

Inferring causality from a mixture of observations and interventions

GT Eclairé

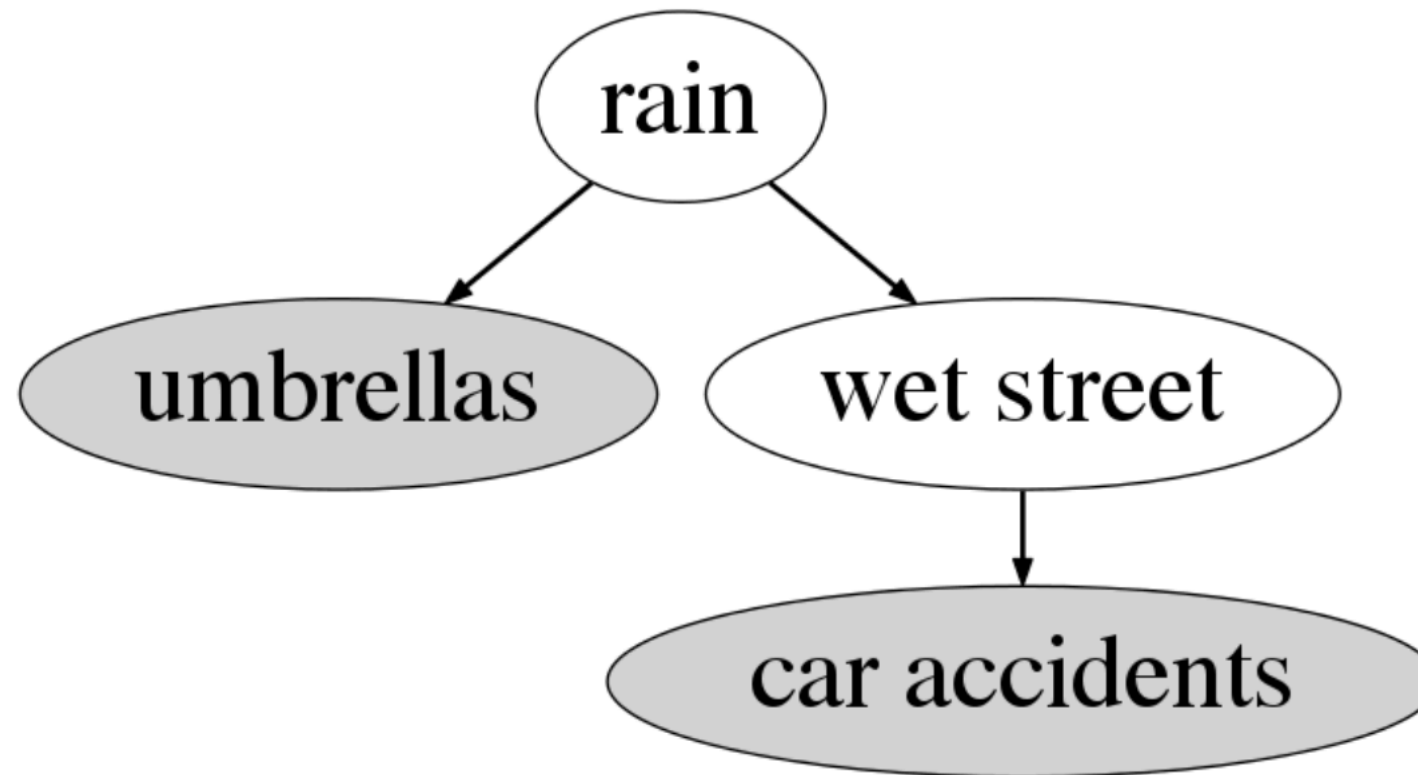
Sorbonne Université, Paris, France

24 mars 2026

G. Nuel



Correlation is not Causation

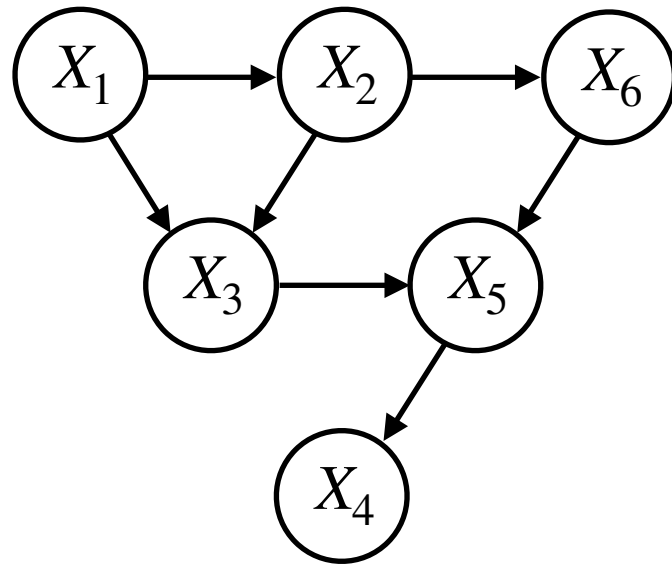


umbrellas and car accidents are correlated

But:

- provoking car accidents does not make appear umbrellas
- distributing umbrellas in the street does not provoke car accidents

Directed Acyclic Graphs



Definition (DAG):

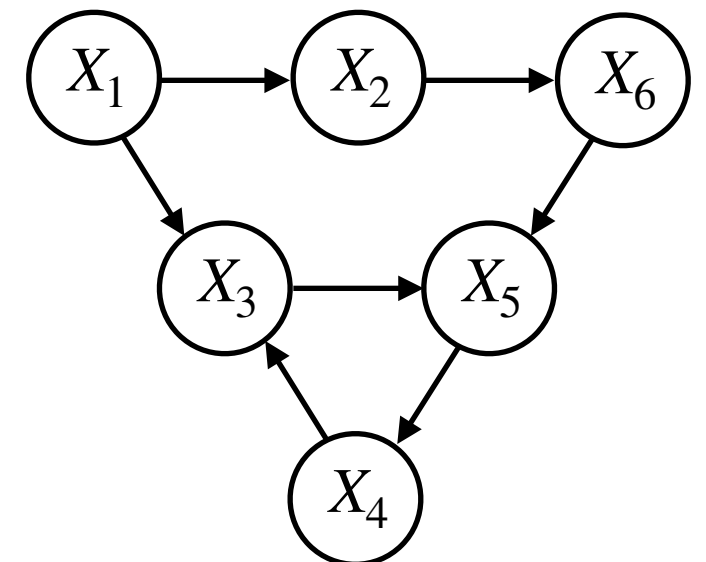
$G = (V = \{1, \dots, n\}, E \subset V \times V)$ is a *Directed Acyclic Graph* if and only if it has no cycle

Remark: loops are ok (e.g. $X_1 - X_2 - X_3$)

Theorem (topological ordering):

$G = (V = \{1, \dots, n\}, E \subset V \times V)$ is a Directed Acyclic Graph if and only if it exists a *topological ordering* $\sigma_1, \dots, \sigma_n$ such that $\forall i, j \in V$ such that $(i, j) \in E$ we have $\sigma_i < \sigma_j$

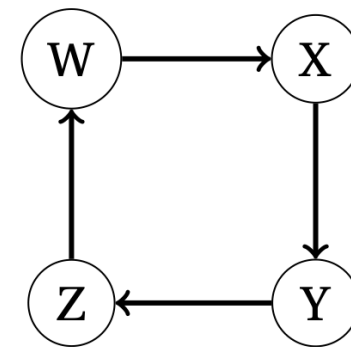
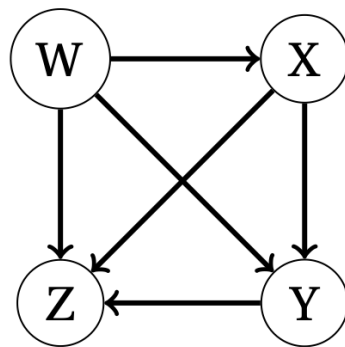
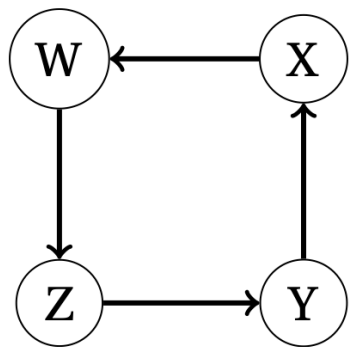
Example: $\sigma = (1, 2, 3, 6, 5, 4)$ is a topological ordering



not a DAG

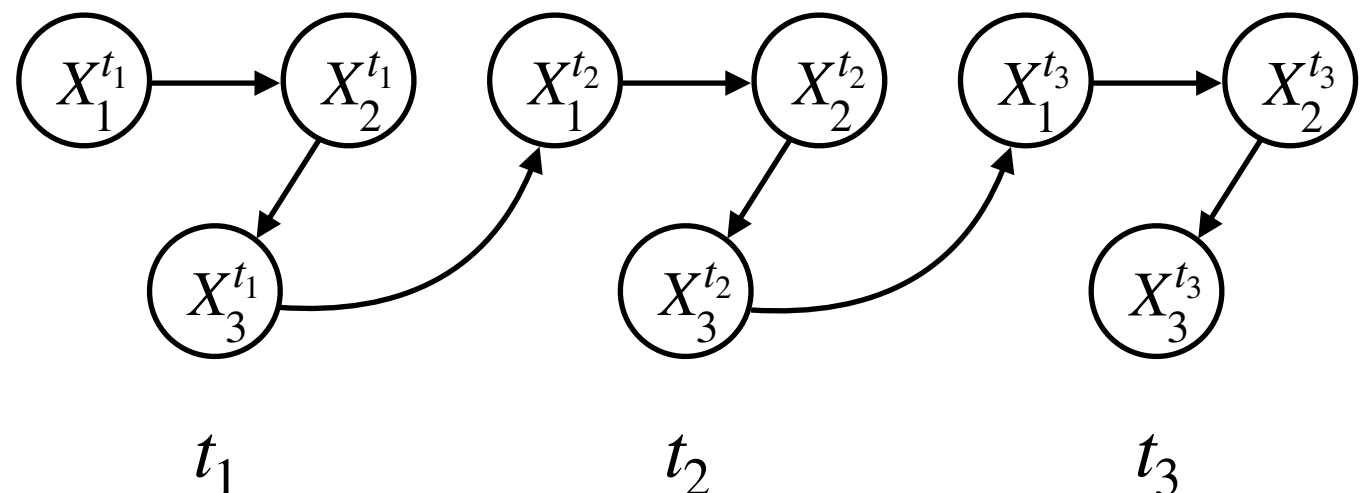
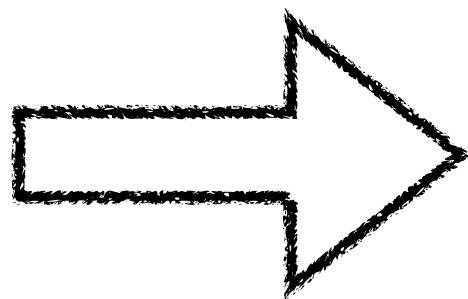
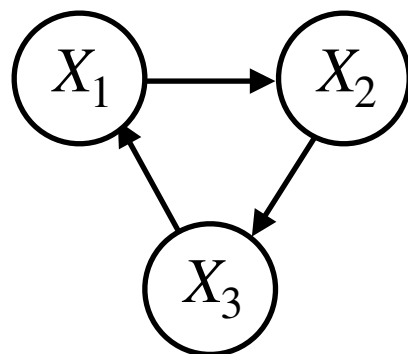
Directed Acyclic Graphs

Why not cyclic directed graph ?

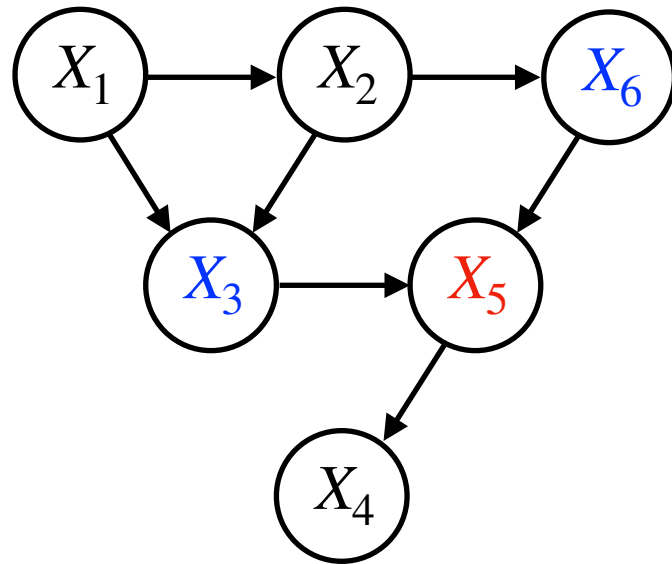


Three models with the same distribution (Monneret 2019)

What about feedback loops ?



Bayesian Networks



Definition (Bayesian Network):

$(G, \mathbb{P})$ is a Bayesian Network if and only if $G = (V, E)$ is a DAG and

$$\mathbb{P}(X) = \prod_{j=1}^n \mathbb{P}(X_j | X_{\text{pa}_j})$$

with $\text{pa}_j = \{i \in V, (i, j) \in E\}$

Example:

$$\mathbb{P}(X_1, \dots, X_6) = \mathbb{P}(X_1) \mathbb{P}(X_2 | X_1) \mathbb{P}(X_3 | X_1, X_2) \mathbb{P}(X_6 | X_2) \mathbb{P}(X_5 | X_3, X_6) \mathbb{P}(X_4 | X_5)$$

NB: the topological ordering provide a generative procedure

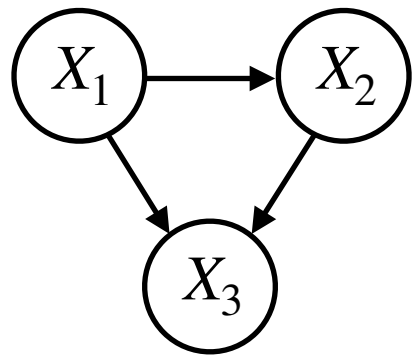
Example of conditional distribution:

With *Gaussian Bayesian Networks* $\mathbb{P}(X_i | X_{\text{pa}_i} = Z) \sim \mathcal{N}(Z\beta, \sigma^2)$ but

we can use any GLM: binomial $\mathcal{B}(n, \text{softmax}(Z\beta))$, Poisson $\mathcal{P}(e^{Z\beta})$, etc.

Markov Equivalence Class

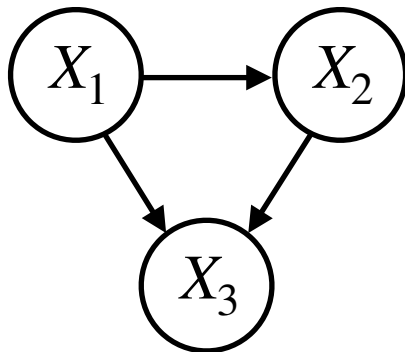
Simulation



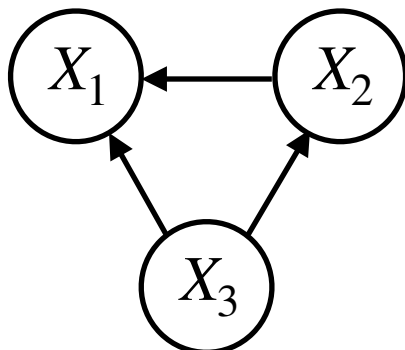
```
set.seed(42)
x1=rnorm(1000)
x2=-0.3*x1+rnorm(1000)
x3=1.2*x1+0.5*x2+rnorm(1000)
```

	x1	x2	x3
[1,]	1.3709584	1.9137710	2.85261368
[2,]	-0.5646982	0.6935316	-0.60879604
[3,]	0.3631284	0.8617949	-0.85808419
[4,]	0.6328626	0.1871146	-1.15371251
[5,]	0.4042683	-1.1172139	-1.36529329
[6,]	-0.1061245	-0.5656456	-0.04433397

Estimations



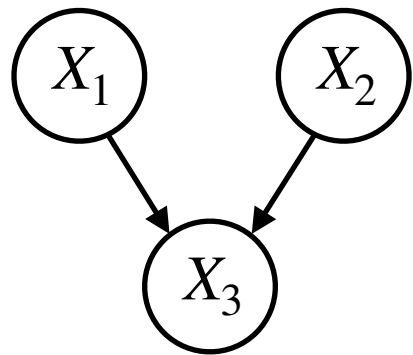
```
> reg=list(lm(x1~1),lm(x2~x1),lm(x3~x1+x2))
> sum(sapply(reg,logLik))
[1] -4272.506
```



```
> reg=list(lm(x1~x2+x3),lm(x2~x3),lm(x3~1))
> sum(sapply(reg,logLik))
[1] -4272.506
```

Markov Equivalence Class

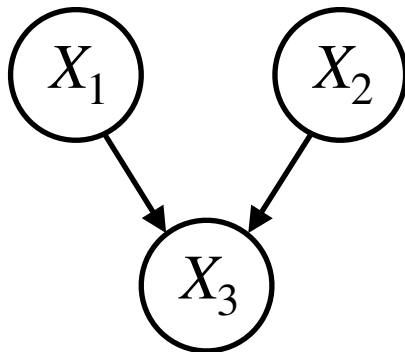
Simulation



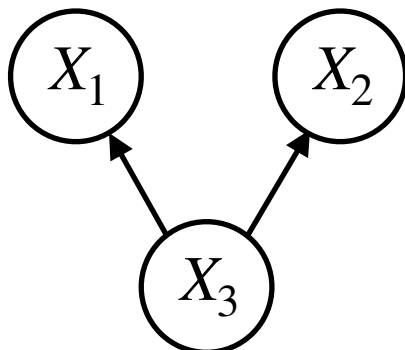
```
set.seed(42)
x1=rnorm(1000)
x2=rnorm(1000)
x3=1.2*x1+0.5*x2+rnorm(1000)
```

	x1	x2	x3
[1,]	1.3709584	1.9137710	2.85261368
[2,]	-0.5646982	0.6935316	-0.60879604
[3,]	0.3631284	0.8617949	-0.85808419
[4,]	0.6328626	0.1871146	-1.15371251
[5,]	0.4042683	-1.1172139	-1.36529329
[6,]	-0.1061245	-0.5656456	-0.04433397

Estimations



```
> reg=list(lm(x1~1),lm(x2~1),lm(x3~x1+x2))
> sum(sapply(reg,logLik))
[1] -4314.239
```



```
> reg=list(lm(x1~x3),lm(x2~x3),lm(x3~1))
> sum(sapply(reg,logLik))
[1] -4405.149
```

Markov Equivalence Class

Definition (skeleton) :

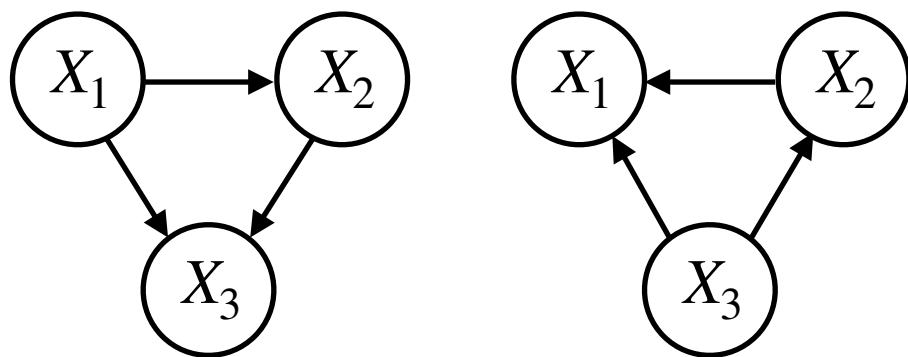
The *skeleton* of a DAG is the *undirected* graph induced by its (directed) edges

Definition (v-structure):

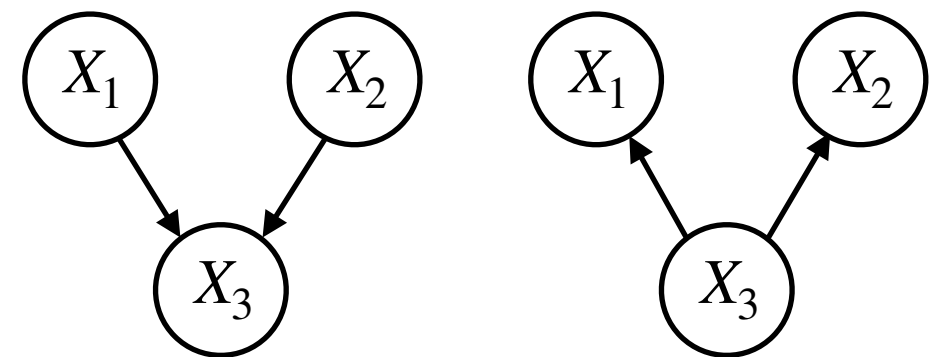
(A, B, C) is a *v-structure* of a DAG iff:
 $A \rightarrow B \leftarrow C$ without $A \rightarrow C$ nor $A \leftarrow C$

Theorem (2.1 in Andersson *et al* 1997):

Two DAGs are Markov equivalent if and only if they have the same *skeleton* and the same *v-structures* (also called *immoralities*).



