Bayesian inference for a partial order from random linear extensions

Ometse S. Mogapi[1], Tom A.B. Snijders[2]
Geoff K. Nicholls[2], Bernard W. Silverman[2],
David Johnson[2] and Nicholas E. Karn[3]

University of [1]Botswana, [2]Oxford, [3]Southampton

Thanks R, Rgraphviz, snow, nem, and dependants. GKN ISBN/Valencia 7/6/10

Data (Karn & Johnson)

$\sim$3000 Royal Acta 1066-mid 12th C: witness lists

```
[1] William, Archbishop of Canterbury
[2] Roger, Bishop of Salisbury                                 (2)
[3] William, Giffard, bishop of Winchester, 1100-1129          (1)
[4] Bernard, Bishop of St David's                              (6)
[5] William, de Warelwast, bishop of Exeter                    (5)
[6] Urban, bishop of Llandaff                                  (8)
[7] Geoffrey, Rufus, Bishop of Durham
[8] Robert, de Sigillo, Bishop of London                       (10)*
   (Richard, de Belmeis I, bishop of London, 1108-1127)
[9] Miles, of Gloucester, earl of Hereford 1141-1143
[10] Henry, de Port
[11] Walter, de Amfreville
[12] William, de Folis
```

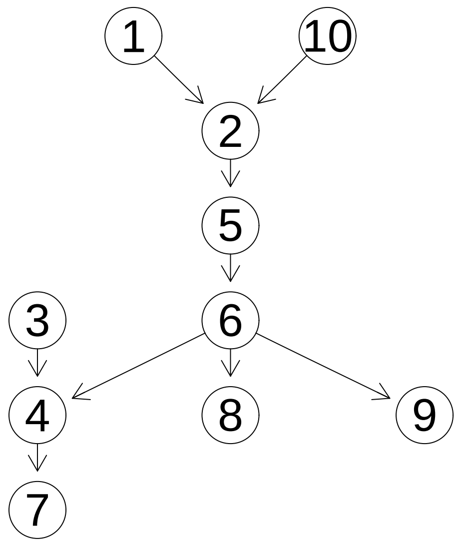
Covariates for Acta (not Bishops): where and when (date range) signed,...

Social hierarchy of bishops $\Leftrightarrow$ partial order on bishops.

Example (heavily and selectively thinned)

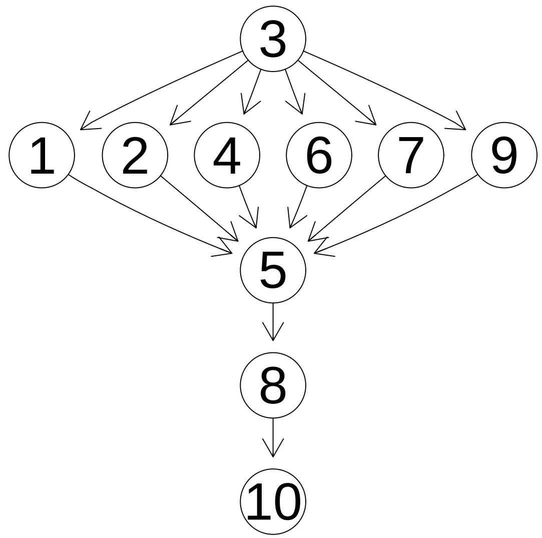
Witness lists (bishops of interest, 1119-1121)

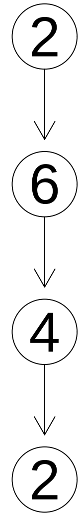
[1119]	5	6	4	7		
[1120]	3	4				
[1121]	1	2				
[1121]	10	1	2	5	6	8
[1121]	1	10	2	5	6	9
[1121]	1	2				
[1121]	10	1	2			



Witness list (bishops of interest, 1127-1129)

[1127]	9	10				
[1127]	2	9	10			
[1127]	2	1	6	5	8	10
[1127]	2-	-6				
[1127]	2	9	6	10		
[1127]	2	9	6	10		
[1129]	7	10				
[1129]	6-	7	-4	10		
[1129]	3	4<->2				

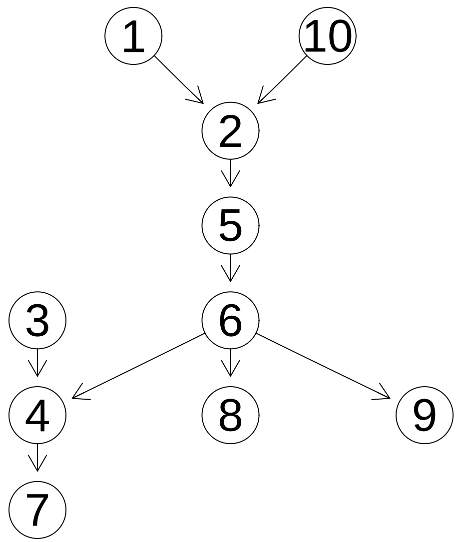




Example (heavily and selectively thinned)

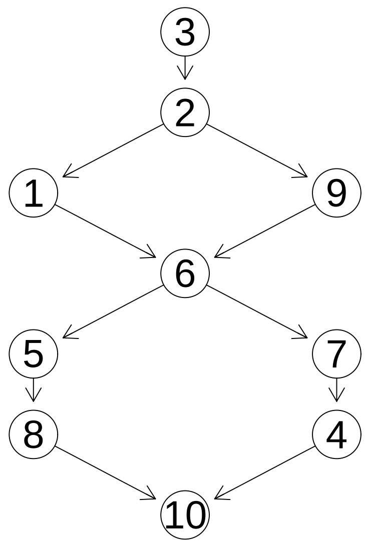
Witness lists (bishops of interest, 1119-1121)

```
[1119]    5    6    4    7
[1120]    3    4
[1121]    1    2
[1121]   10    1    2    5    6    8
[1121]    1   10    2    5    6    9
[1121]    1    2
[1121]   10    1    2
```



