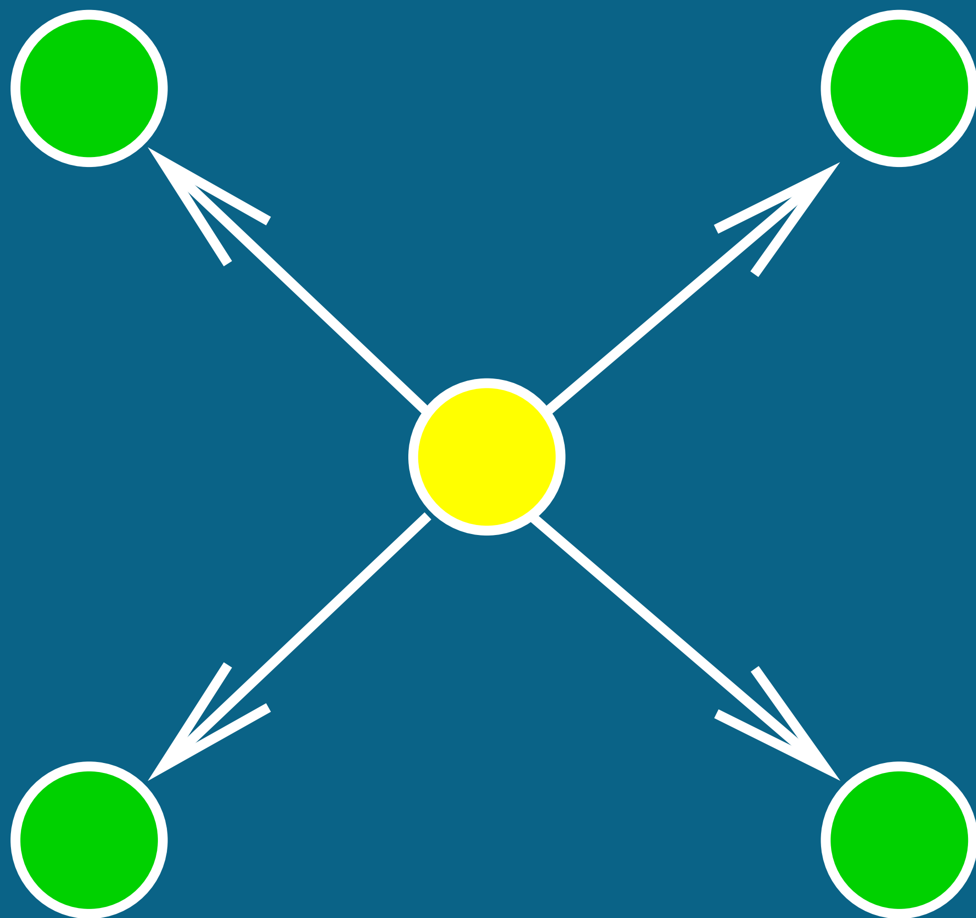


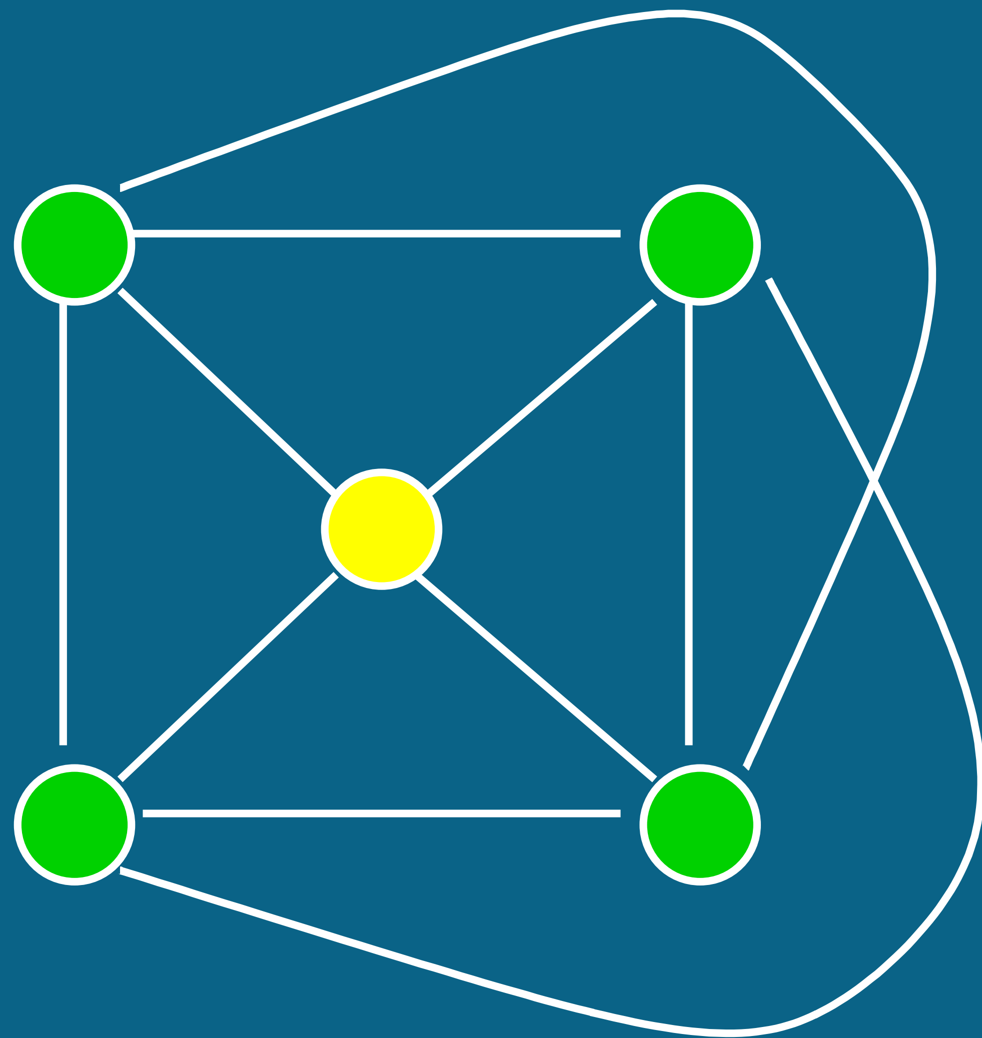
From Spellman et al., http://cellcycle-www.stanford.edu/

Shortcomings of clustering

- Global correlation measure.
- This obscures relationships that exist over only a subset of the data.

- Global correlation measure.
- This obscures relationships that exist over only a subset of the data.
- Partition of the genes into disjoint sets.
- Association of each gene with a single biological function or process.
- Oversimplification of the biological system.





Bayesian networks

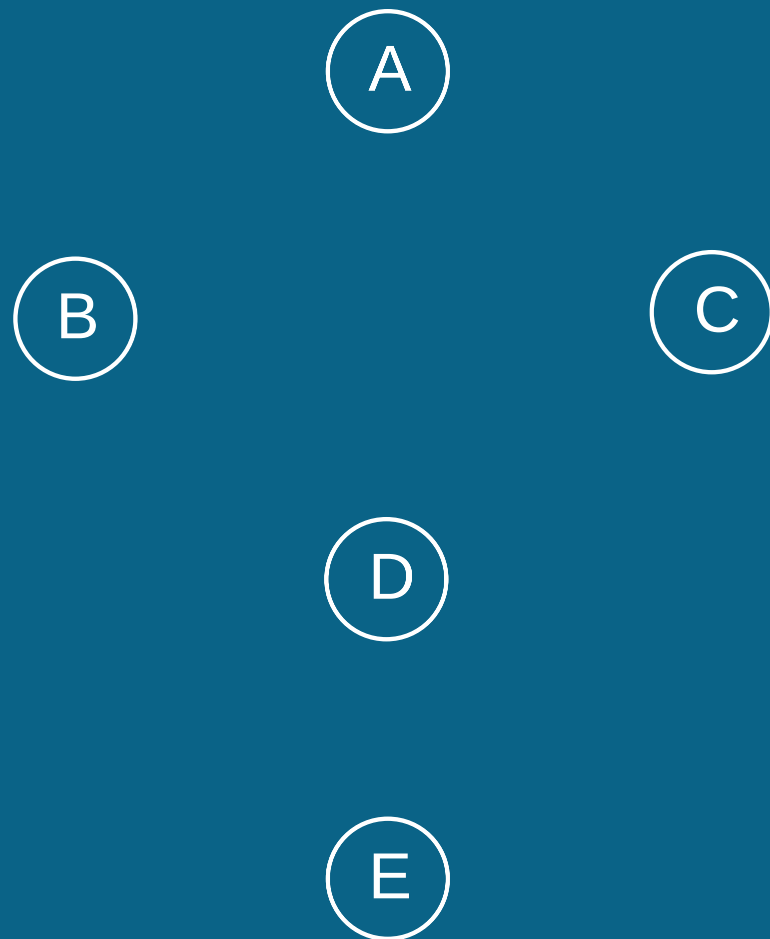
or

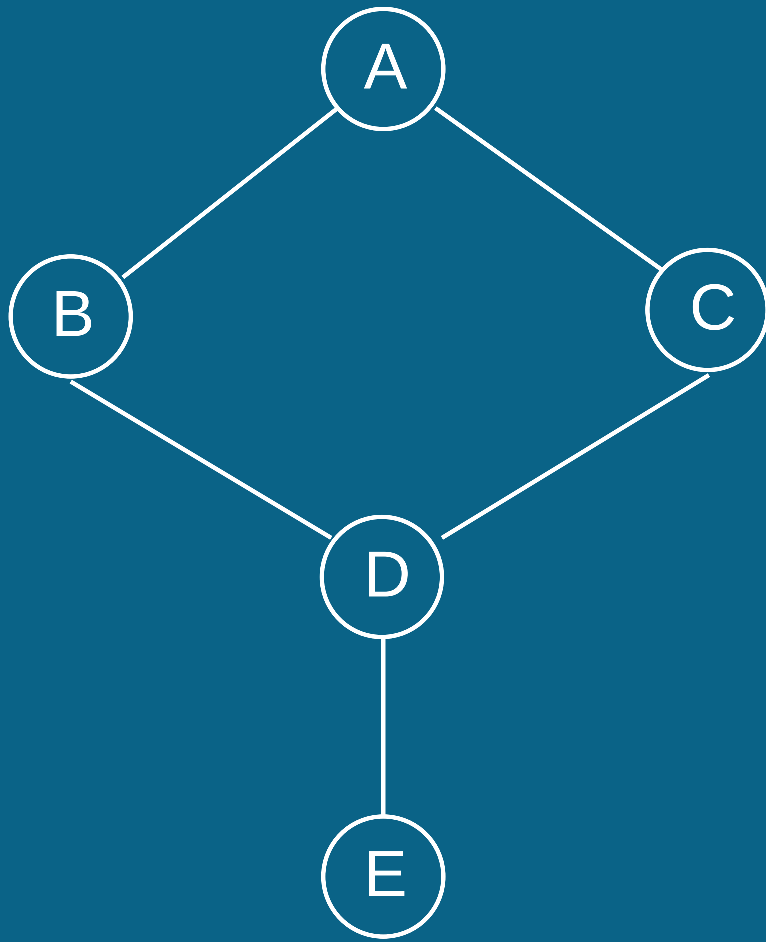
Directed acyclic graphs (DAGs)

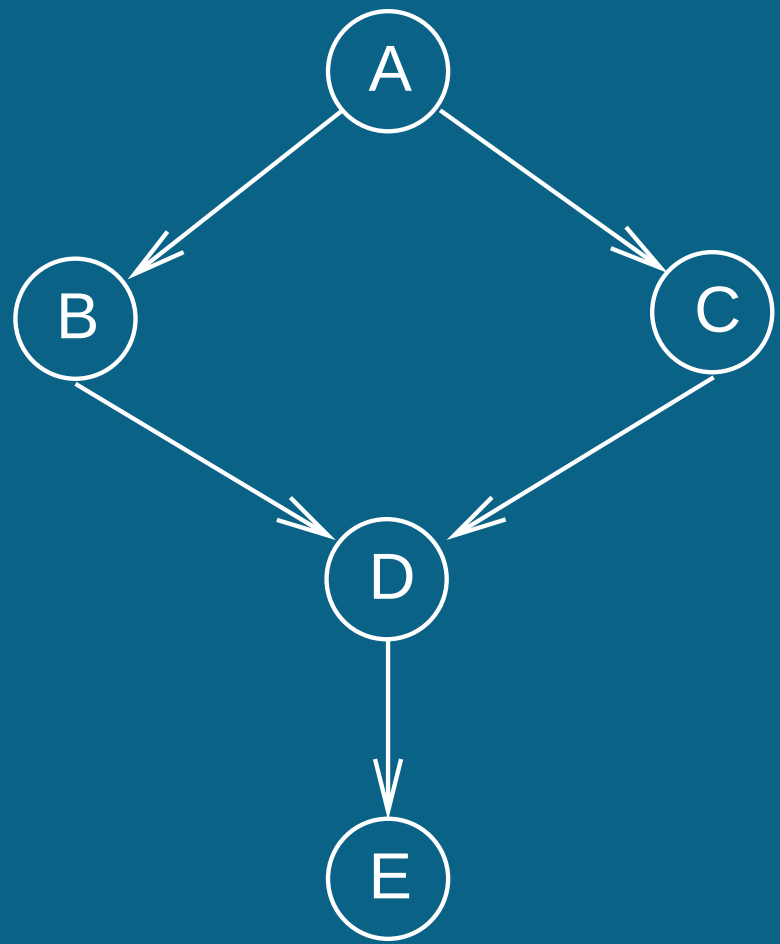
Probabilistic framework for **robust inference** of **interactions** in the **presence** of **noise**

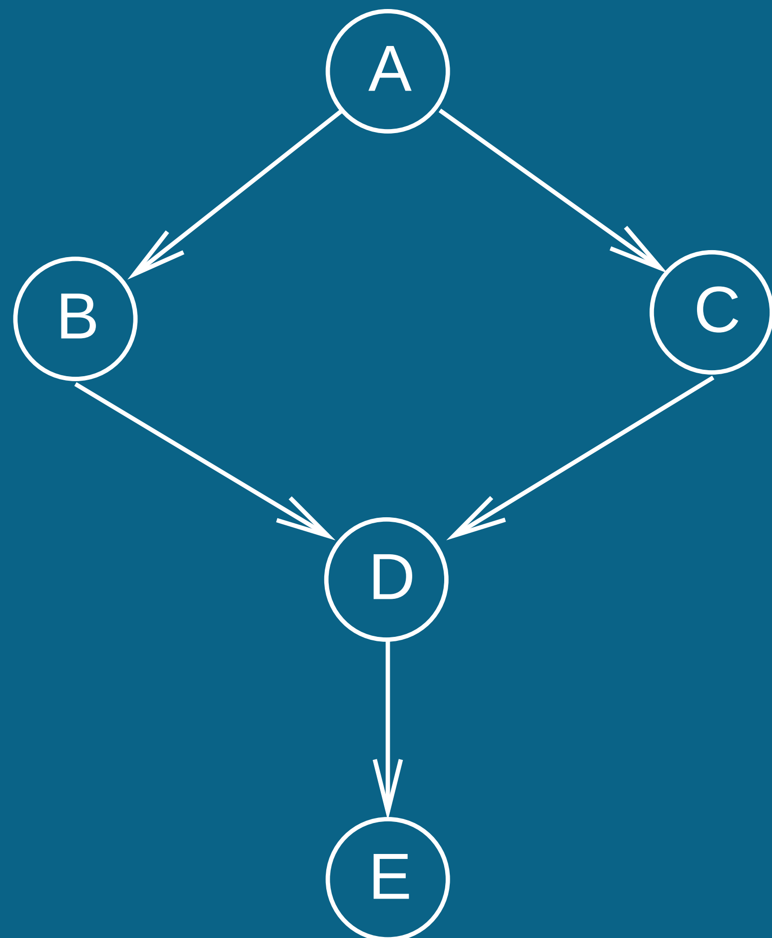
Probabilistic framework for **robust inference** of **interactions** in the **presence** of **noise**