Same skeleton no v-structures



Same skeleton different v-structures

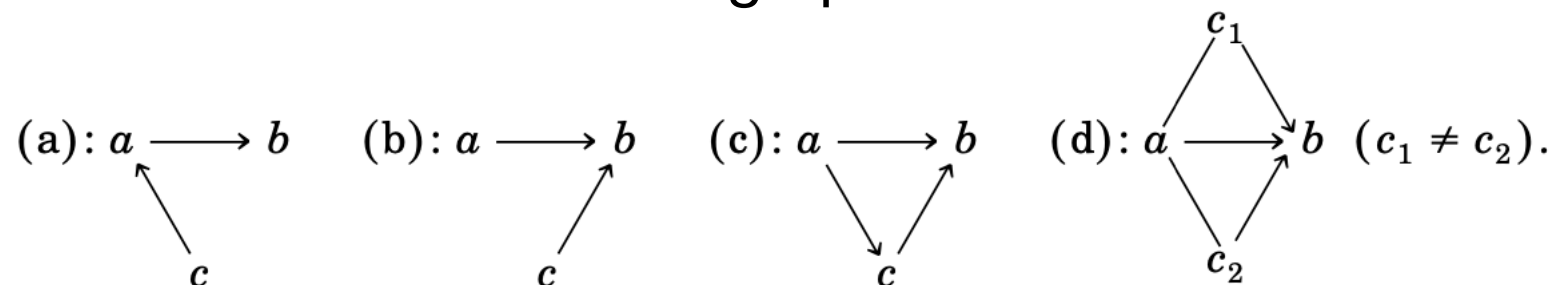
CPDAG: Completed Partially Directed Acyclic Graph

Definition (CPDAG):

The CPDAG (also called essential graph) is a PDAG representing the MEC of a DAG. Directed edge iff shared by all DAGs, undirected otherwise.

Definition (strongly protected arrows):

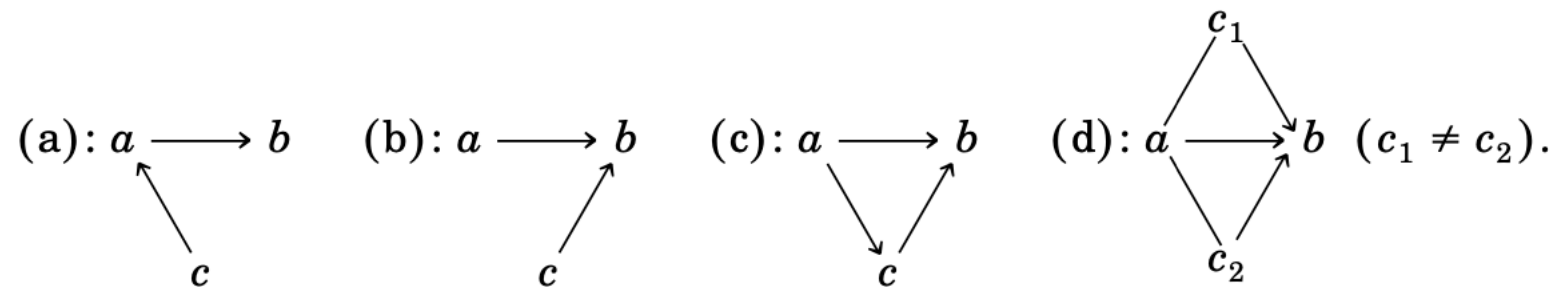
$a \rightarrow b$ is *strongly protected* in G if $a \rightarrow b$ occurs in at least one of the following configurations in the induced subgraph



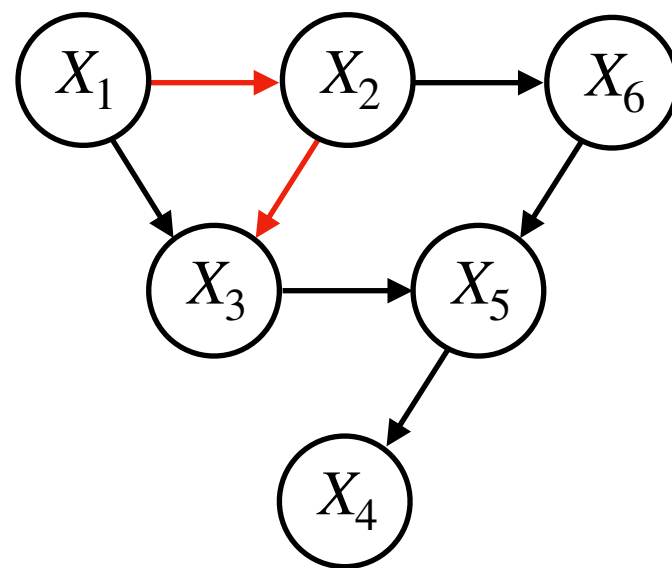
Definition 3.3 from Andersson *et al* (1997)

Algorithm (Algo 1, Hauser & Bühlmann, 2012): we can build a CPDAG from a DAG by dropping all arrows not strongly protected, updating the edges, and repeat until convergence

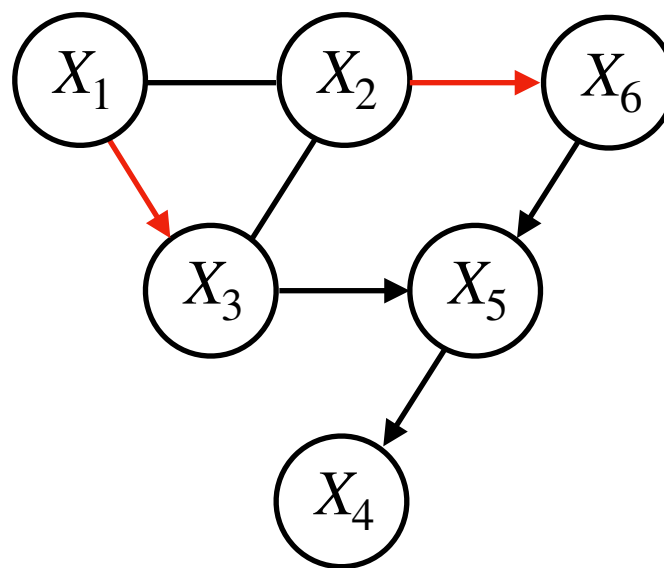
CPDAG: Completed Partially Directed Acyclic Graph



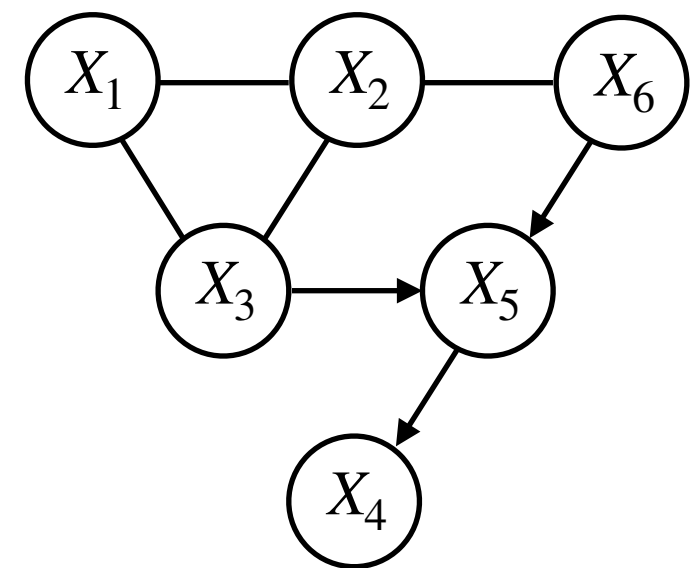
$a \rightarrow b$ strongly protected (Andersson *et al*, 1997)



Initial DAG



Intermediary PDAG



CPDAG

Posterior DAG Distribution

$$\mathbb{P}(G|\text{data}) \propto \underbrace{\mathbb{P}(G)}_{\text{DAG prior}} \times \overbrace{\int_{\theta} \underbrace{\mathbb{P}(\text{data}|G, \theta)}_{\text{likelihood}} \times \underbrace{\mathbb{P}(\theta|G)}_{\text{param. prior}} d\theta}_{\text{integrated likelihood}}$$

Rather than integrating the likelihood, we use the following approximation:

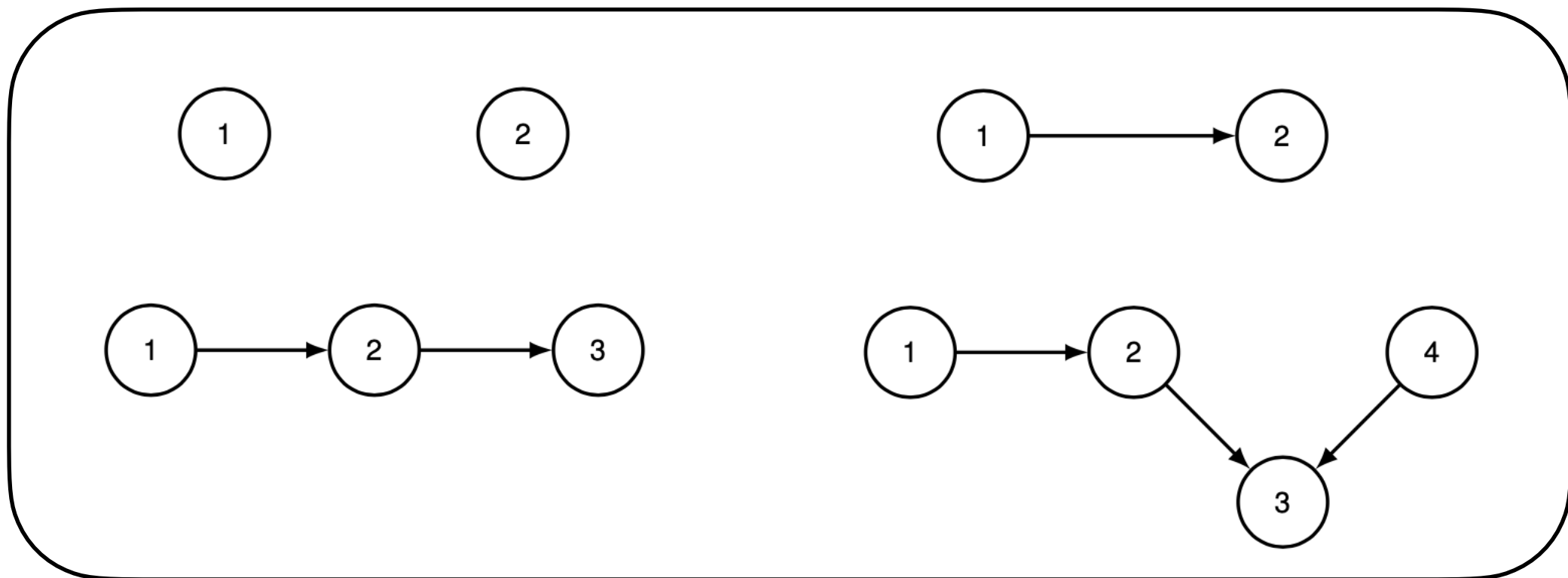
$$\log \mathbb{P}(G|\text{data}) \simeq \text{Cst.} + \log \mathbb{P}(G) + \log \text{lik}(\hat{\theta}|G) - \text{pen}(G)$$

where the penalty function can be either:

- $\text{pen}(G) = \frac{1}{2} \sum_j (|\text{pa}_j| + 2) \log n_j$ (BIC)
- $\text{pen}(G) = \frac{1}{2} \sum_j \left\{ (|\text{pa}_j| + 2) \log n_j + \log \binom{|\text{pa}_j|}{p-1} \right\}$ (eBIC)

Toy-examples

Four reference DAGs

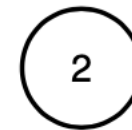
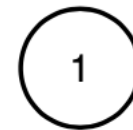


Experiments:

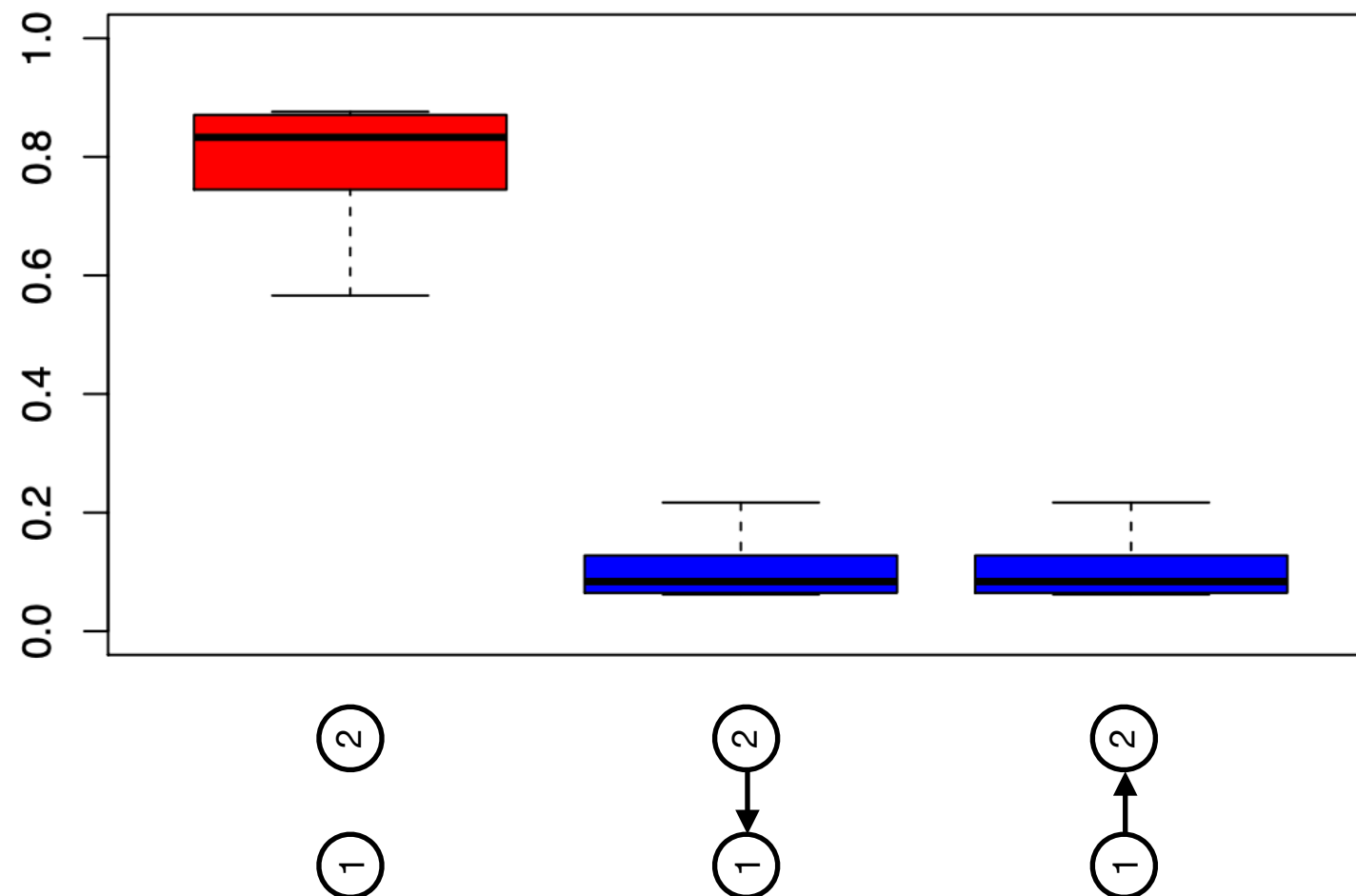
- Simulate 200 observations using a GBN
- Exhaustive search over the DAG space
- Posterior $\mathbb{P}(G \mid \text{data})$ over 100 replicates

$p = 2$ search over 3 DAGs

dag = [] []

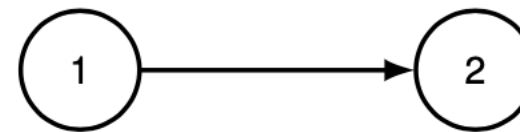


200 WT

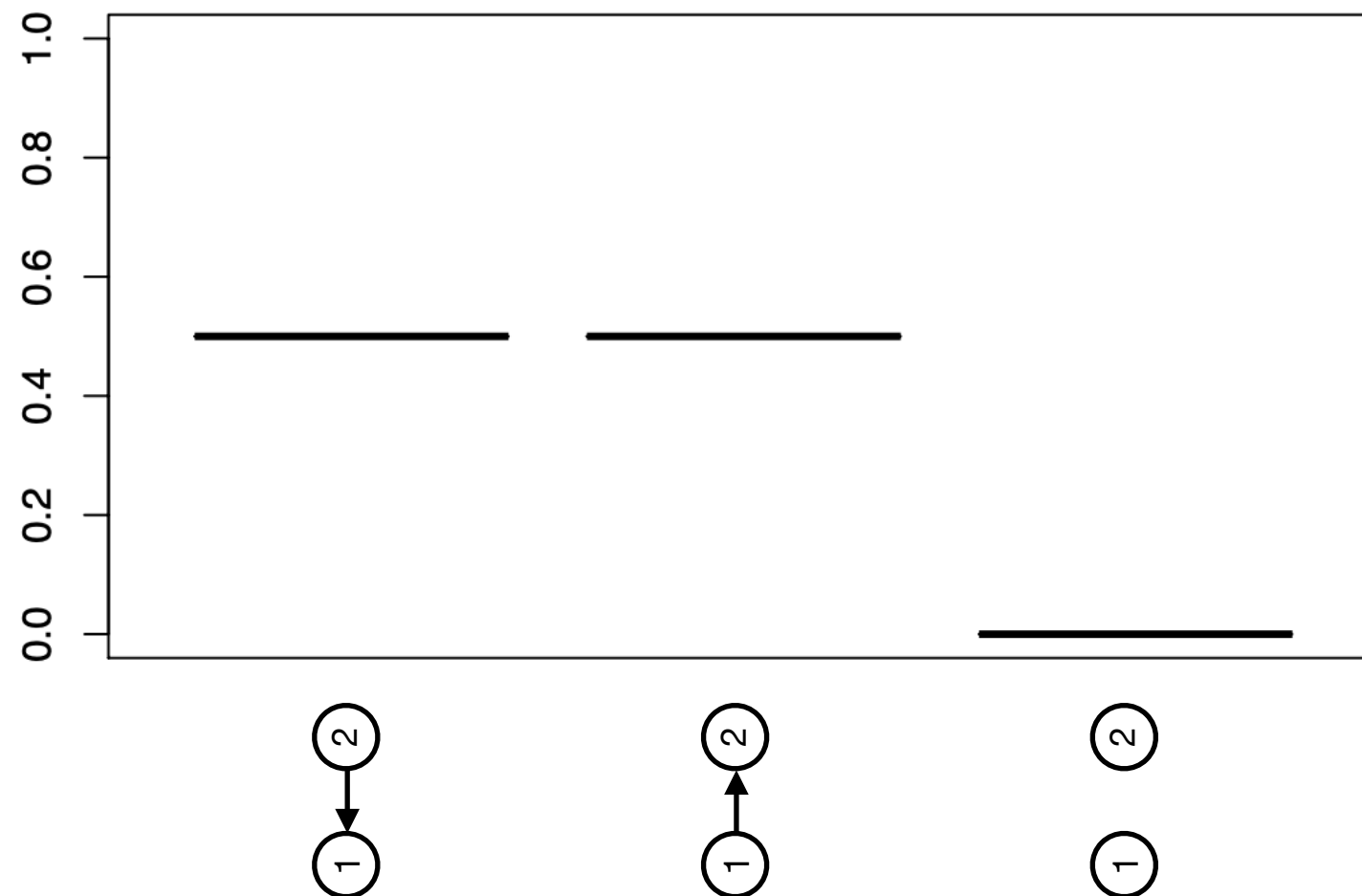


$p = 2$ search over 3 DAGs

dag = [] [1]