Witness list (bishops of interest, 1127-1129)

```
[1127]    9   10
[1127]    2    9   10
[1127]    2    1    6    5    8   10
[1127]    2-  -6
[1127]    2    9    6   10
[1127]    2    9    6   10
[1129]    7   10
[1129]    6-   7   -4   10
[1129]    3    2<->4
```



It's on my TODO list

- quantify and visualise uncertainty in the evolving hierarchy
- date undated lists, fit undated list between other dated lists
- two sets of lists (windowed on time): same underlying hierarchy or distinct?
- outliers, covariates (clerk, location, beneficiary)?

Now *that*, I did do

- ignore time-series aspect (window on time), covariates
- observation model Pr(lists|PO) queue jumping or late arrival outliers
- simple priors P(PO) (depth distribution, consistent priors)
- MCMC algorithms for Pr(PO,parameters|lists)
- summarise posterior dbns on POs *via* "consensus POs"

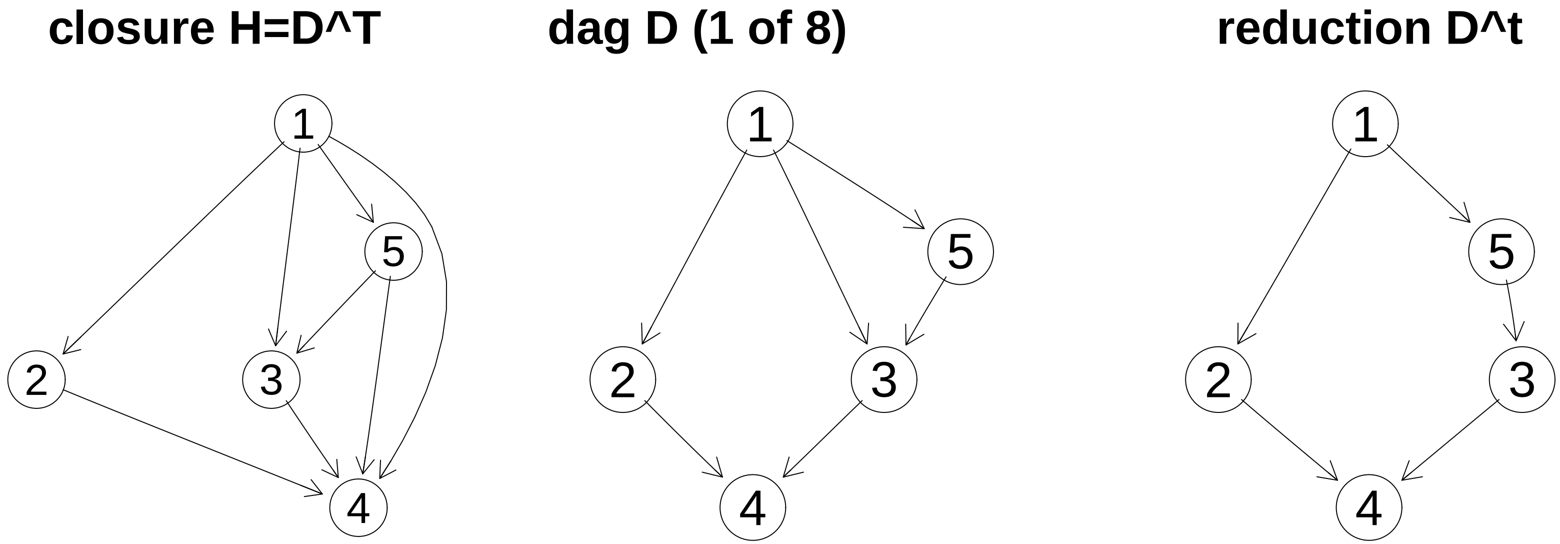
Berrenwinkel, Eriksson Sturmfels (2007), Berrenwinkel and Sullivant (2009)
MLE for PO from linear extension-related data in 'conjunctive Bayesian networks'.

Fröhlich, Beißbarth *et al.* (2007), (2009)
Minimise AIC-weighted partial orders (annealing, signaling-network data).

Manilla (2006) best bucket order (ordered random partition) for seriation;
Numata *et al.* (2008) parsimony PO.

Gupta and Damien (2002) ranking, conjugacy;
Halekoh & Vlach (2004), Puolamäki, Manilla *et al.* (2006); ranking, seriation;
Lerche & Sørensen (2003); ranking, visualisation

Brightwell (1993) 'Models for Random Partial Orders'
Three families of probability distributions for random PO's (properties, $n \to \infty$).



Observation: $\#\{$DAGs in same PO as D$\} = 2^{|D^T|-|D^t|}$ above $8 = 2^{8-5}$.

