



Bayesian inference over model structure

Nicos Angelopoulos

`nicos.angelopoulos@imperial.ac.uk`

Imperial College

London, UK

work done with James Cussens

overview



- basic syntax of DLPs
- Bayesian learning of model structure
- learning classification and regression trees
- learning Bayesian networks
- the bims library

note on nomenclature



DLPs Distributional logic programs

Bims Bayesian inference of model structure

DLPs - description

We extend LP's clausal syntax with probabilistic guards that associate a resolution step using a particular clause with a probability whose value is computed on-the-fly.

The intuition is that this value can be used as the probability with which the clause is selected for resolution.

Thus in addition to the logical relation a clause defines over the objects that appear as arguments in its head, it also defines a probability distribution over aspects of this relation.

DLPs - influences



SLPs labels on clauses

PCCP computed on-the-fly guards

DLPs example

member(*H*, [*H* | _]).

member(*El*, [*H* | *T*]) :- (C₁)

member(*El*, *T*).

L :: *length*(*List*, *L*) ~ *El* :: *umember*(*El*, *List*)

$\frac{1}{L}$:: *L* :: *umember*(*El*, [*El* | *Tail*]). (C₂)

$1 - \frac{1}{L}$:: *L* :: *umember*(*El*, [*H* | *Tail*]) :- (C₃)
umember(*El*, *Tail*).

DLPs abbreviated example



$$\frac{1}{L} :: \text{length}(\text{List}, L) :: \text{umember}(El, [El|Tail]). \quad (C_4)$$

$$1 - \frac{1}{L} :: L :: \text{umember}(El, [H|Tail]) :- \quad (C_5)$$

$$K \text{ is } L - 1,$$

$$K :: \text{umember}(El, Tail).$$

DLPs query



? – $umember(X, [a, b, c])$.

$X = a$ (1/3 of the times = 1/3)

$X = b$ (1/3 of the times = 2/3 * 1/2)

$X = c$ (1/3 of the times = 2/3 * 1/2 * 1)

Bims



MCMC over single program defined model spaces

- classification trees
- structure of Bayesian networks

model averaging inference

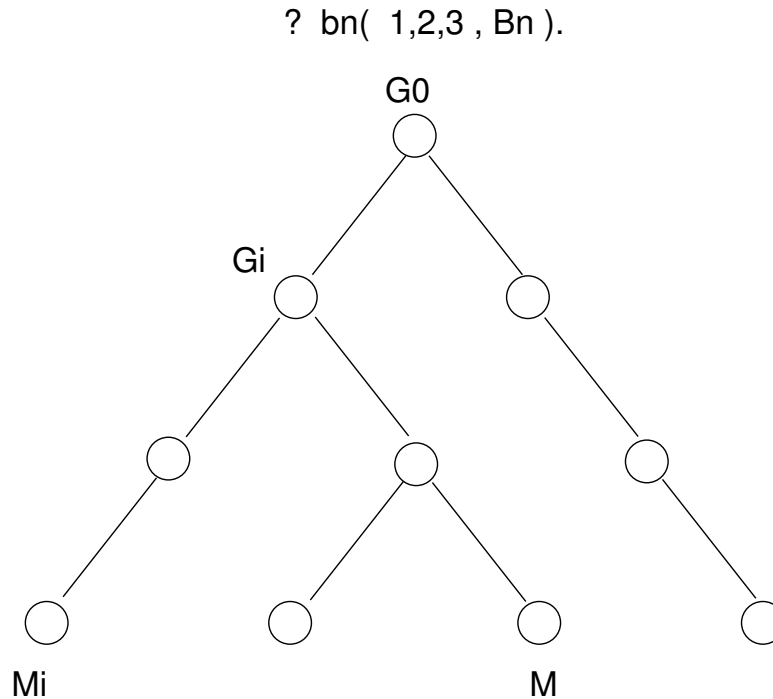
Bayes' Theorem

$$p(M|D) = \frac{p(D|M)p(M)}{\sum_M p(D|M)p(M)}$$

Metropolis-Hastings

$$\alpha(M_i, M_*) = \min \left\{ \frac{q(M_*, M_i)P(D|M_*)P(M_*)}{q(M_i, M_*)P(D|M_i)P(M_i)}, 1 \right\}$$

DLP defined model space



From M_i identify G_i then sample forward to M_\star .