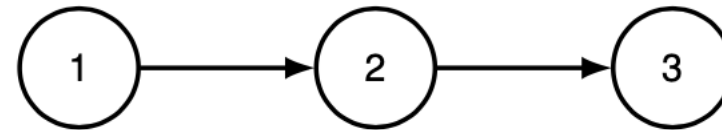


200 WT

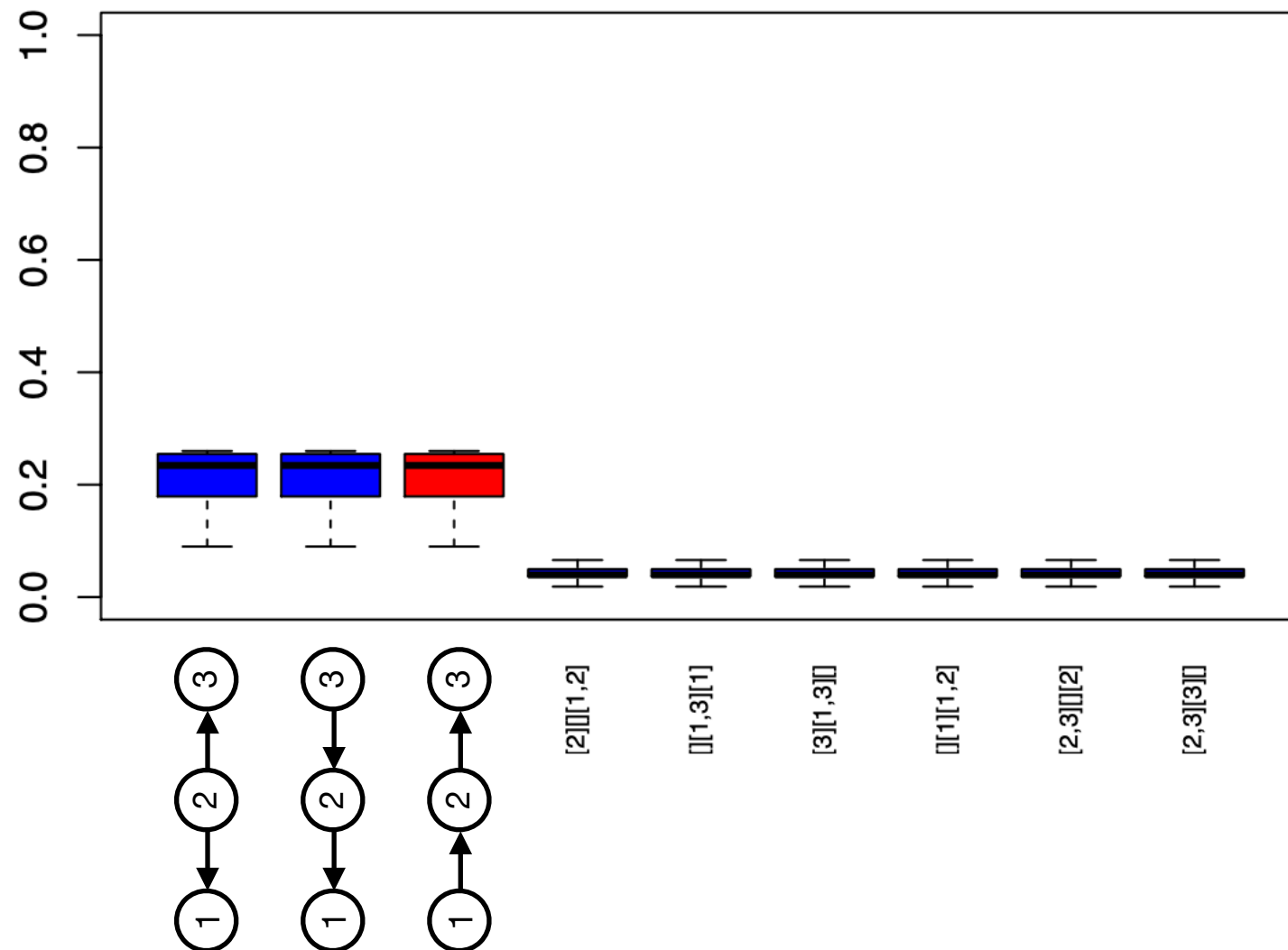


$p = 3$ search over 25 DAGs

dag = [] [1] [2]

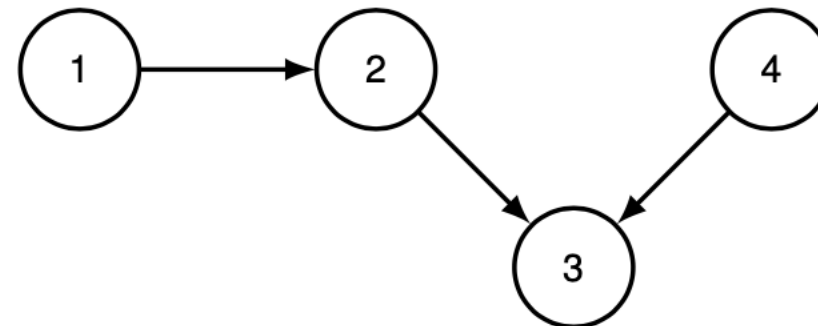


200 WT

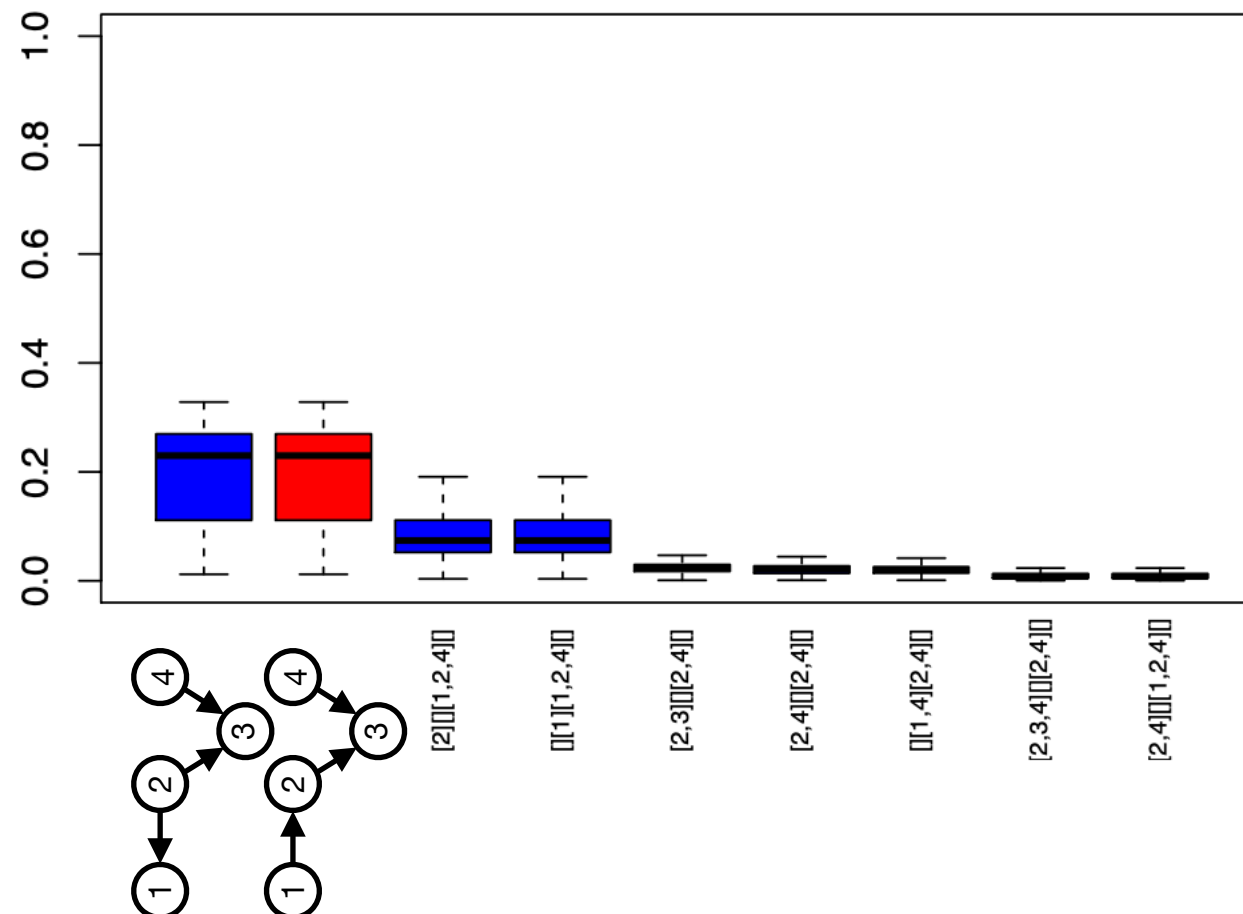


$p = 4$ search over 543 DAGs

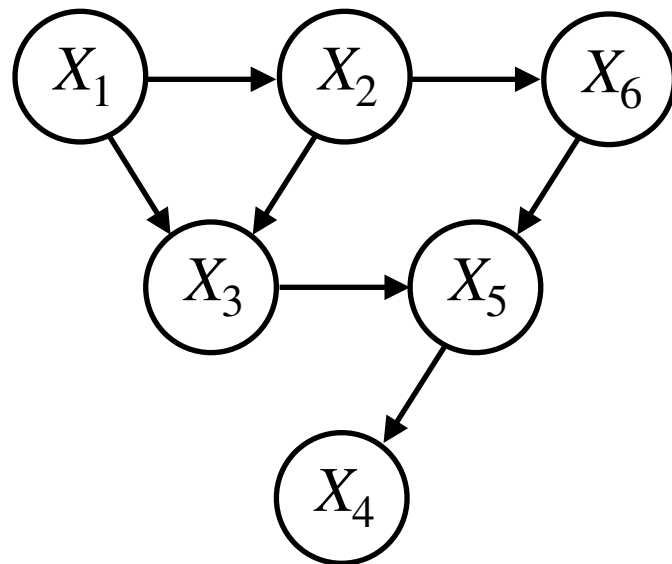
dag = [] [1] [2, 4] []



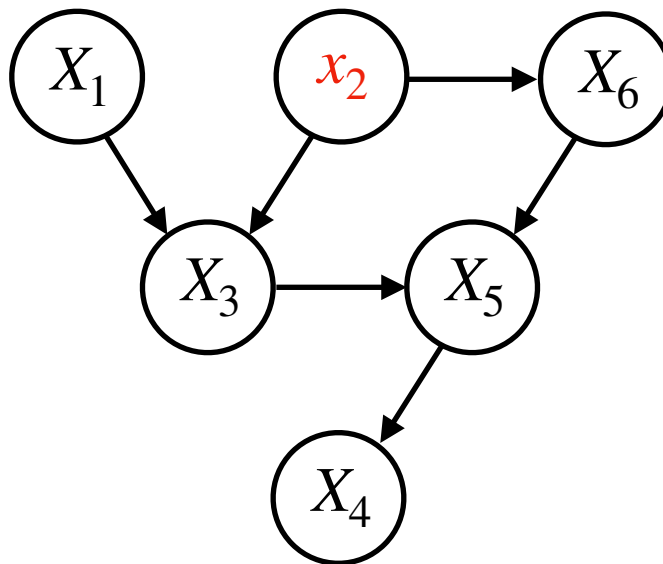
200 WT



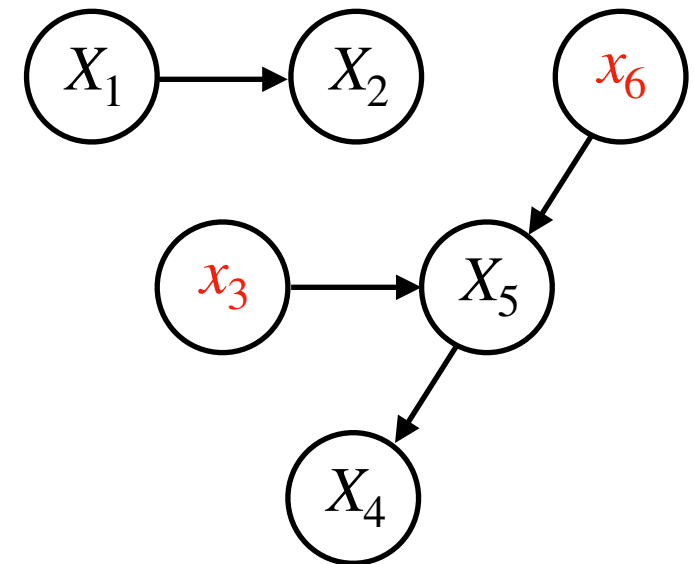
Interventions: Do operator



Observation



$\text{Do}(X_2 = x_2)$



$\text{Do}(X_3 = x_3, X_6 = x_6)$

Example of interventions:

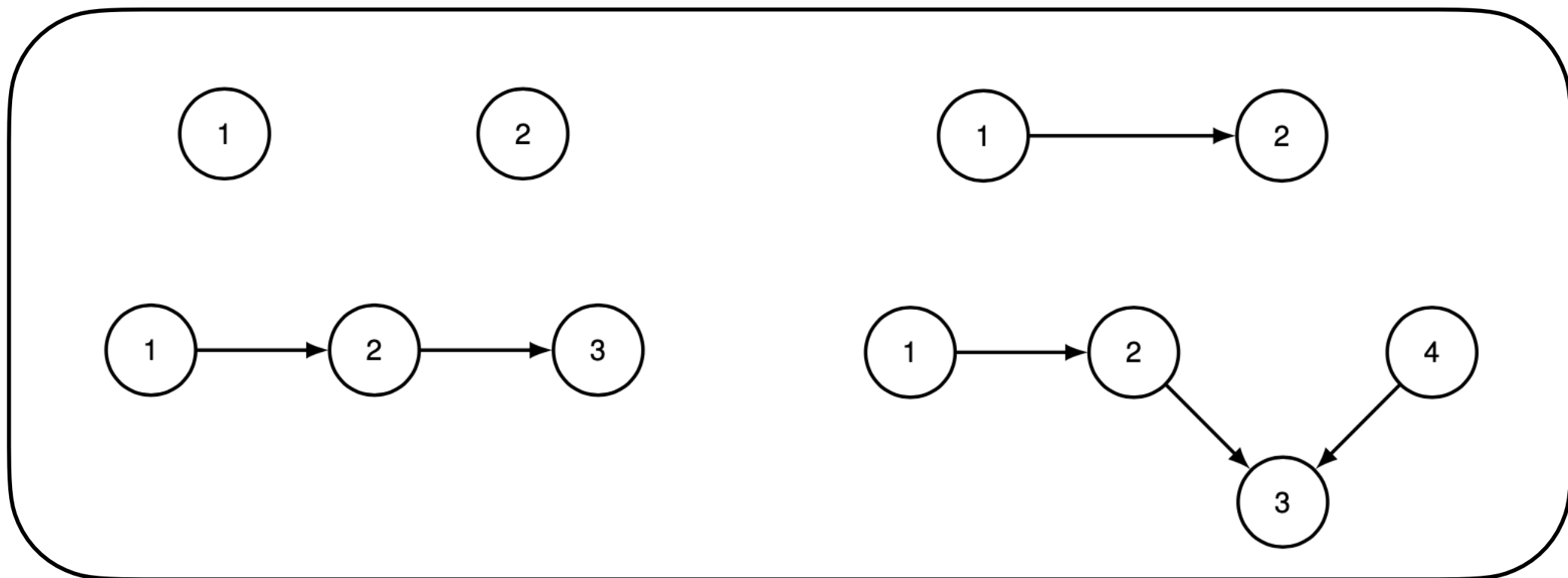
- Clinical randomization $\text{Do}(T = t)$
- Gene knock-out $\text{Do}(G = 0)$
- Knock-down/up
- Functional knock-out



$$\mathbb{P}(X | \text{Do}(Y = y)) \neq \mathbb{P}(X | Y = y)$$

Back to the Toy-examples

Four reference DAGs

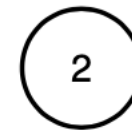
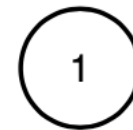


Experiments:

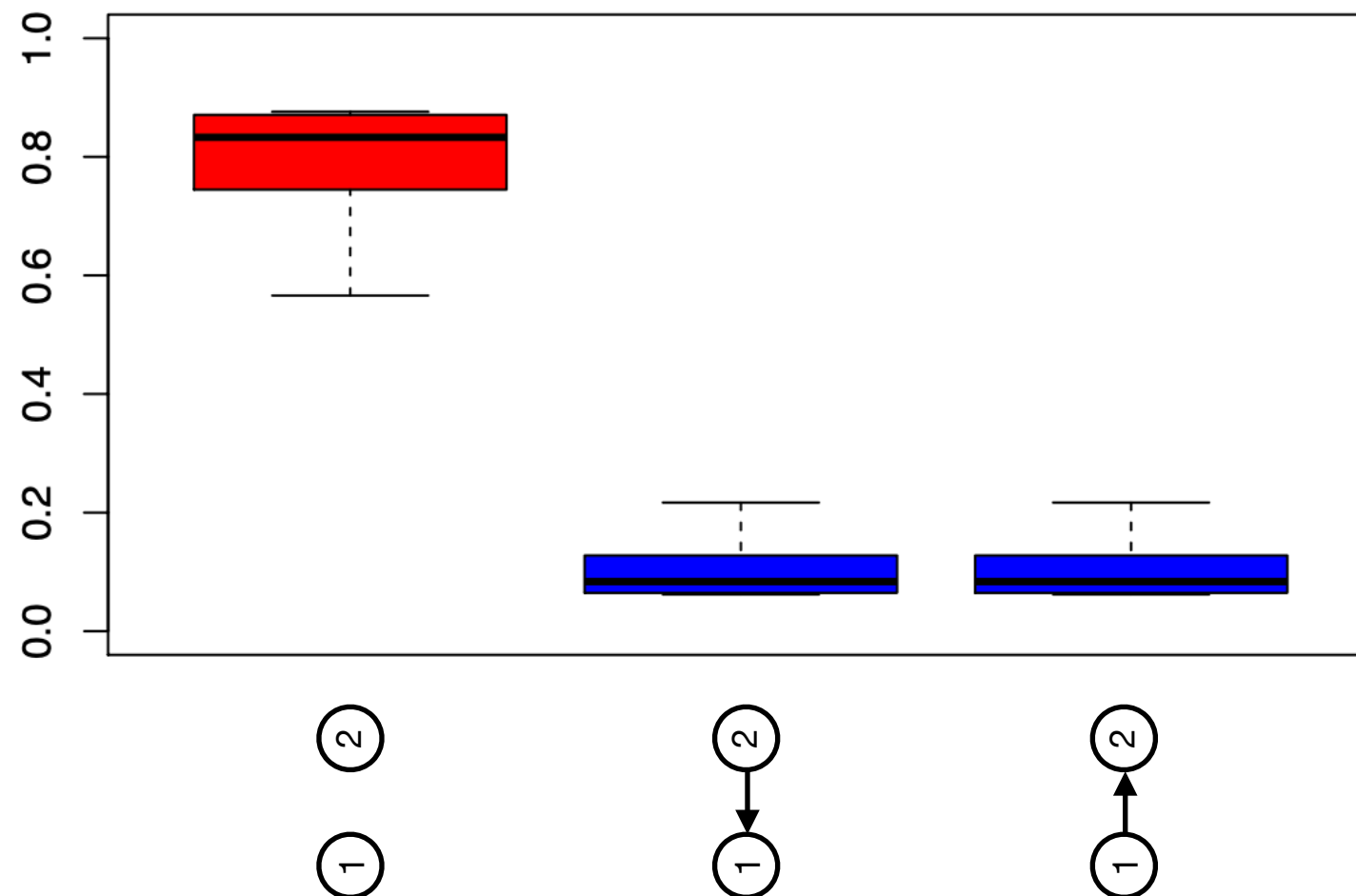
- Simulate 200 observations using a GBN
- **Plus interventions !**
- Exhaustive search over the DAG space
- Posterior $\mathbb{P}(G \mid \text{data})$ over 100 replicates