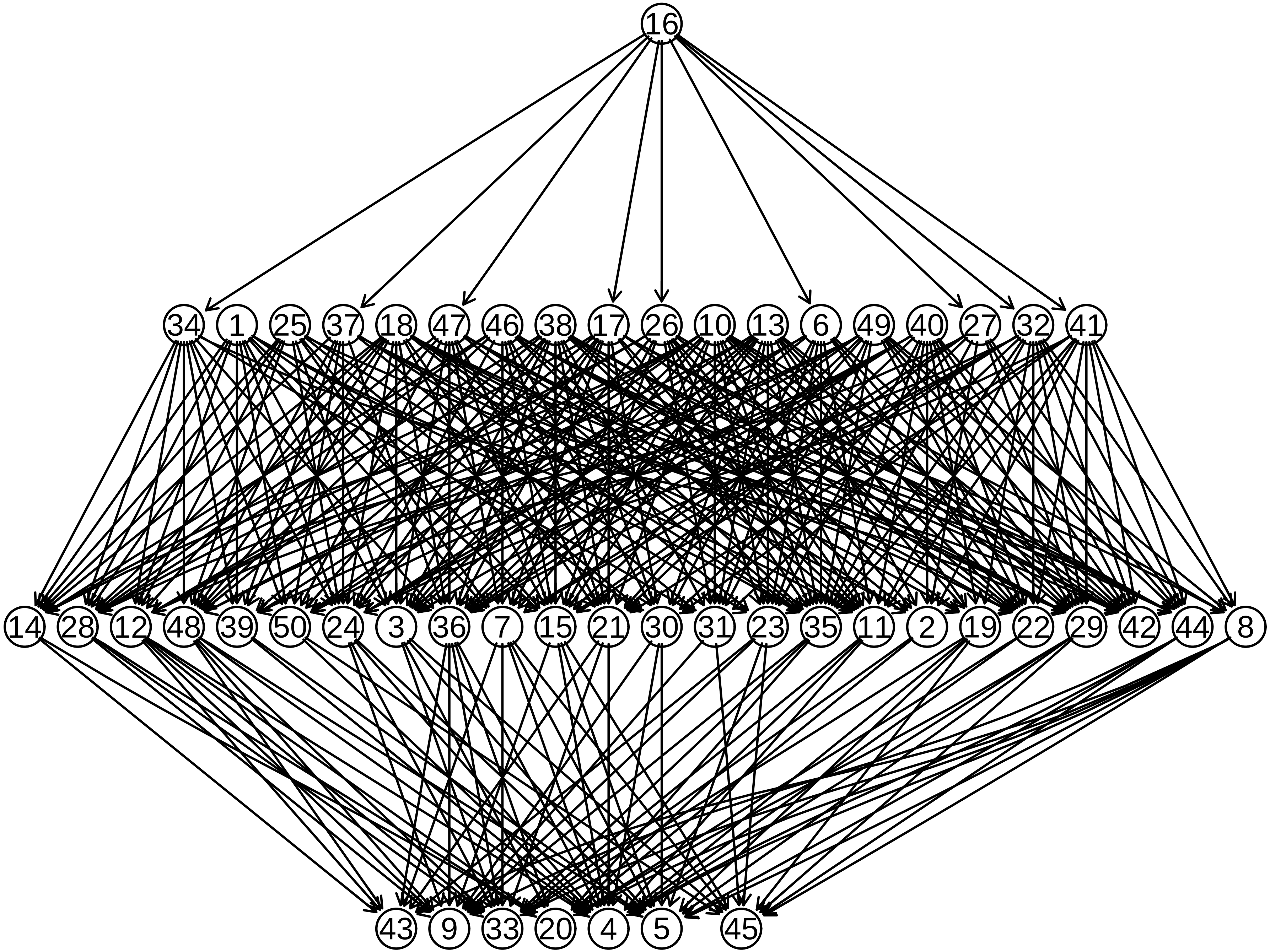
MCMC for $H \sim$ Uniform$(\mathcal{P}^T_{[n]})$. Suppose DAG $D_{(m)} = c$.