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

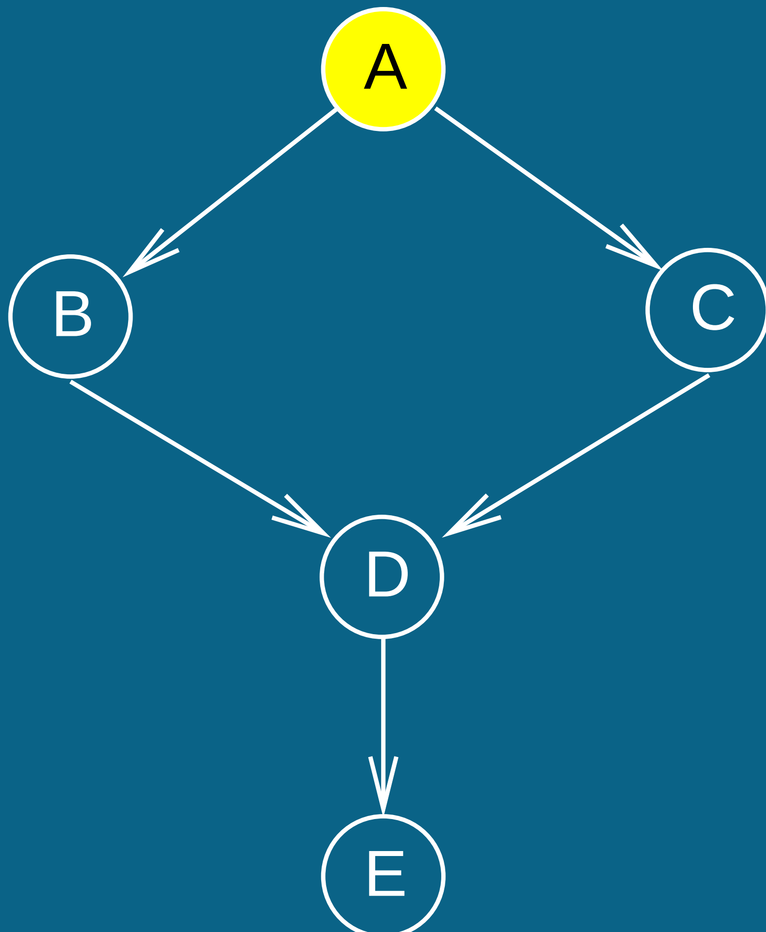

A
B
C
D
E


A
B
C
D
E


A
B
C
D
E

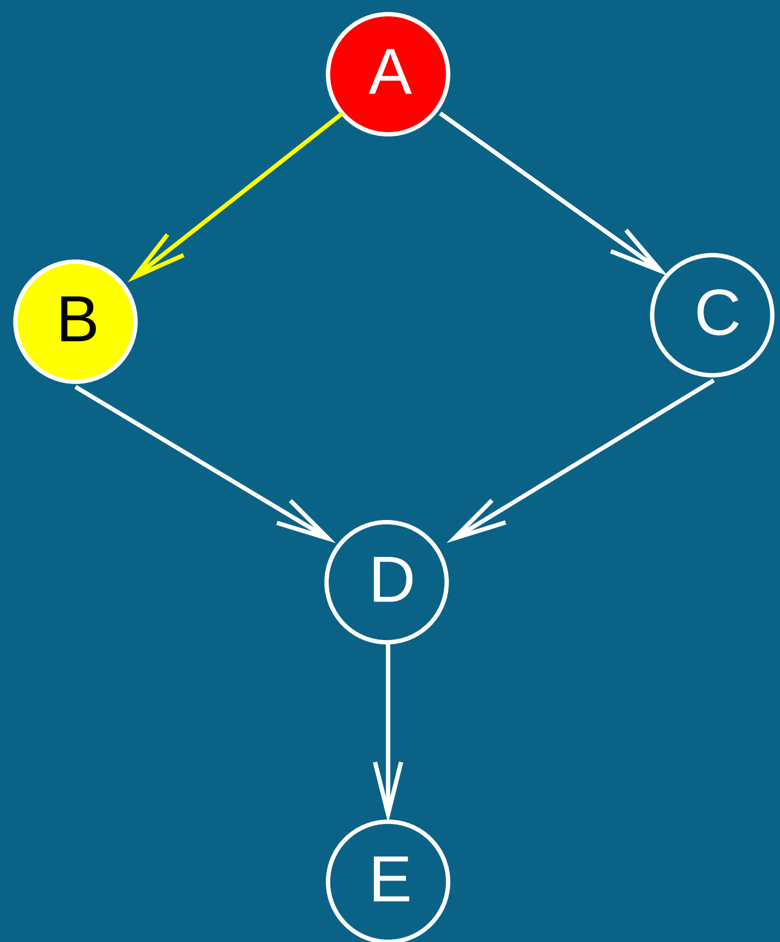


$$P(A,B,C,D,E) = \prod_i P(\text{node}_i|\text{parents}_i)$$



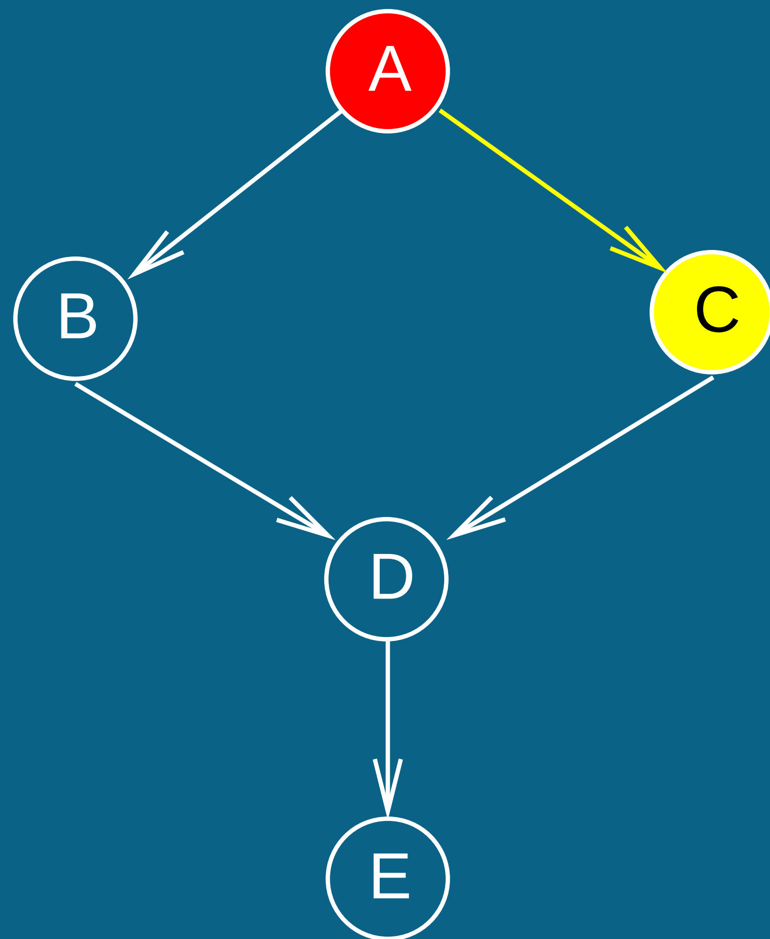
$$P(A, B, C, D, E) =$$

$$P(A)$$



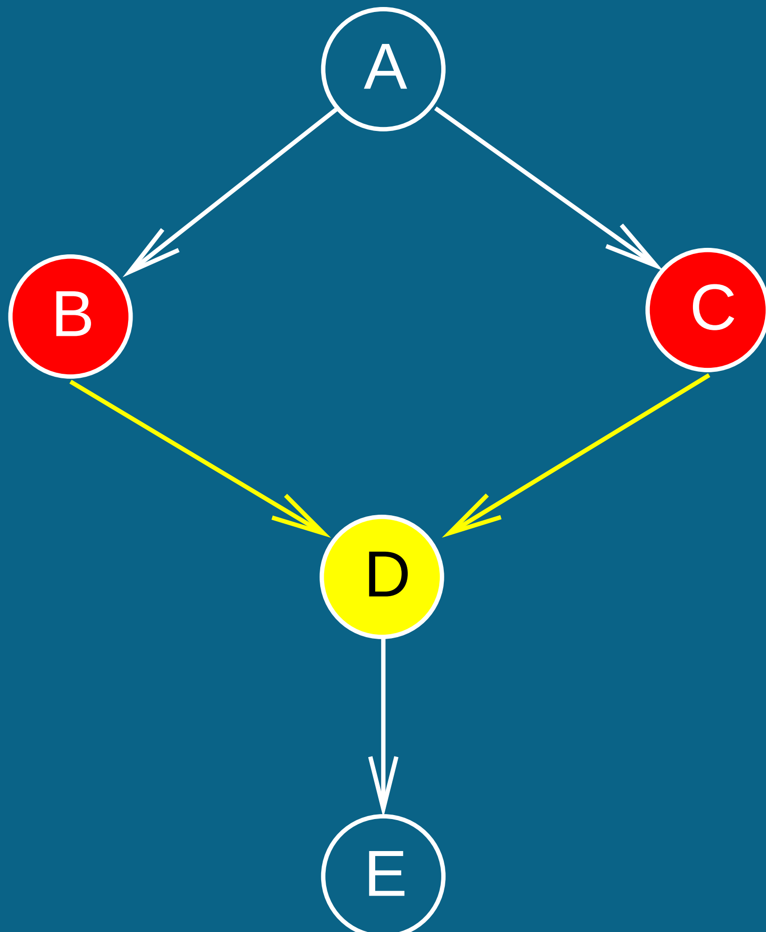
$$P(A, B, C, D, E) =$$

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



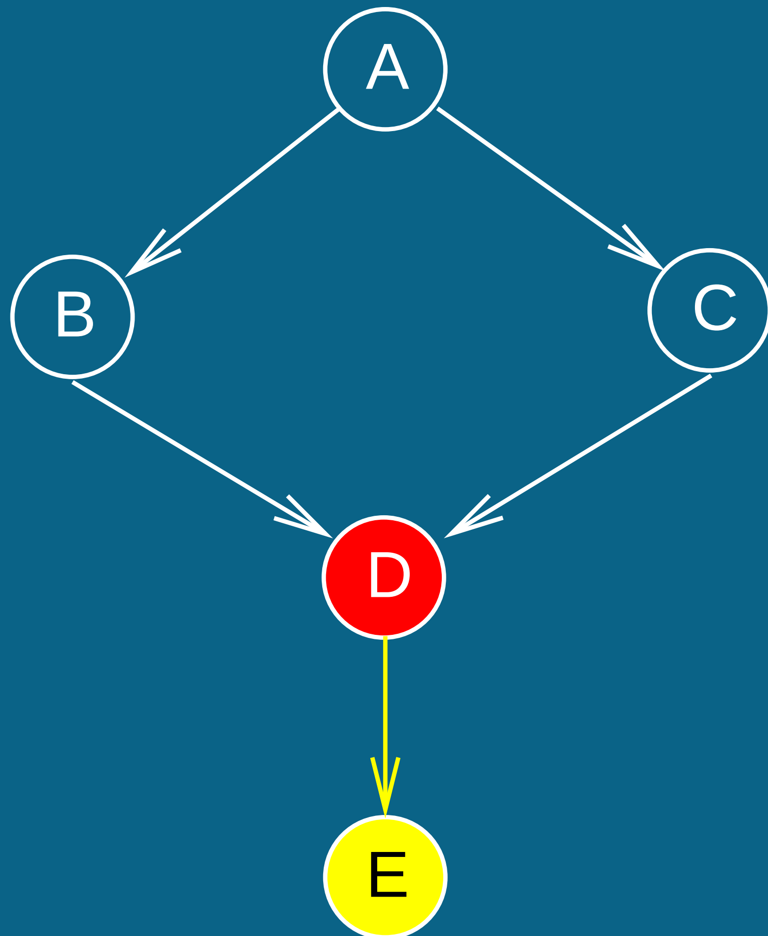
$$P(A, B, C, D, E) =$$

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



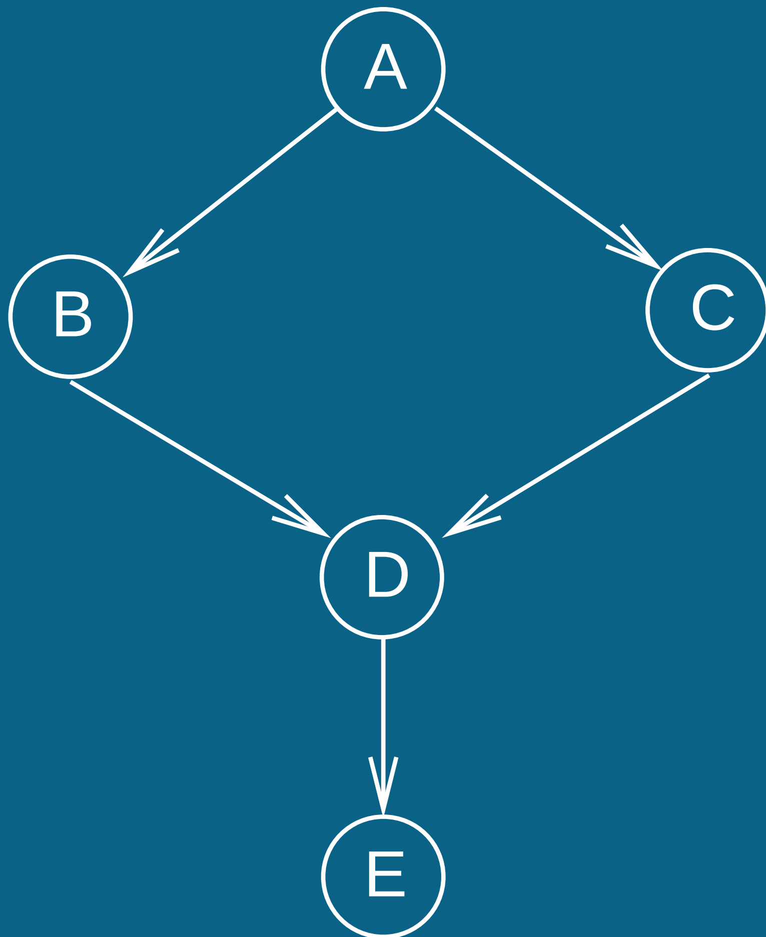
$$P(A, B, C, D, E) =$$

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



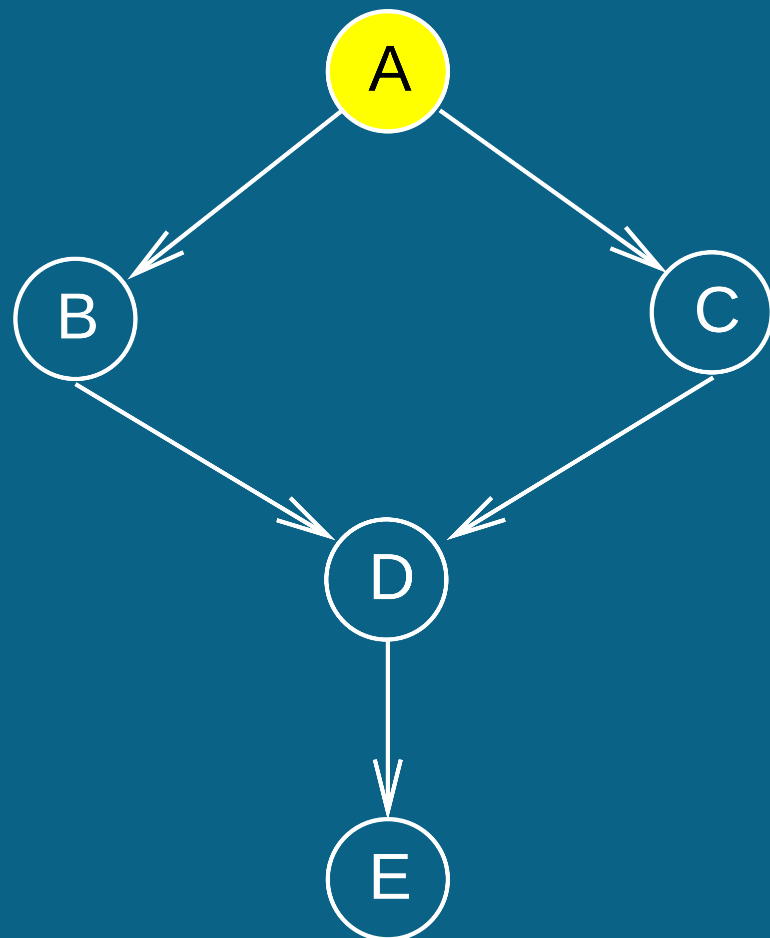
$$P(A, B, C, D, E) =$$

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

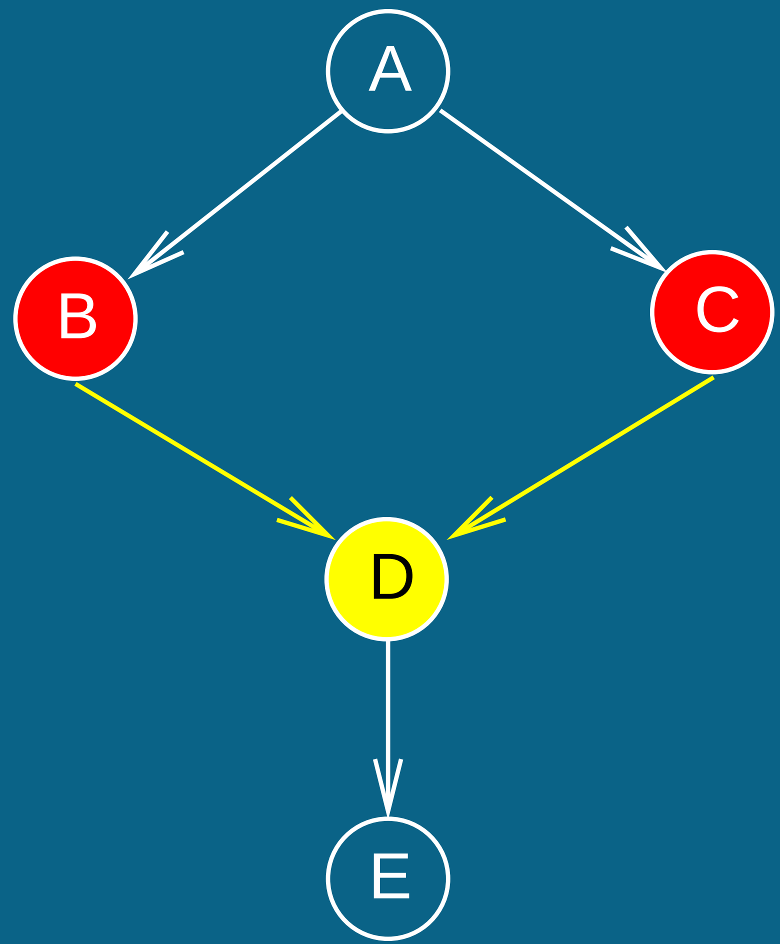


No cycles !

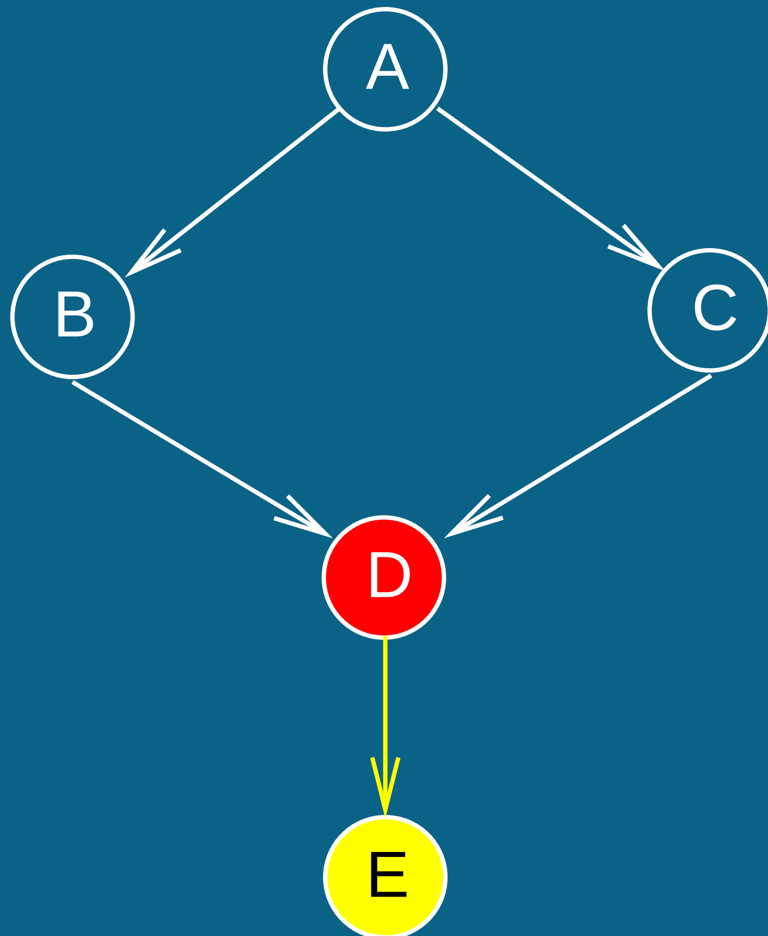
Biological interpretation



Initiation of cell (sub-)cycle



Co-regulation

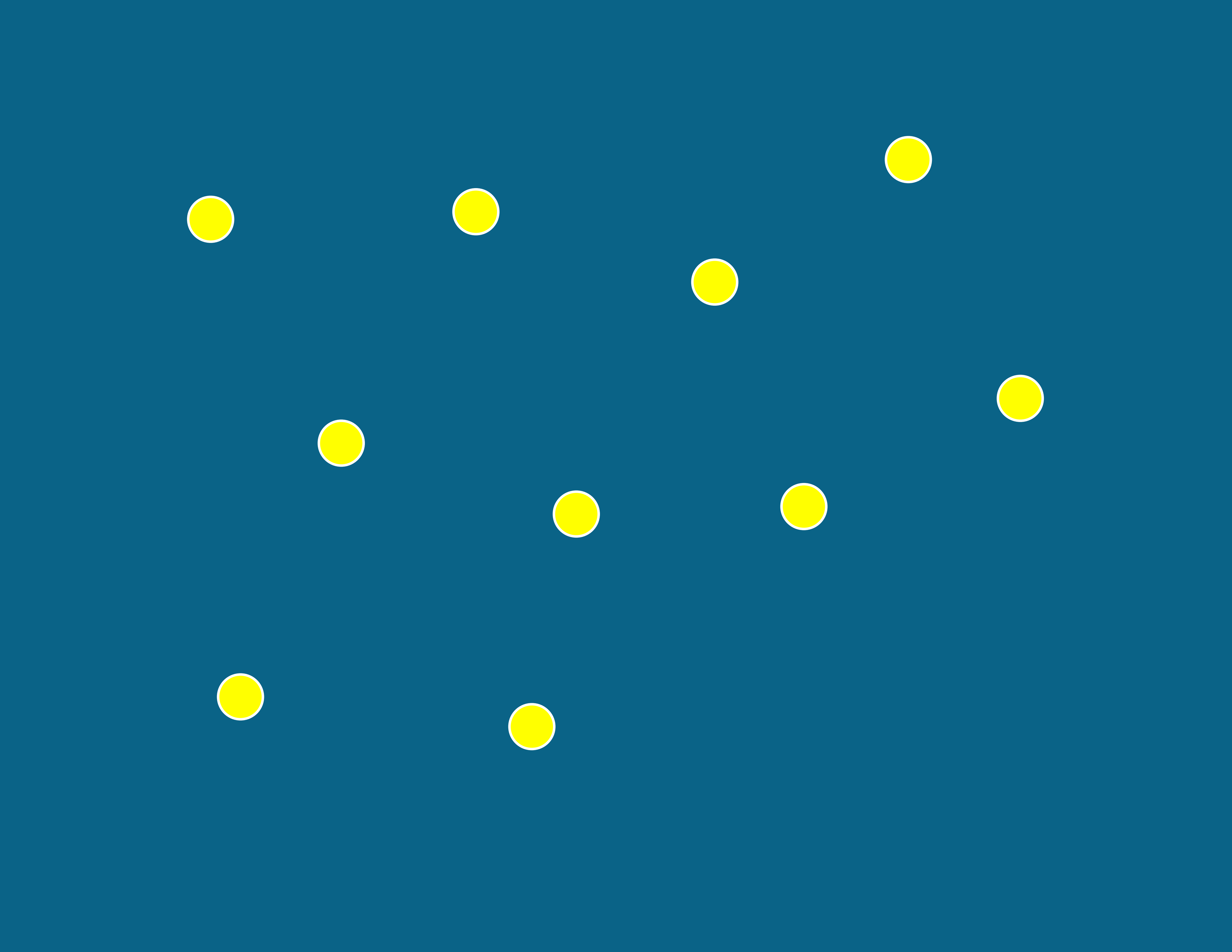


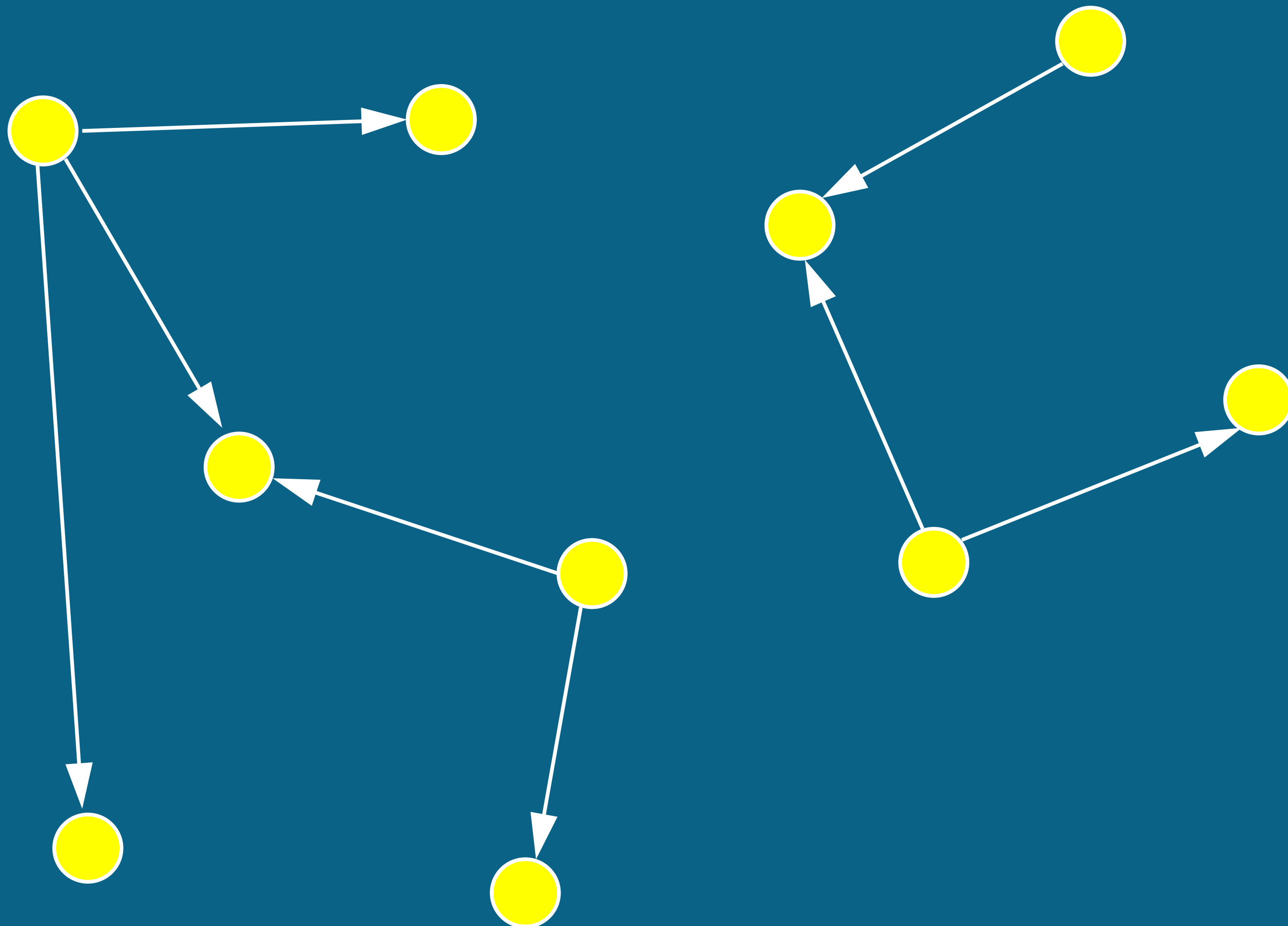
Mediation

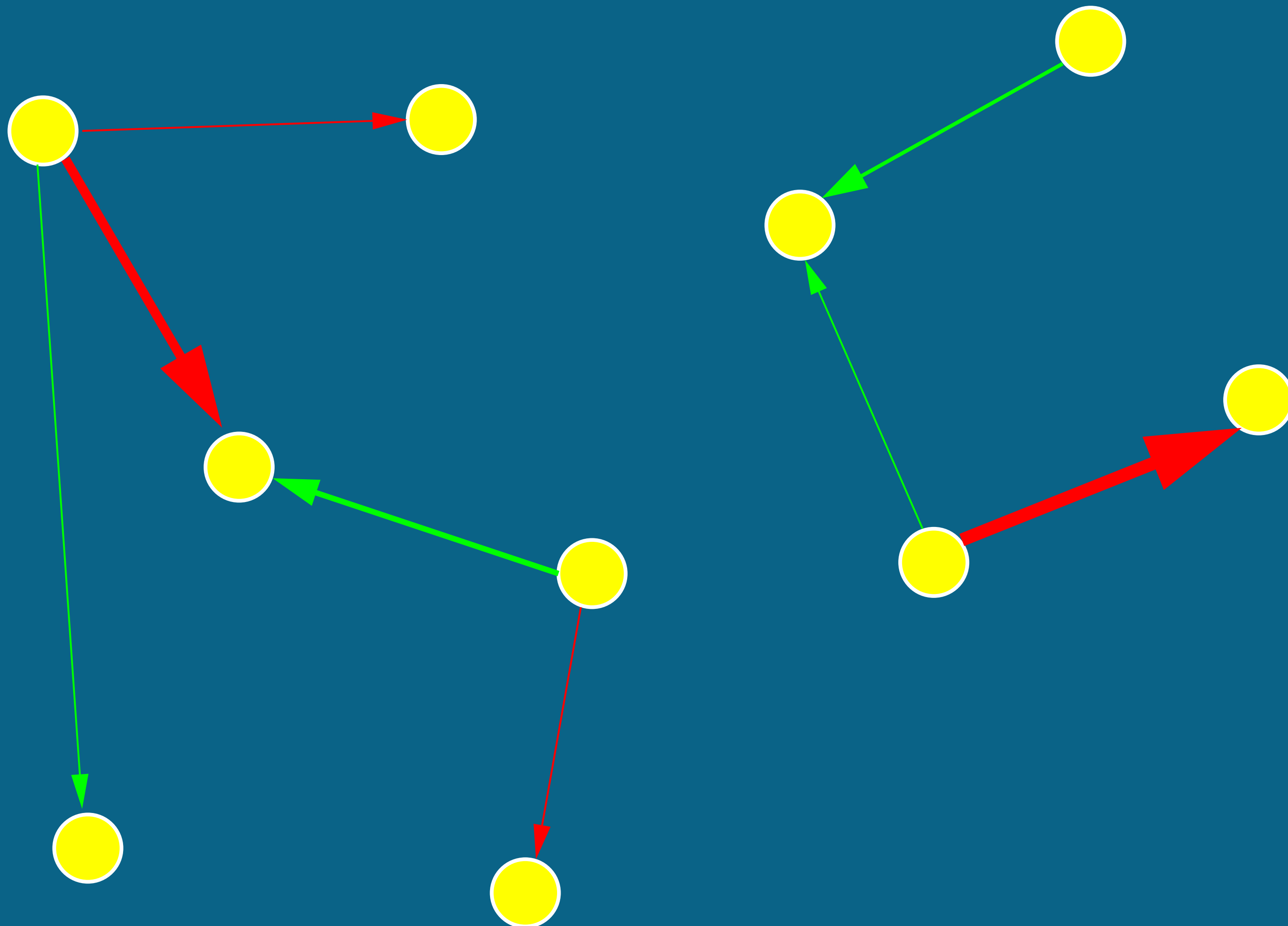
Application: Genetic Networks

Application: Genetic Networks

Can we learn structure from data?







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.

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.

Naive approach

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

Naive approach

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

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}

Objective: Sample from the posterior distribution

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

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

Objective: Sample from the posterior distribution

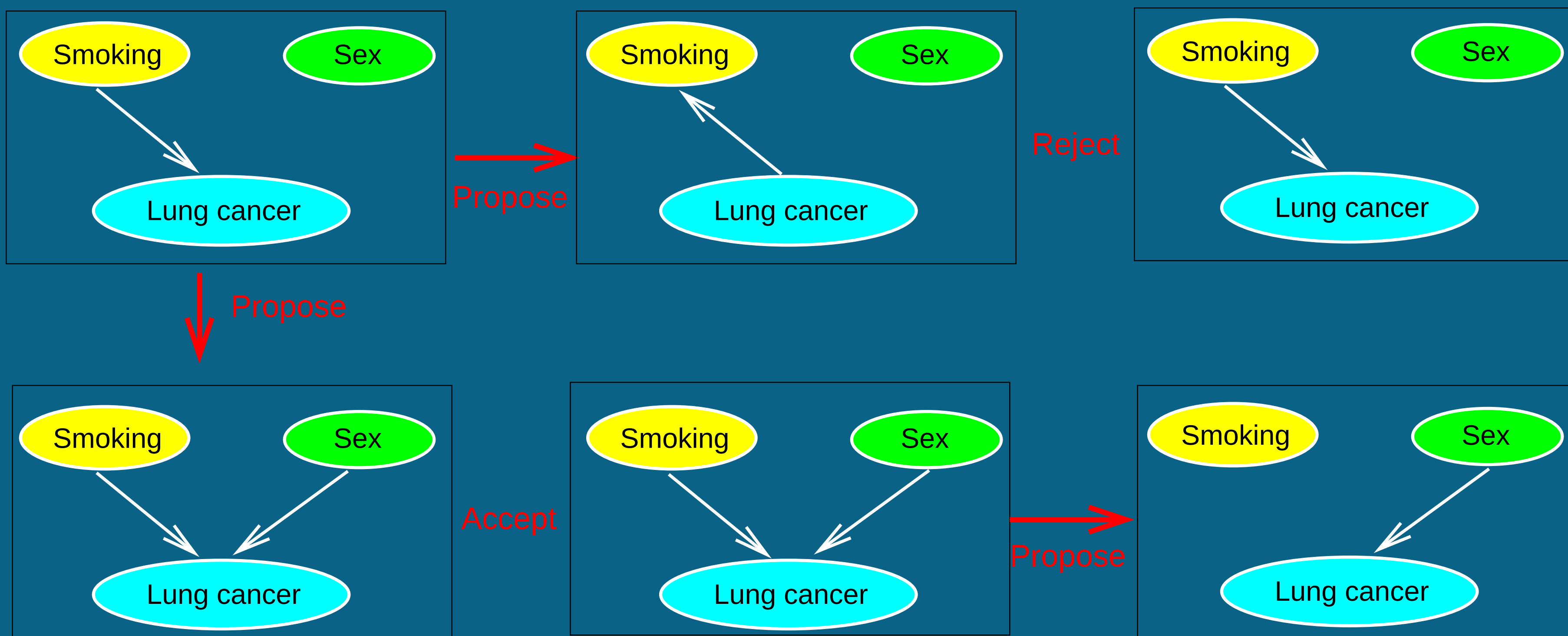
$$P(M_k|D) = \frac{P(D|M_k)P(M_k)}{\sum_i P(D|M_i)P(M_i)}$$

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

Markov chain Monte Carlo (MCMC):

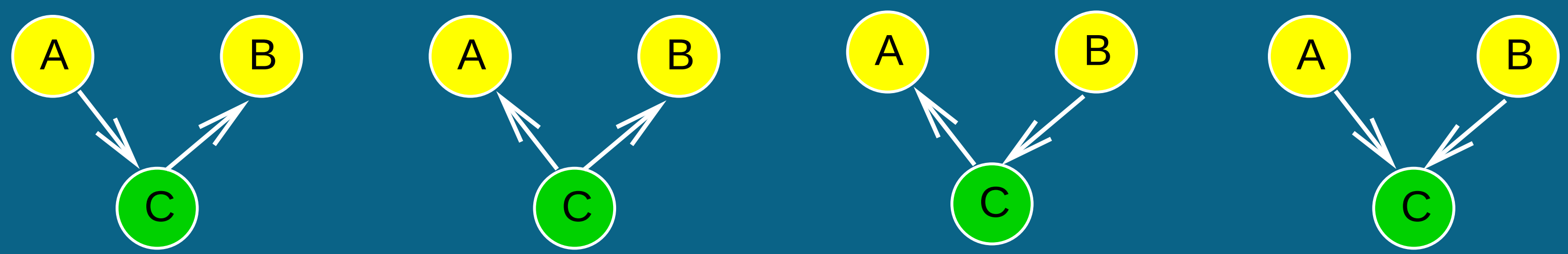
- Proposal move: Given network M_{old}, propose a new network M_{new} with probability $Q(M_{new}|M_{old})$.
- Acceptance/Rejection: Accept this new network with probability $\min\left\{1, \frac{P(D|M_{new})P(M_{new})}{P(D|M_{old})P(M_{old})} \times \frac{Q(M_{old}|M_{new})}{Q(M_{new}|M_{old})}\right\}$

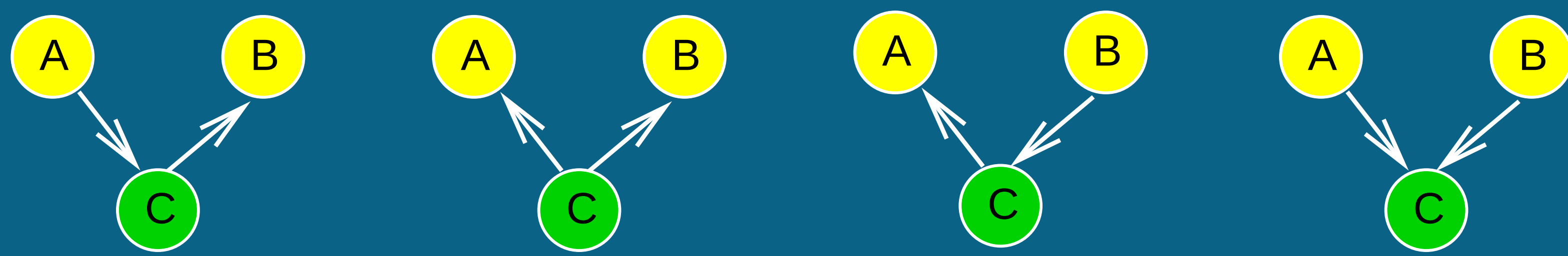
Markov chain Monte Carlo (MCMC)