$p = 2$ search over 3 DAGs

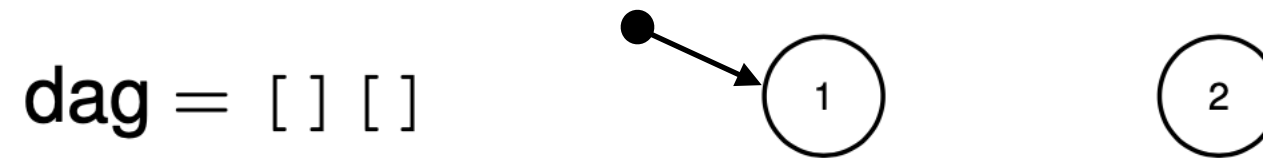
dag = [] []



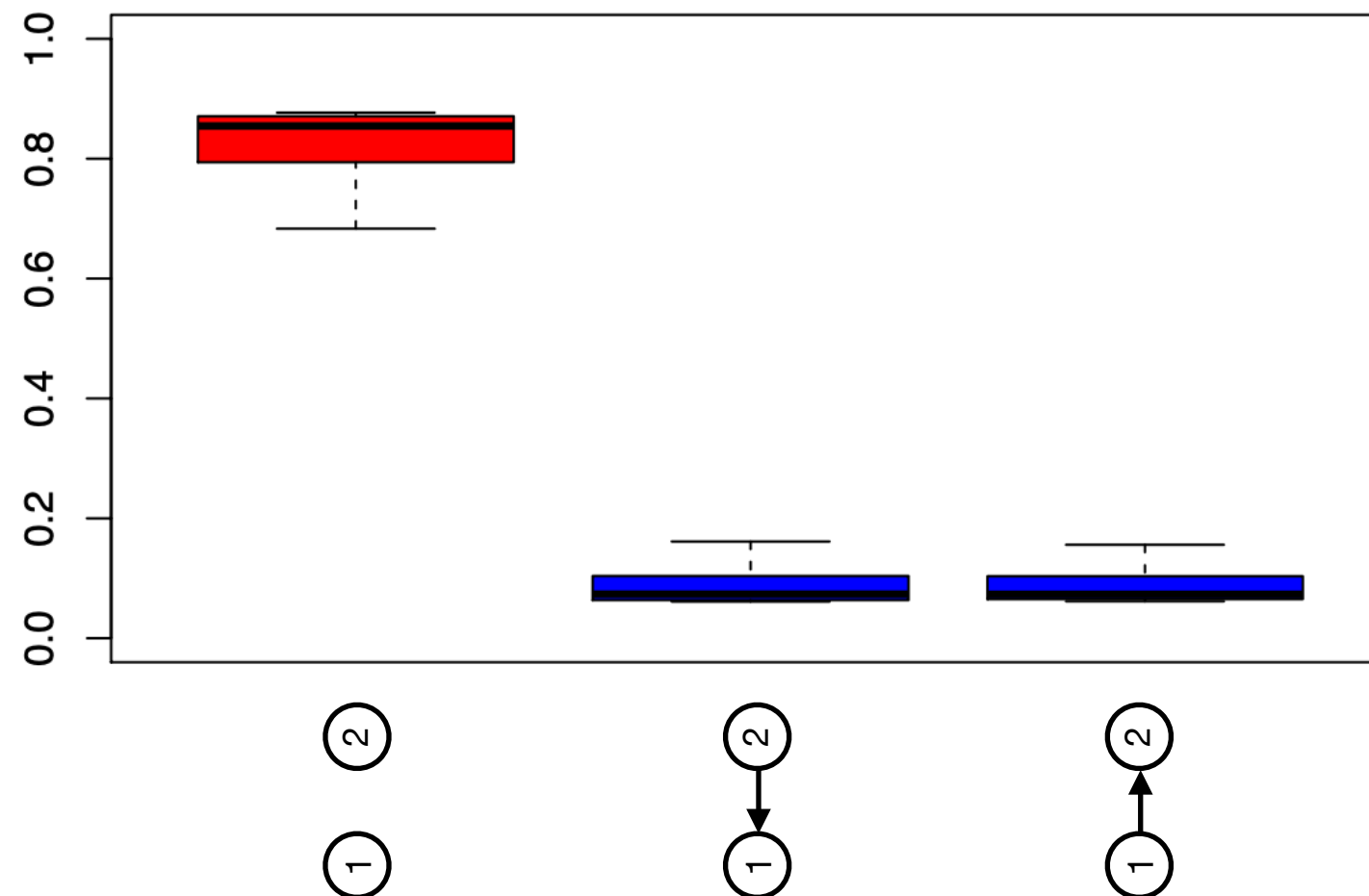
200 WT



$p = 2$ search over 3 DAGs

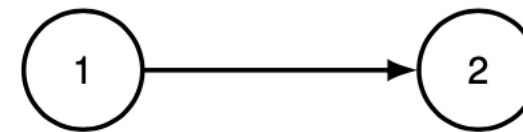


200 WT + 5 KO(1)

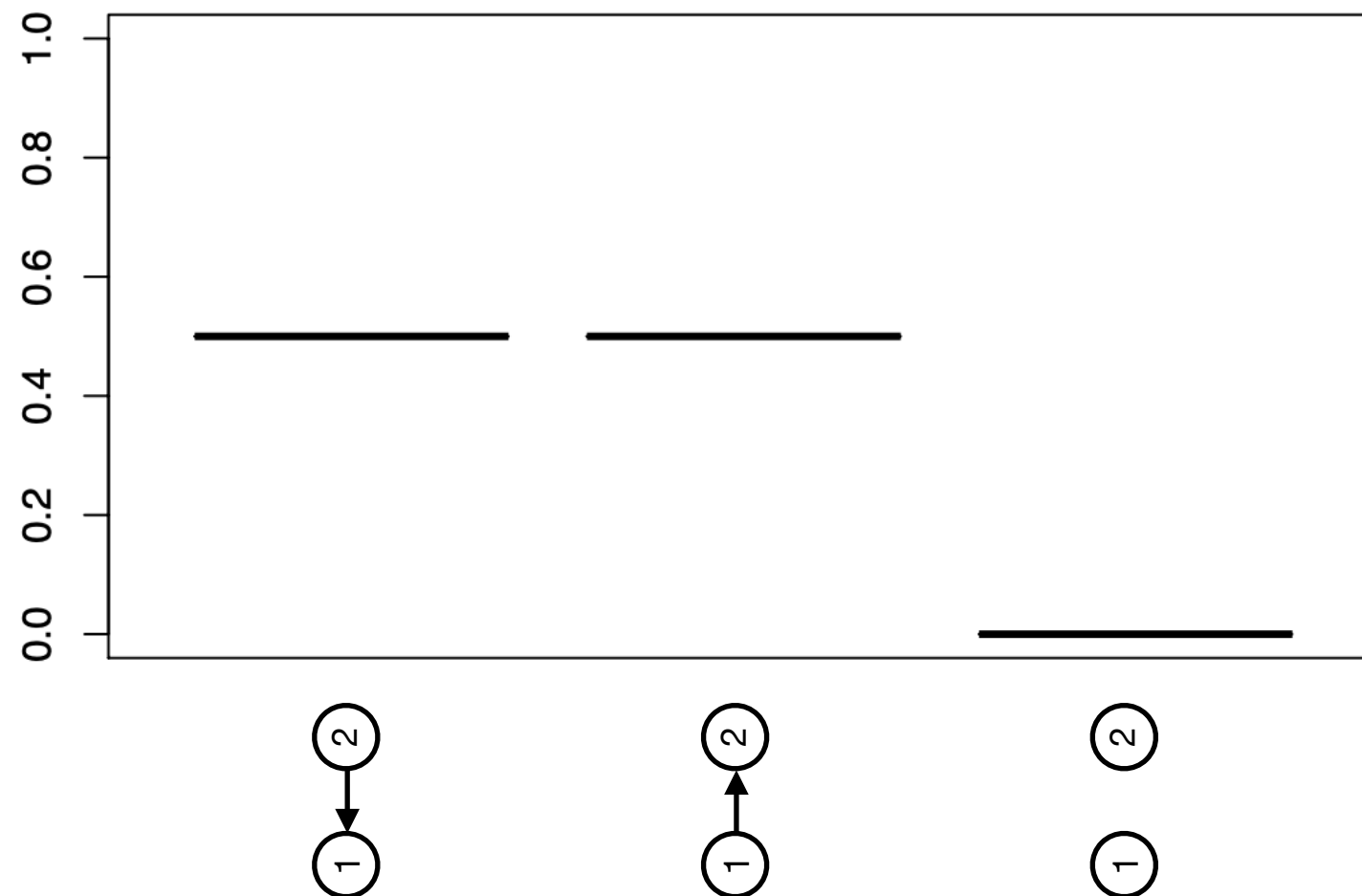


$p = 2$ search over 3 DAGs

dag = [] [1]

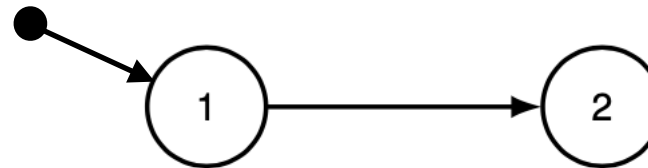


200 WT

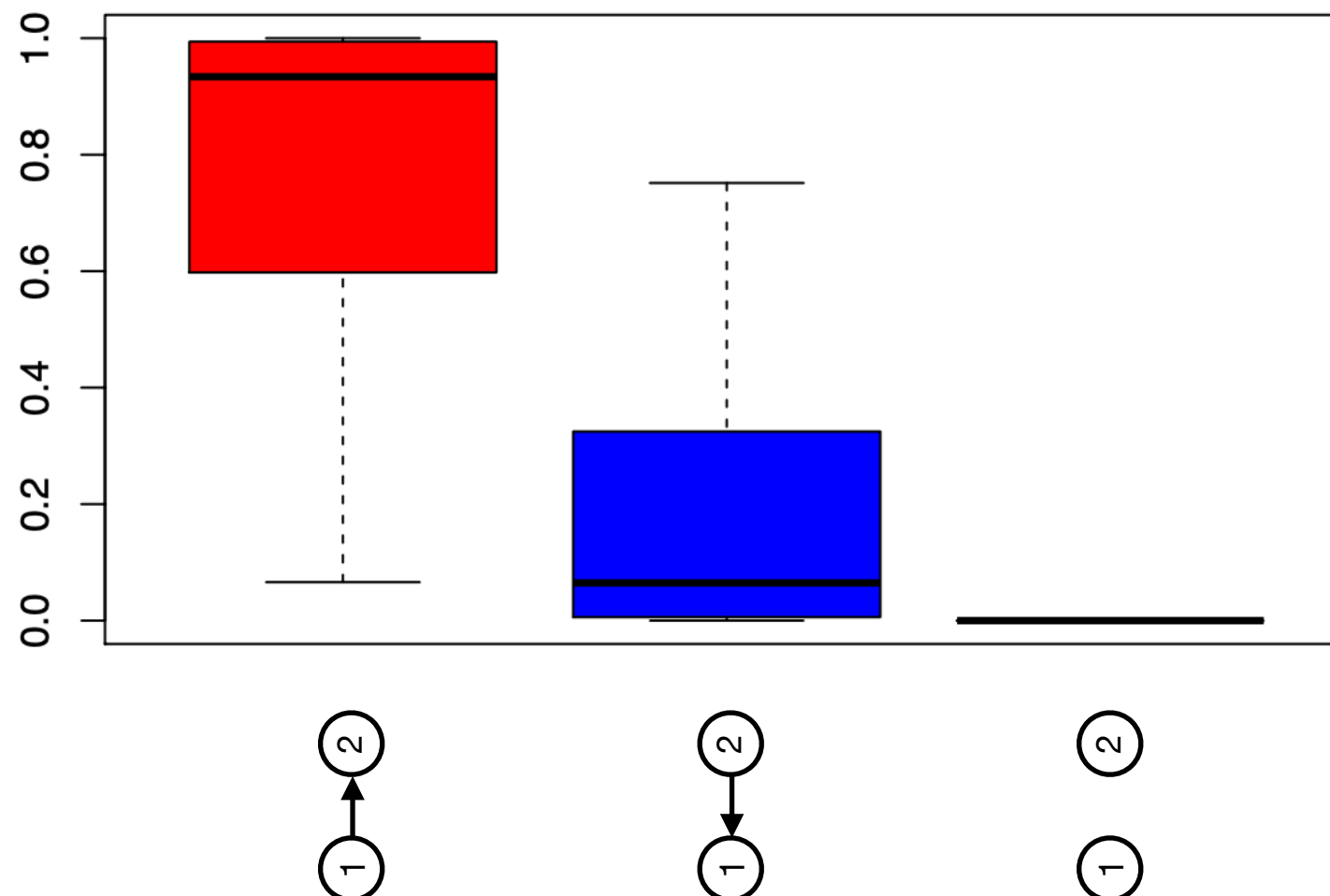


$p = 2$ search over 3 DAGs

dag = [] [1]