$q(M_i, M_\star)$ is the probability of proposing M_\star when M_i is the current model.

CART prior

$$P_{\text{split}}(\eta) = \alpha(1 + d_\eta)^{-\beta}$$

(C₀) $\text{cart}(\zeta, \xi, M, \text{Cart}) : -$

ψ_0 is ζ ,

ψ_0 : $\text{split}(0, \zeta, \xi, M, \text{Cart})$.

(C₁) ψ_D : $\text{split}(D, \zeta, \xi, M_B, \text{nd}(F, \text{Val}, L, R)) : -$

ψ_{D+1} is $\zeta * (1 + D)^{-\xi}$,

D_1 is $D + 1$,

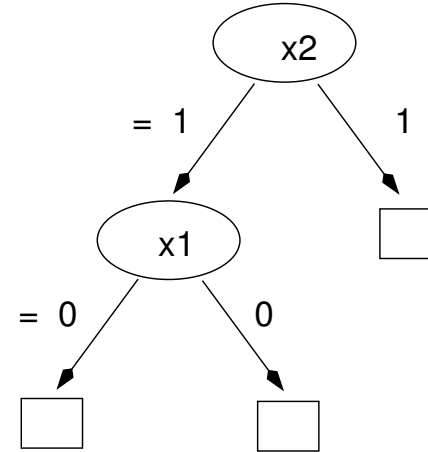
$r_select(F, \text{Val}, M_B, L_B, R_B)$,

ψ_{D+1} : $\text{split}(D_1, \zeta, \xi, L_B, L)$,

ψ_{D+1} : $\text{split}(D_1, \zeta, \xi, R_B, R)$.

(C₂) $1 - \psi_D$: $\text{split}(D, \zeta, \xi, M_B, \text{lf})$.

?- $\text{cart}(\zeta, \xi, A, M)$.



$M = \text{nd}(x2, 1, \text{nd}(x1, 0, \text{lf}, \text{lf}), \text{lf})$

virtual screening



objective

improve chances of discovering binding molecules based on examples from screened chemical libraries.