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

Limitations of Friedman's approach

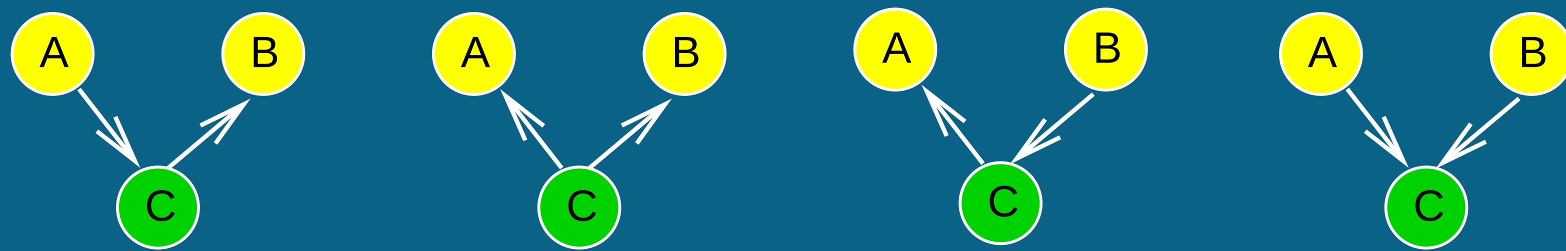

A
B
C
A
B
C
A
B
C
A
B
C



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

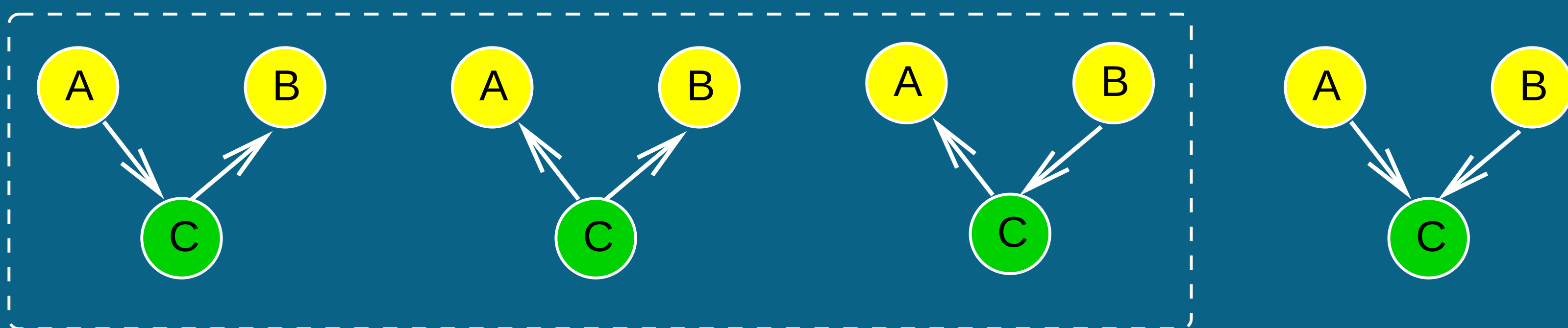


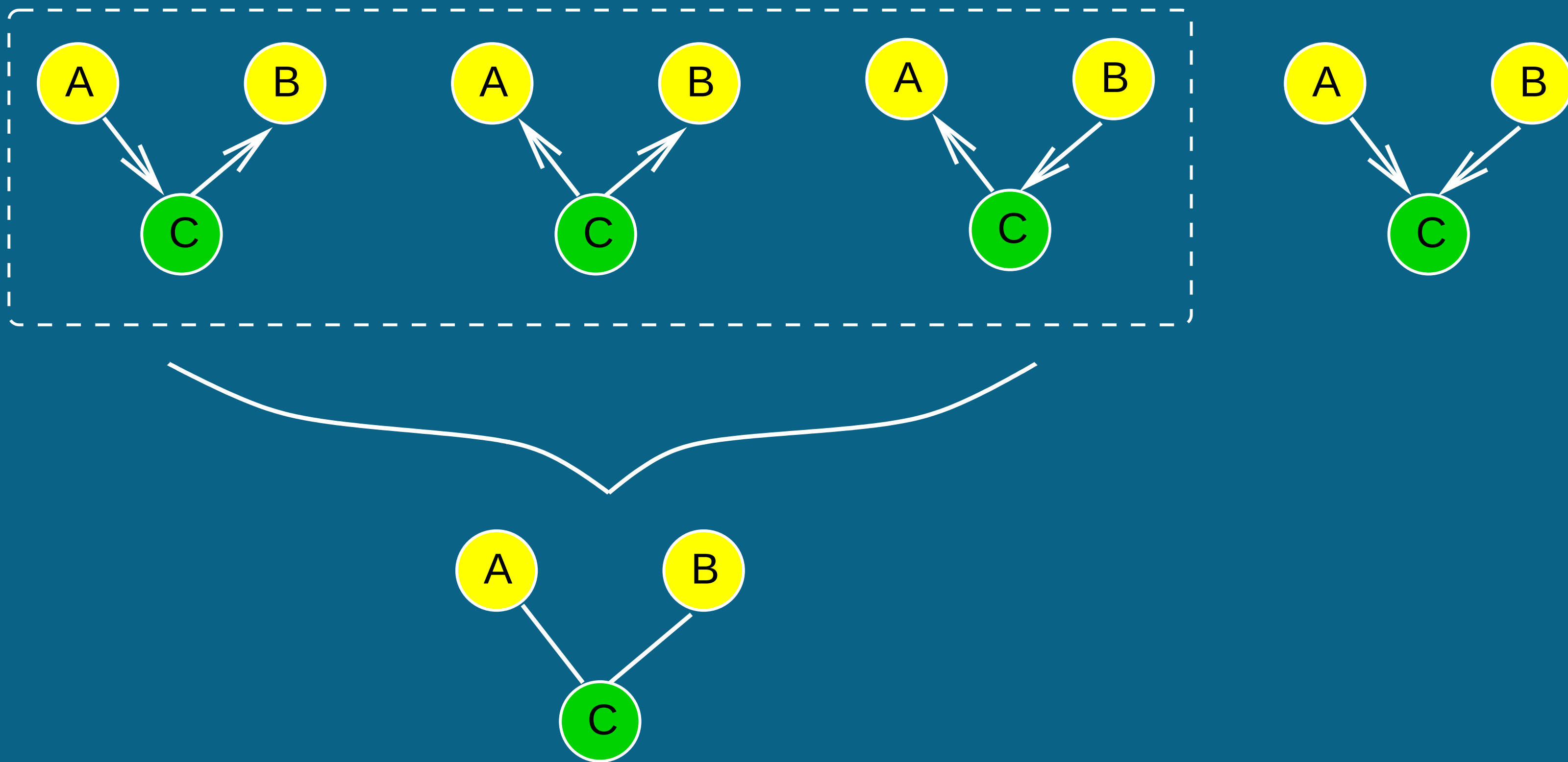
$$P(A,B,C) =$$

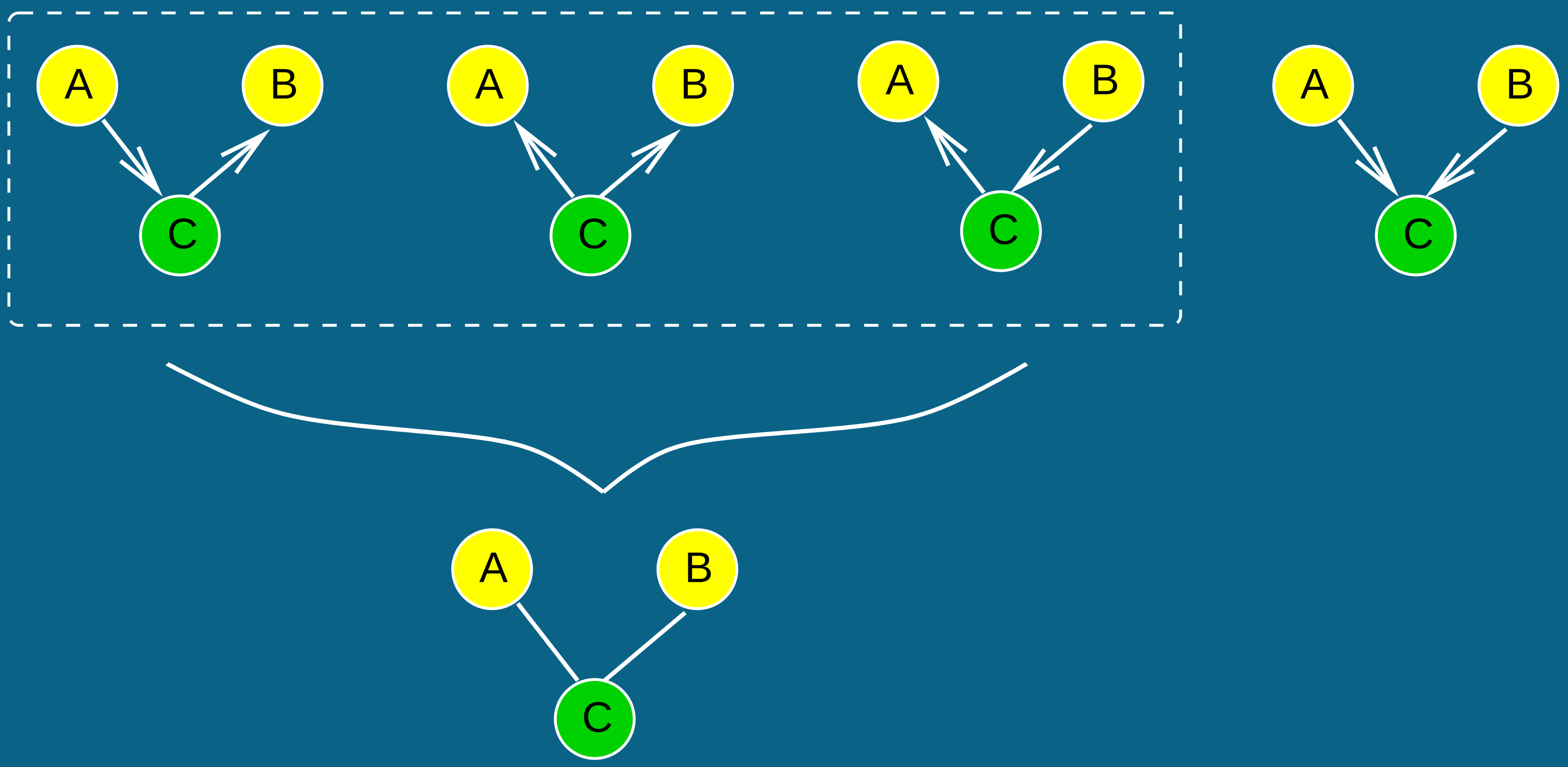
$P(B|C)\ P(C|A)\ P(A)$ | $P(A|C)\ P(B|C)\ P(C)$ | $P(A|C)\ P(C|B)\ P(B)$ | $P(C|A,B)\ P(A)\ P(B)$

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

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

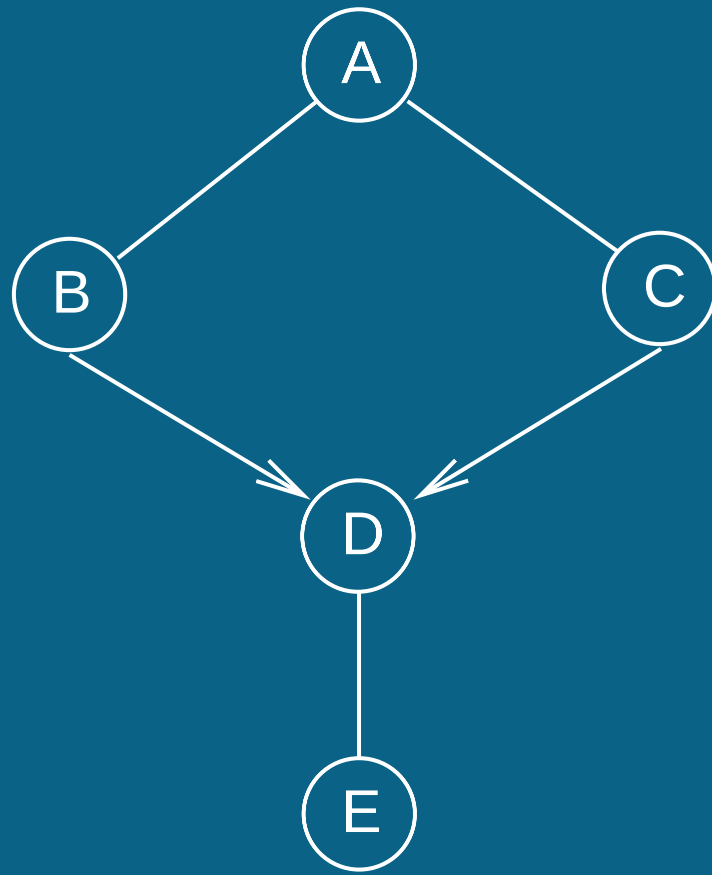



A
B
C
A
B
C
A
B
C
A
B
C
A
B
C

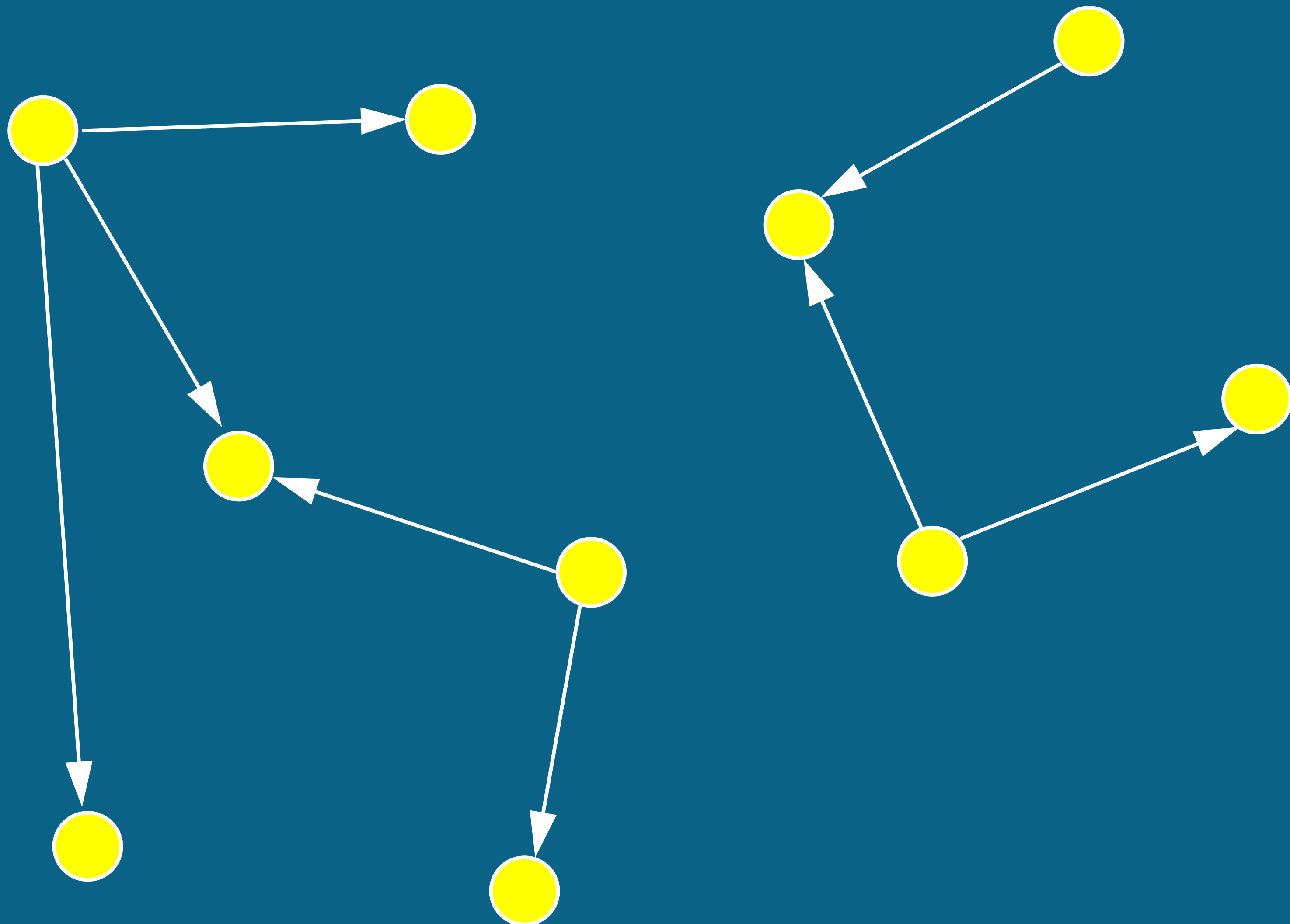


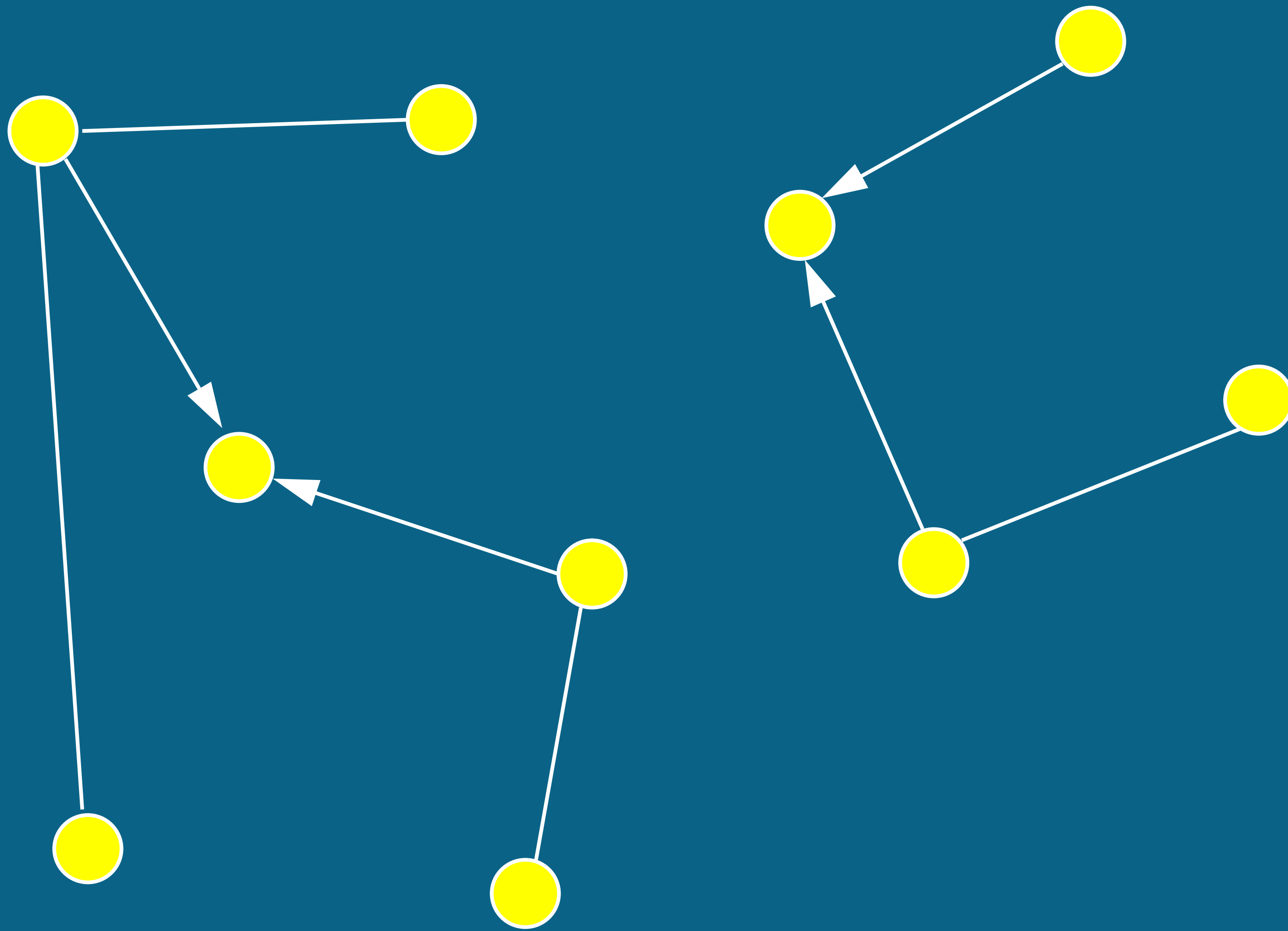
- Two DAGs are equivalent iff they have the same skeleton (= the underlying undirected graph) and the same v-structure.
- v-structure: Converging directed edges into the same node without an edge between the parents.

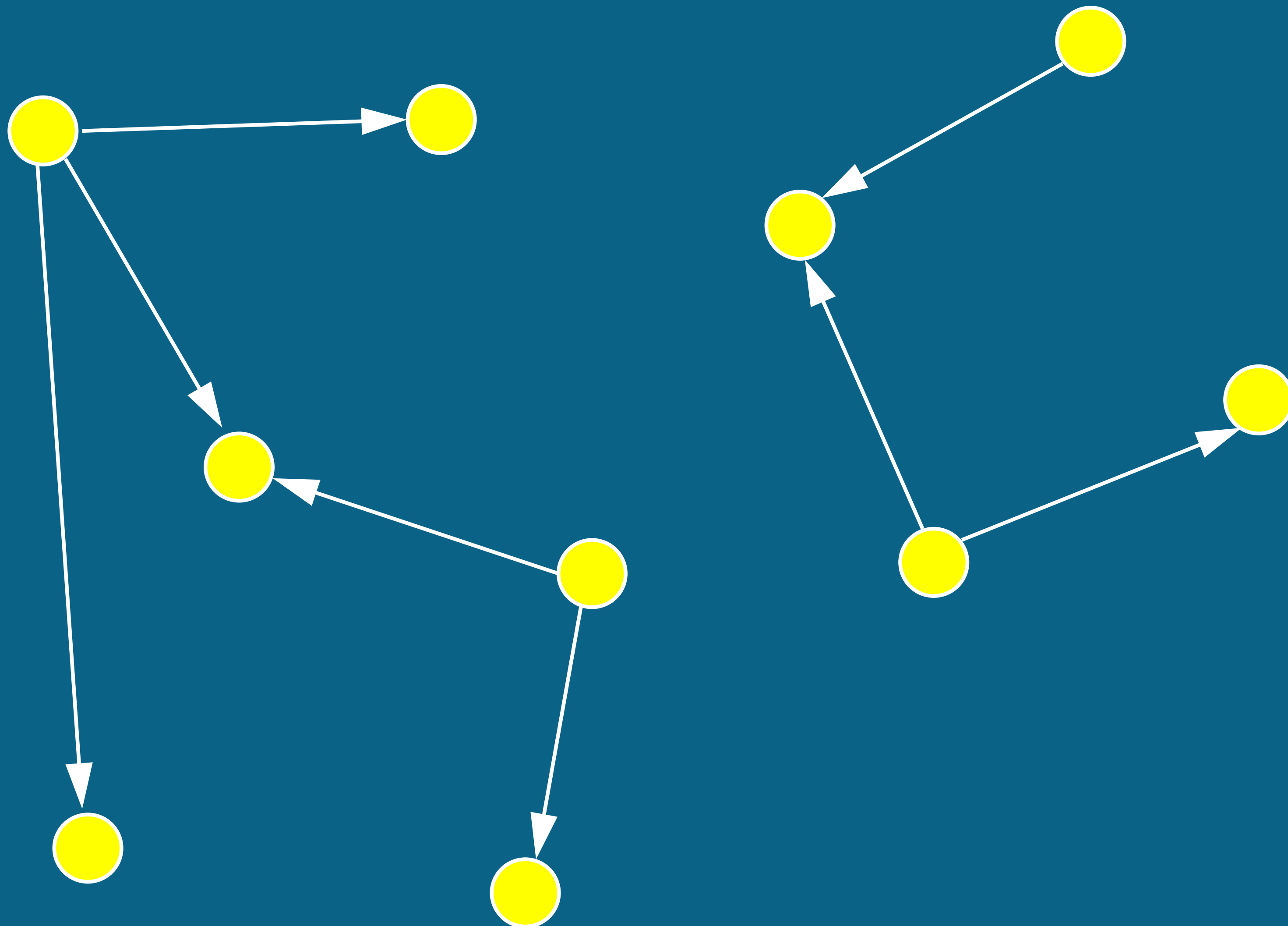
An equivalence class of DAGs can be represented by a PDAG (partially directed acyclic graph).

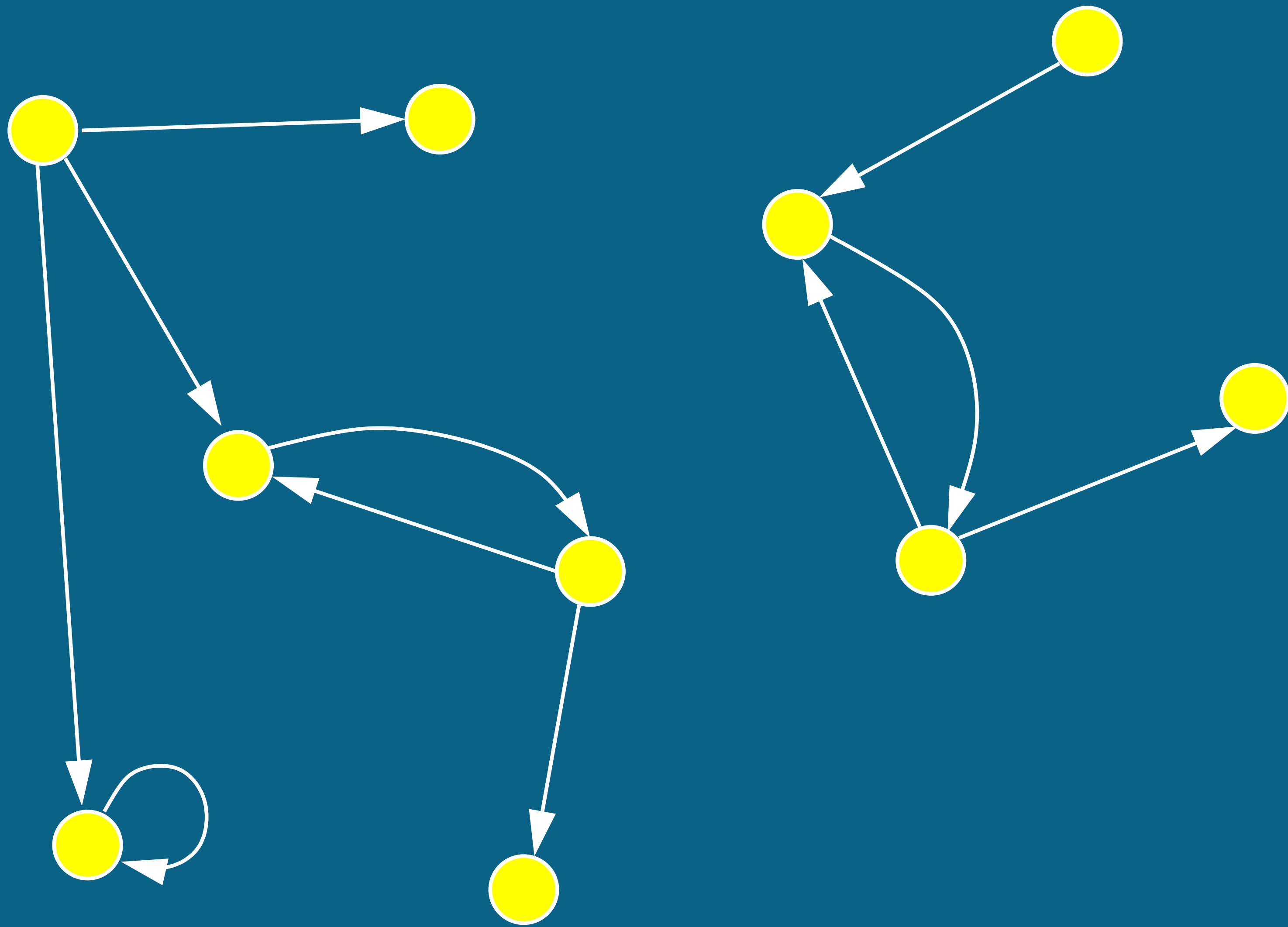


We can only learn PDAGs from the data!