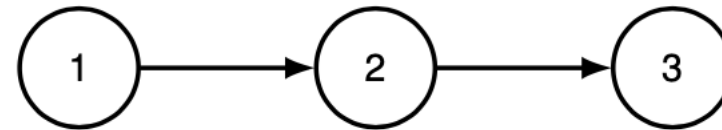


200 WT + 5 KO(1)

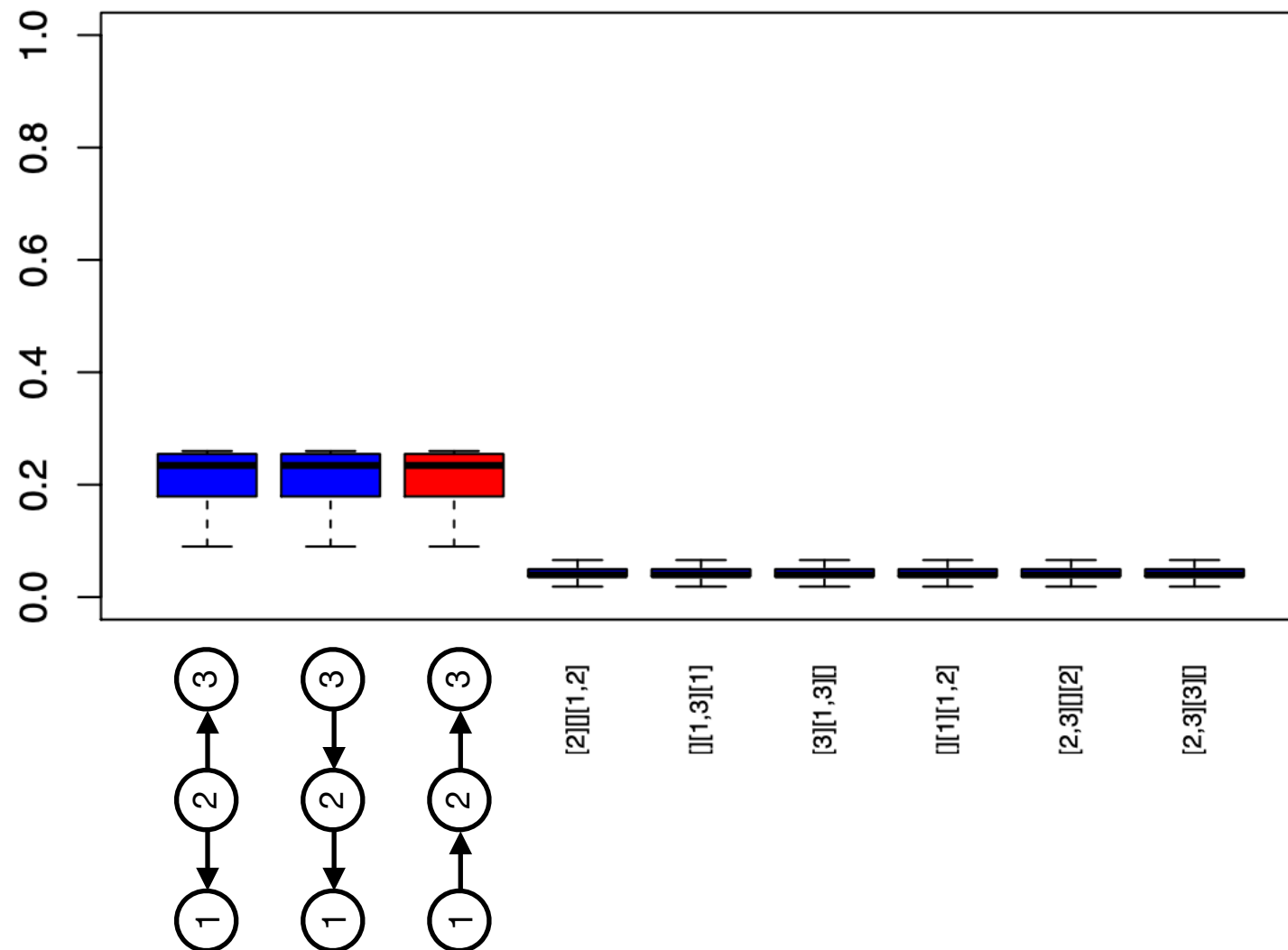


$p = 3$ search over 25 DAGs

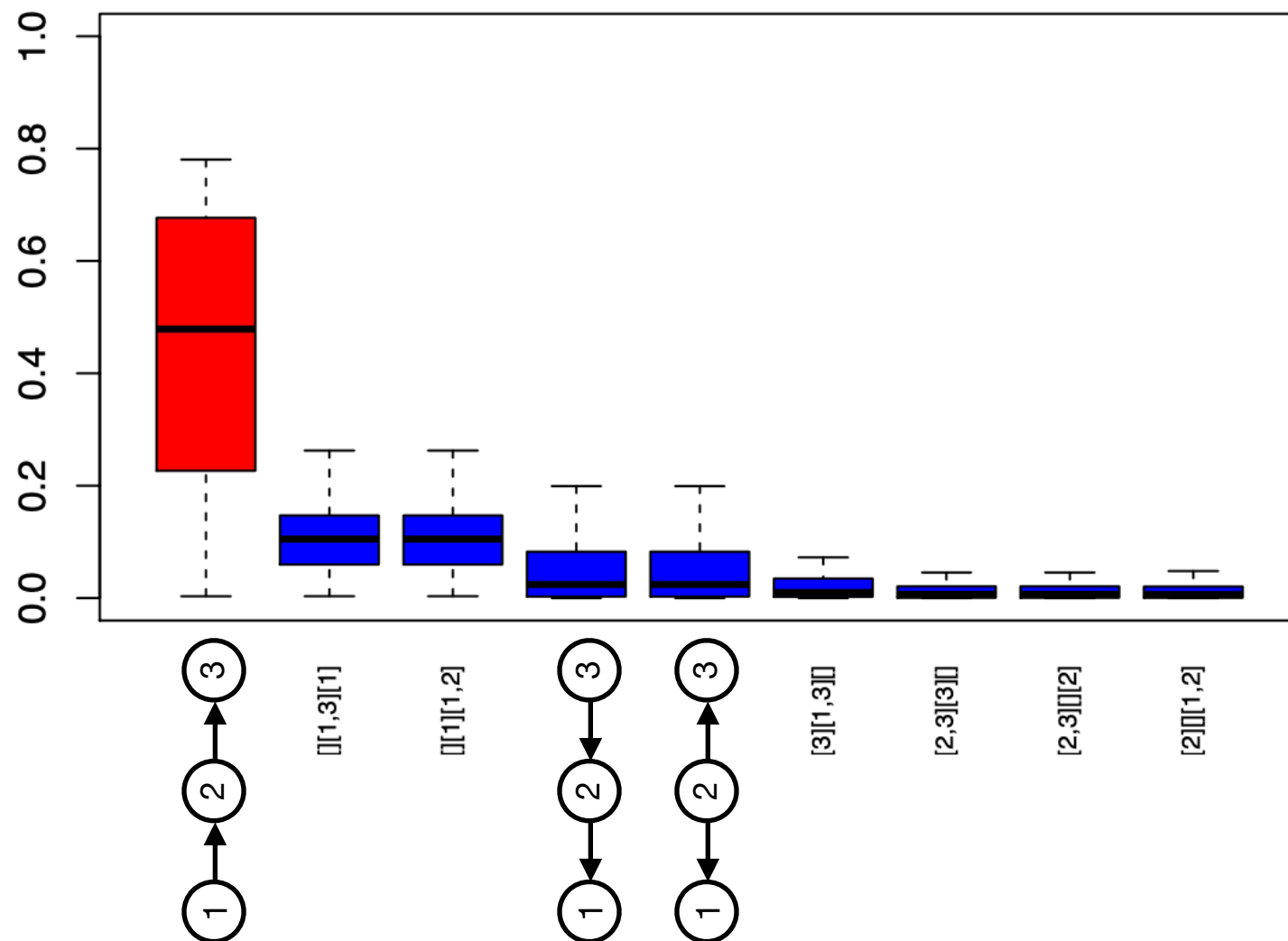
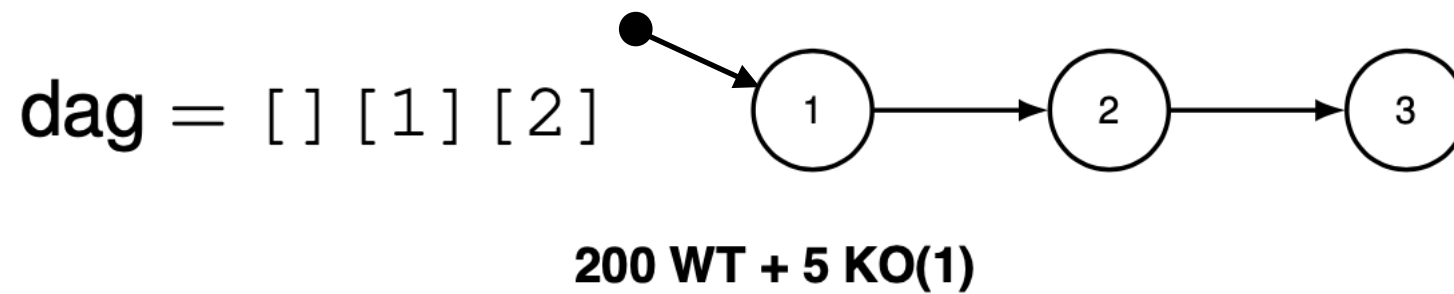
dag = [] [1] [2]



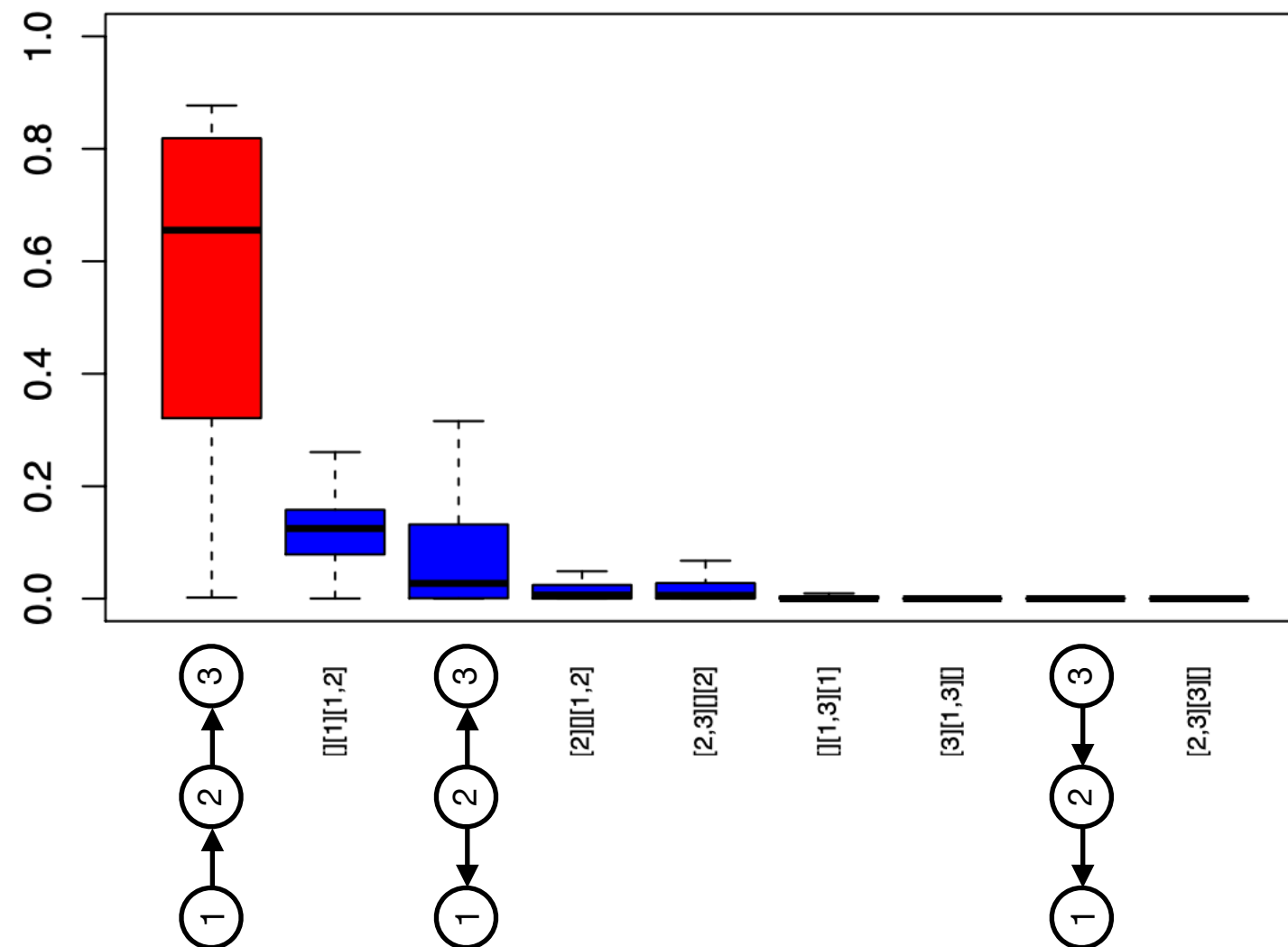
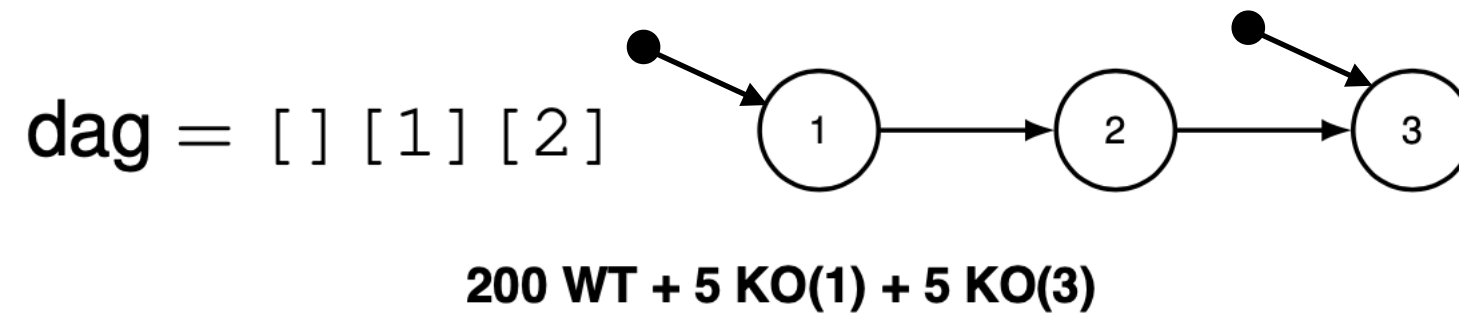
200 WT



$p = 3$ search over 25 DAGs

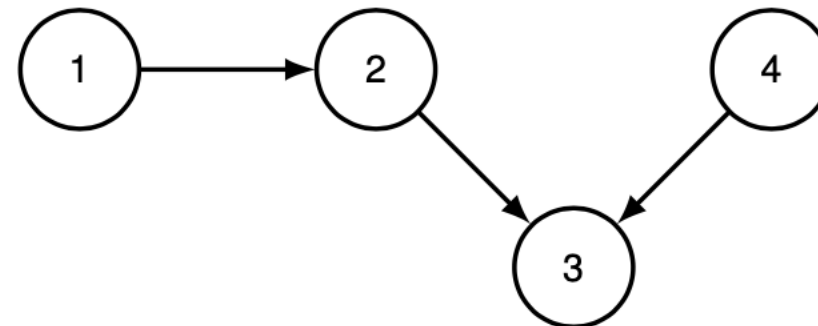


$p = 3$ search over 25 DAGs

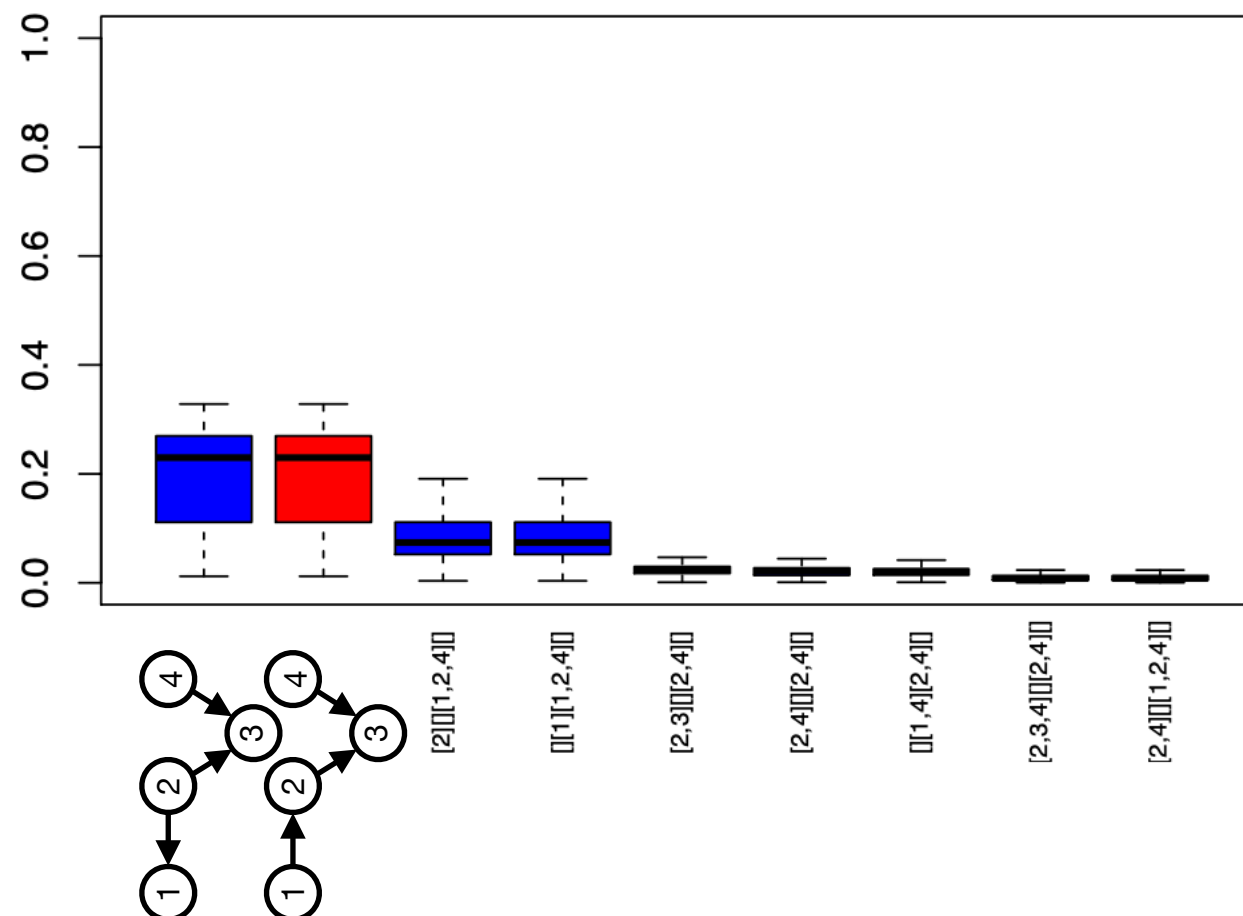


$p = 4$ search over 543 DAGs

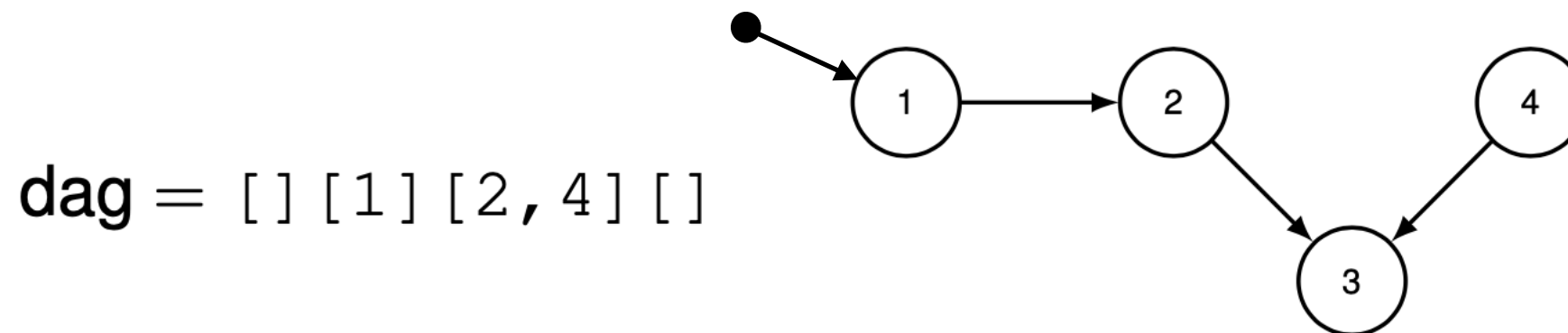
dag = [] [1] [2, 4] []