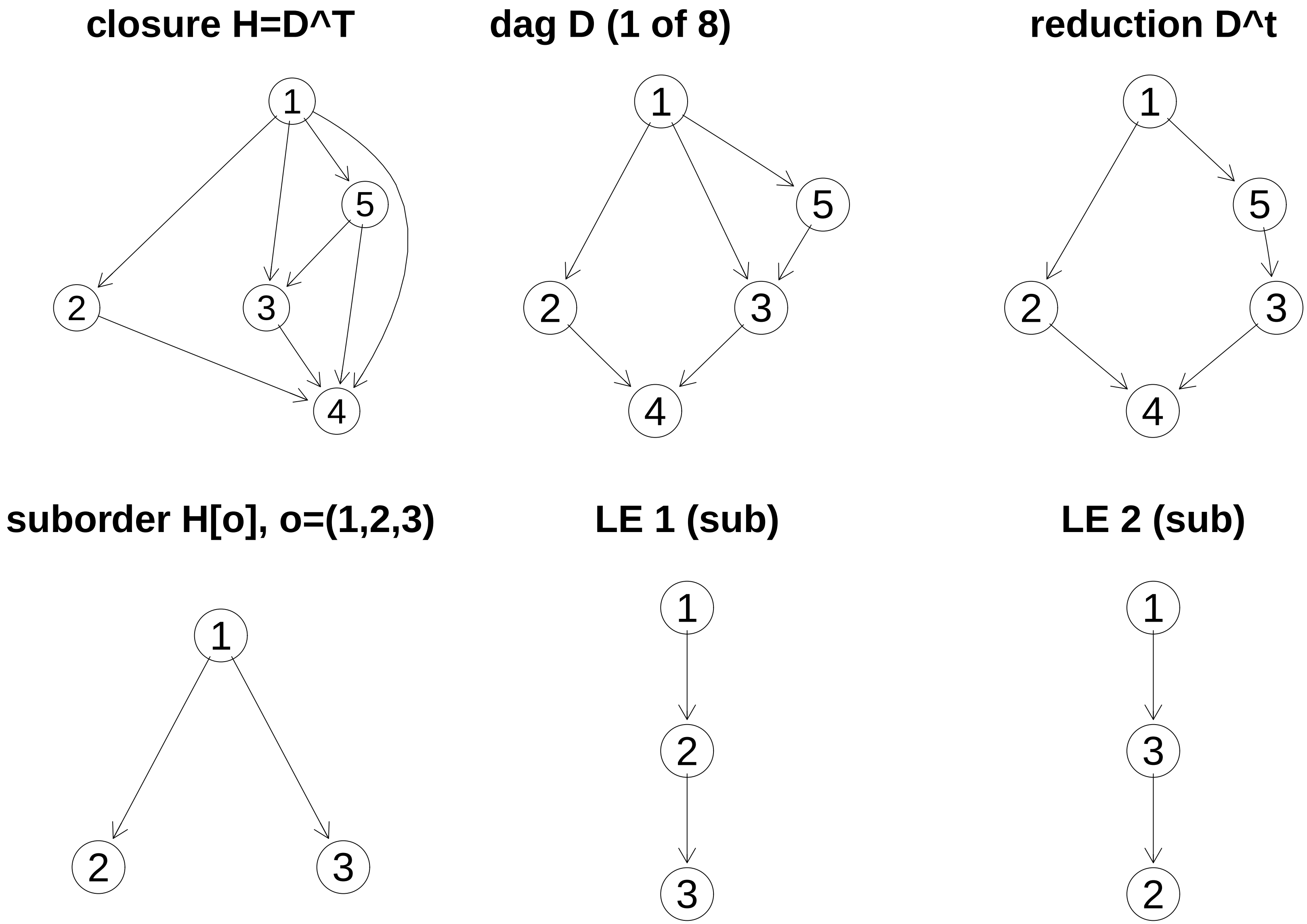
1. Set $d = c$. Pick a pair $\langle x, y \rangle$ and toggle it $d_{xy} = 1 - c_{xy}$.
2. Set $D_{(m+1)} = d$ w.p. $\alpha_D(d|c)$ and otherwise $D_{(m+1)} = c$.

$$\alpha_D(d|c) = \min\left(1, \frac{2^{|c^T|-|c^t|}}{2^{|d^T|-|d^t|}}\right) \qquad \text{and } \alpha = 0 \text{ if } d \text{ cyclic}$$

Then $D^T_{(m)}$ is uniform in $\mathcal{P}^T_{[n]}$.

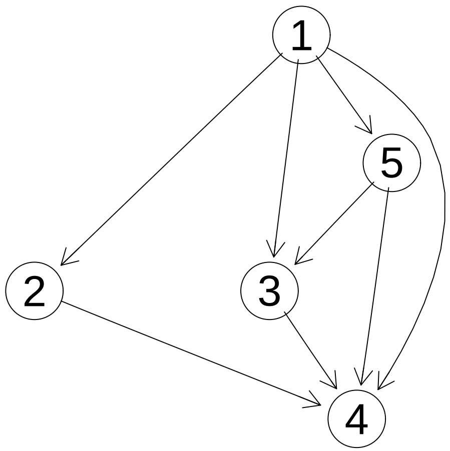
Aho *et al.* (1972) reduction unique, La Poutre *et al.* (1988) efficient $(D \pm \langle x, y \rangle)^t$.


16
34
1
25
37
18
47
46
38
17
26
10
13
6
49
40
27
32
41
14
28
12
48
39
50
24
3
36
7
15
21
30
31
23
35
11
2
19
22
29
42
44
8
43
9
33
20
4
5
45

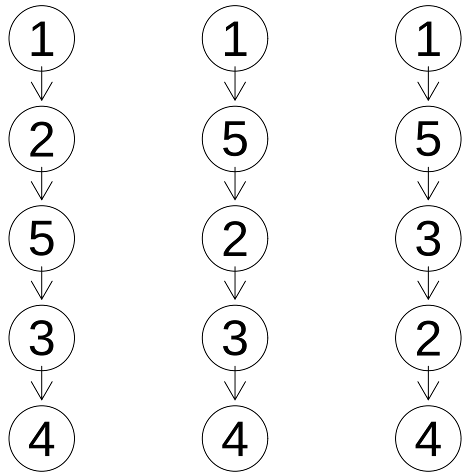


R nem-package (for D^t, D^T) and RGraphviz (for drawing)

closure H=D^T



LE 1 LE 2 LE 3



Dimension (of a PO)