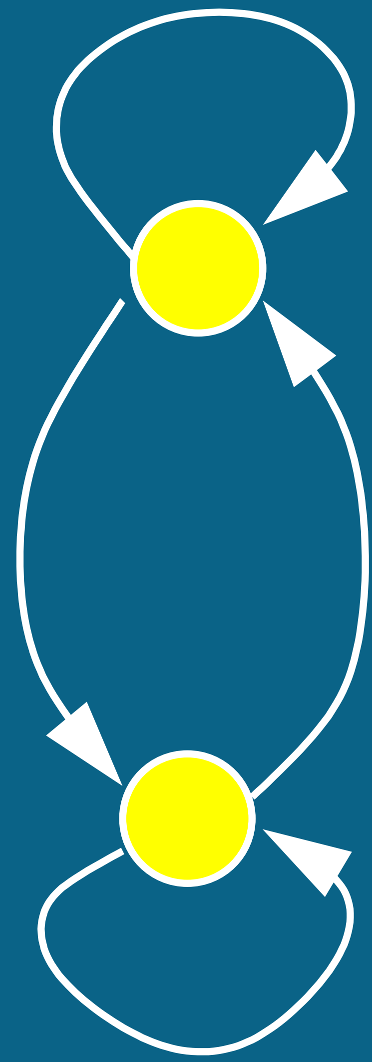


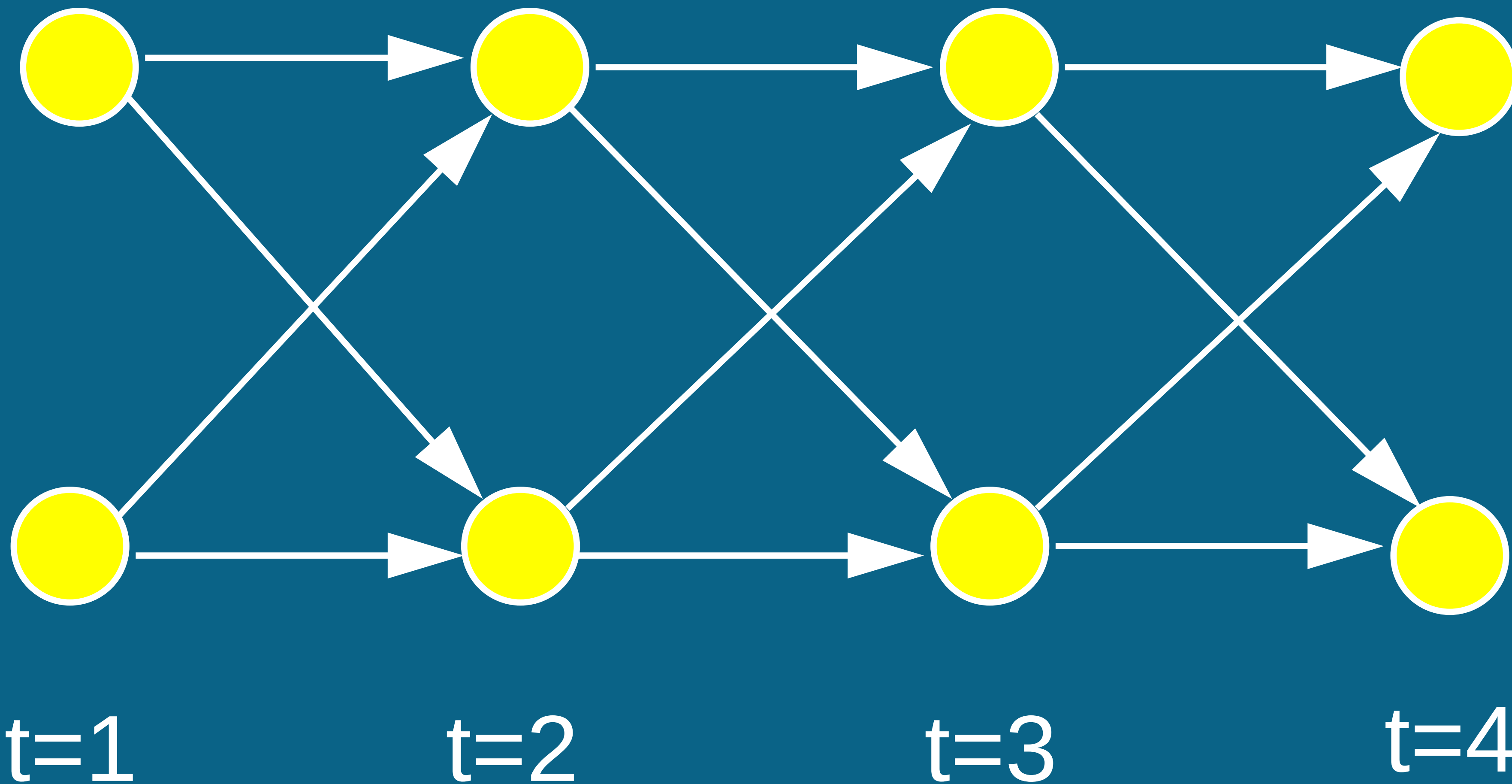


Causality and feedback

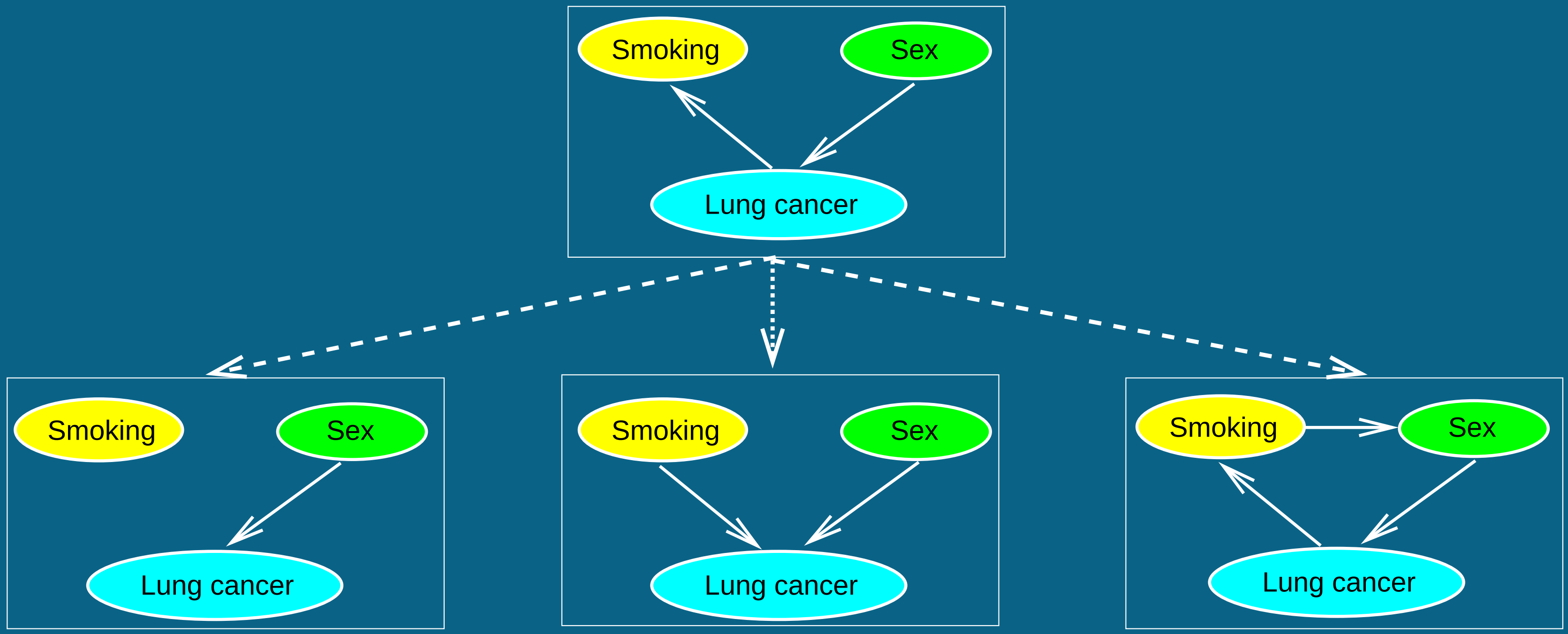
Causality and feedback

Dynamic Bayesian networks

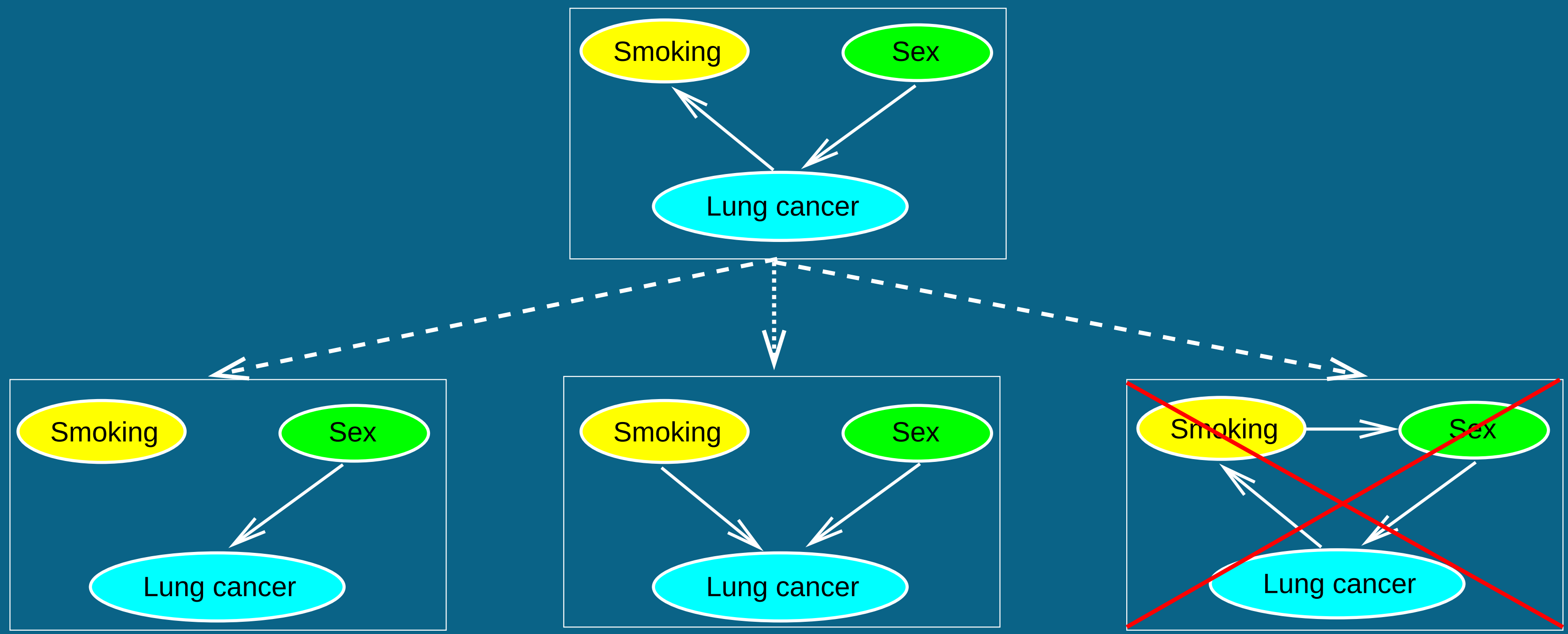


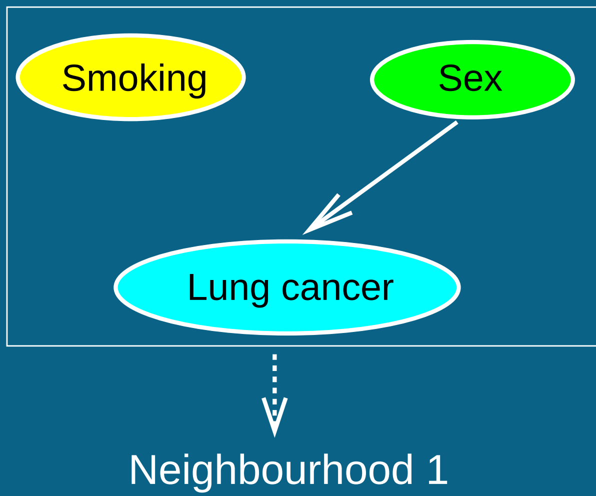

t=1
t=2
t=3
t=4

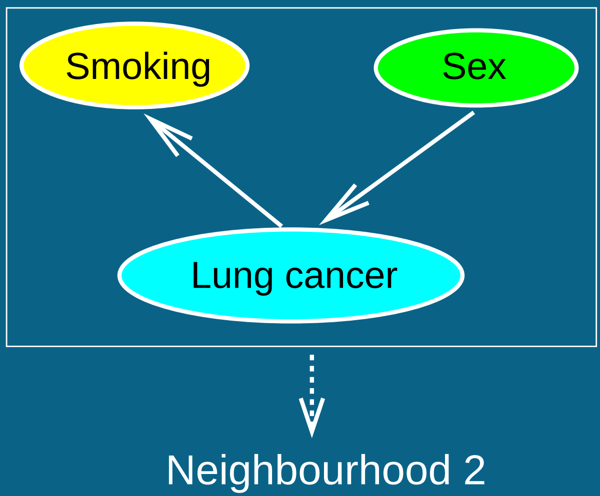
Hastings ratio

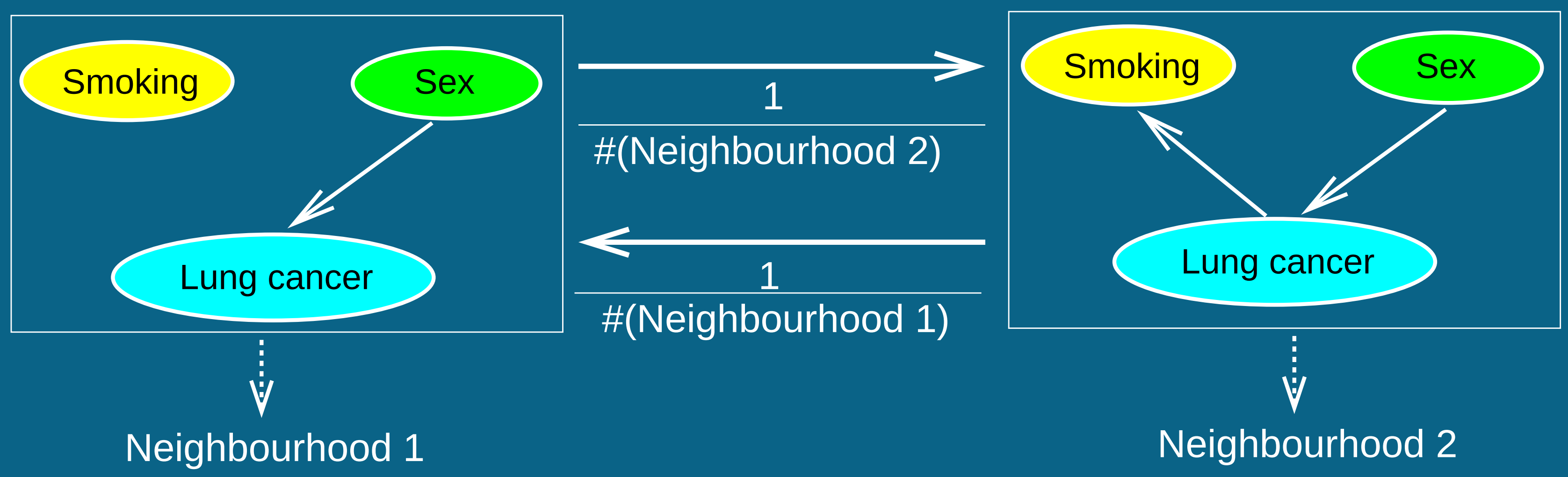


Smoking
Sex
Lung cancer
Smoking
Sex
Lung cancer
Smoking
Sex
Lung cancer
Smoking
Sex
Lung cancer


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


Smoking
Sex
Lung cancer
Neighbourhood 1

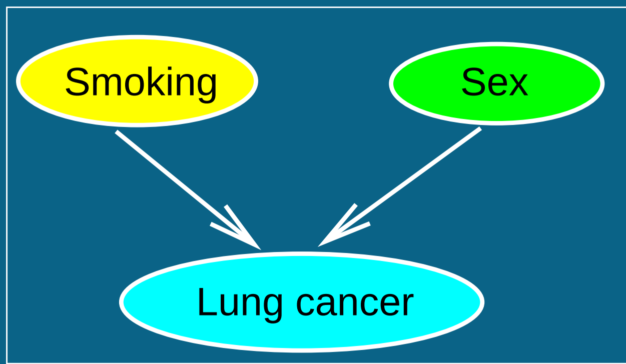

Smoking
Sex
Lung cancer
Neighbourhood 2

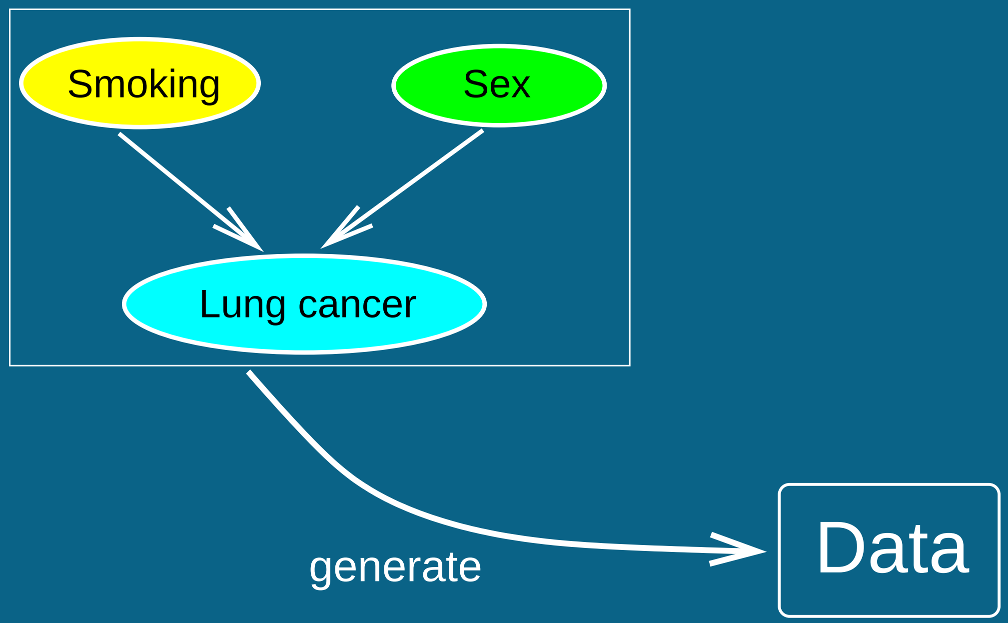

Smoking
Sex
Lung cancer
Neighbourhood 1
1
#(Neighbourhood 2)
1
#(Neighbourhood 1)
Smoking
Sex
Lung cancer
Neighbourhood 2

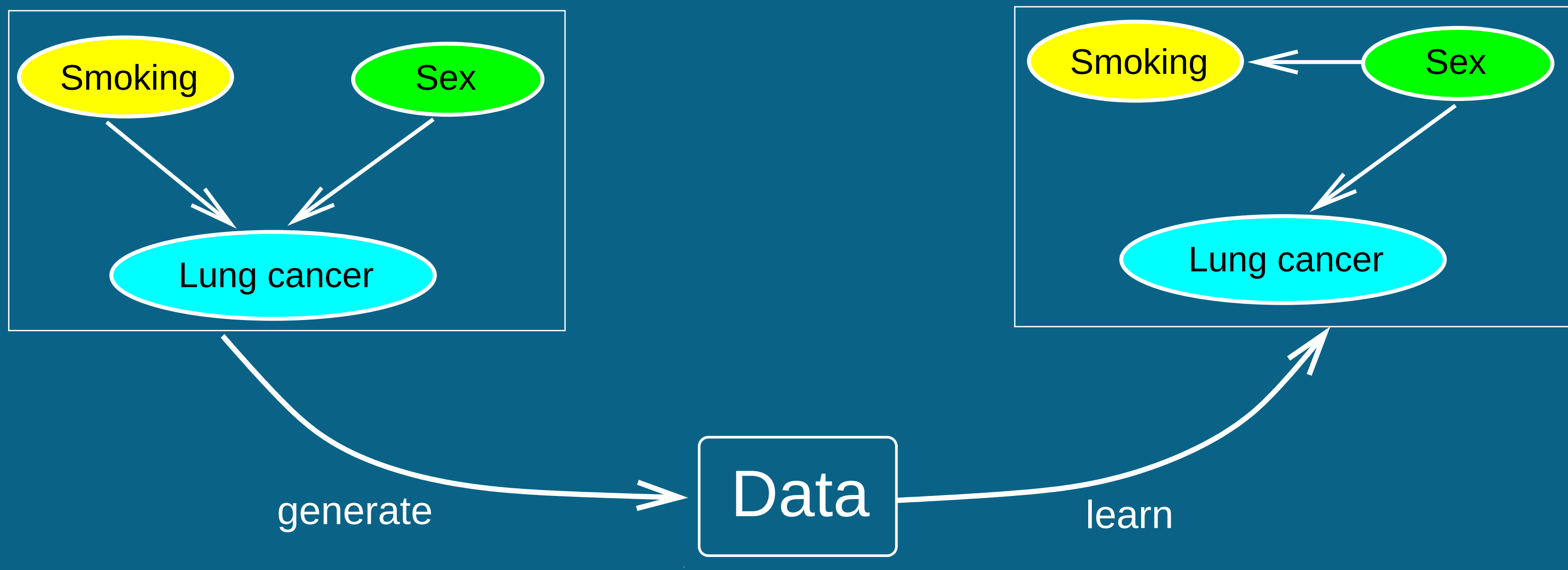
Dynamic Bayesian network

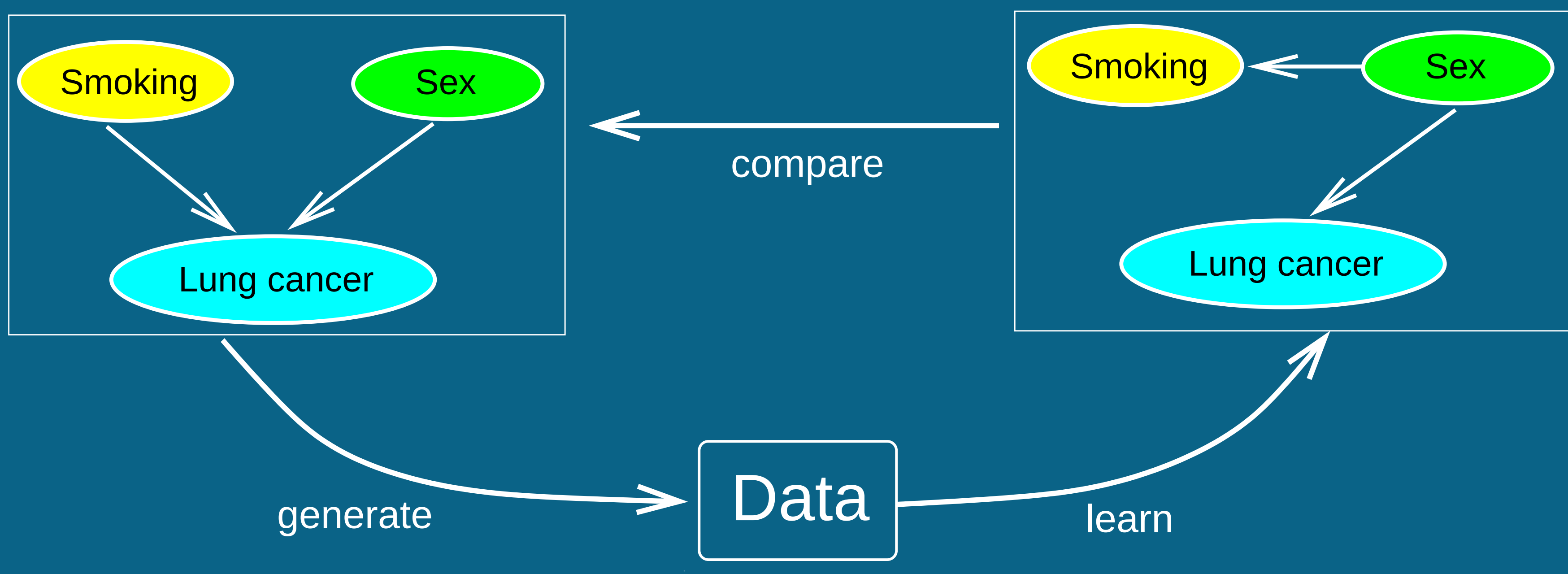
Hastings ratio $\longrightarrow$ trivial

How many samples are needed ?