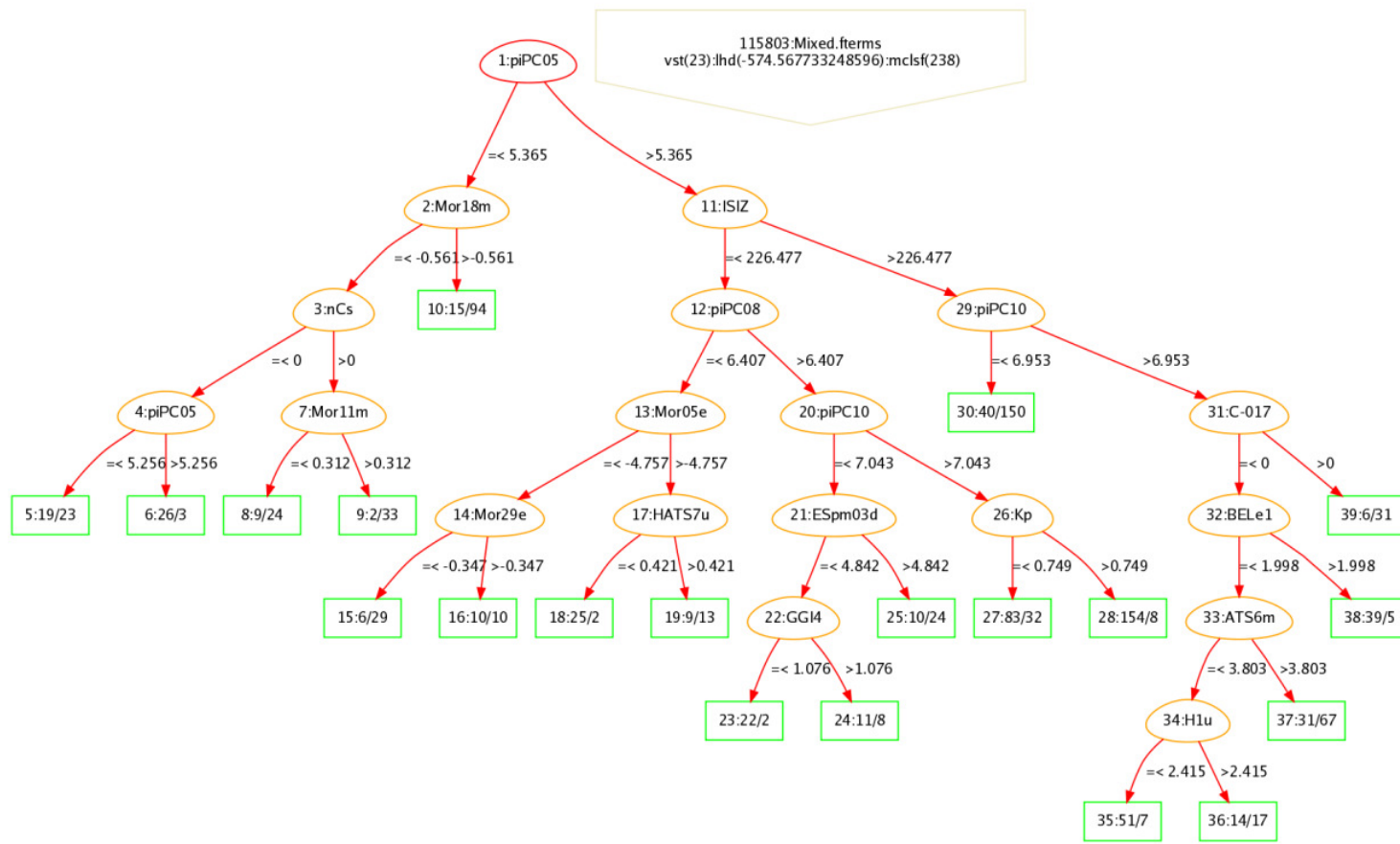
pyruvate kinase affinity data

582 active and 582 inactive molecules with 1100 descriptors

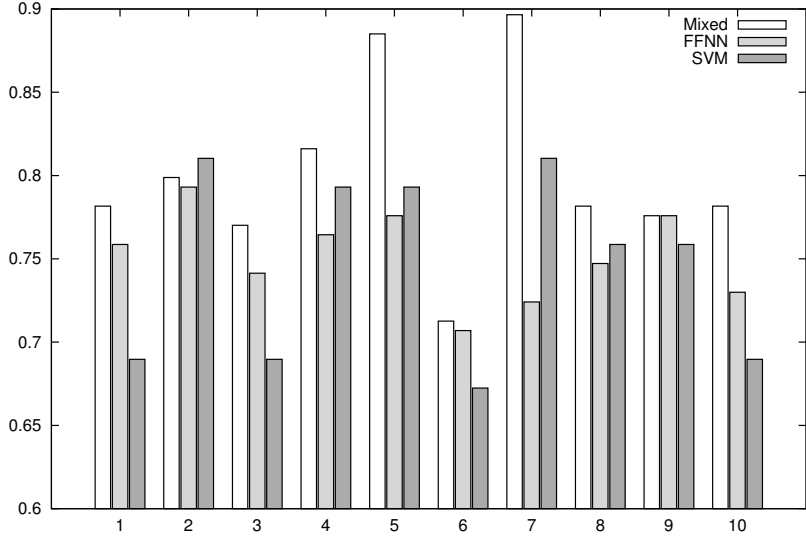
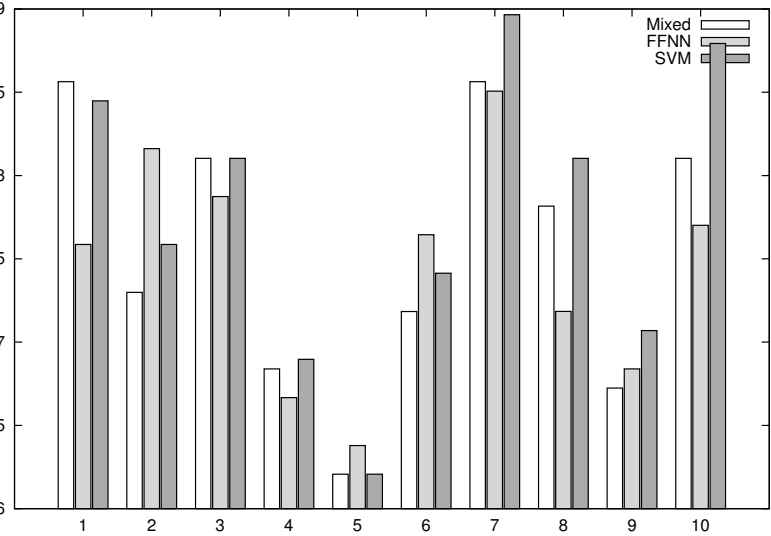
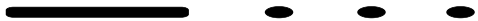
ten fold cross-validation