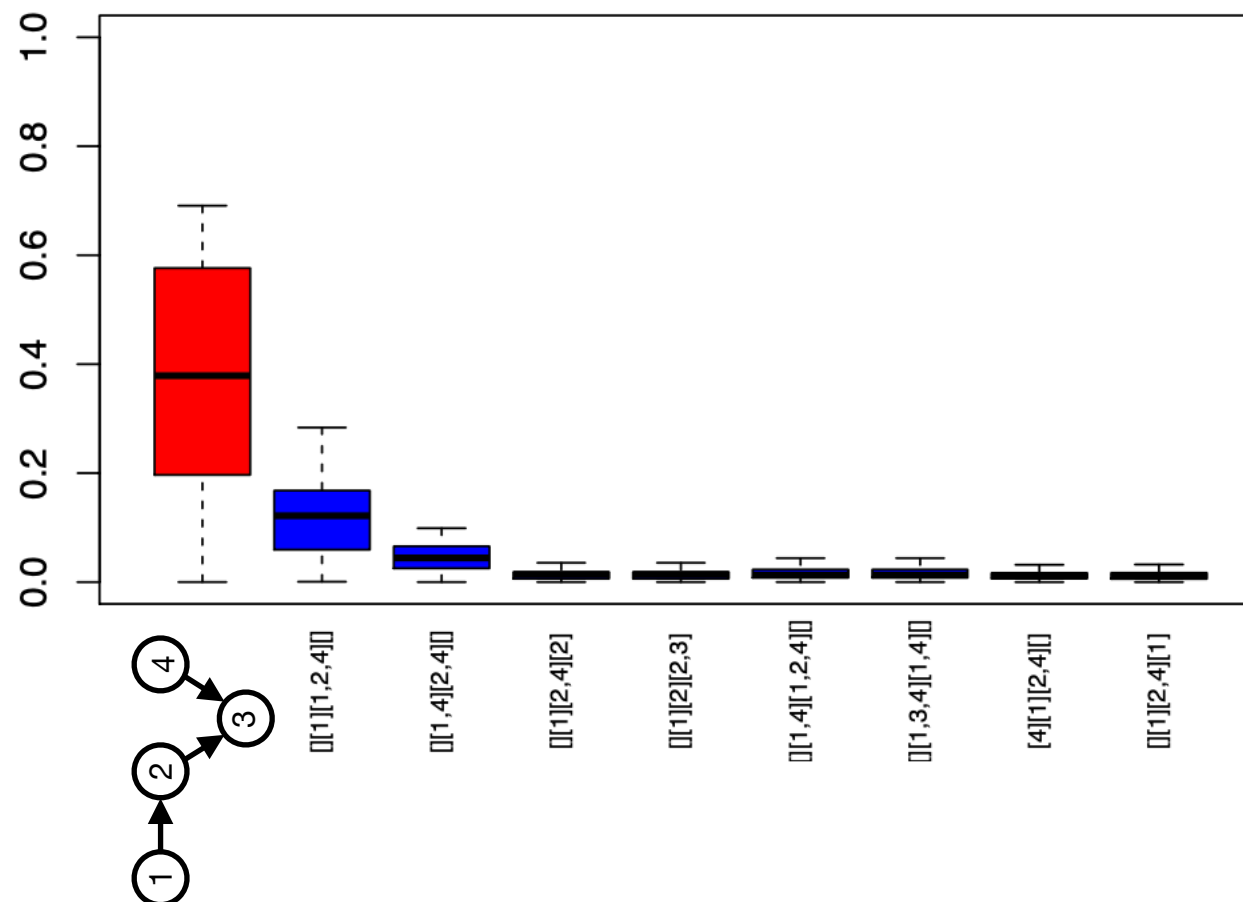
200 WT



$p = 4$ search over 543 DAGs

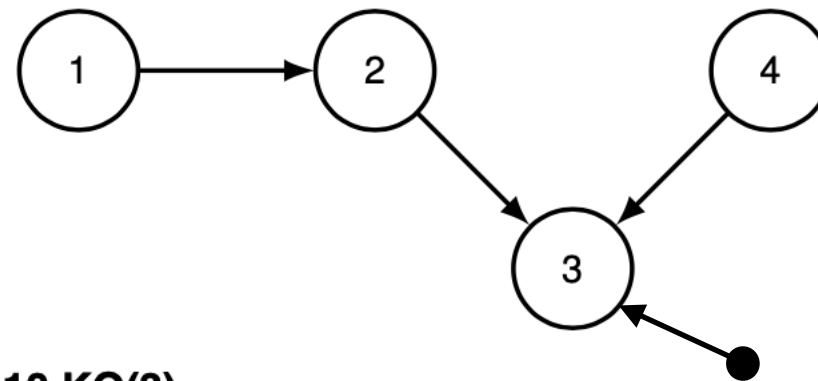


200 WT + 10 KO(1)

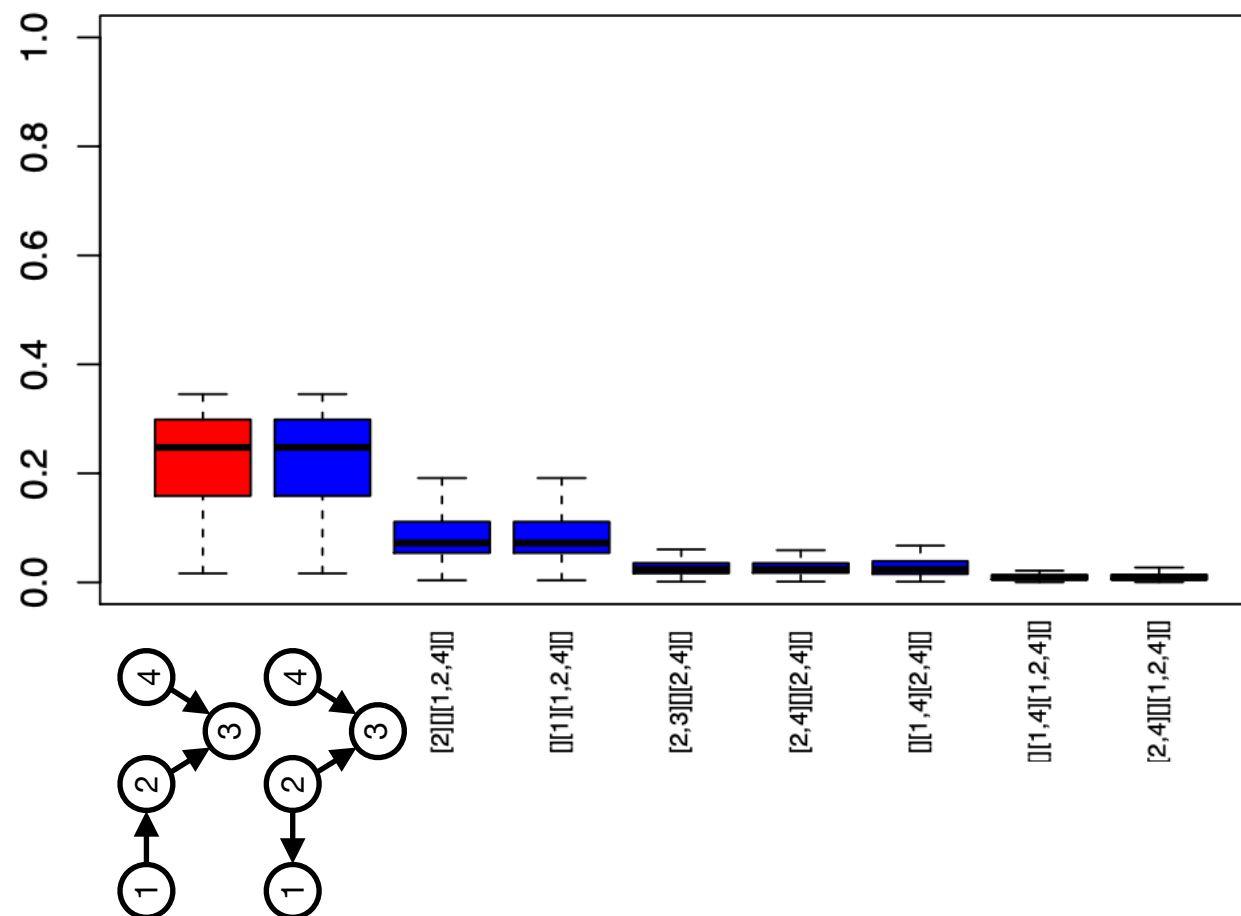


$p = 4$ search over 543 DAGs

dag = [] [1] [2, 4] []



200 WT + 10 KO(3)

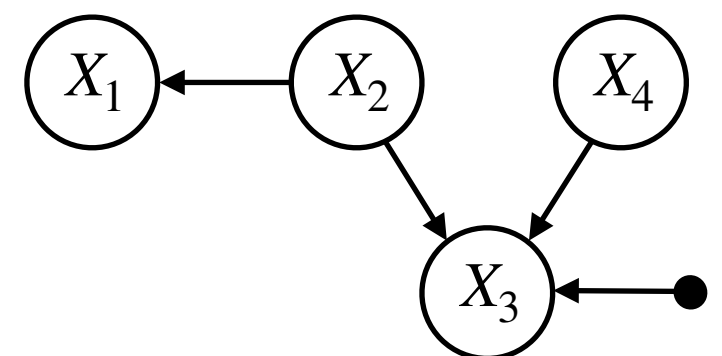
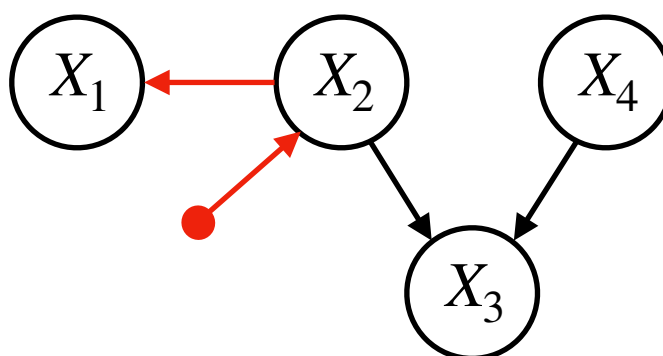
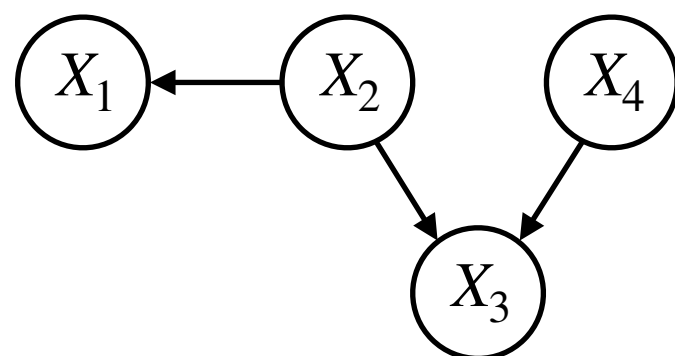
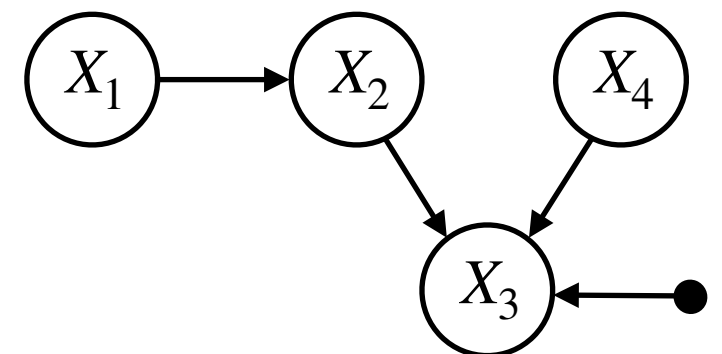
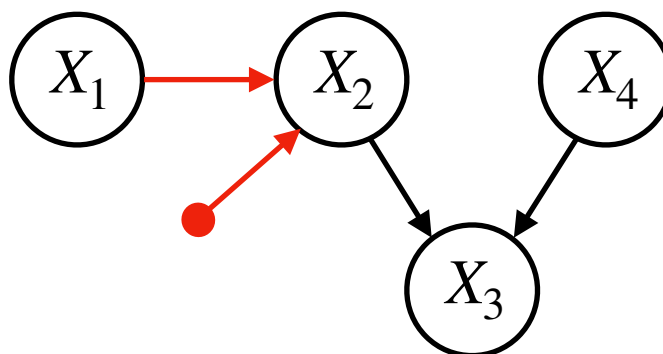
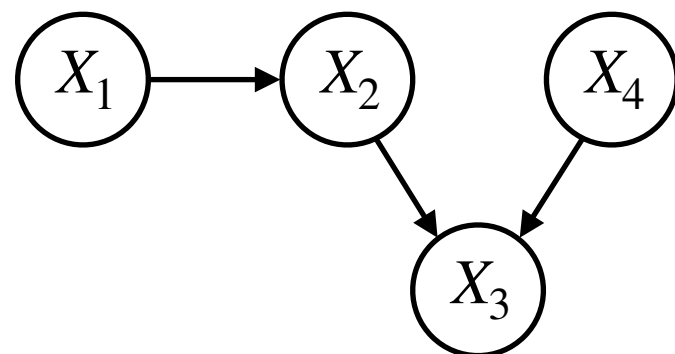


$\mathcal{I}$ -Markov Equivalence Class

Theorem (3.9 in Yang *et al* 2018):

Two DAGs are $\mathcal{I}$ -Markov equivalent with $\emptyset \in \mathcal{I}$ if and only if they have the same *skeleton* and the same *v-structures*.

NB: strongly protected arrows (not from intervention node) of a $\mathcal{I}$ -DAG with $\emptyset \in \mathcal{I}$
Are *exactly* the strongly $\mathcal{I}$ -protected arrows of the DAG (Hauser & Bühlmann, 2012)

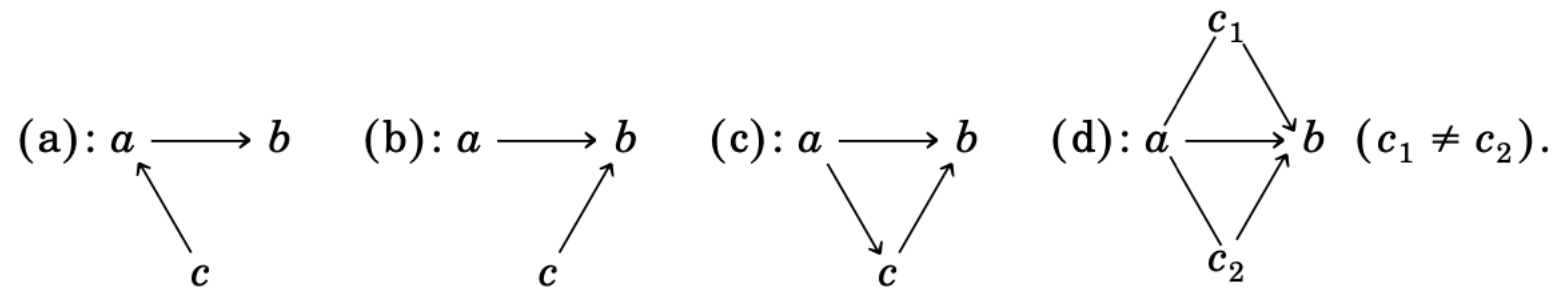


ME DAGs

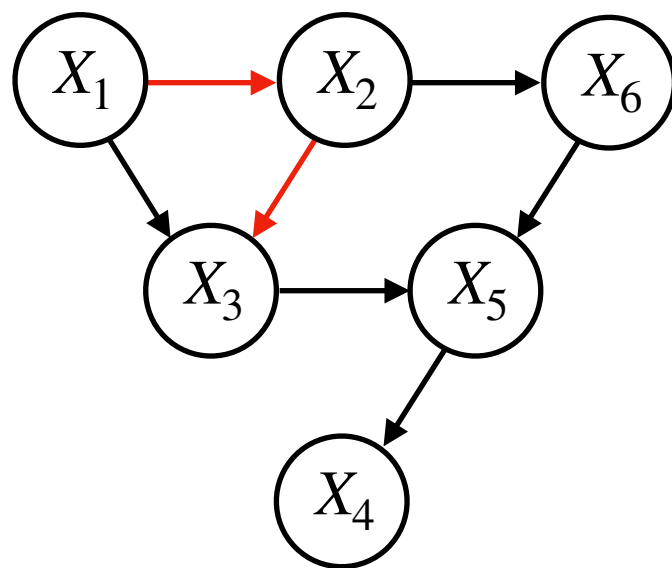
not $\mathcal{I}$ -ME DAGs

$\mathcal{I}$ -ME DAGs

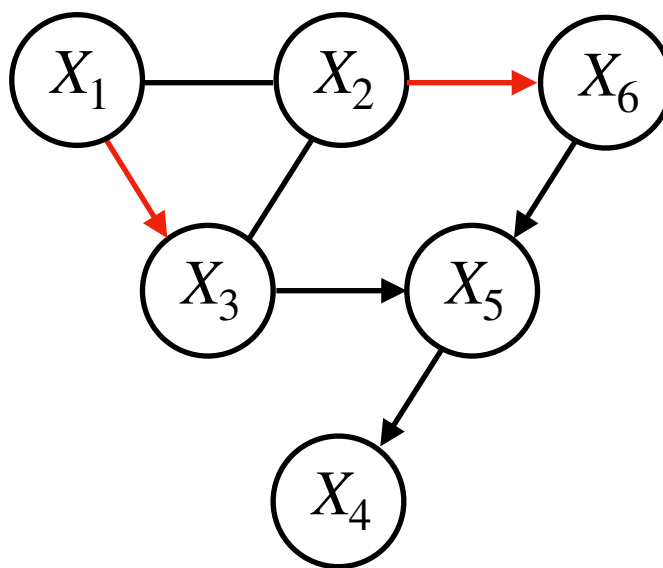
$\mathcal{I}$ -CPDAG: $\mathcal{I}$ -Completed Partially Directed Acyclic Graph



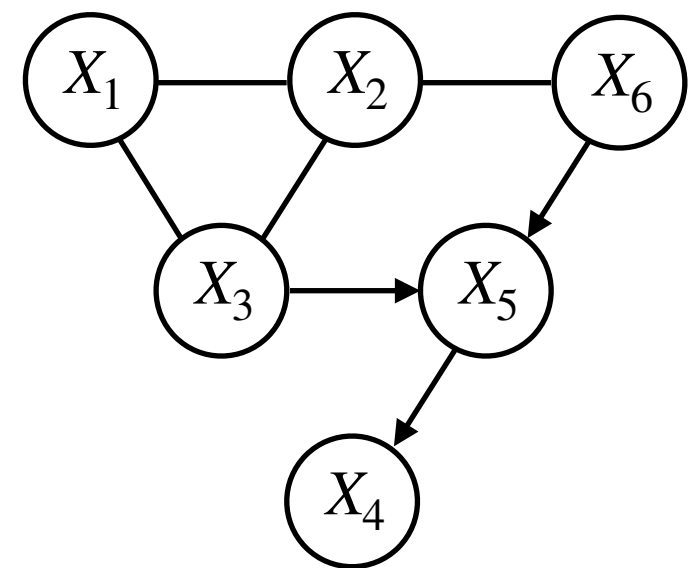
$a \rightarrow b$ strongly protected (Andersson *et al*, 1997)



Initial DAG

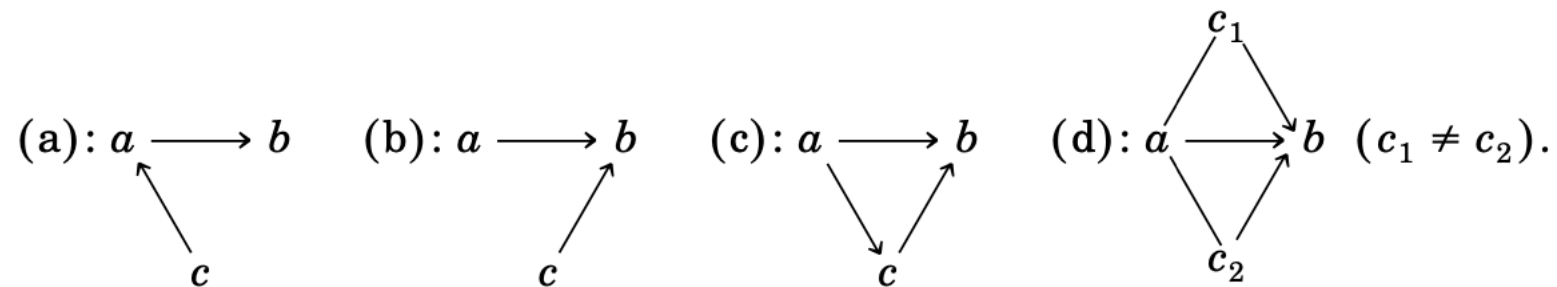


Intermediary PDAG

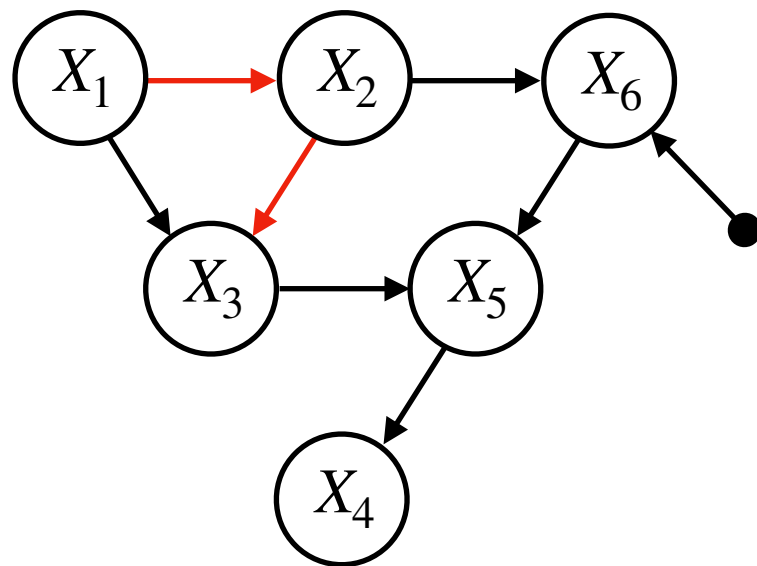


CPDAG

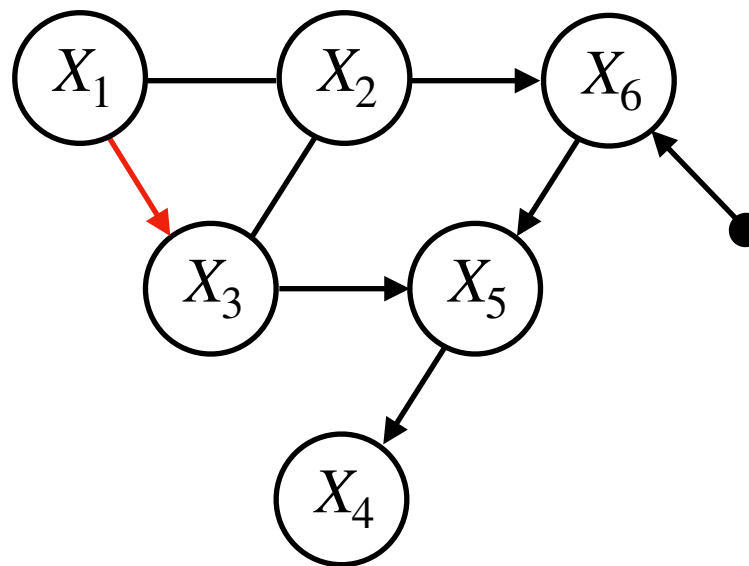
$\mathcal{I}$ -CPDAG: $\mathcal{I}$ -Completed Partially Directed Acyclic Graph



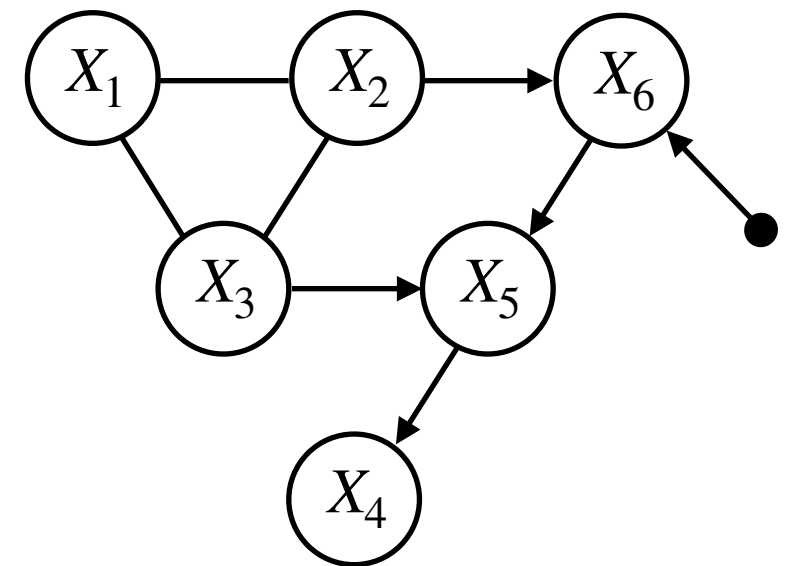
$a \rightarrow b$ strongly protected (Andersson *et al*, 1997)