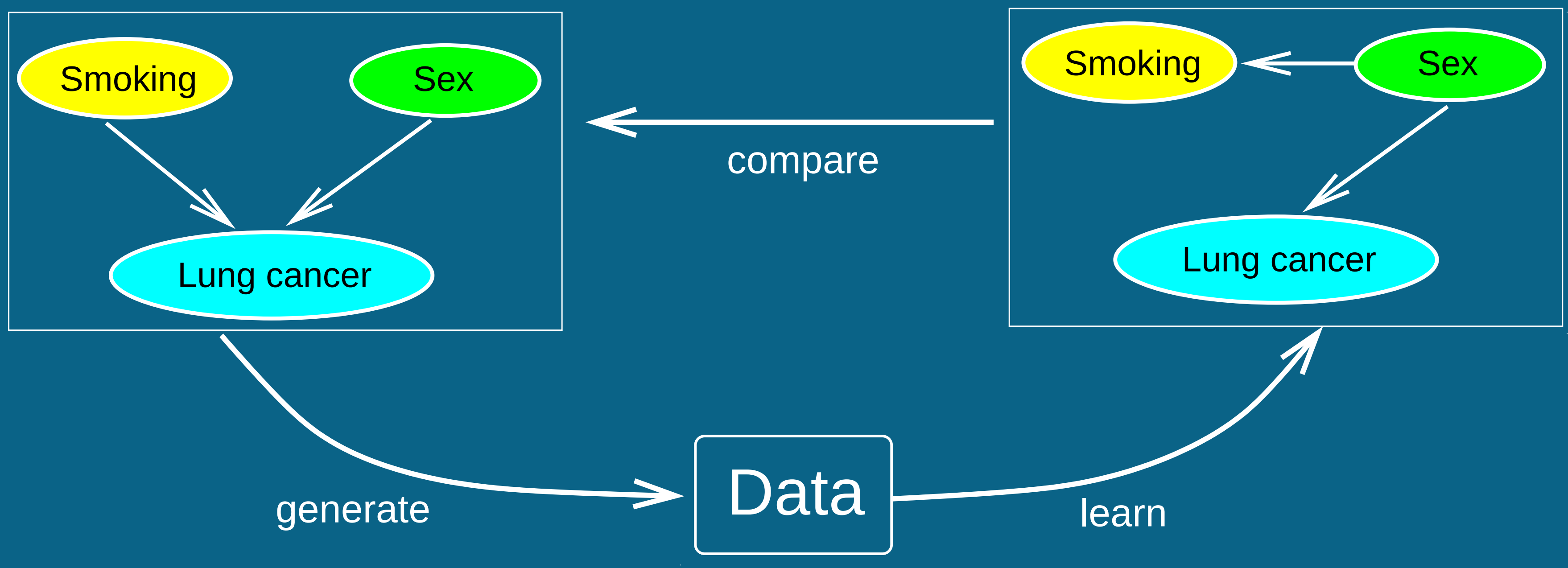

Smoking
Sex
Lung cancer


Smoking
Sex
Lung cancer
generate
Data


Smoking
Sex
Lung cancer
generate
Data
learn
Smoking
Sex
Lung cancer

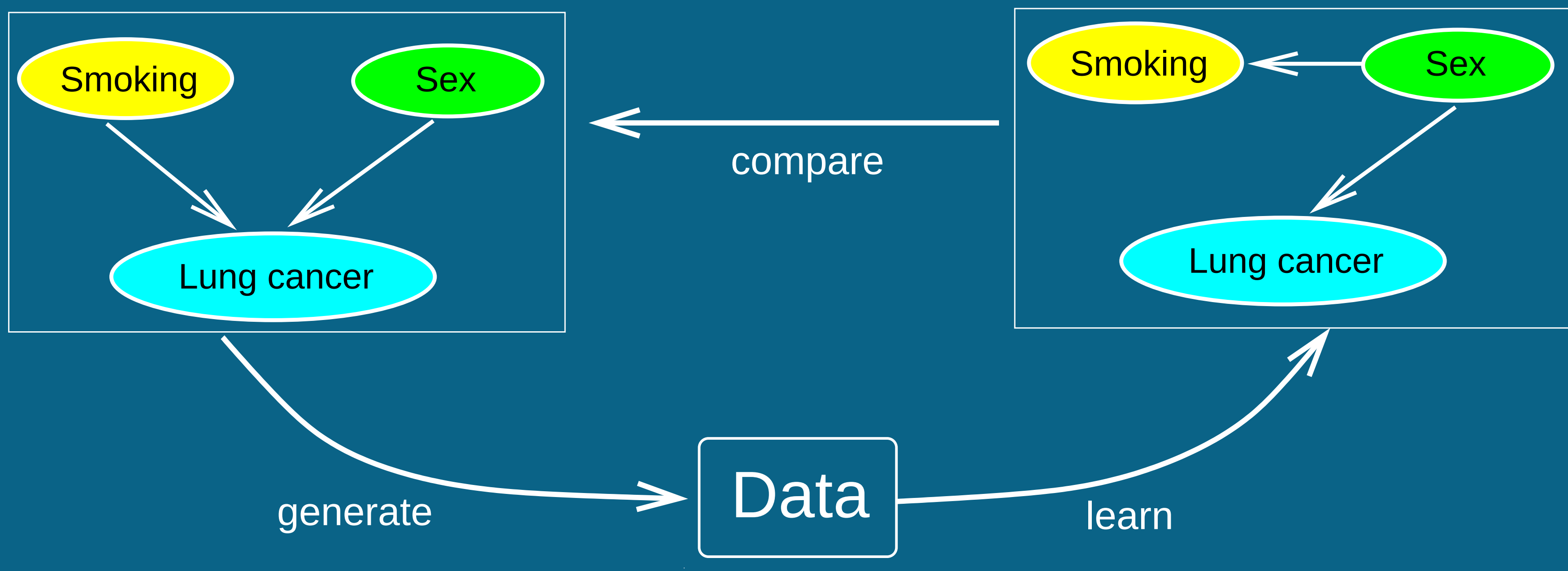


Smoking
Sex
Lung cancer
compare
Smoking
Sex
Lung cancer
generate
Data
learn



Sensitivity:

Probability of recovering a true edge

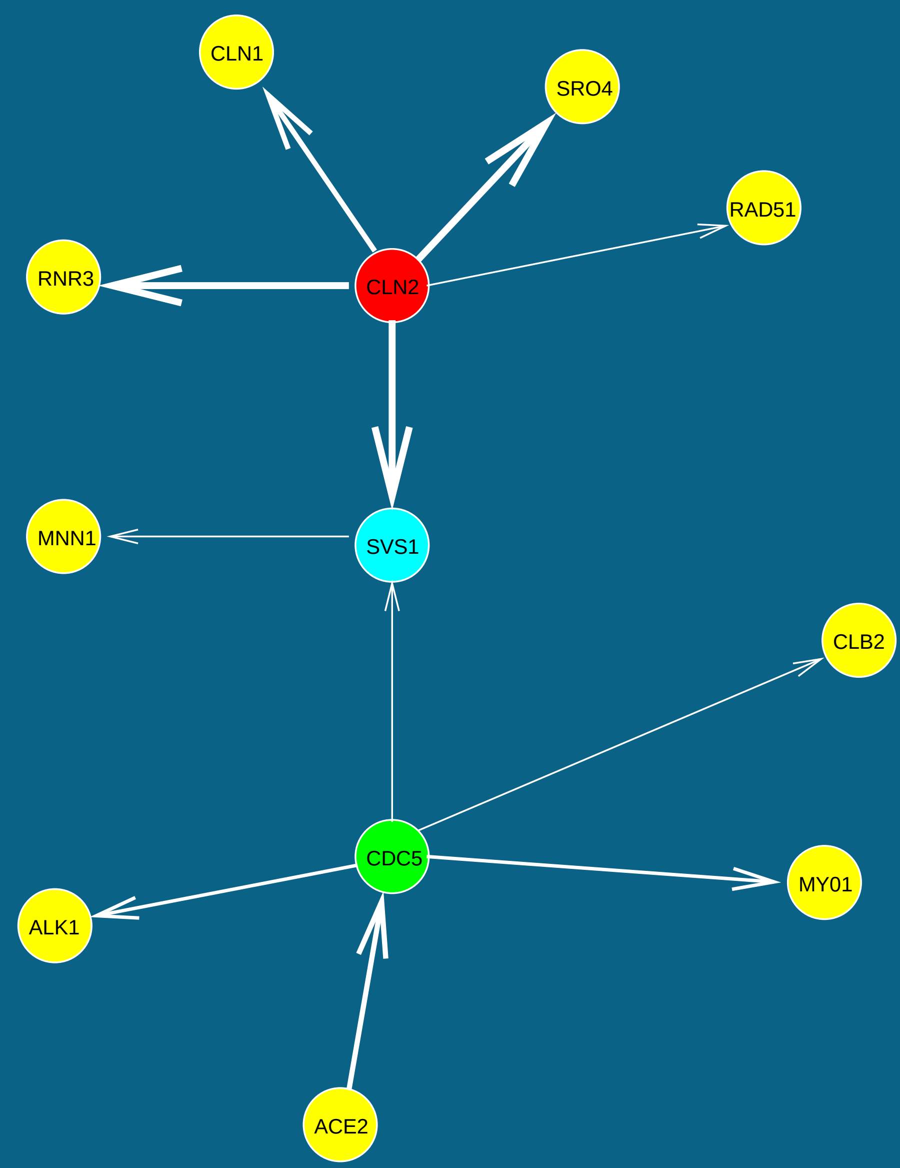


Sensitivity:

Probability of recovering a true edge

Specificity:

Probability of suppressing a spurious edge


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

- Parameters:
  - Binomial, quasi deterministic
  - Binomial, stochastic
  - Trinomial, stochastic
- Training set size: 100, 30, 7
- Repetitions: 5 times with different random number generator seeds.
- MCMC: 50,000 steps equilibration; 50,000 steps sampling

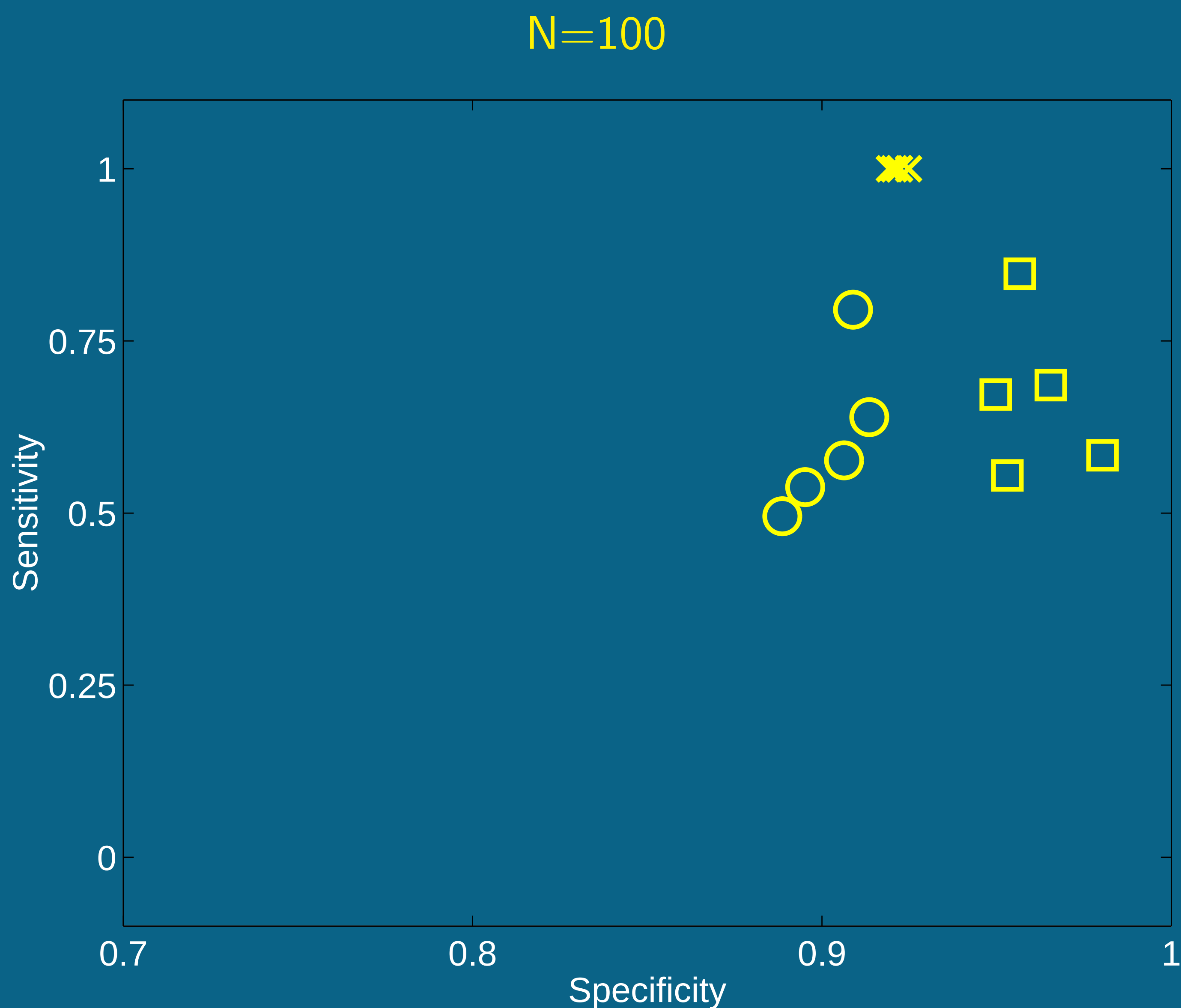
Prior

Genes co-regulated by only a few factors:

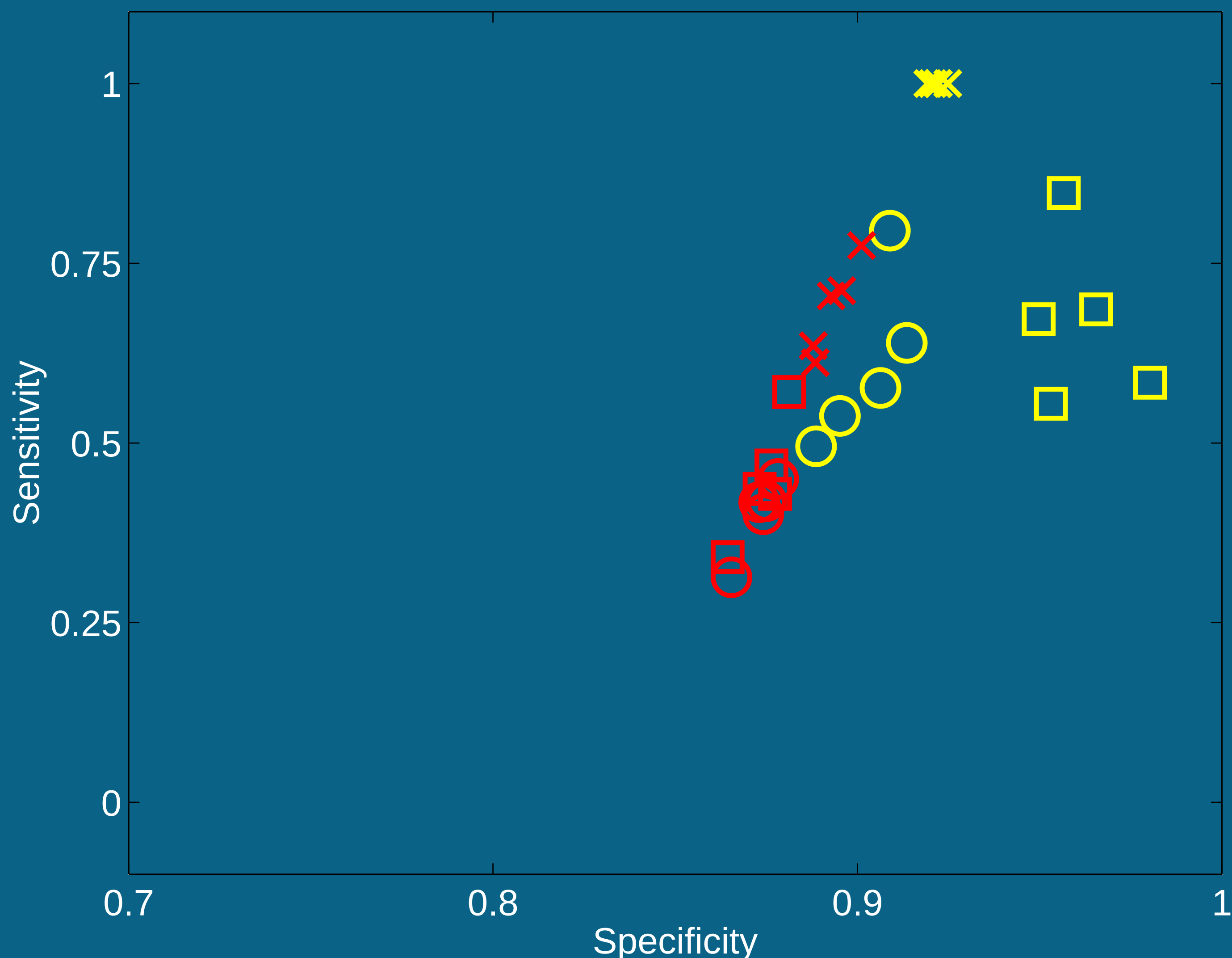
Maximum fan-in of 2

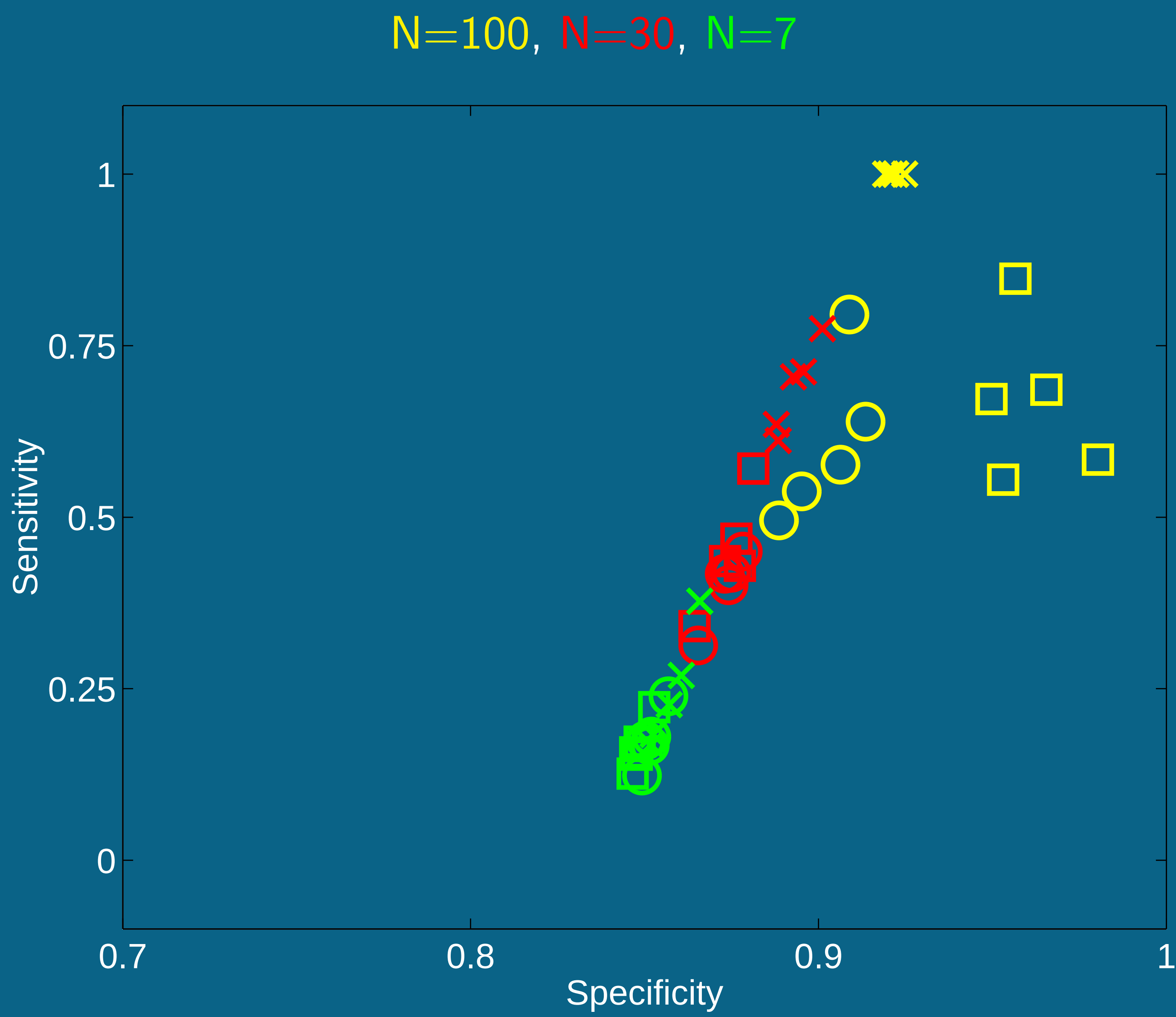
Small number of activators that control the cell cycle:

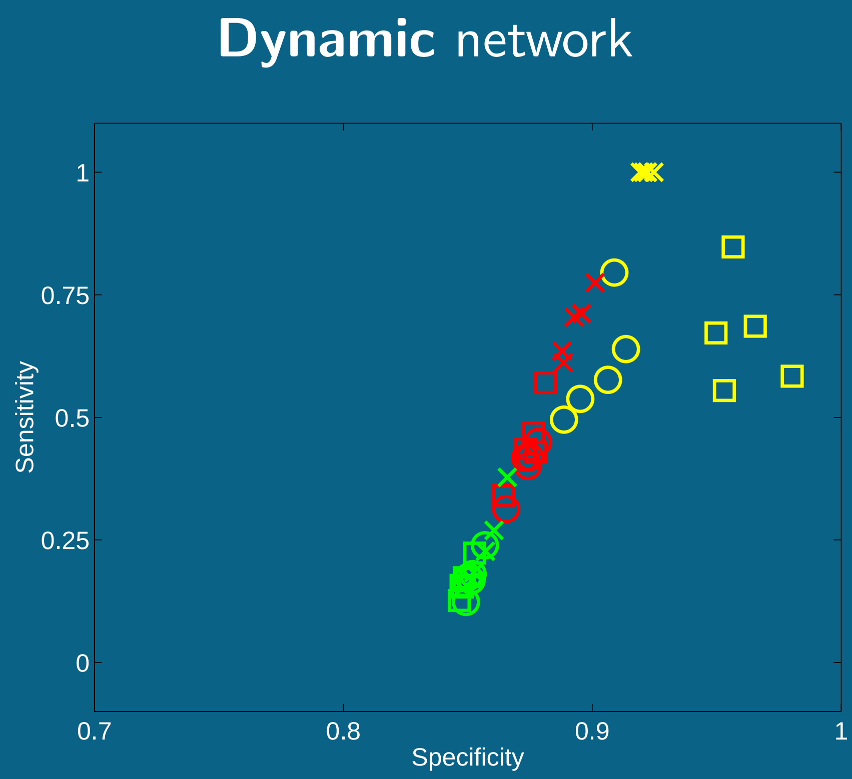
Geometric distribution on the number of nodes with non-zero fan-out (expectation value $= 2$)