compared to Feed Forward Neural Networks and SVMs

highest likelihood model



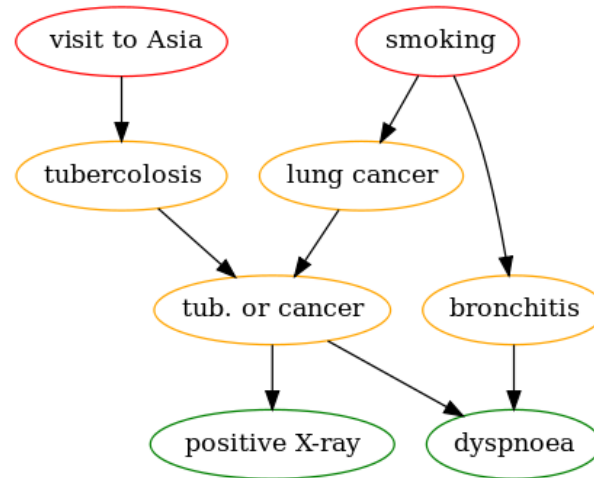
ten fold cross validation



$$Sensitivity = \frac{T^+}{T^+ + F^-}$$

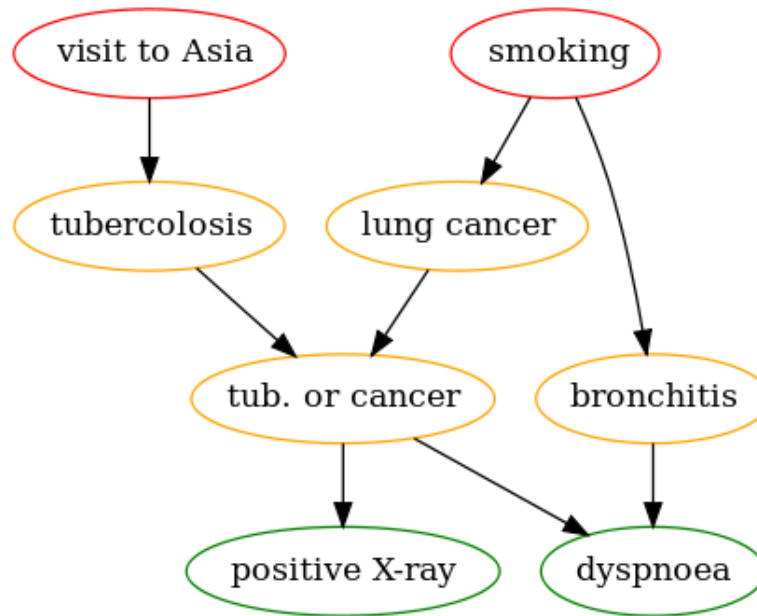
$$Specificity = \frac{T^-}{T^- + F^+}$$

Bayes nets



- directed acyclic graphs (DAGs)
- nodes are random variables
- directed edges identify all direct influences to a child
- each node has an associated conditional probability table

The Asia Bayes net



$$P(A, T, E, L, S, B, D, X) =$$
$$P(X|E)P(D|E, B)P(E|T, L)P(B|S)P(L|S)P(S)P(T|A)P(A)$$

BN prior

```
bn( [], _Nodes, [] ).
bn( [Nd|Nds], PotPa, [Nd-[]|TBn] ) :-
    orphan( Nd ),
    !,
    bn( Nds, [Nd|PotPa], TBn ).
bn( [Nd|Nds], PotPa, [Nd-Pa|TBn] ) :-
    length( PotPa, Len ),
    parents_limit( Lim ),
    Gaps is min( Len, Lim ) + 1,
    [Gaps] : choose_pa( PotPa, Pa ),
    bn( Nds, [Nd|PotPa], TBn ).

1/G : [G] : choose_pa( _PotPa, [] ).
1 - 1/G : [G] : choose_pa( PotPa, [Pa|TPa] ) :-
    pselect( PotPa, Pa, RemPa ),
    NxtG is G - 1,
    [NxtG] : choose_pa( RemPa, TPa ).
```