Initial $\mathcal{I}$ -DAG

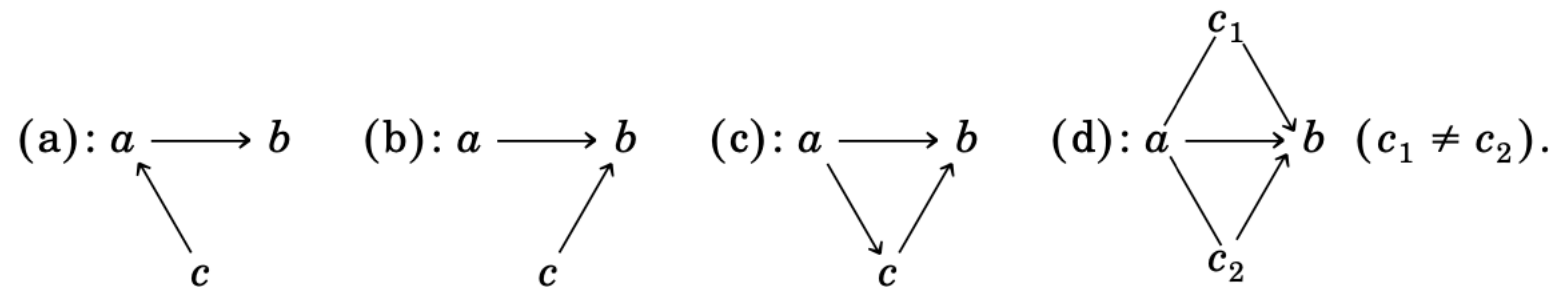


Intermediary $\mathcal{I}$ -PDAG

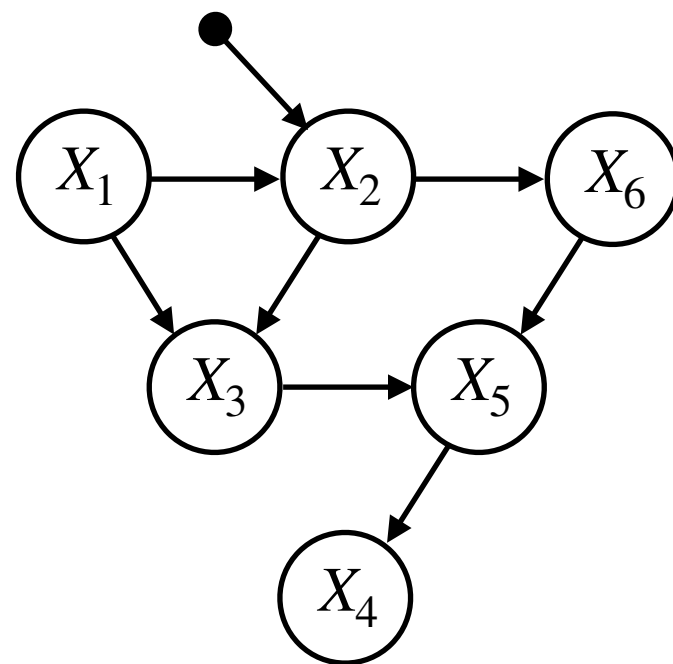


$\mathcal{I}$ -CPDAG

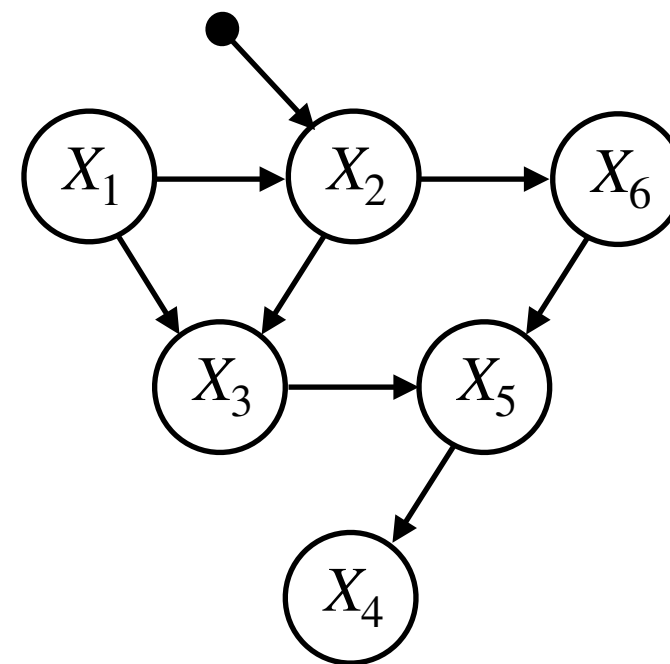
$\mathcal{I}$ -CPDAG: $\mathcal{I}$ -Completed Partially Directed Acyclic Graph



$a \rightarrow b$ strongly protected (Andersson *et al*, 1997)

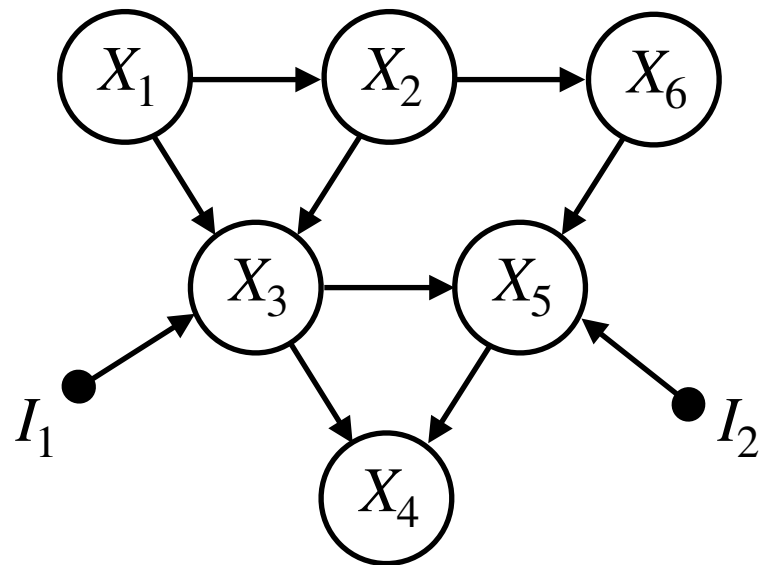


Initial $\mathcal{I}$ -DAG

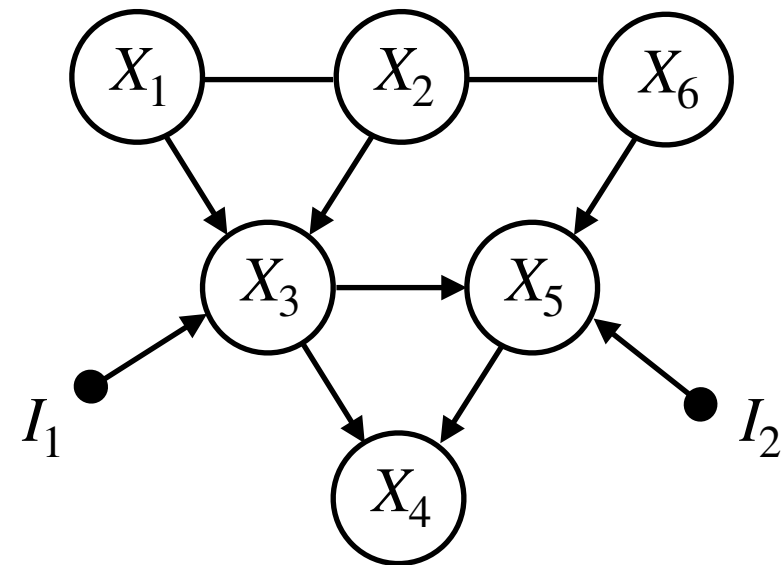


$\mathcal{I}$ -CPDAG

Likelihood of a $\mathcal{F}$ -CPDAG

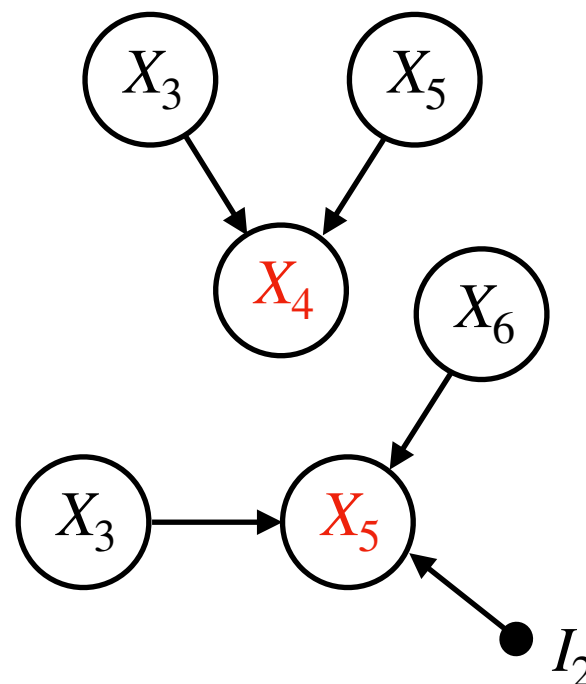
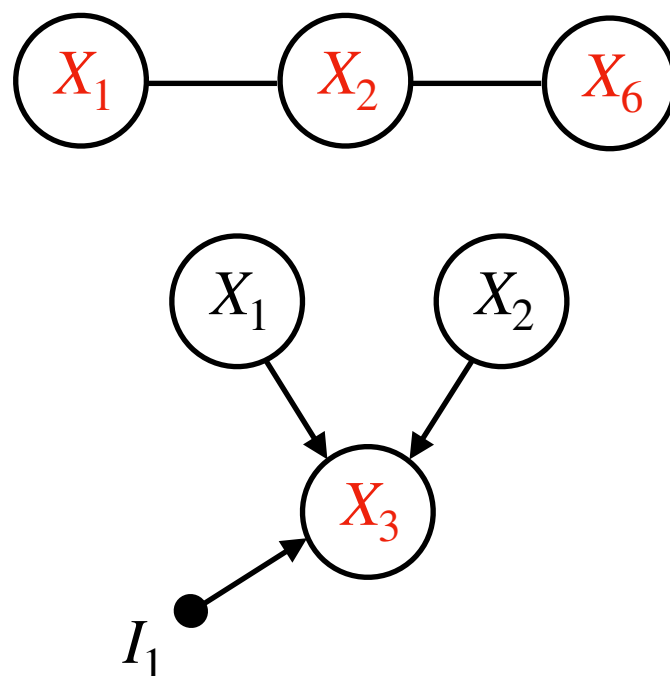


Initial $\mathcal{F}$ -DAG



$\mathcal{F}$ -CPDAG

Chain components



$$\mathbb{P}(X_1, X_2, X_6)$$

$$\mathbb{P}(X_3 | X_1, X_2, I_1)$$

$$\mathbb{P}(X_4 | X_3, X_5)$$

$$\mathbb{P}(X_5 | X_3, X_6, I_2)$$

Real life implementation

The problem: compute for all G

$$\mathbb{P}(G|\text{data}) \propto \exp \left(\log \mathbb{P}(G) + \text{loglik}(\hat{\theta}|G) - \text{pen}(G) \right)$$

A simple solution: by enumerating all DAGs.

p	number of DAGs
1	1
2	3
3	25
4	543
5	29,281
6	3,781,503
7	1,138,779,265
8	783,702,329,343

A better idea: through MCMC.

MCMC framework

The problem: sample DAG G from

$$\mathbb{P}(G|\text{data}) \propto \exp \left(\log \mathbb{P}(G) + \text{loglik}(\hat{\theta}|G) - \text{pen}(G) \right)$$

Metropolis-Hastings: perform iteratively

- propose $G' \sim q(\cdot|G)$
- accept G' with rate $\min(1, \alpha)$ with

$$\alpha = \frac{\mathbb{P}(G'|\text{data})}{\mathbb{P}(G|\text{data})} \times \frac{q(G|G')}{q(G'|G)}$$

MC output: a collection of N DAGs (default: $N = 5000$):

$$\underbrace{\text{DAG}_1, \text{DAG}_2, \dots, \text{DAG}_B}_{\text{burn-in (default: } B = 1000\text{)}}, \underbrace{\text{DAG}_{B+1}, \text{DAG}_{B+2}, \dots, \text{DAG}_N}_{\text{exploitable DAGs}}$$

empirical posterior:

$$\mathbb{P}(G = g|\text{data}) = \frac{1}{N - B} \sum_{i=B+1}^N \mathbb{1}_{\text{DAG}_i=g}$$

Proposals & Implementation

DAG space: MC3 (Madigan & Raftery, 1995)

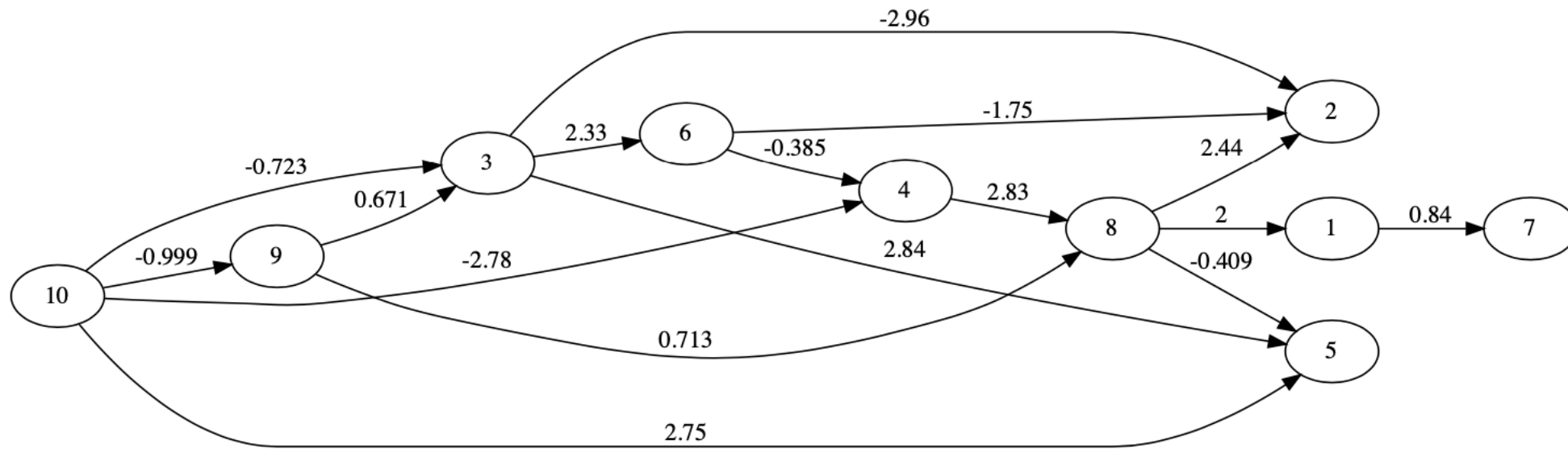
- Add/remove/flip arrow uniformly
- DAG constraint need smart update of route tables
- Available in `structmcmc` R package (Goudie, 2016)
- More constraints: max number of parents, fixed arrows

CPDAG space: He *et al* (2013), Castelletti *et al* (2018)

- Six moves: InsertU, DeleteU, InsertD, DeleteD, MakeV, RemoveV
- Plus one: ReverseD (Chickering 2002)
- Multiple theoretical conditions, asymmetric proposal
- https://github.com/FedeCastelletti/obayes_learn_essential_graphs

10 genes example

A random DAG with $p = 10$ genes



j	1	2	3	4	5	6	7	8	9	10
m	-0.61	-0.41	1.14	-1.84	1.00	0.71	-1.31	-0.96	0.06	0.70
σ	1.90	1.10	0.77	1.30	0.81	0.72	0.98	1.20	0.91	0.41

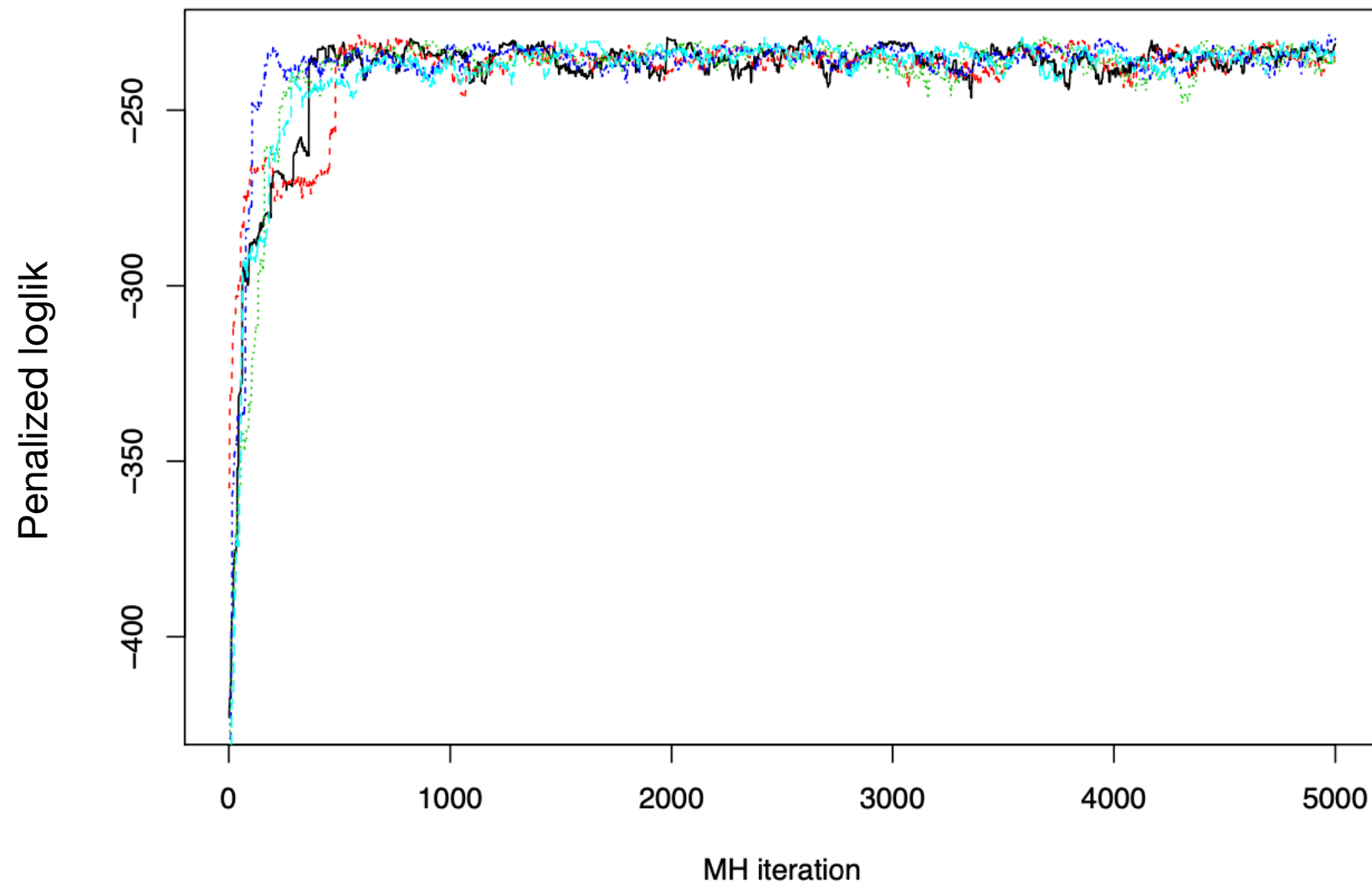
Some values (a causal ordering 10, 9, 3, 6, 4, 8, 2, 5, 1, 7):

$$\text{pa}(1) = \{8\} \quad \text{pa}(4) = \{6, 10\} \quad \text{pa}(10) = \emptyset$$

$$w_{6,2} = -1.75 \quad \ell_{6,2} = w_{6,2} + w_{6,4} \times w_{4,8} \times w_{8,2} = -4.41$$