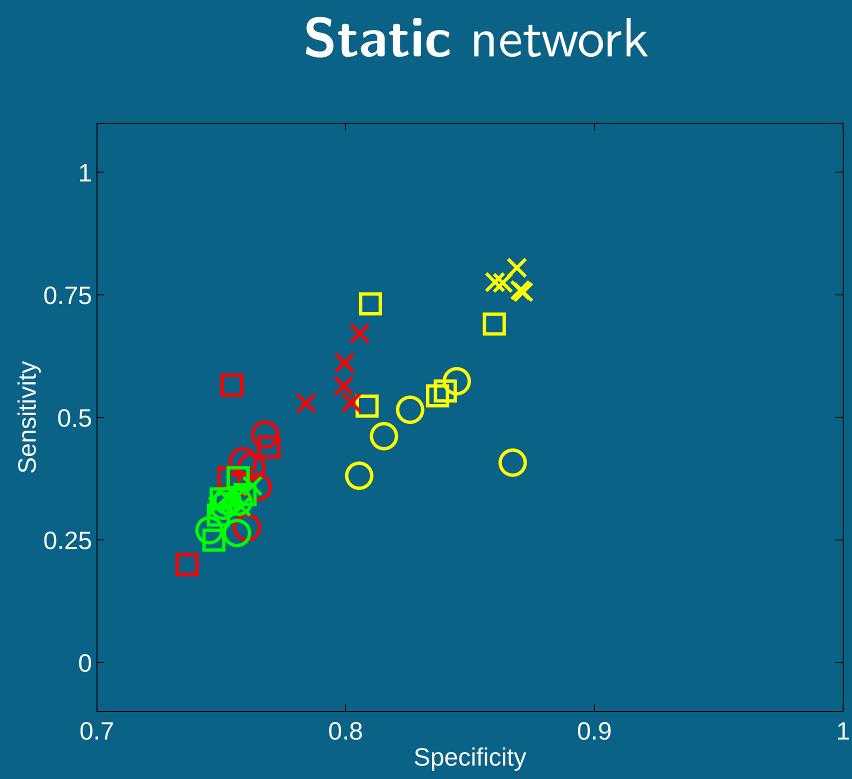

N=100
1
0.75
0.5
0.25
0
Sensitivity
0.7
0.8
0.9
1
Specificity

N=100, N=30

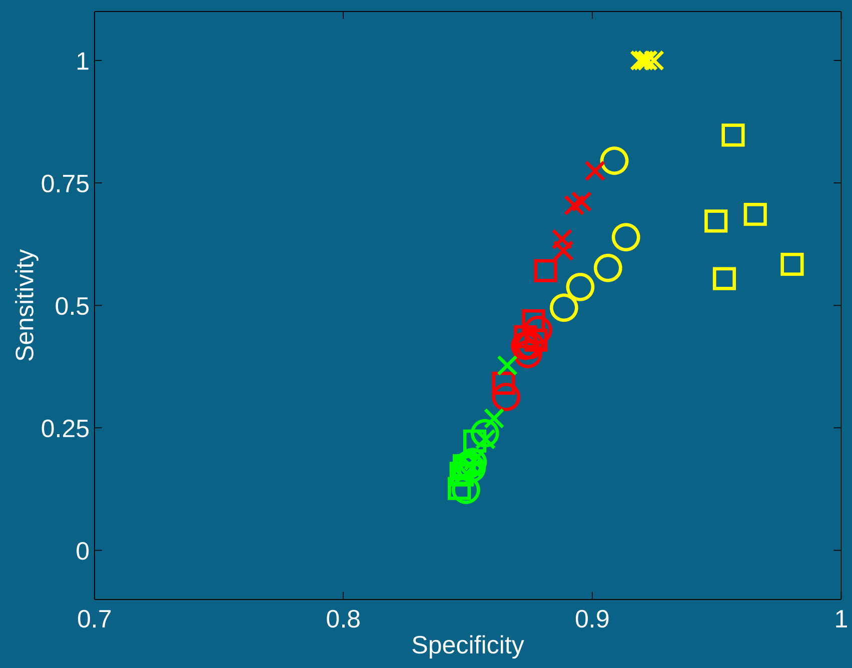



N=100, N=30, N=7
Sensitivity
Specificity
0
0.25
0.5
0.75
1
0.7
0.8
0.9
1

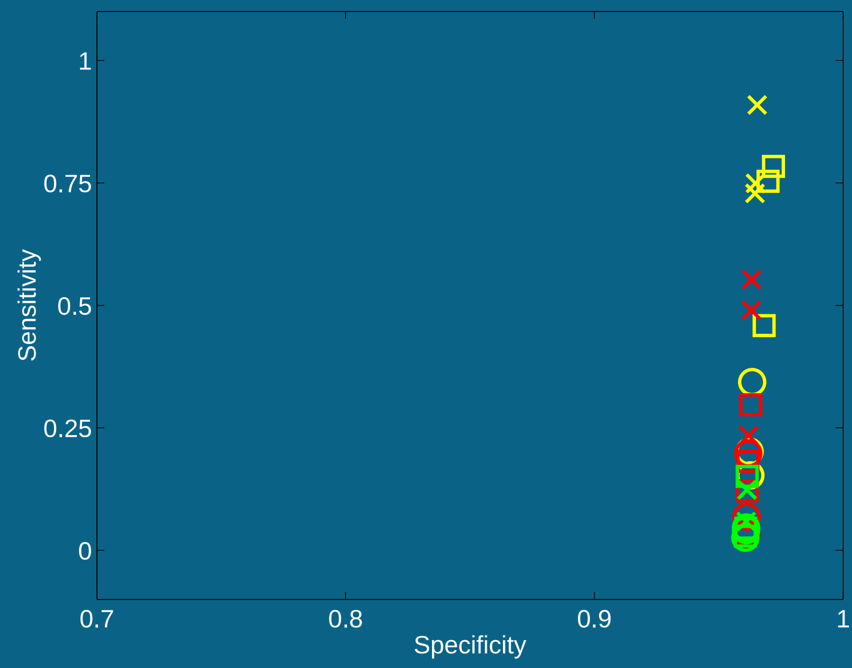

Dynamic network
Sensitivity
Specificity
0
0.25
0.5
0.75
1
0.7
0.8
0.9
1

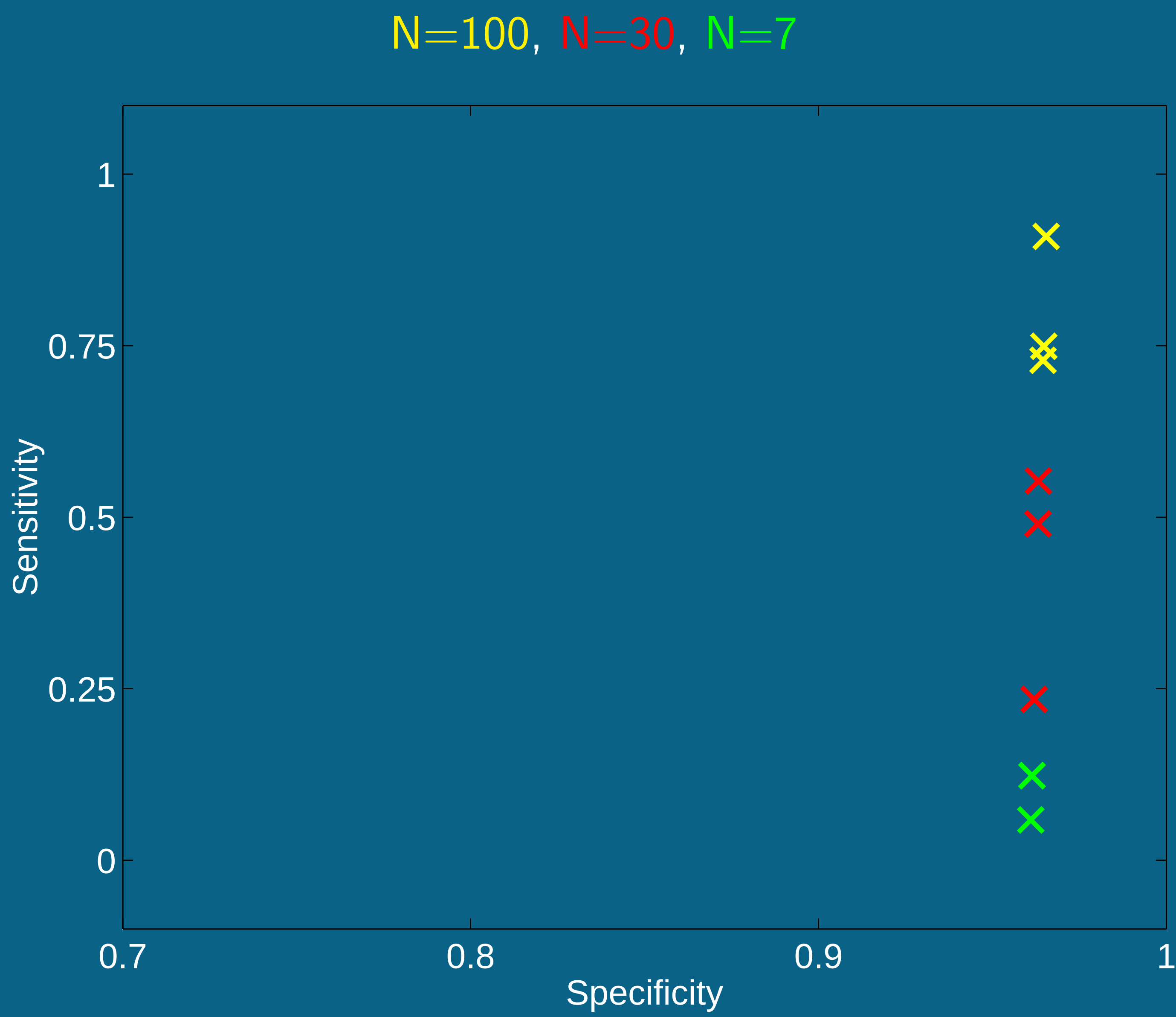

Static network
Sensitivity
Specificity
0
0.25
0.5
0.75
1
0.7
0.8
0.9
1

12 nodes

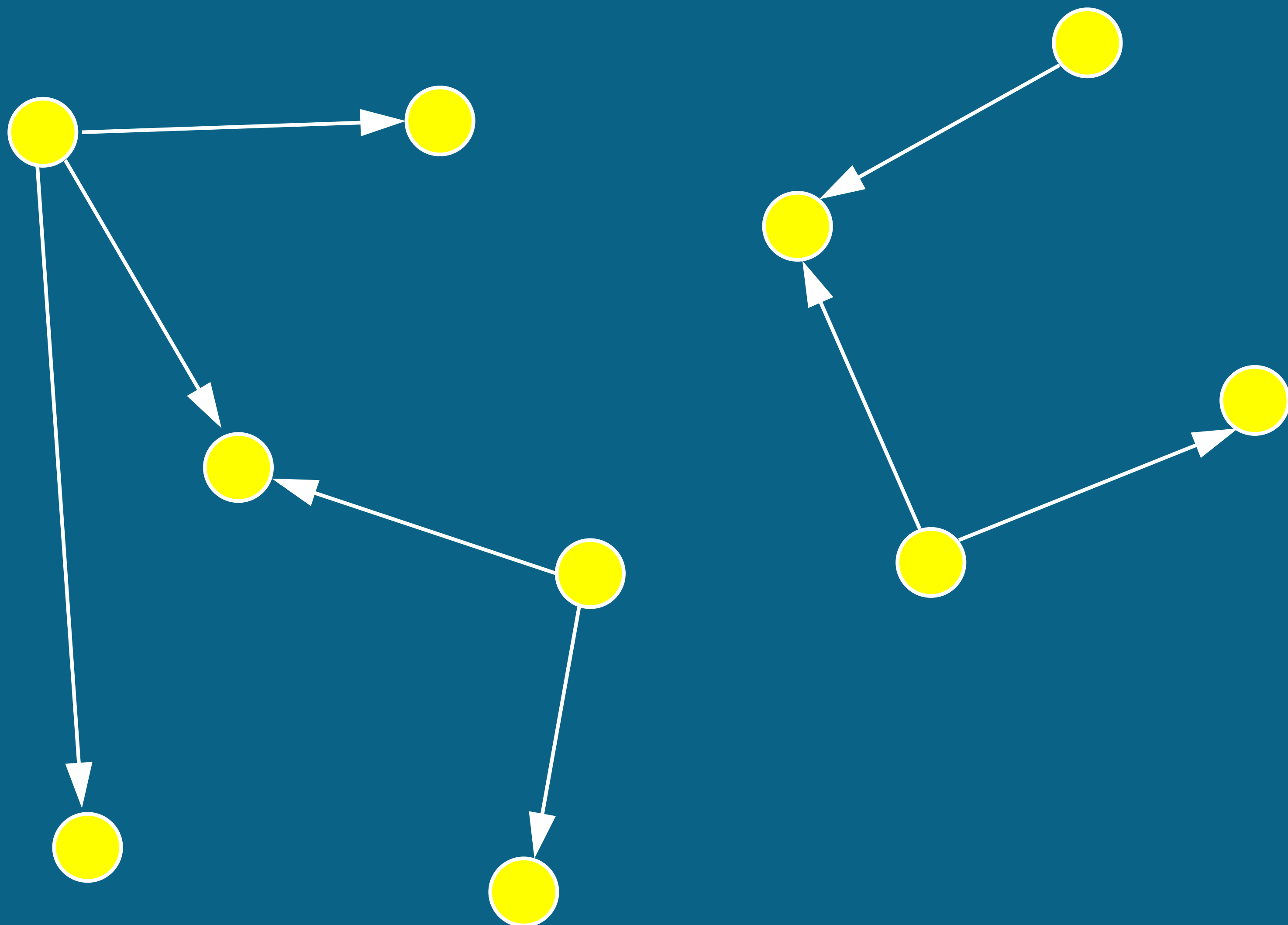


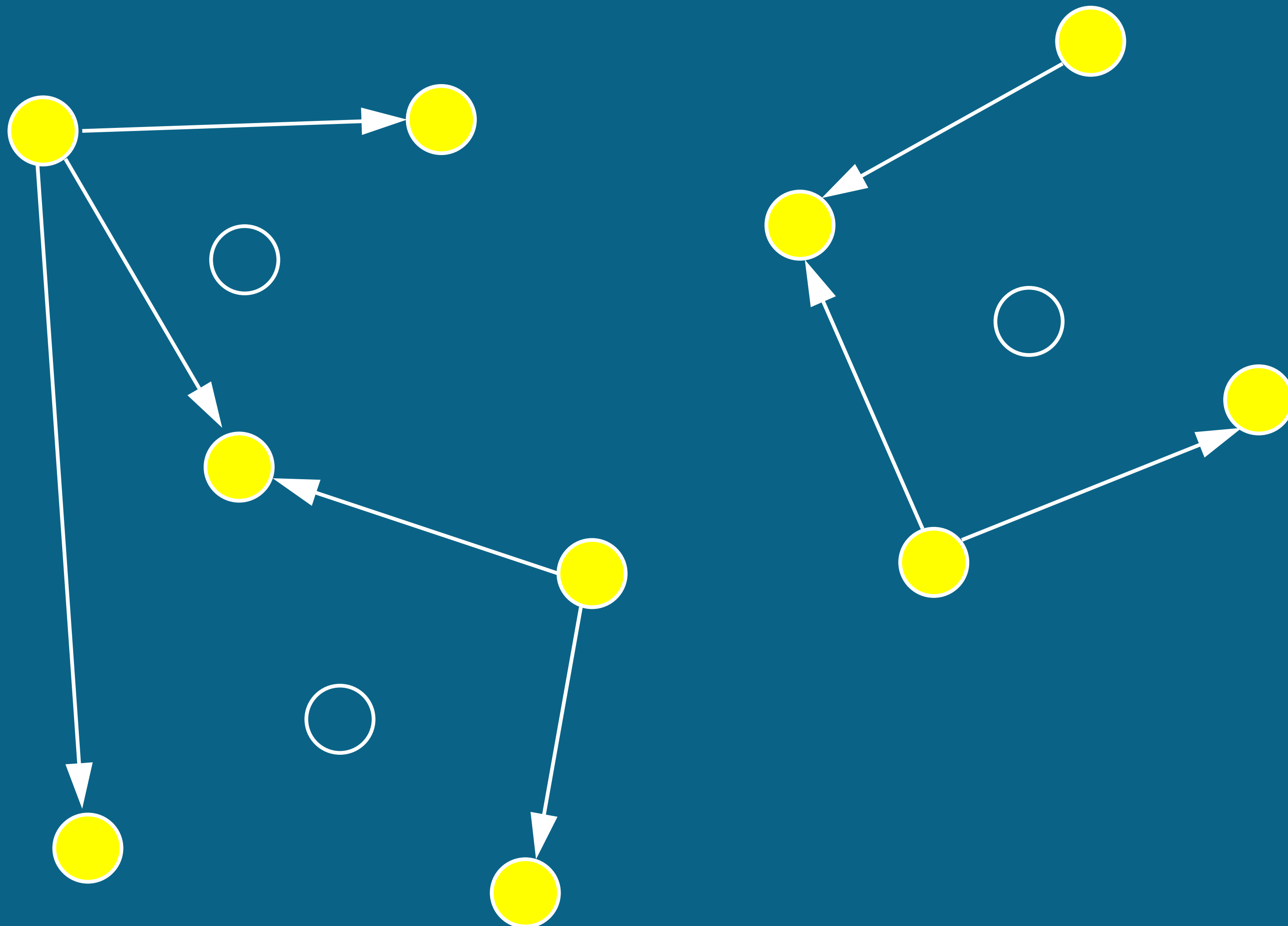
50 nodes

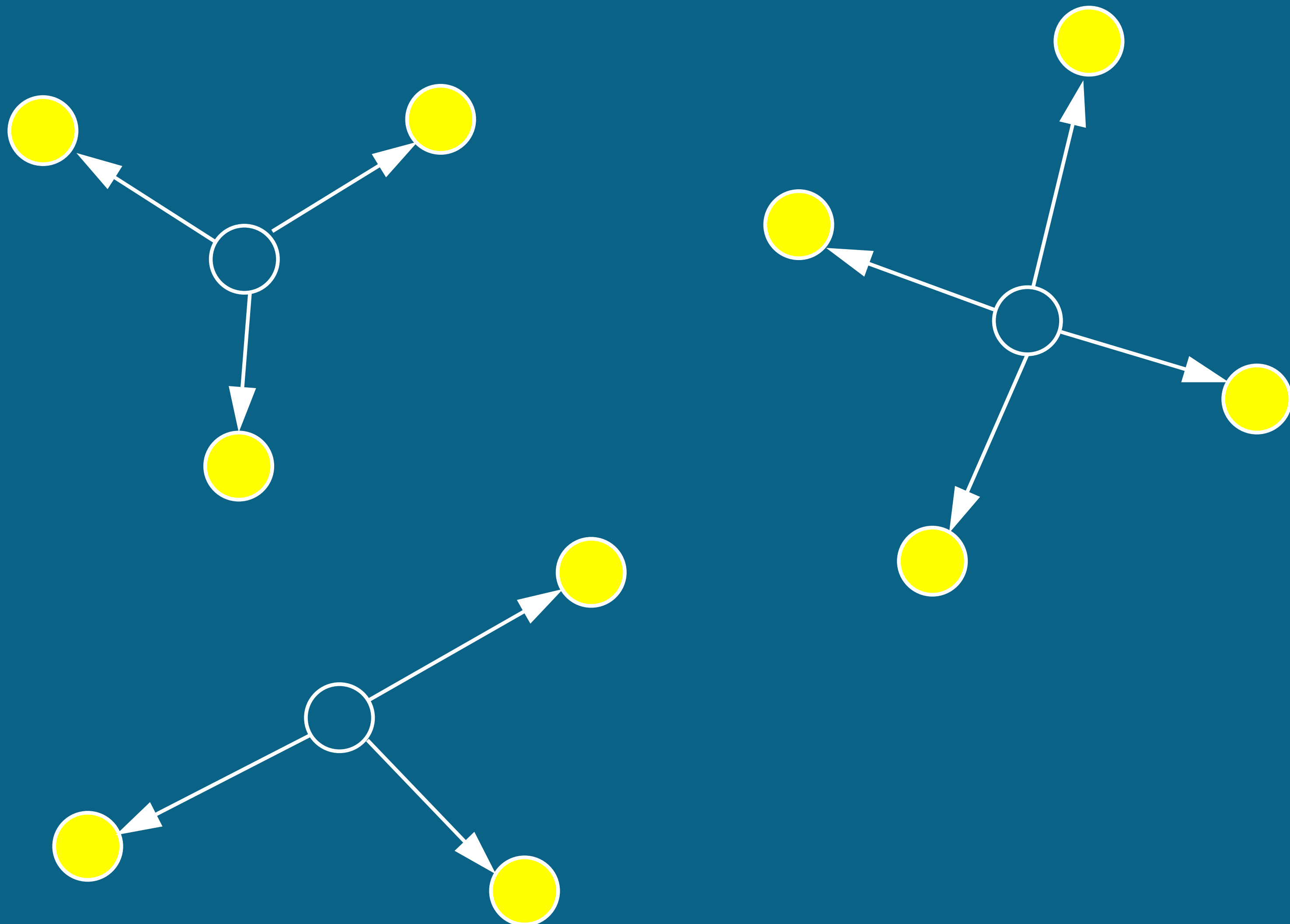



N=100, N=30, N=7
Sensitivity
Specificity
1
0.75
0.5
0.25
0
0.7
0.8
0.9
1

$N{=}7 \longrightarrow$ We need a different approach







Data

Ed Wagner, Peter Ghazal (GTI)

Viral gene expression data.

68 genes

7 time points

Data

Ed Wagner, Peter Ghazal (GTI)

Viral gene expression data.

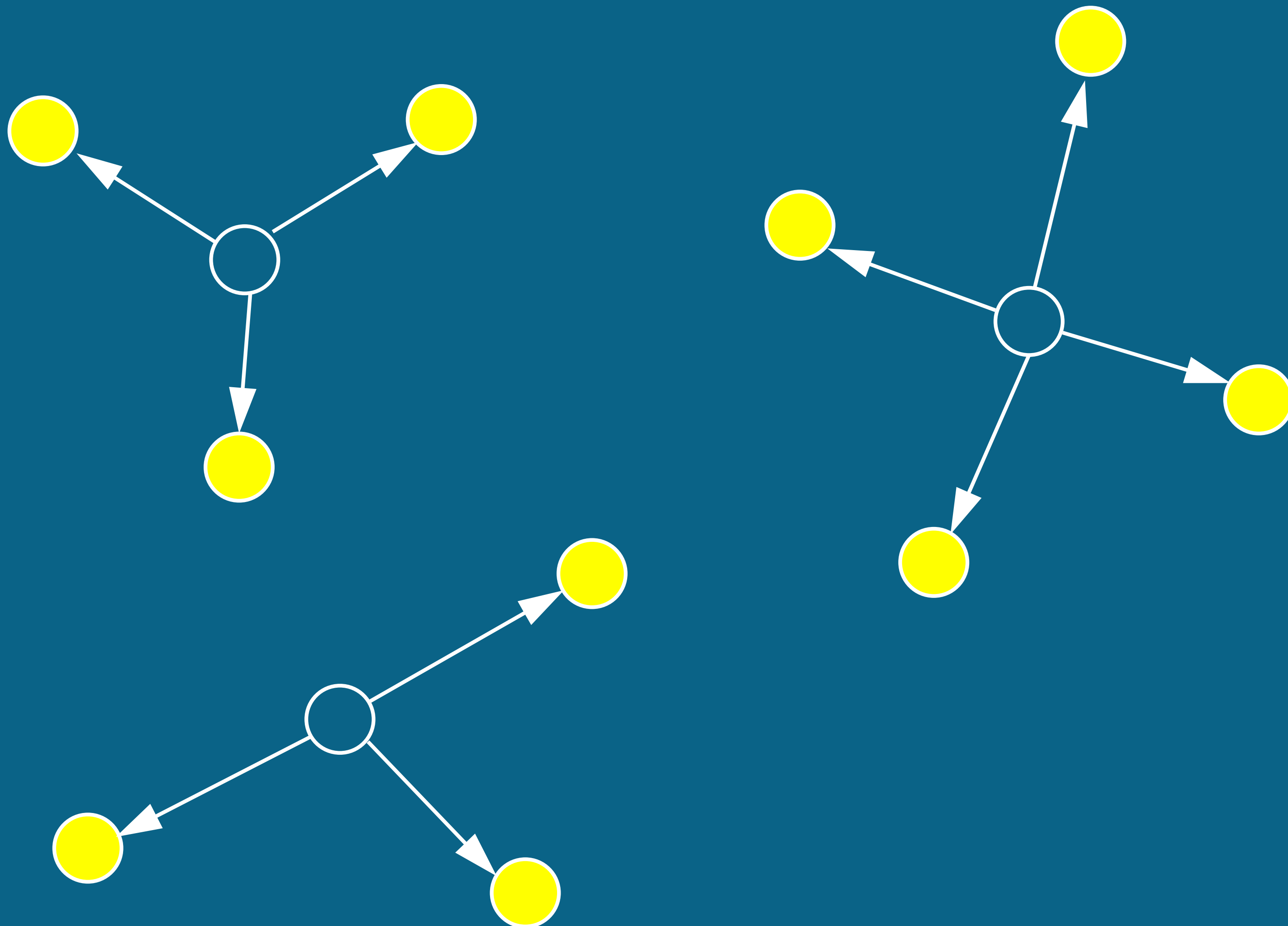
68 genes

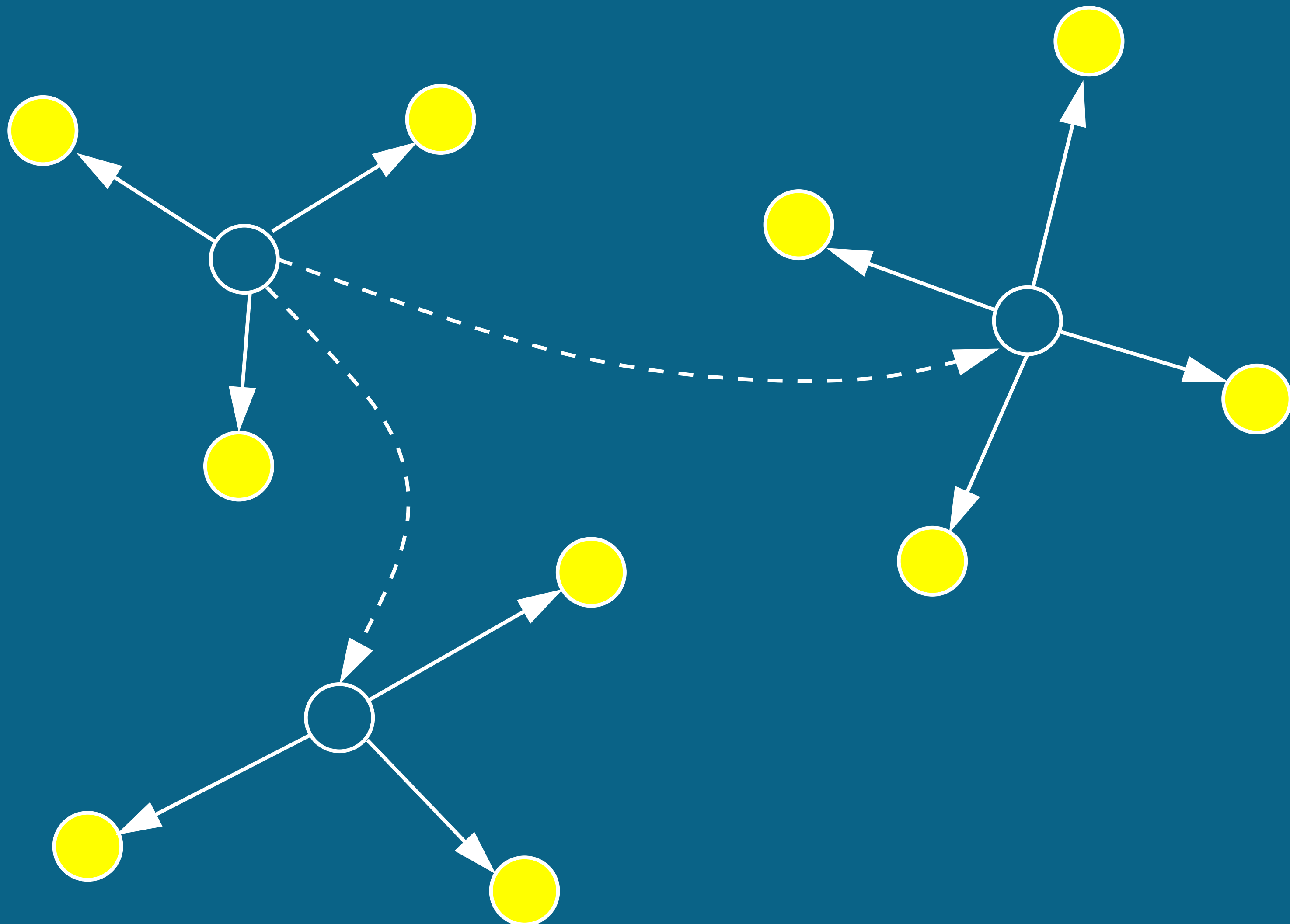
7 time points

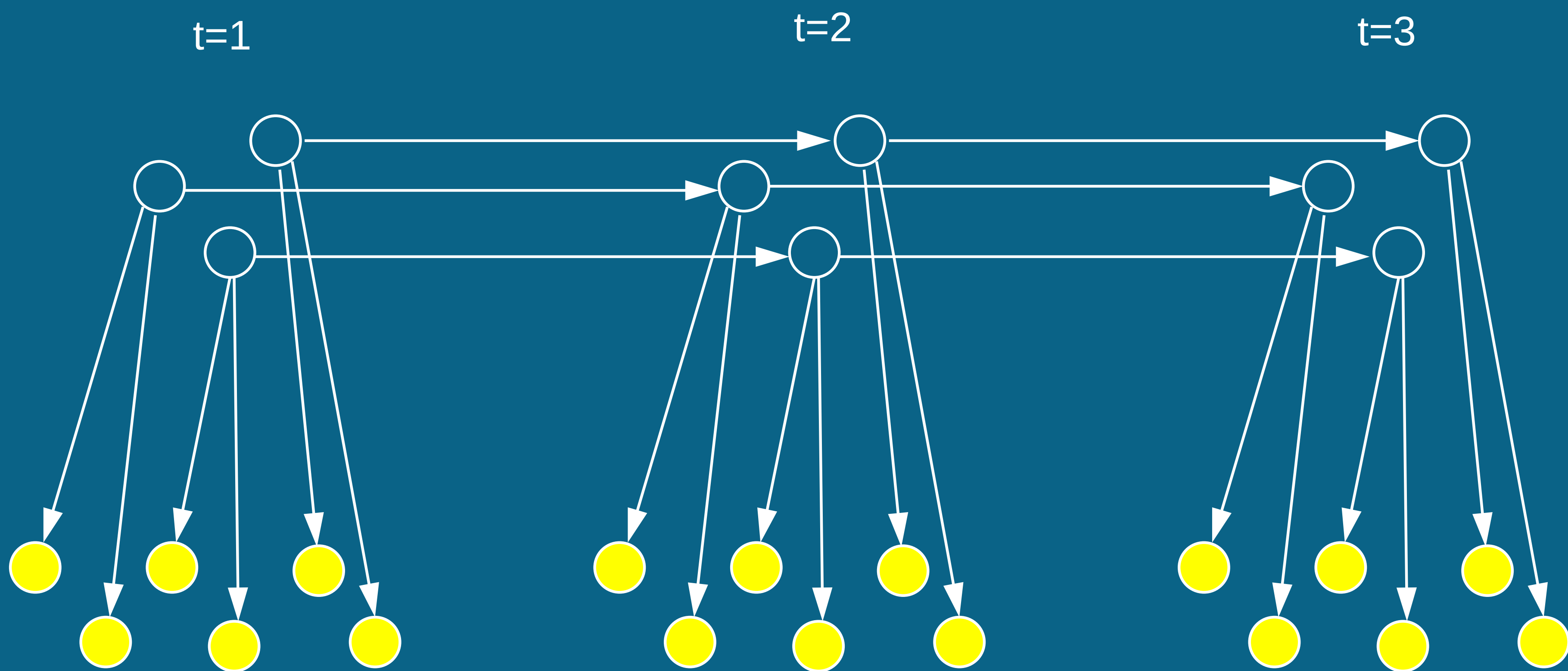
Prior knowledge

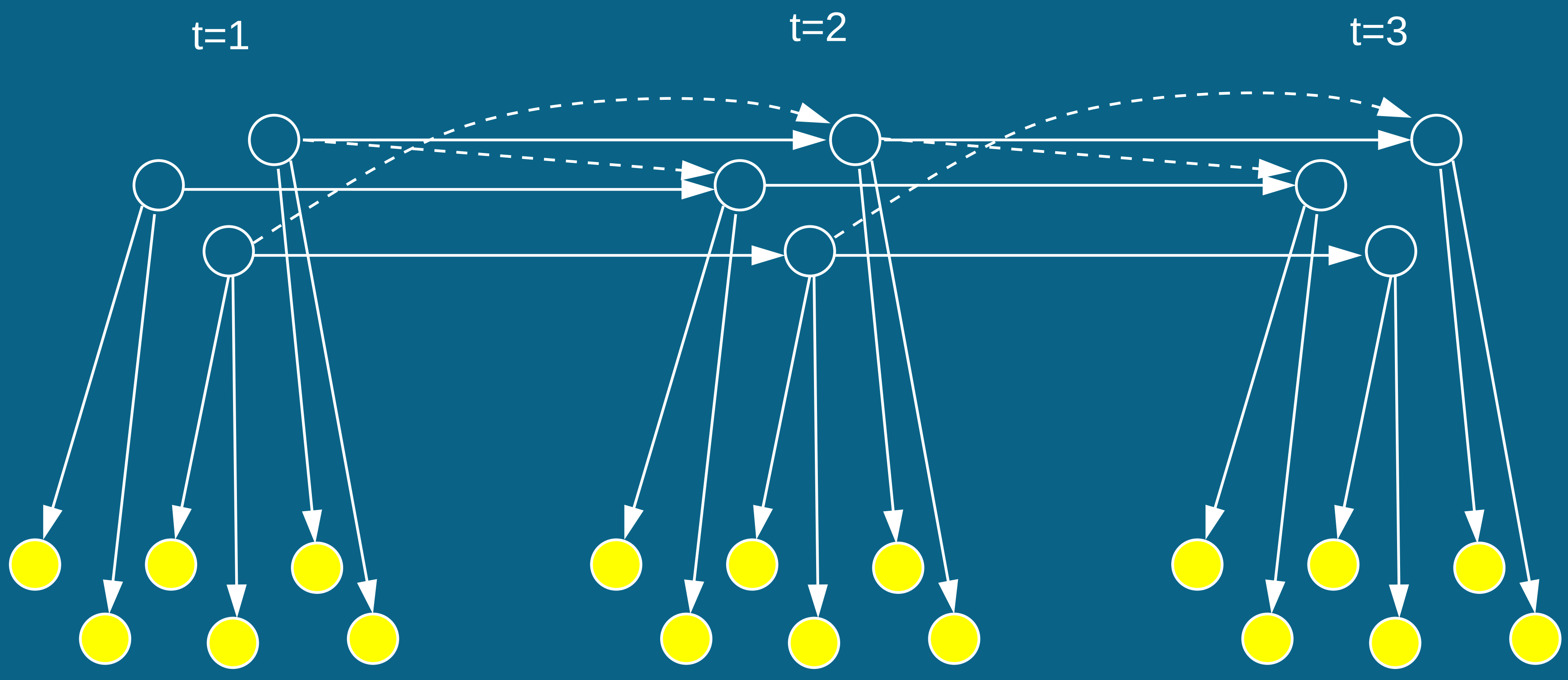
5 groups of co-transcribed genes

'Operons'






t=1
t=2
t=3


t=1
t=2
t=3

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

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

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

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.