Bayesian networks



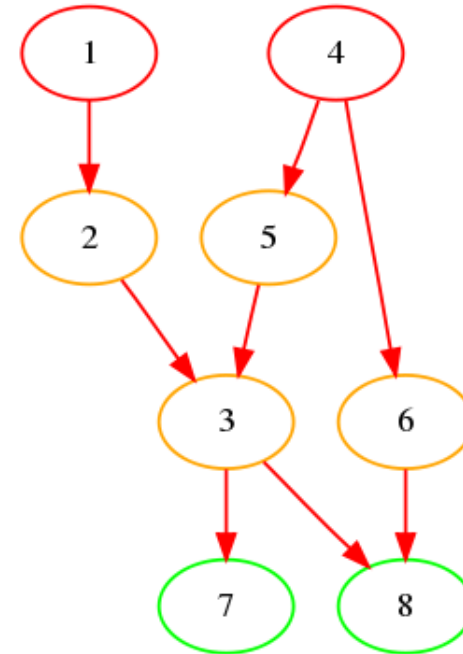
ordering

absence/presence of specific arcs

preference distributions over
parental group sizes

orphans

independence relation



Bayesian networks



BN	Nodes	Arcs	Ref
ASIA	8	8	Lauritzen and Spiegelhalter (1988)
ALARM	37	46	Beinlich et al. (1989)
HAILFINDER	56	66	Abramson et al. (1996)
INSURANCE	27	52	Binder et al. (1997)
PARITY1	22	31	Koivisto and Sood (2004)
PARITY2	100	53	Koivisto and Sood (2004)

Priors and backtracking combinations

Comb	Prior	backtracking
1	1/2 on each poss. edge	all
4	$\propto Pa(x) $	parents
5	$\propto Pa(x) $	par. no orphans
8	$= Pa(x) $	parents
9	$= Pa(x) + \text{indep}$	single component

Averaged results

ASIA									
	ID			NE			CK		
1	1.00	1.00	1.00	0.1215	0.1252	0.1286	0.1474	0.1500	0.1539
4	1.00	1.00	1.00	0.0934	0.1221	0.0933	0.1396	0.1552	0.1399
5	0.99	0.99	0.99	0.0433	0.0423	0.0432	0.0654	0.0642	0.0649
8	0.16	0.16	0.16	0.0052	0.0051	0.0050	0.0053	0.0051	0.0051
9	0.16	0.16	0.16	0.0051	0.0052	0.0051	0.0051	0.0052	0.0052

HAILFINDER									
	ID			NE			CK		
1	1.00	1.00	1.00	0.2401	0.2384	0.2300	0.2433	0.2416	0.2326
4	1.00	1.00	1.00	0.0271	0.0251	0.0235	0.0290	0.0273	0.0259
5	1.00	1.00	1.00	0.0110	0.0115	0.0113	0.0143	0.0149	0.0145
8	1.00	1.00	1.00	0.0123	0.0115	0.0115	0.0151	0.0147	0.0148
9	1.00	1.00	1.00	0.0118	0.0117	0.0116	0.0151	0.0149	0.0148