H $=$ LE 1 $\cap$ LE 3 so we have $\dim(H) = 2$

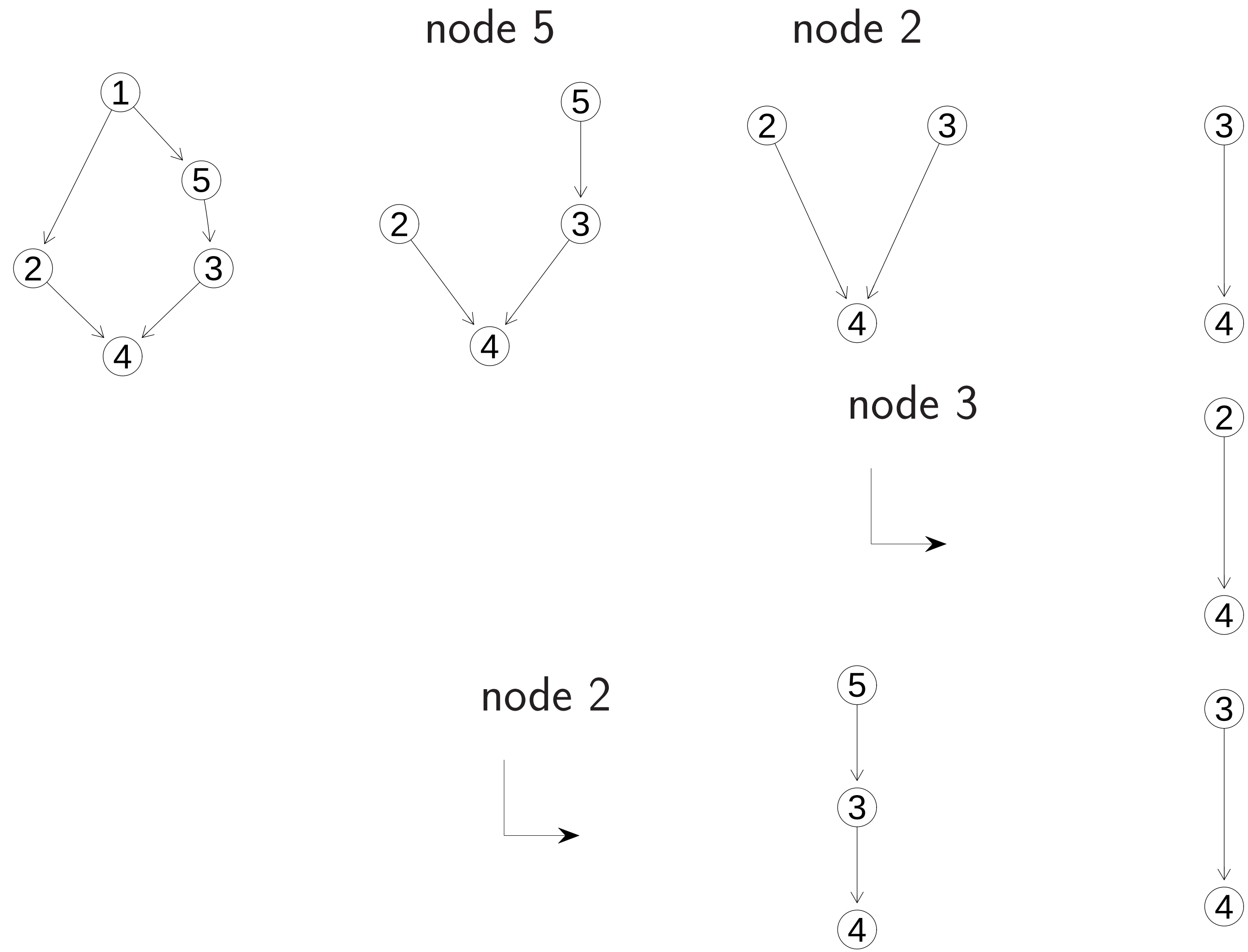
Depth and Width

$d(H) = 4$, $w(H) = 2$.

$\dim(H) \leq w(H) \leq n$ (if $n < \infty$)

Counting linear extensions

$$C(H) = \sum_{i:d(i)=1} C_i(H) \quad \text{where} \quad C_i(H) = H[-i] \qquad \text{Edelman } et\ al \text{ (1989)}$$



so $C(H) = 3 \quad C_1(H) = 3, \quad C_5(H[-1]) = 2.$

Observation process in act i

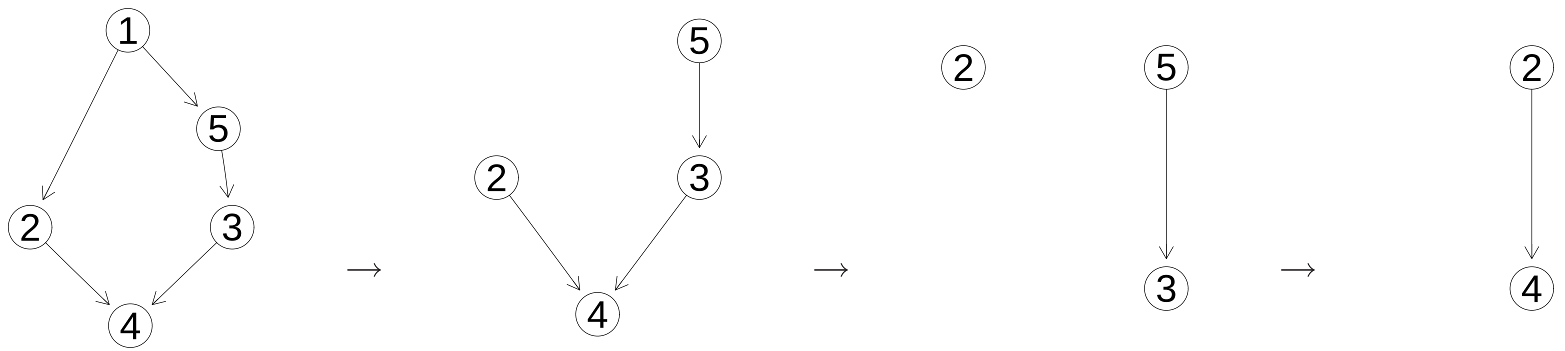
n_i witnesses labels $o_i = \{y_{i,1}, y_{i,2}, \dots, y_{i,n_i}\}$, $o_i \subseteq [n]$