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

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

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

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.

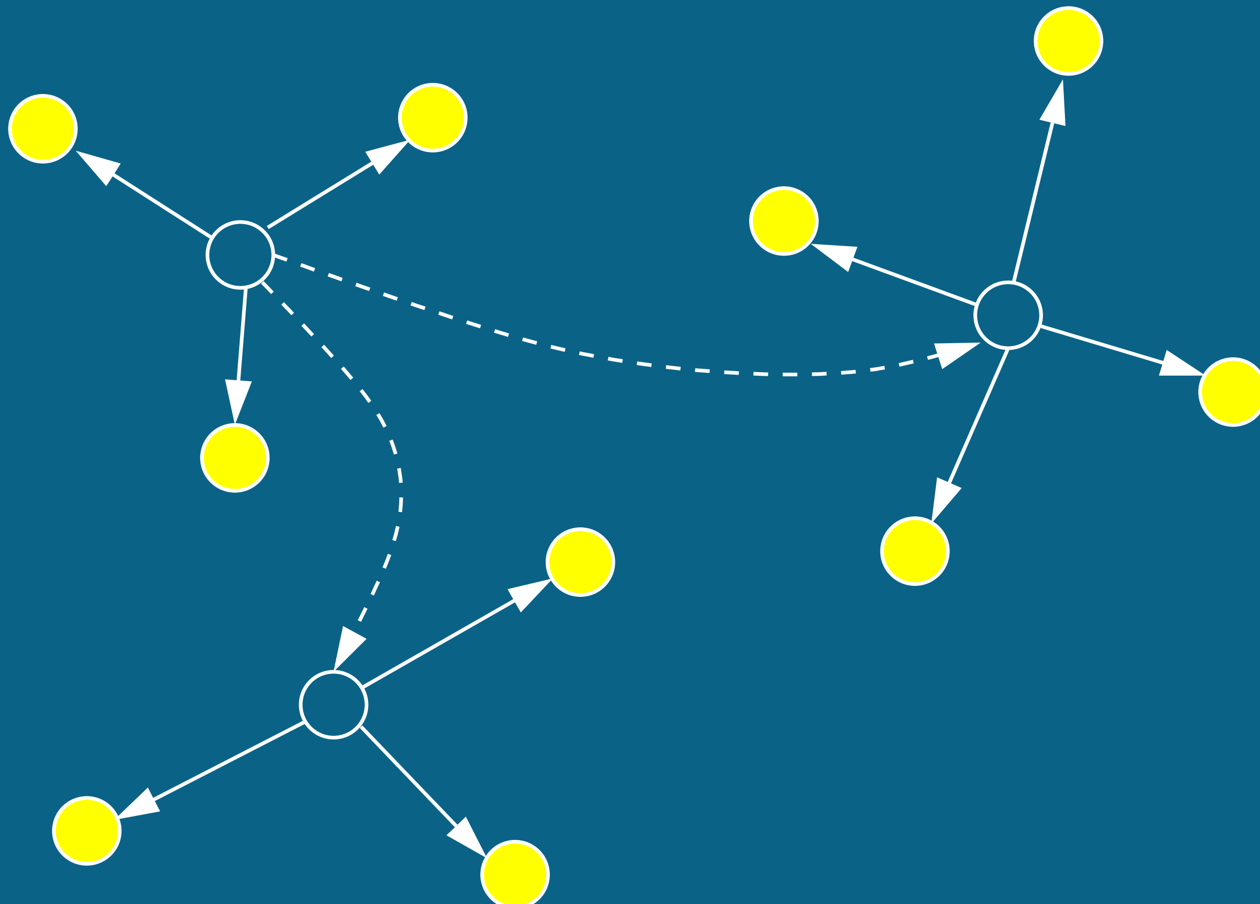
BIC Approximation (Schwarz 1978)

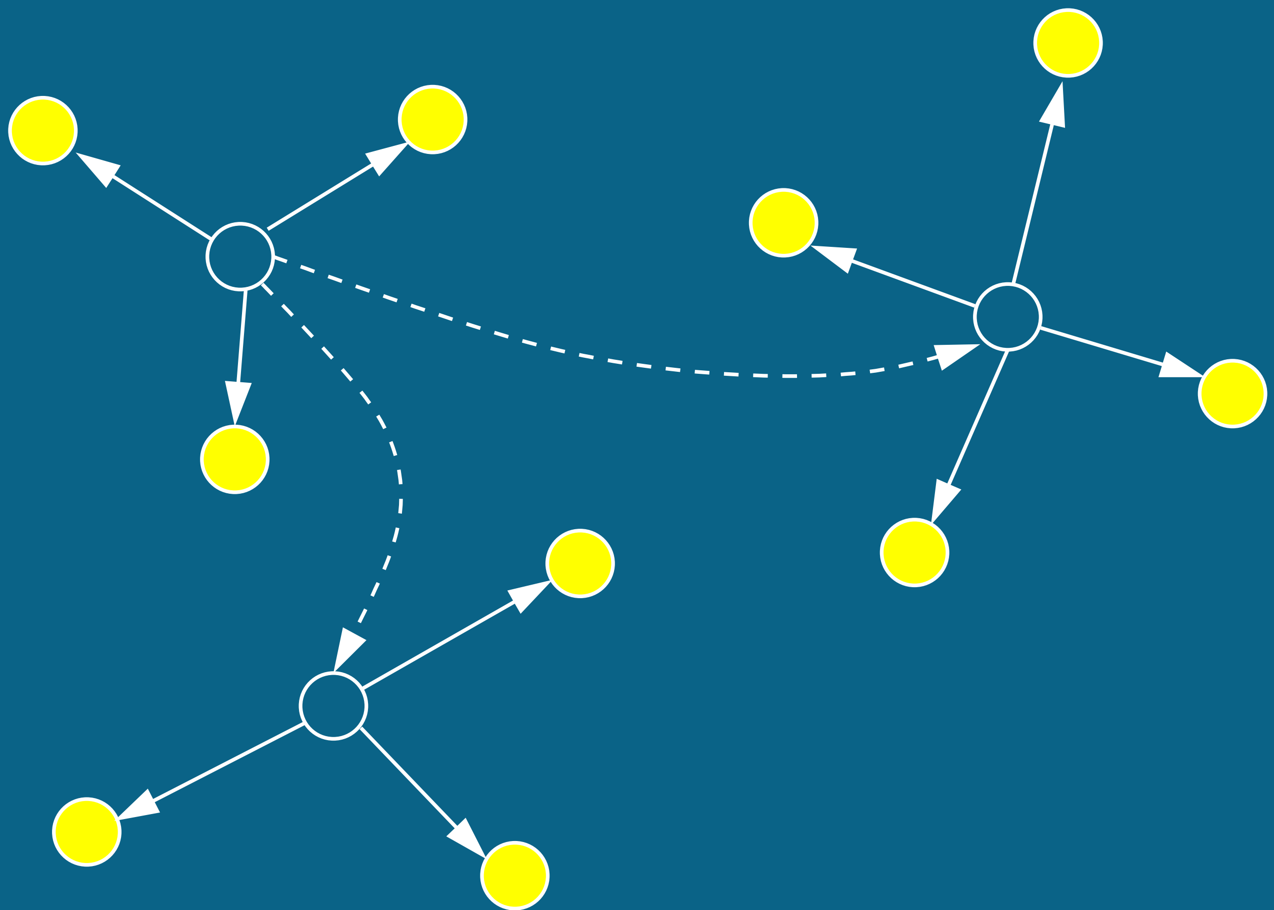
$$BIC = 2\hat{L}_M \; - \; \nu \ln N$$

$$\begin{aligned}
P(D|M) &= \int P(D|\mathbf{q},M)P(\mathbf{q}|M)d\mathbf{q} \;\approx\; P(D|\hat{\mathbf{q}},M)P(\hat{\mathbf{q}}|M)\sqrt{\frac{(2\pi)^\nu}{\det \mathbf{H}}} \\
\mathbf{H} &= -\nabla_{\mathbf{q}}\nabla_{\mathbf{q}}^\dagger \left[\ln P(D|\mathbf{q},M) + \ln P(\mathbf{q}|M) \right]_{\mathbf{q}=\hat{\mathbf{q}}} \\
\ln P(D|M) &= \ln P(D|\hat{\mathbf{q}},M) + \ln P(\hat{\mathbf{q}}|M) - \frac{1}{2}\ln\det \mathbf{H} + \frac{\nu}{2}\ln(2\pi) \\
&\approx \ln P(D|\hat{\mathbf{q}},M) - \frac{1}{2}\ln\det \mathbf{H} + \frac{\nu}{2}\ln(2\pi) \\
&= \hat{L}_M - \frac{1}{2}\sum_{i=1}^{\nu} \ln\frac{\varepsilon_i}{2\pi}, \qquad \varepsilon_i \approx 2\pi N \;\; \forall i \\
&\approx \hat{L}_M - \frac{\nu}{2}\ln N
\end{aligned}$$

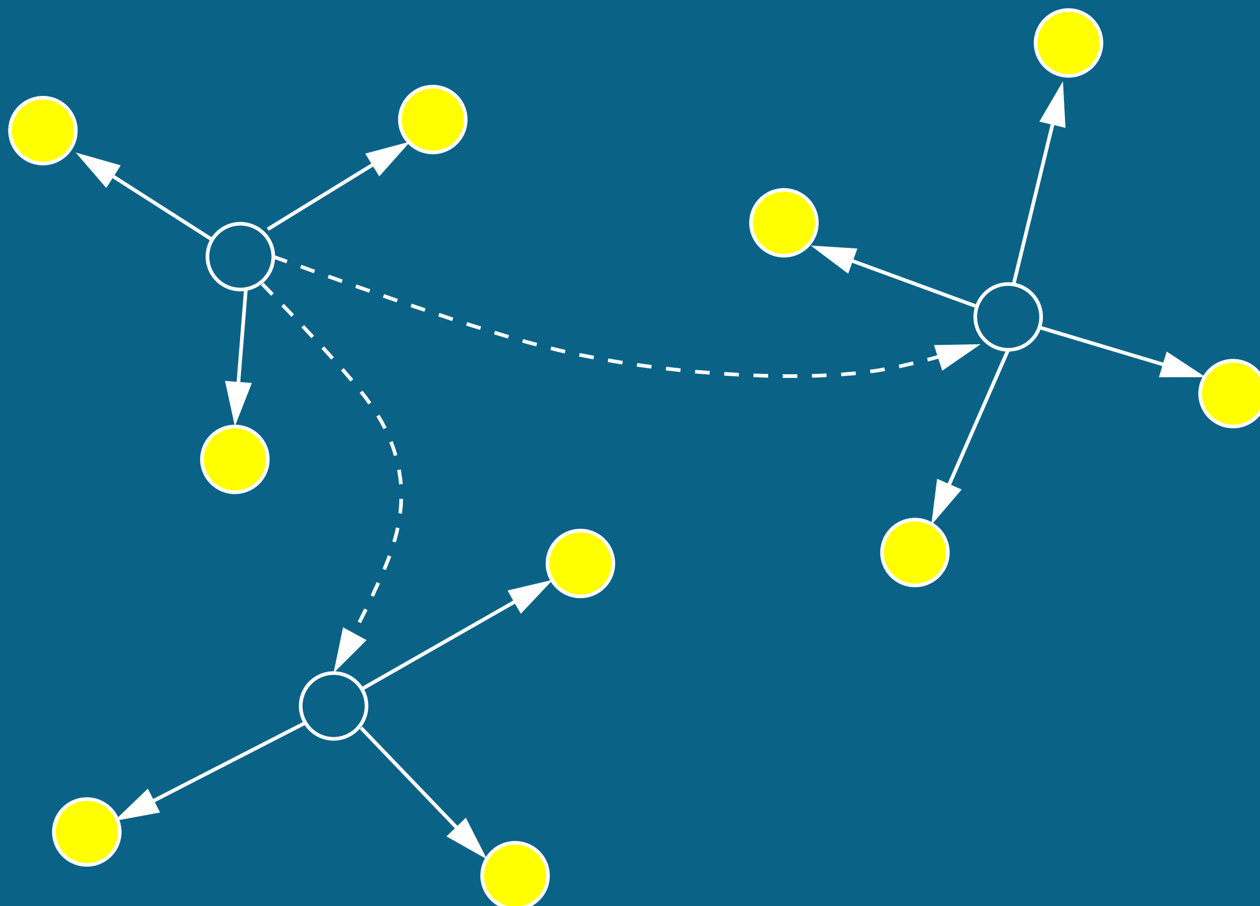
$$BIC = 2\hat{L}_M \; - \; \nu \ln N$$

Maximum likelihood optimisation with the EM algorithm.



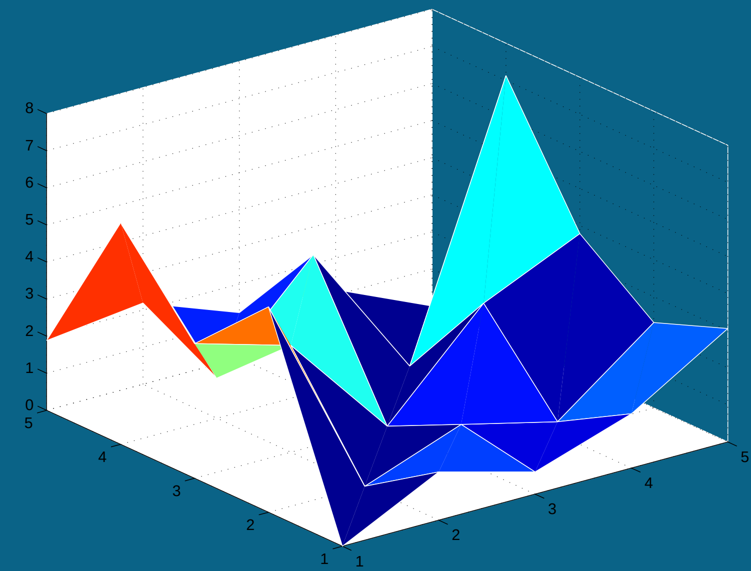


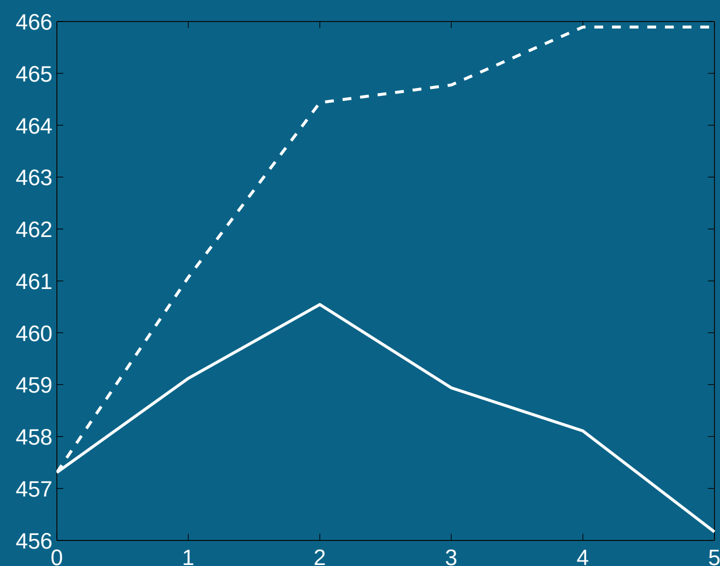
Exhaustive search impossible: ca $30,000$ configurations (for $H = 5$)

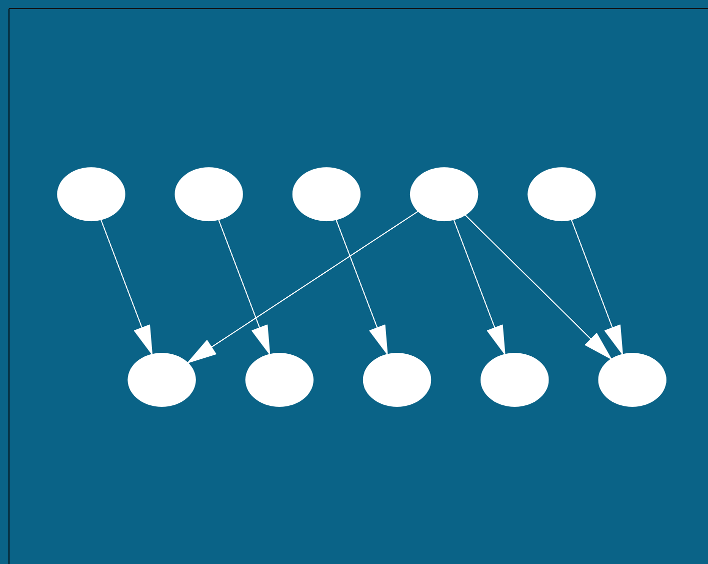


Exhaustive search impossible: ca $30,000$ configurations (for $H = 5$)

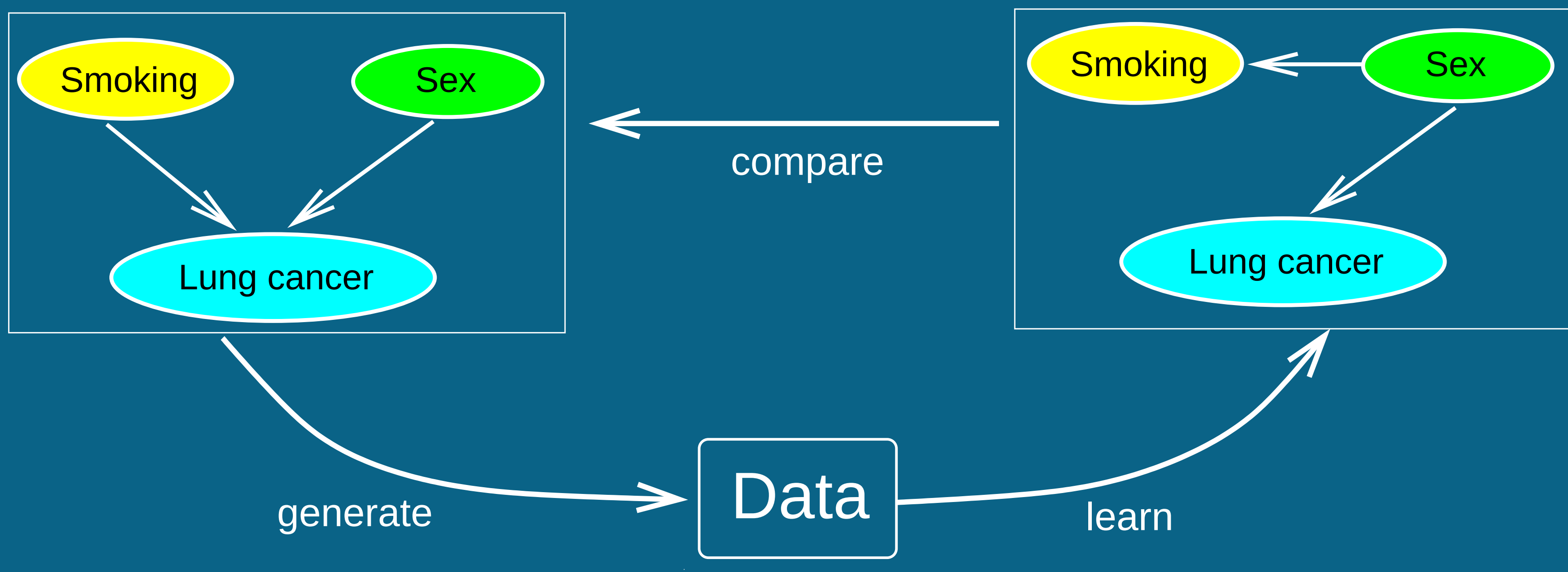
Heuristic method: Individual edge scoring: $5 \times 5 = 25$ configurations

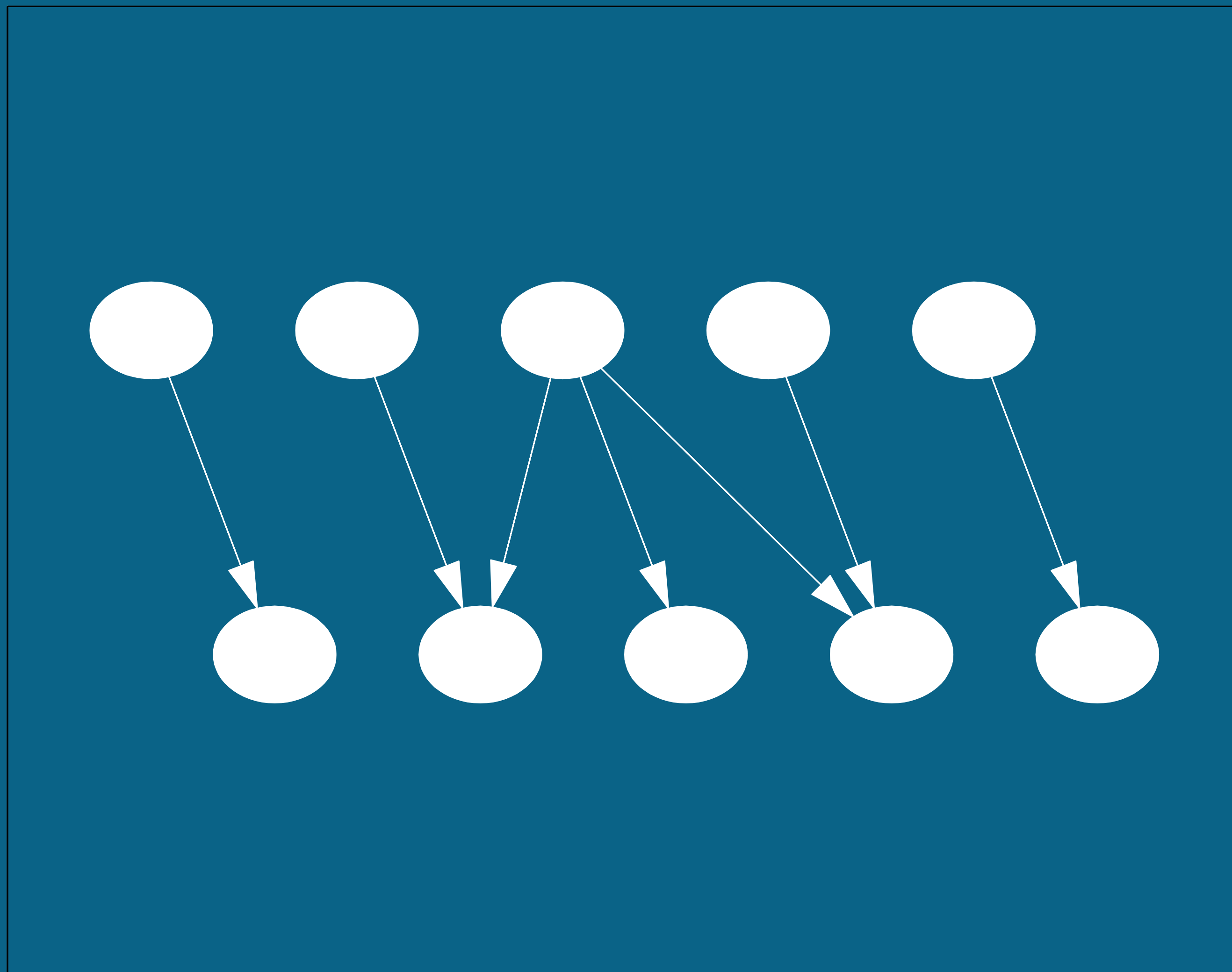

8
7
6
5
4
3
2
1
0
5
4
3
2
1
1
2
3
4
5


466
465
464
463
462
461
460
459
458
457
456
0
1
2
3
4
5

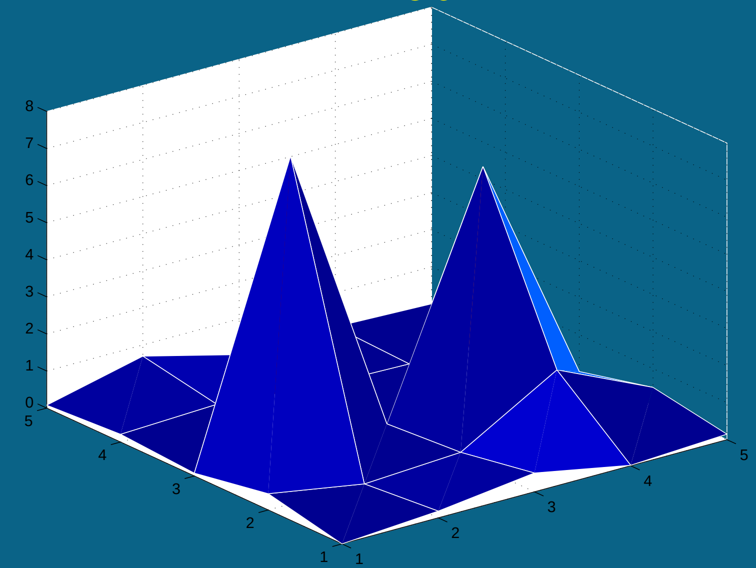


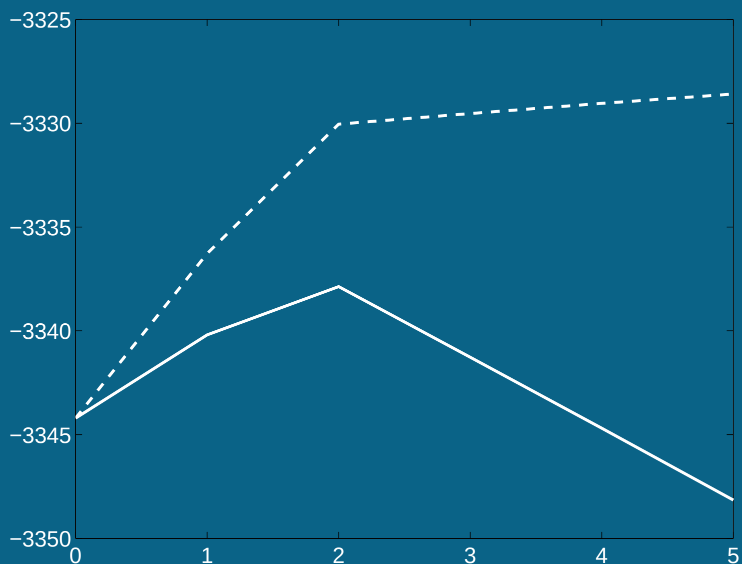
Can we trust these results ?