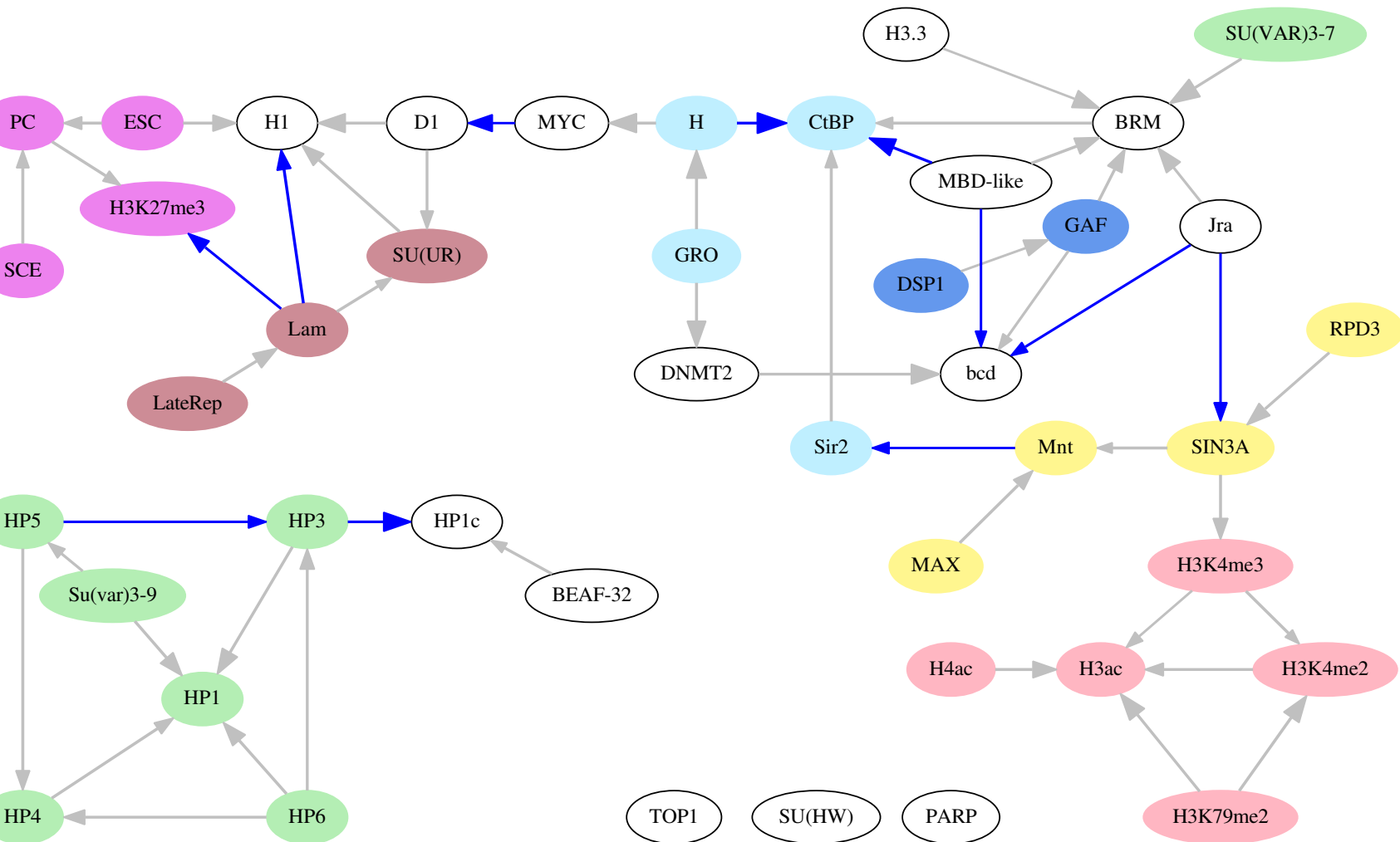
MAP and MLE



HAILFINDER

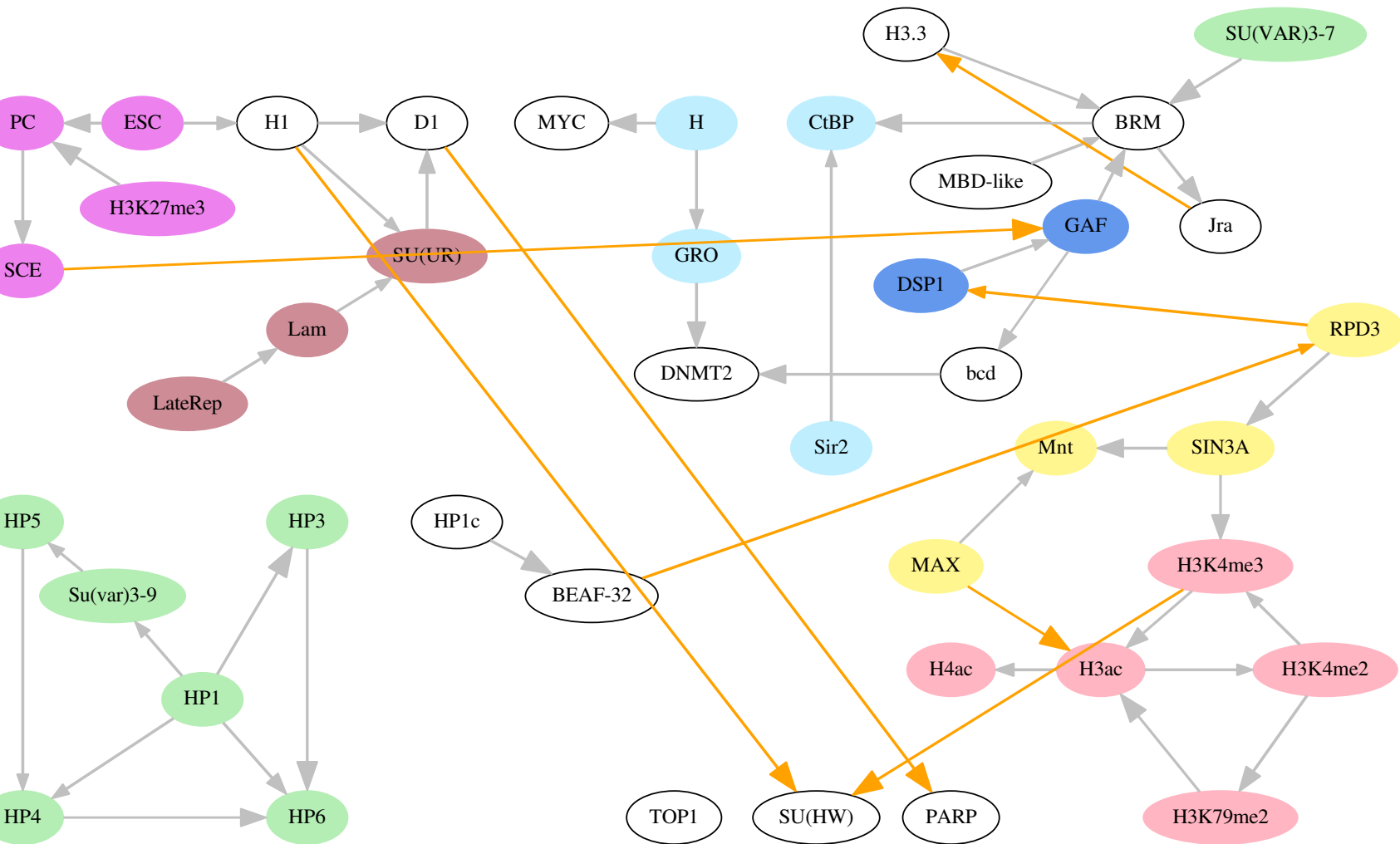
	BN	ID			NE		
1	MAP	1	1	1	0.2397	0.2382	0.2292
1	MLE	1	1	1	0.2397	0.2382	0.2302
4	MAP	1	1	1	0.0277	0.0258	0.0239
4	MLE	1	1	1	0.0277	0.0232	0.0229
5	MAP	1	1	1	0.0102	0.0092	0.0095
5	MLE	1	1	1	0.0117	0.0092	0.0098
8	MAP	1	1	1	0.0140	0.0108	0.0108
8	MLE	1	1	1	0.0133	0.0121	0.0102
9	MAP	1	1	1	0.0127	0.0102	0.0108
9	MLE	1	1	1	0.0114	0.0121	0.0114