Data $y_i = (y_{i,1}, y_{i,2}, \dots, y_{i,n_i})$ respects $H[o_i]$

$$L(h,p;y_i) = \Pr(Y_i = y_i | H = h, O = o_i, P = p)$$

$$= \begin{cases} 1/C(H) & \text{no errors, } p = 0, \\ \prod_{j=1}^{n_i-1} \left(\frac{p}{n_i - j} + (1-p)\frac{C_j(h[y_{j:n_i}])}{C(h[y_{(j:n_i)}])} \right) & \text{queue jumping, } 0 < p < 1. \end{cases}$$

$$1 \to 4 \to 2 \to 5 \to 3$$



$$L(h,p;y_i) = \left(\frac{p}{5} + (1-p)\right)\left(\frac{p}{4} + (1-p)\frac{0}{3}\right)\left(\frac{p}{3} + (1-p)\frac{1}{3}\right)\left(\frac{p}{2} + (1-p)\frac{1}{1}\right)$$

Prior distributions for partial orders (I)

$Uniform(\mathcal{P}_{[n]}^T)$:

almost surely depth 3 (and width about $n/2$), Klietman et al 1975

is not consistent for marginalisation (we observe on suborders).

$H \sim$ K-dim(k, β) with $k = n$:

from Winkler (1985) random K-dimensional orders

adds correlation parameter β, controls dbn of $d(H)$

(remains) consistent for marginalisation

$$\text{for all } 0 < m < n, \quad \sum_{h'[1:m]=h} \pi_{H,n}(h') = \pi_{H,m}(h)$$

TPO(α, θ): random order on Pitman (1995) random partition

random total pre-order (BWS),

consistent for marginalisation

Prior distributions for partial orders (II)

Bishops latent attributes wealth, time in office, weight.

$A > B$ if A beats B on all attributes.

Random K-dimensional orders

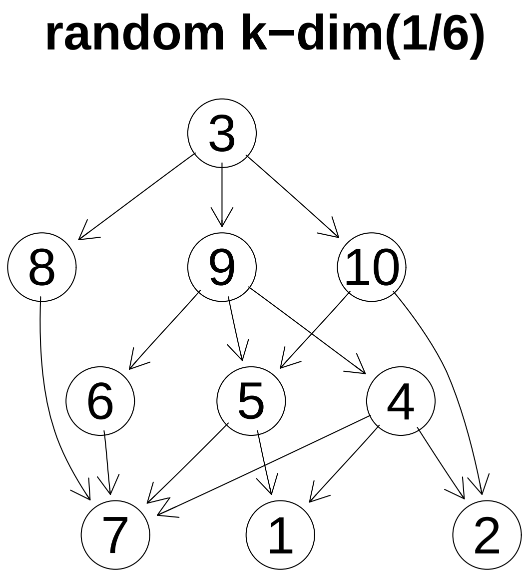
Bishop $i = 1, 2, ...n$ latent attributes $Z_i \sim \text{Uniform}(0,1)^K$

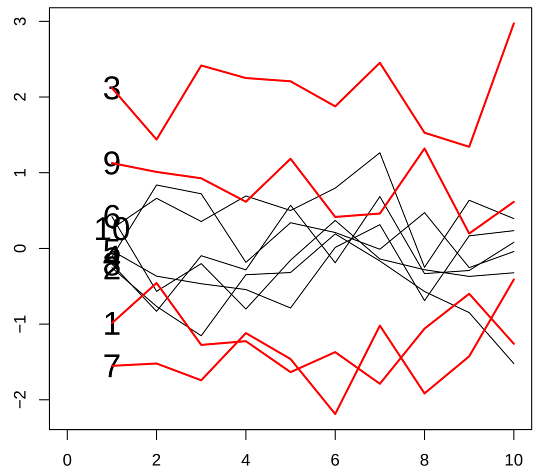
$$H : \{x \to y \text{ if } Z_{x,j} > Z_{y,j} \text{ on all attributes } j = 1, 2, ..., K\}$$

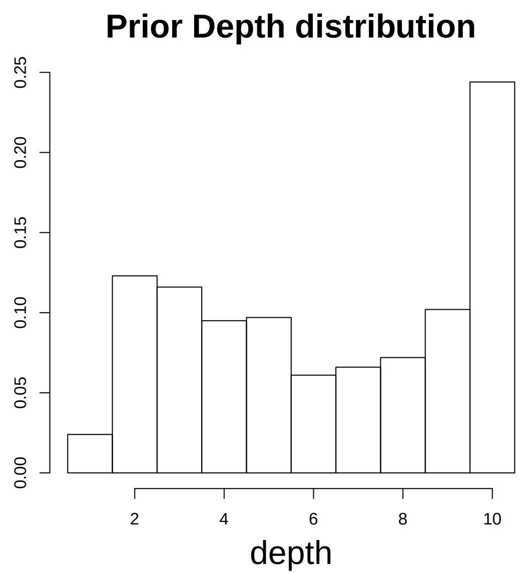
Dimension at most K, very shallow PO's.