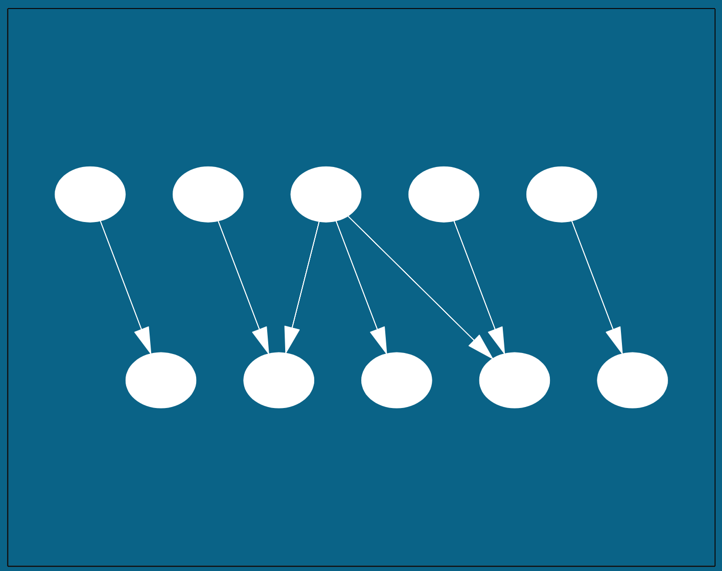

Smoking
Sex
Lung cancer
compare
Smoking
Sex
Lung cancer
generate
Data
learn

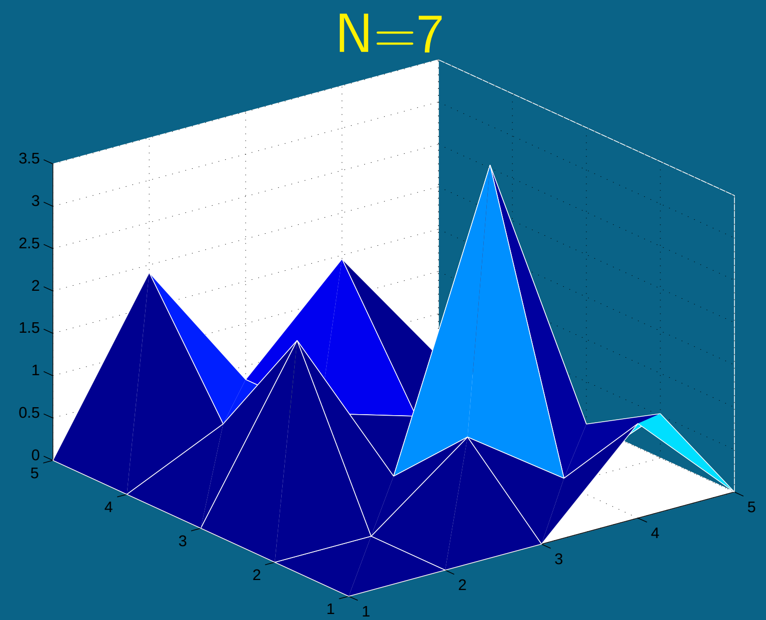


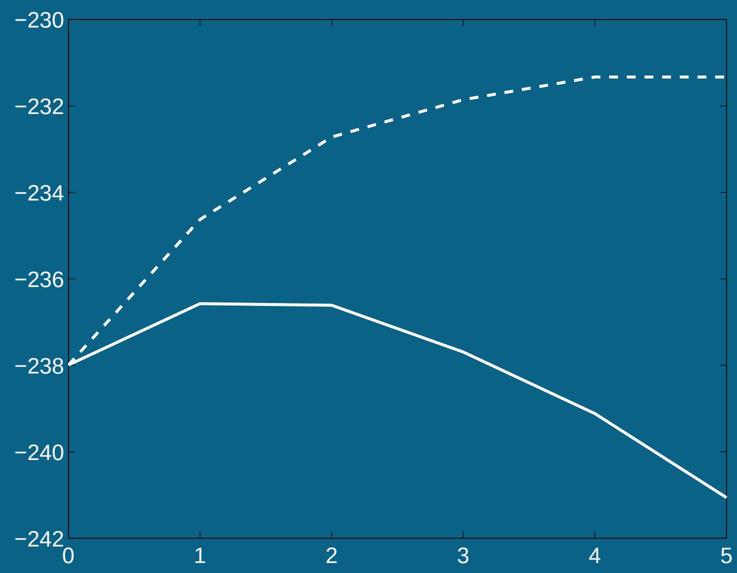
N=50

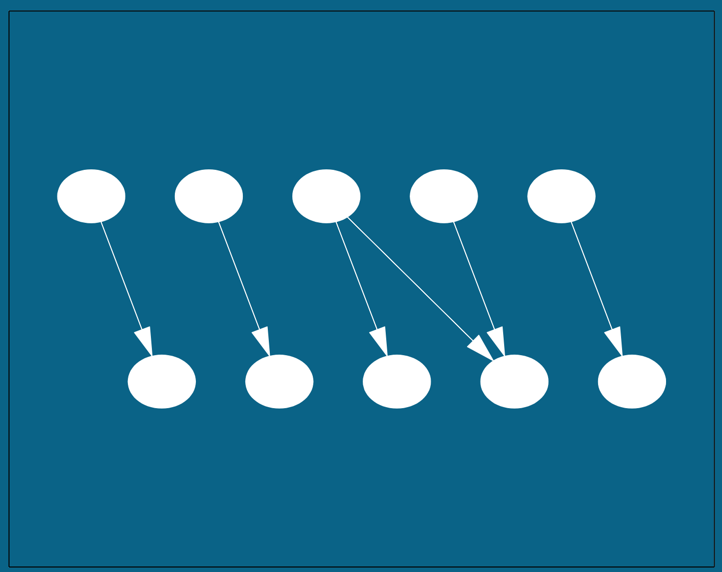







N=7
3.5
3
2.5
2
1.5
1
0.5
0
5
4
3
2
1
1
2
3
4
5


−230
−232
−234
−236
−238
−240
−242
0
1
2
3
4
5

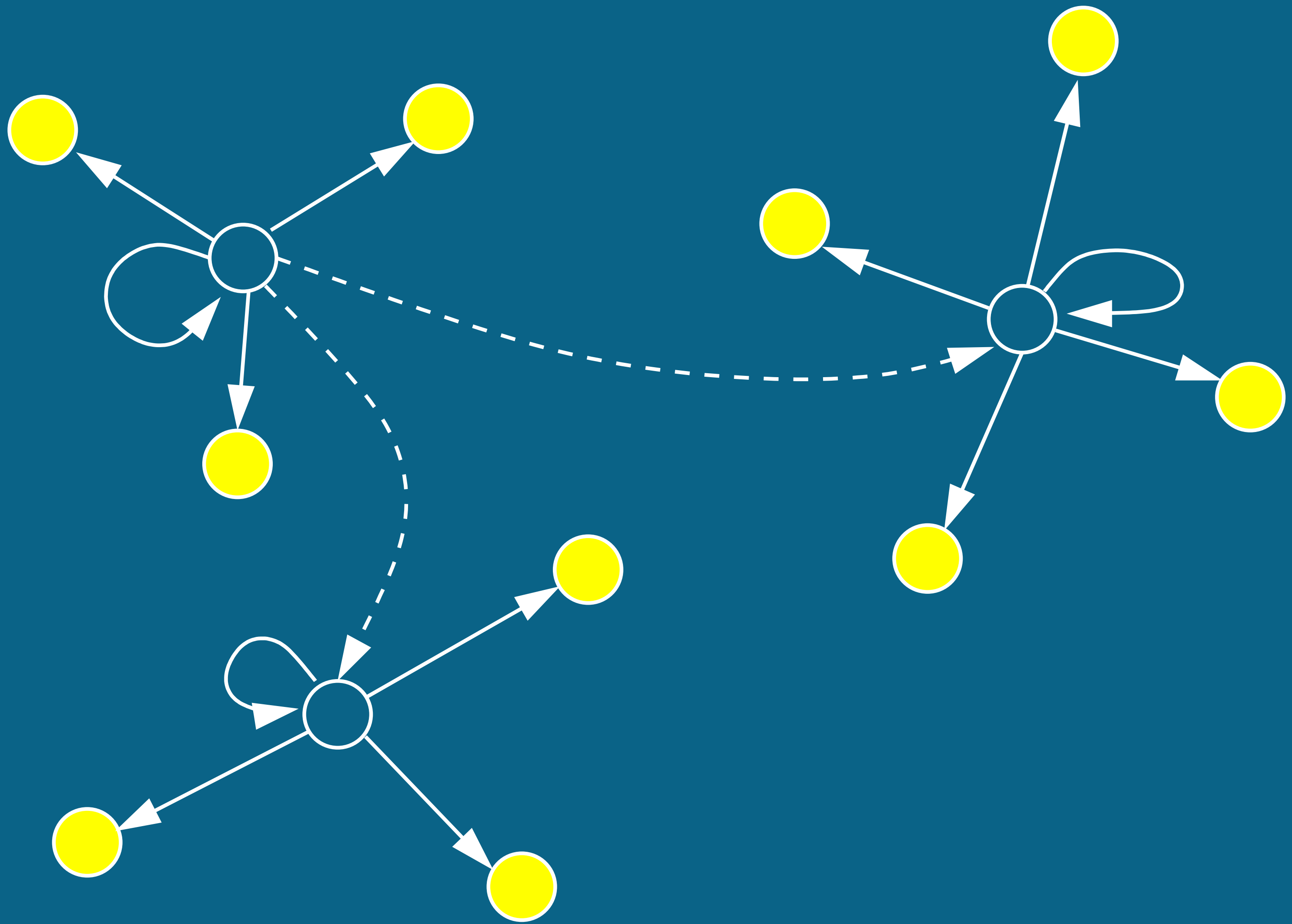


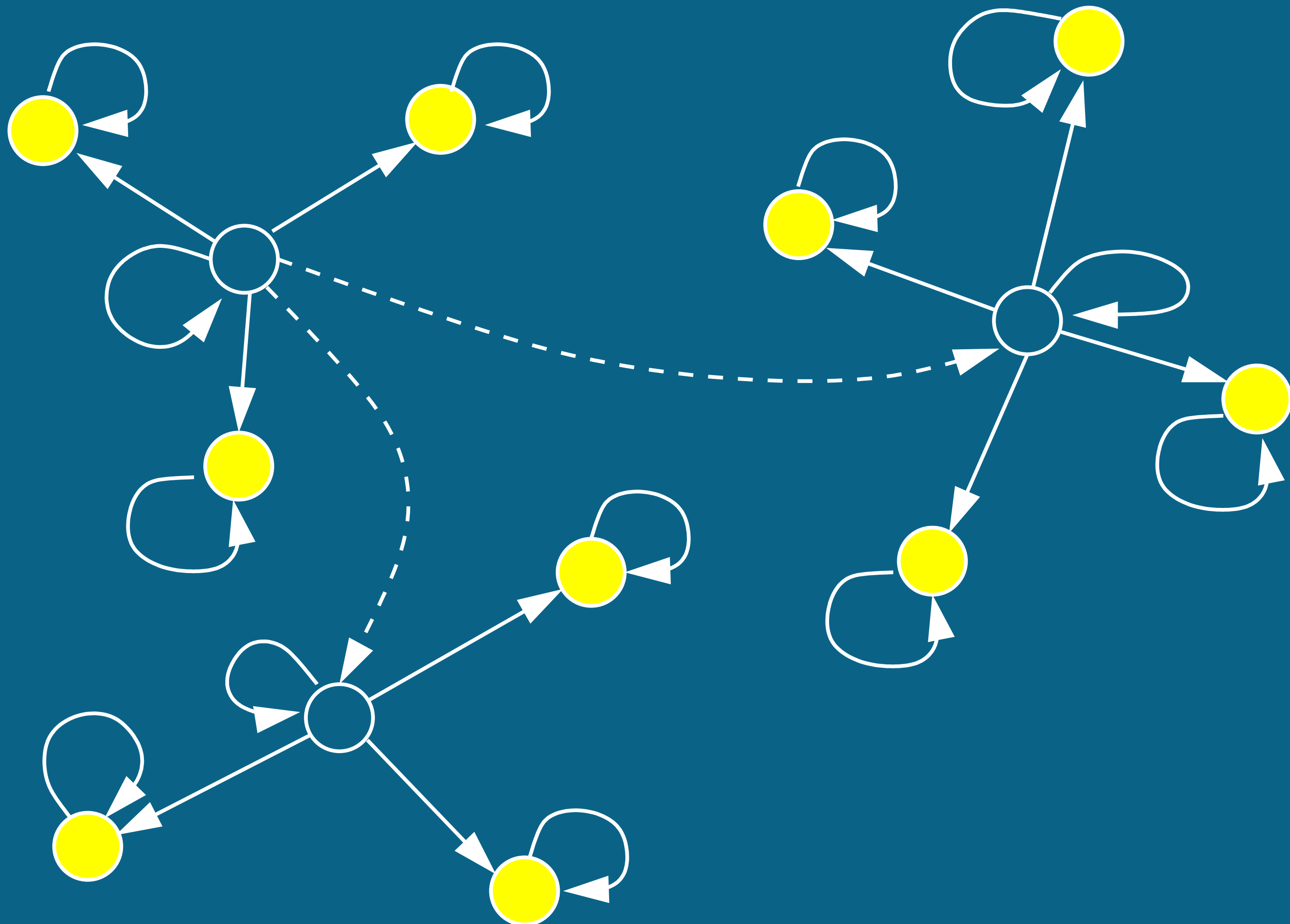
Conclusion

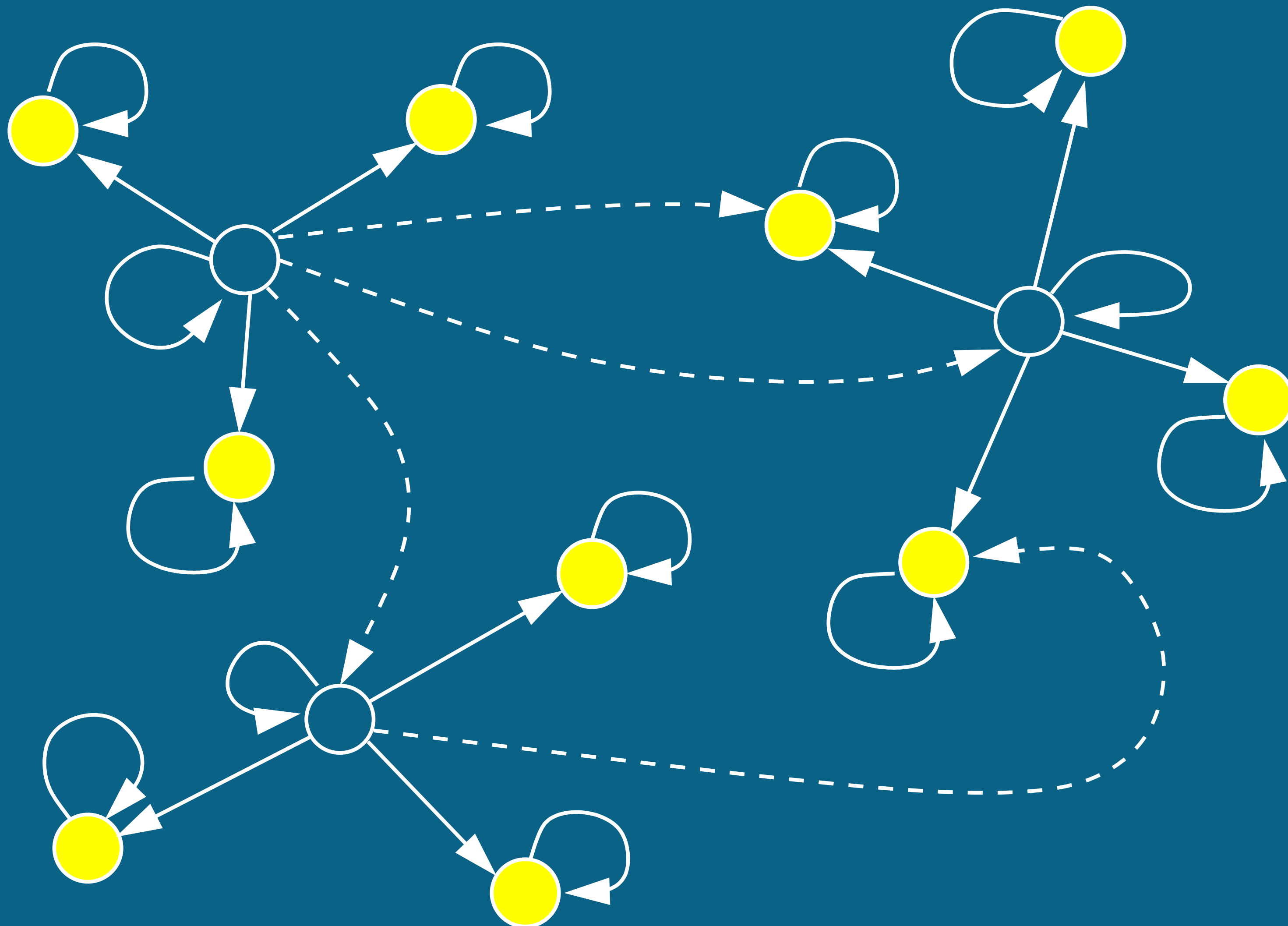
Conclusion

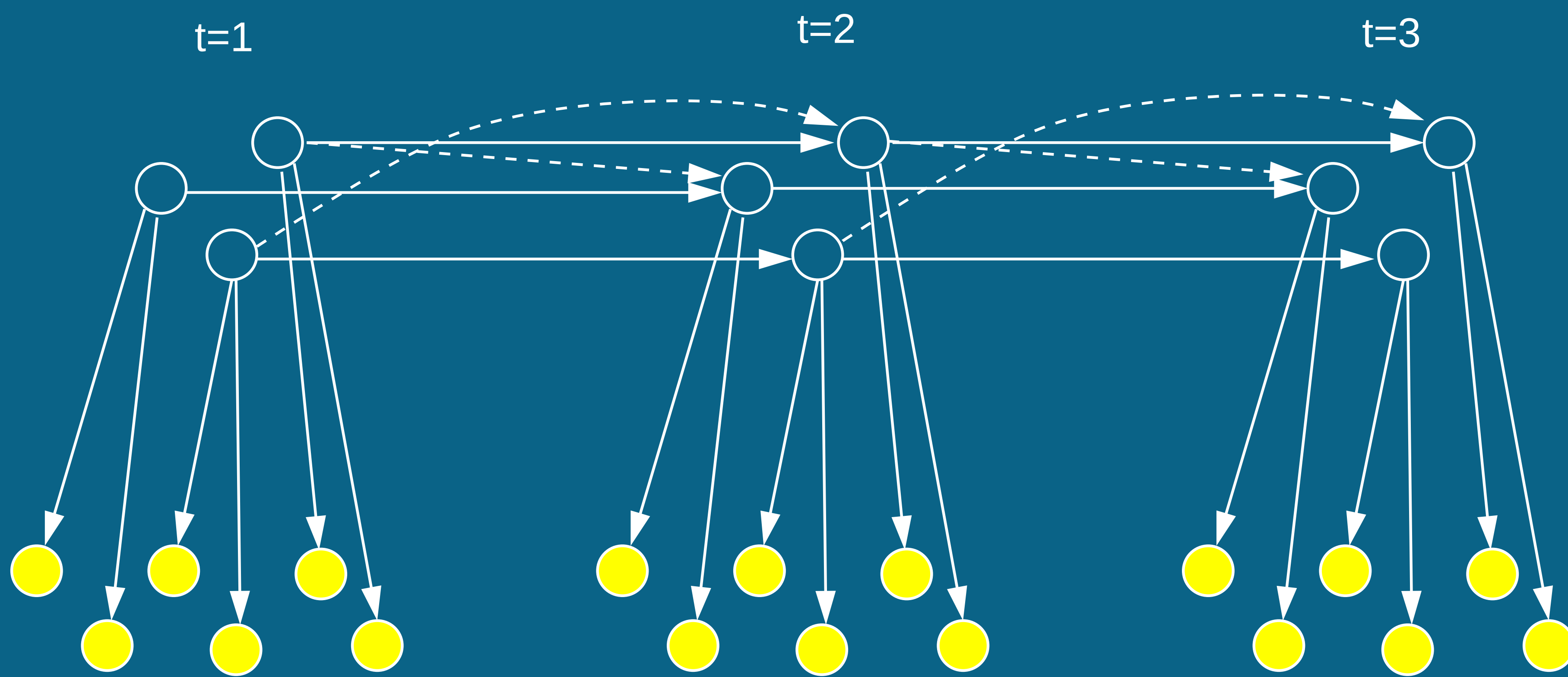
More data !

Future work

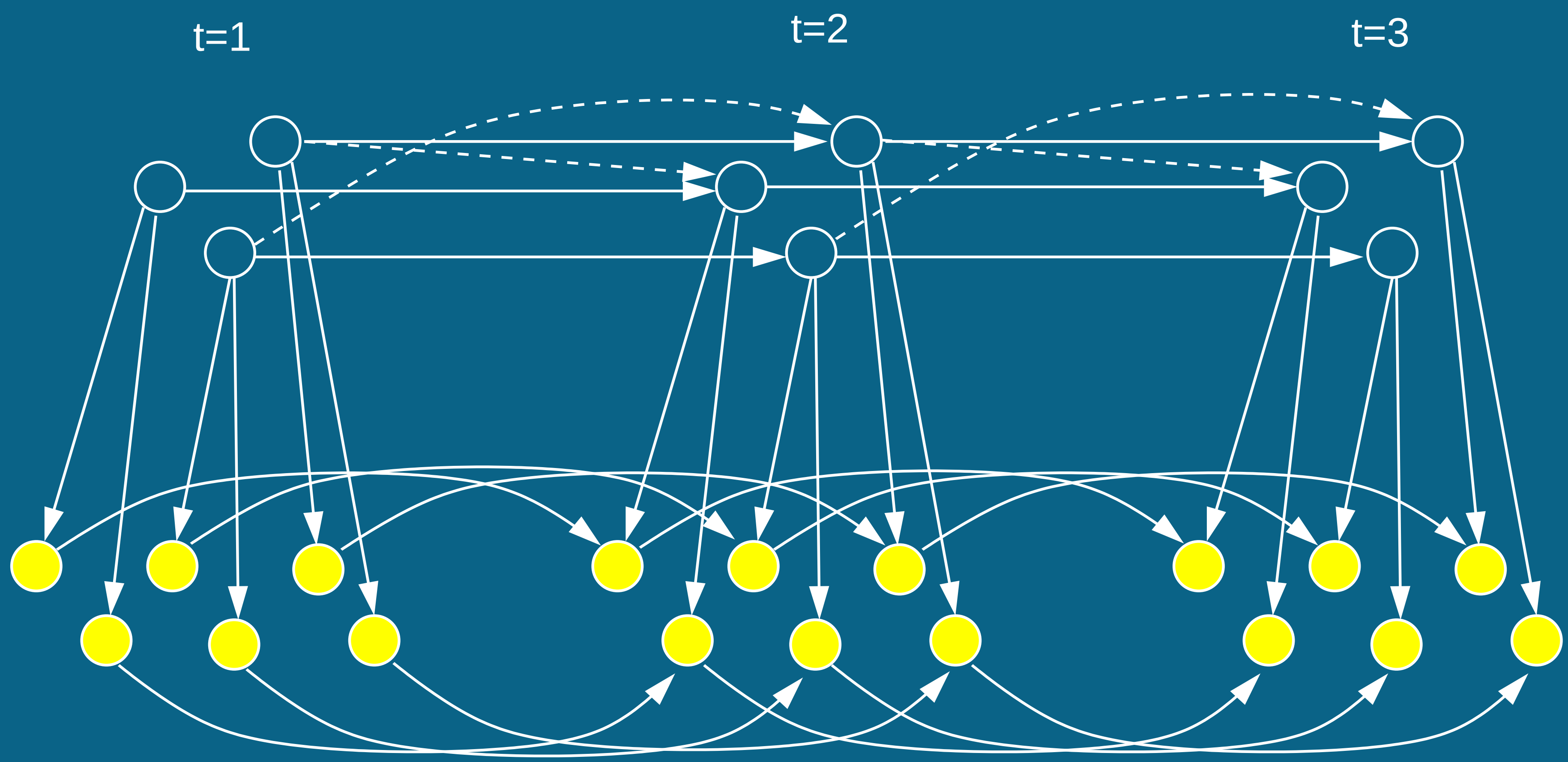


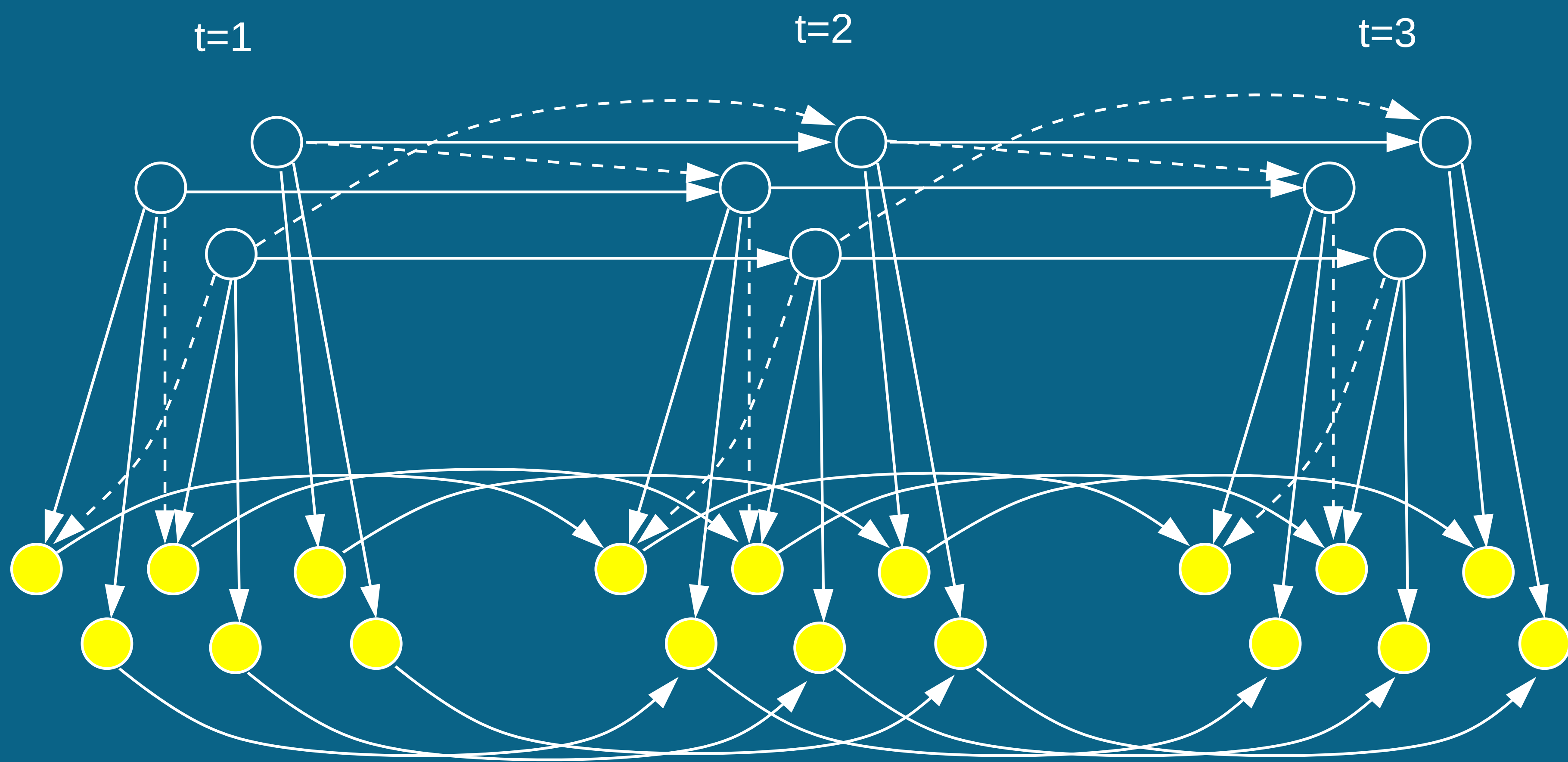






t=1
t=2
t=3


t=1
t=2
t=3



t=1
t=2
t=3

Acknowledgements

Acknowledgements

Chris Glasbey

Peter Ghazal

Acknowledgements

Chris Glasbey

Peter Ghazal

Chris Williams

Felix Agakov