The original chomatin model: BN_{80}



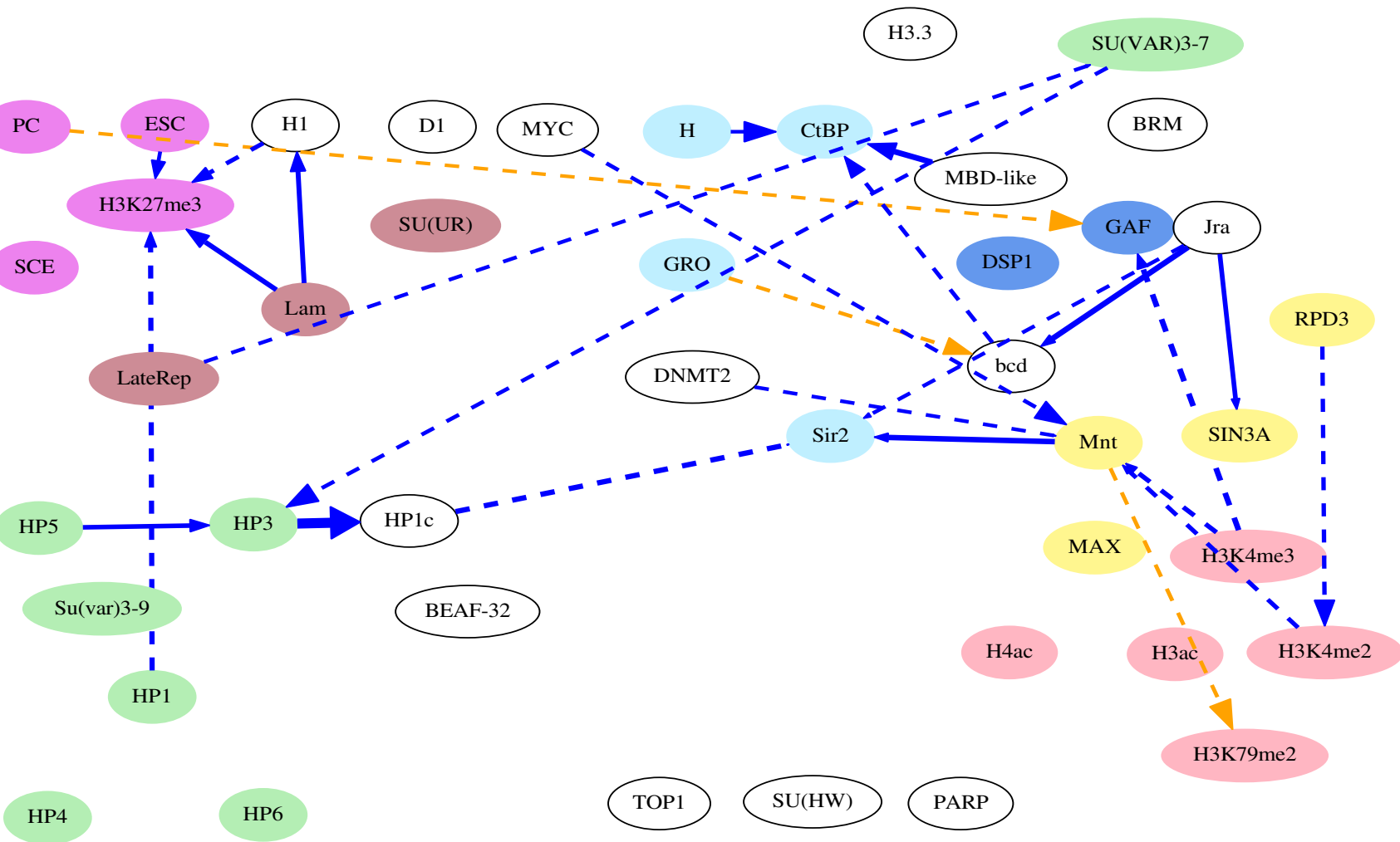
Bas van Steensel, et al. Feb. 2010, Genome Research

The MCMC model



At 80%, 41 common, 8 new, 11 missing. Tot = 52.

Differences (0.3) / (0.2)



BN_{80}/M_{80} dir.: dom., width : $abs(BN_{80} - M_{80})$, head: \rightarrow / \leftrightarrow , $\Rightarrow > 0.2 \rightarrow > 0.3$

availability



Easily installed in SWI-Prolog

```
? – pack_install(bims).
```

```
? – library(bims).
```

```
? – bims([]).
```

```
? – bims([data(carts), models(carts), likelihood(carts)]).
```

bottom line- Inference



MCMC

posterior probability is more versatile than point estimates

expensive and tricky to implement in practice

hard to actually encode priors

Bims

can represent detailed priors

inference over BNs, CARTs and other models

dependencies in priors pose difficulties