Choose $K = n$ and correlate $Z_i \sim N(0, R)$, $R_{jj} = 1$, $R_{ij} = \rho$

(I took $\rho \sim \text{Beta}(1, 1/6)$ for flat-ish prior depth dbn)







Sample Results

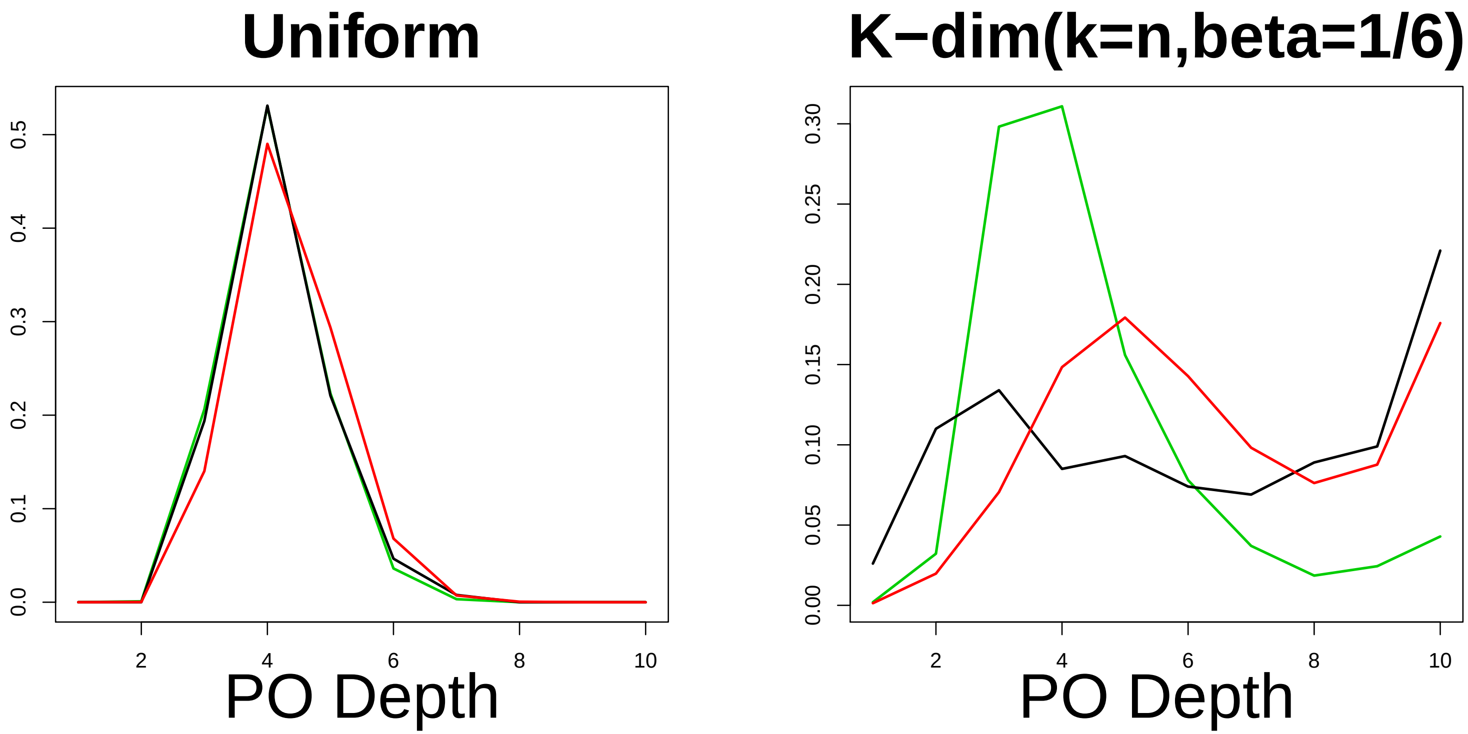
Priors

$H \sim$ K-dim(n, β) and $\rho \sim$ Beta$(1, \beta = 1/6)$ in posterior $\pi(h, \rho, p|y)$

$H \sim Uniform(\mathcal{P}^{T}_{[n]})$ in posterior $\pi(h, p|y)$.

Observation Models

Q-jumping allowed (with prior $p \sim Beta(1/2, 2)$) at each signature.



Prior (black) 1119-1121 (red) 1127-1129 (green)

Conclusion

- have ignored time-series aspect (window on time), covariates

- observation model Pr(lists|PO) with queue jumping

- simple priors P(PO) (depth distribution, consistent priors)

- MCMC algorithms for Pr(PO,parameters|lists)