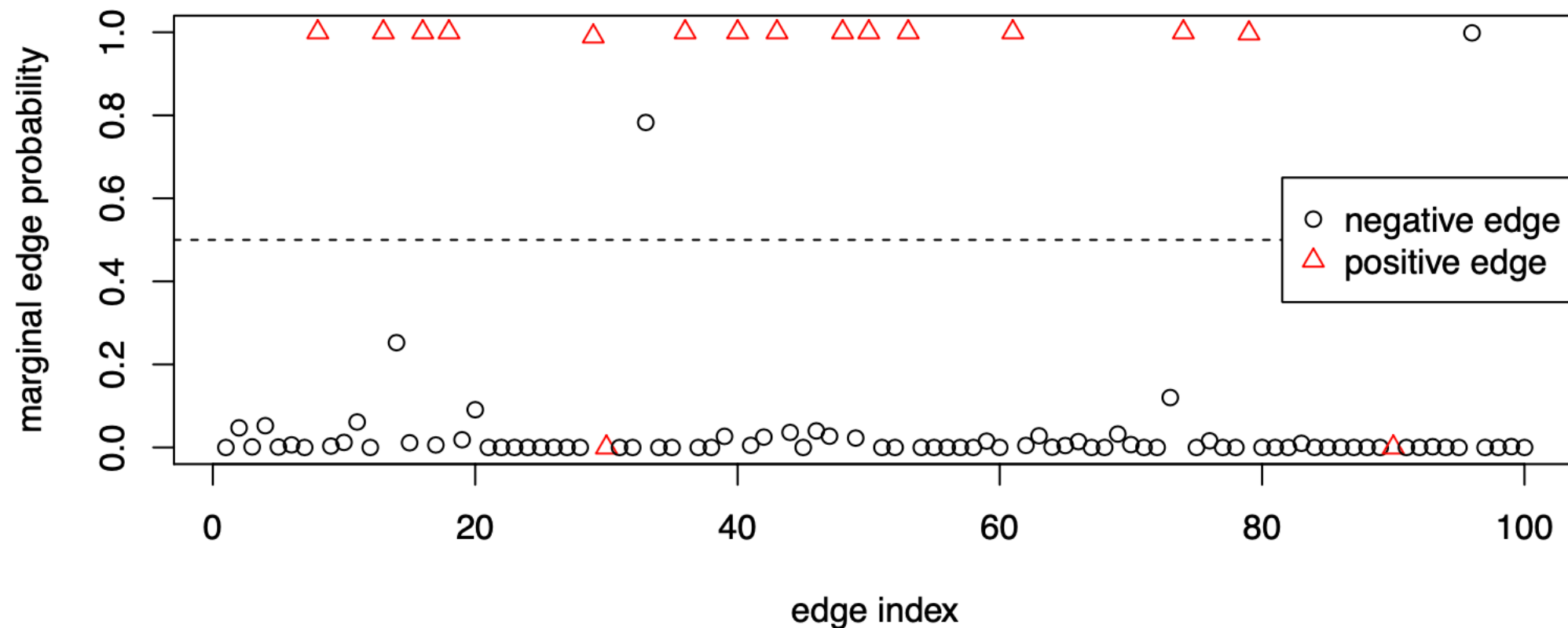
MCMC convergence

Design: fixed DAG and data (50 WT + 50 KO), 5000 MCMC iterations, unconstrained search, acceptance rate $\simeq 40\%$



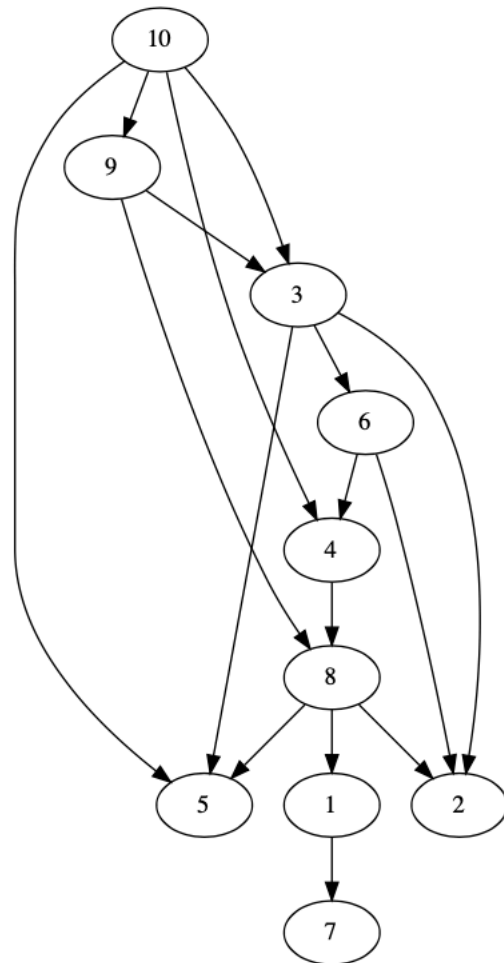
5 replications

Marginal edge probability

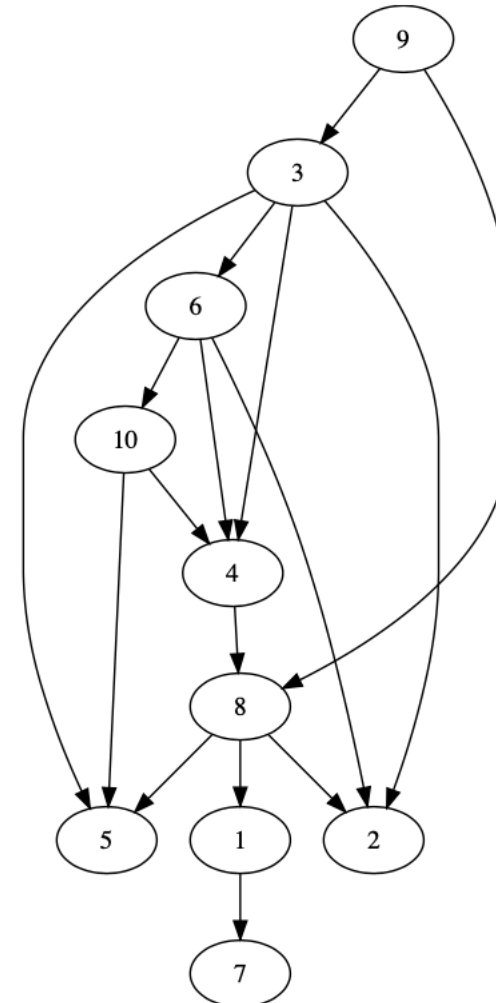


marginal posterior distribution on edges
threshold at 0.5 – AUROC=0.91 [0.78-1.00]

Consensus DAG



reference



consensus MC3

Direct effects

i	j	$w_{i,j}^*$	mean	sd
3	2	-2.96	-2.59	0.03
10	4	-2.78	-2.39	0.04
6	2	-1.75	-1.84	0.03
10	9	-1.00	0.00	0.00
10	3	-0.72	0.00	0.00
8	5	-0.41	-0.41	0.01
6	4	-0.39	-0.75	0.13
4	2	0.00	0.06	0.10
3	4	0.00	0.63	0.33
6	10	0.00	-0.06	0.00

i	j	$w_{i,j}^*$	mean	sd
9	3	0.67	0.82	0.08
9	8	0.71	0.62	0.07
1	7	0.84	0.84	0.00
8	1	2.00	2.03	0.02
3	6	2.33	2.32	0.01
8	2	2.44	2.42	0.03
10	5	2.75	2.93	0.02
4	8	2.83	2.67	0.01
3	5	2.84	2.82	0.03

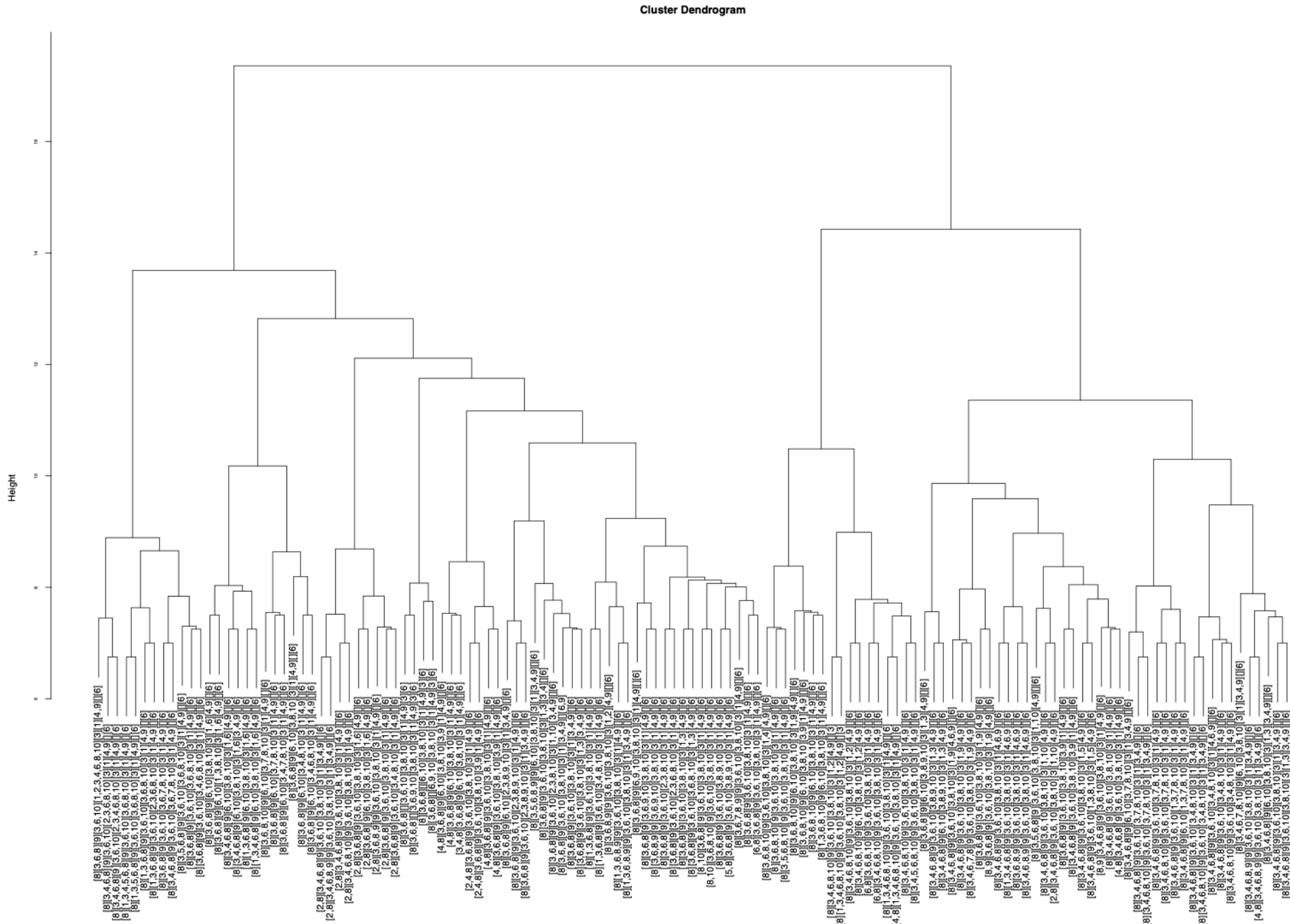
In/out degrees

i	$\text{in}_i^{g^*}$	mean	sd
1	1	1.12	0.36
2	3	3.44	0.67
3	2	0.99	0.10
4	2	2.81	0.43
5	3	3.16	0.42
6	1	1.02	0.12
7	1	1.09	0.29
8	2	2.13	0.34
9	1	0.01	0.10
10	0	1.00	0.05

i	$\text{out}_i^{g^*}$	mean	sd
1	1	1.07	0.25
2	0	0.08	0.27
3	3	3.94	0.61
4	1	1.34	0.52
5	0	0.02	0.13
6	2	3.07	0.27
7	0	0.03	0.18
8	3	3.00	0.00
9	2	2.11	0.35
10	4	2.11	0.32

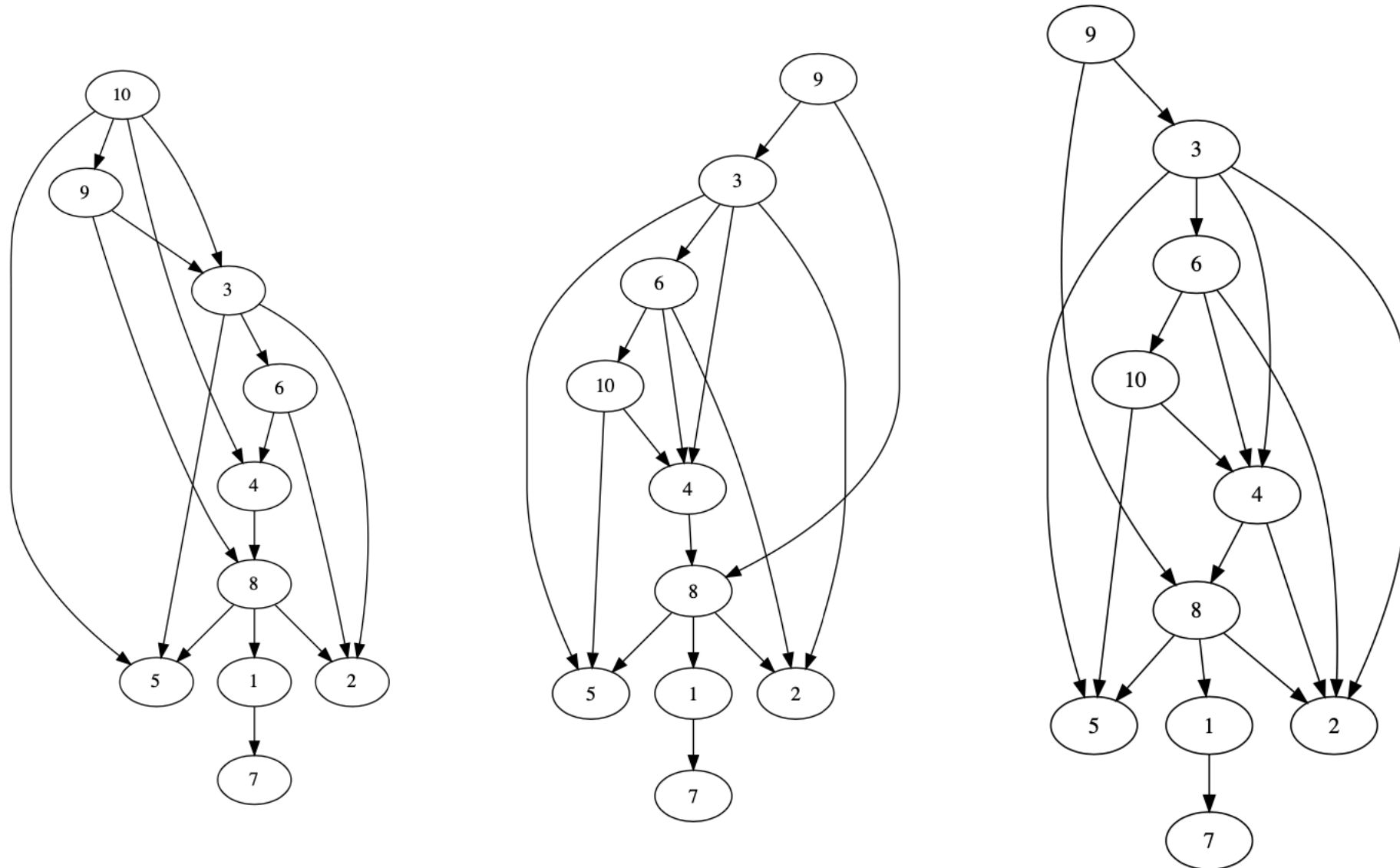
DAG clustering

Literature: Kendall tau and greedy centroids (Malmi, 2015).



Centroid DAG

Literature: Kendall tau and greedy centroids (Malmi, 2015).



reference

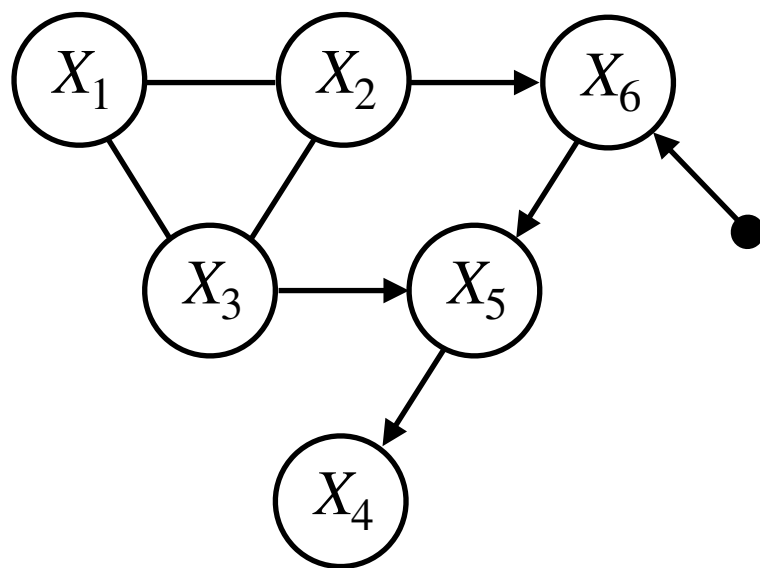
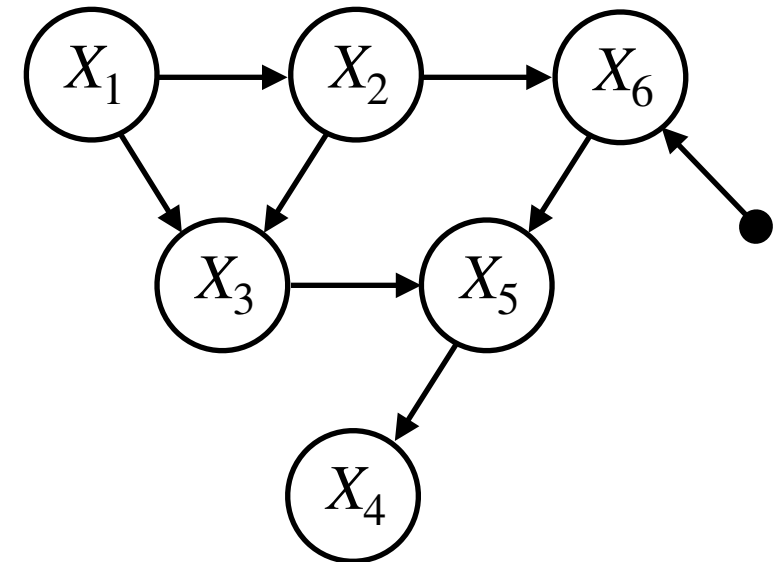
consensus

centroid

Conclusions & Perspectives

Take-home messages:

- Correlation is not causation
- CPDAGs = Markov equivalence class of DAGs
- Extension with interventions $\mathcal{I}$ -CPDAGs
- Relatively « simple » with intervention nodes
- MCMC over DAG or CPDAG spaces



What Next ?

- MCMC over CPDAG not trivial
- What to do with a collection of DAGs or CPDAGs ?
- Clinical trials: mixing observations and interventions ?
- Gene regulation networks: best interventions ?

Few references

- Andersson, S. A., Madigan, D., & Perlman, M. D. (1997). A characterization of Markov equivalence classes for acyclic digraphs. *The Annals of Statistics*, 25(2), 505-541.
- Castelletti, F., Consonni, G., Della Vedova, M. L., & Peluso, S. (2018). Learning Markov equivalence classes of directed acyclic graphs: an objective Bayes approach. *Bayesian Analysis*, 13(4), 1235-1260.
- Hauser, A., & Bühlmann, P. (2012). Characterization and greedy learning of interventional Markov equivalence classes of directed acyclic graphs. *The Journal of Machine Learning Research*, 13(1), 2409-2464.
- He, Y., Jia, J., & Yu, B. (2013). Reversible MCMC on Markov equivalence classes of sparse directed acyclic graphs. *The Annals of Statistics*, 41(4), 1742-1779.
- Yang, K., Katcoff, A., & Uhler, C. (2018, July). Characterizing and learning equivalence classes of causal DAGs under interventions. In *International Conference on Machine Learning* (pp. 5541-5550). PMLR.