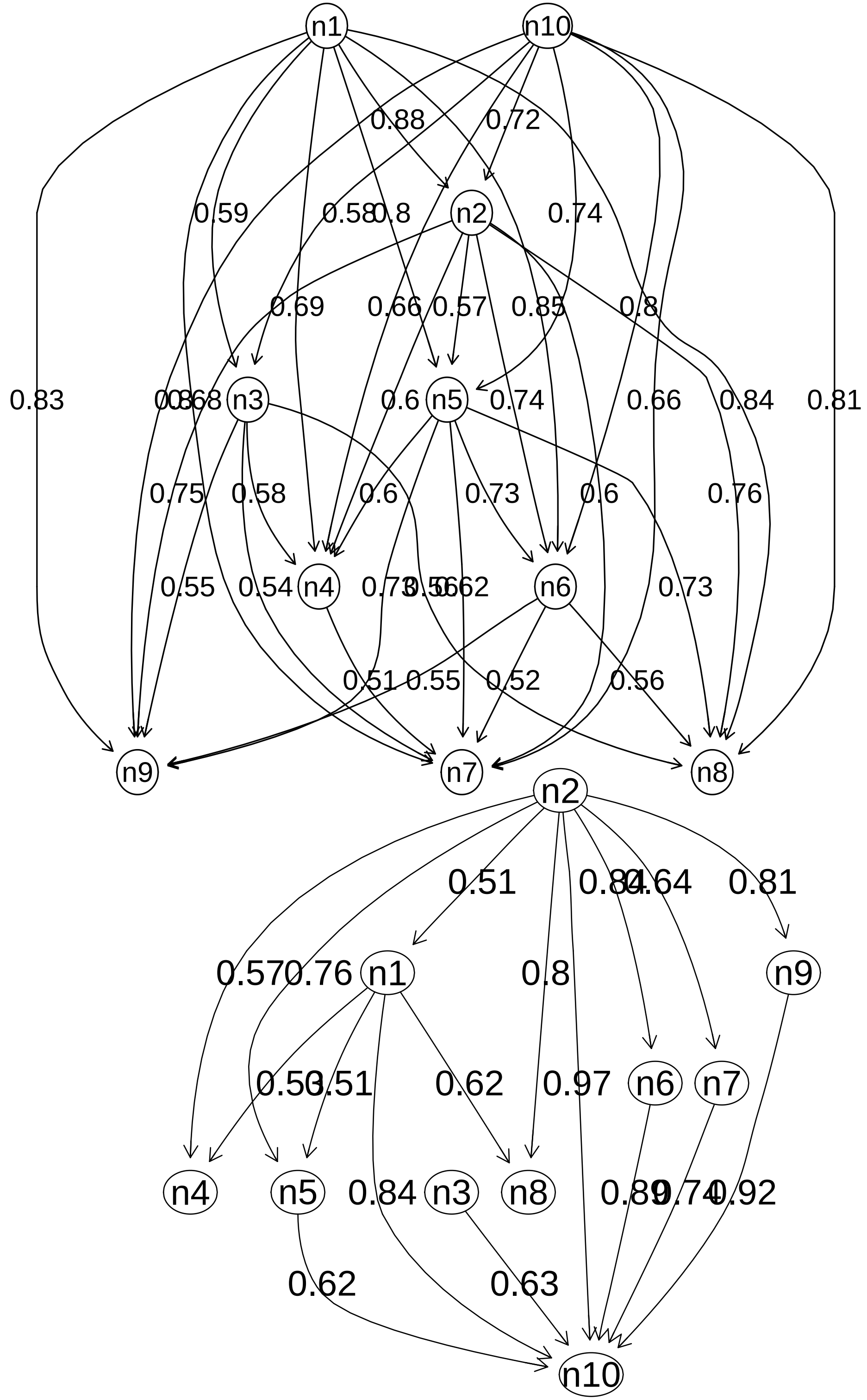
- summarise posterior dbns on POs *via* “consensus POs”

Witness lists (bishops of interest, 1119-1121)

[1119]	5	6	4	7		
[1120]	3	4				
[1121]	1	2				
[1121]	10	1	2	5	6	8
[1121]	1	10	2	5	6	9
[1121]	1	2				
[1121]	10	1	2			

Witness list (bishops of interest, 1127-1129)

[1127]	9	10				
[1127]	2	9	10			
[1127]	2	1	6	5	8	10
[1127]	2	6				
[1127]	2	9	6	10		
[1127]	2	9	6	10		
[1129]	7	10				
[1129]	6	7	4	10		
[1129]	3	4	2			



Likelihood evaluation and generically parallelising a single serial MCMC chain

$L(h; y_i) = 1/|\mathcal{L}(H \downarrow o_i)|$

For $H \in \mathcal{P}^T_{[n]}$, and $V \subset [n]$, let $e_V = |\mathcal{L}(H \downarrow V)|$

Let $U(H)$ be the set of possible 'top' nodes in H.

$$e_{[n]} = \sum_{v \in U(H)} e_{[n]\setminus v}$$

so count LE's recursively. Still time consuming.

Can parallelise accept reject: $D^{(m)} = c$ and k processors.

Let $(d^{(i)}, A^{(i)})$ be candidate (from c) on proc $i = 1, 2, ..., k$ and and 0-reject/1-accept.

1. Compute $(d^{(1)}, A^{(1)}), (d^{(2)}, A^{(2)}), ..., (d^{(k)}, A^{(k)})$ in parallel.
2. Let $j = \min\{i : A^{(i)} = 1, i = 1, 2, ..., k\} \cup \{k + 1\}$.

Set $D^{(m+i)} = c$ for $i = 1, 2, ..., j - 1$. If $j < k + 1$ set $D^{(m+j)} = d^{(j)}$.

`http://www.stats.ox.ac.uk/~nicholls`

for example (in R) and preprint (geophysical app).

A. Sohn, "Parallel N-ary Speculative Computation of Simulated Annealing" IEEE Trans. Parallel Distrib. Syst.,6,997-1005 (1995), and earlier.

Easy error takes $H^{(m)} = g$, $g \in \mathcal{P}_n^T$

1. $(x, y) \sim$ Uniform$([n] \times [n])$ and if $(x, y) \notin g$, and $g + (x, y)$ acyclic, $h = (g + (x, y))^T$ and if $(x, y) \in g^t$ then $h = (g^t - (x, y))^T$.

2. Compute α etc

Not reversible as $g = \{(1, 3), (2, 3)\}$ and $(x, y) = (1, 2)$ since $h = \{(1, 2), (2, 3), (1, 3)\}$ and $(h^t - (x, y))^T = \{(2, 3)\} \neq g$.