Exact Structure Learning of Bayesian Networks

Simon de Givry

MIAT, Université de Toulouse, INRA, Castanet Tolosan, France

October 13, 2016

Outline

Bayesian Network

- Joint probability distribution over random variables
  - Finite set **V** of p variables
  - Finite domains for each discrete variable

  e.g., $p = 3$ genes $\mathbf{V} = \{A, B, C\}$ with domains $\{\mathit{false}, \mathit{true}\}$, construct a Bayesian network that represents $P(A, B, C)$
- Defines a set of conditional (in)dependencies
  e.g., $A \perp\!\!\!\perp B$, $A \not\perp\!\!\!\perp B | C, \ldots$

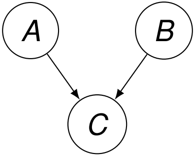
Bayesian Network

- Equivalent to a product of conditional probabilities e.g., $P(A, B, C) = P(A)P(B)P(C|A, B)$
  - One term per variable
  - Must sum to one: $\sum_A \sum_B \sum_C P(A, B, C) = 1$

  - Compact formulation
  - Only some (in)dependencies can be expressed

Bayesian Network

- Equivalent to a compact representation by a
  - Directed acyclic graph (DAG)
  - Conditional probability tables (CPT)



$A = true$	0.2
$A = false$	0.8

$P(A)$

	$A = true$		$A = false$	
	$B = true$	$B = false$	$B = true$	$B = false$
$C = true$	1	0.5	0.5	0
$C = false$	0	0.5	0.5	1

$P(C|A, B)$

$B = true$	0.1
$B = false$	0.9

$P(B)$

Learning Bayesian Network

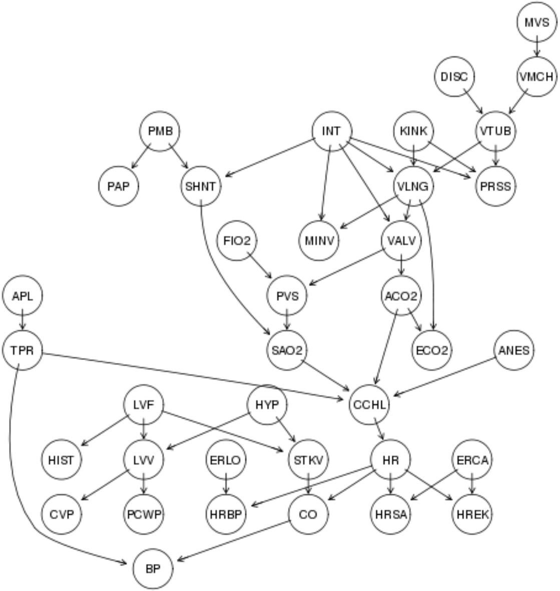
- Expert knowledge
- Learning BN structure from data
  - We will assume complete data (no missing values)
  - Independent and identically distributed sample data

BN structure learning

Different approaches

- Statistical independence tests
  PC [Spirtes, Glymour and Scheines, 1993],...
- **Search and score**
  Hill-Climbing [Chickering et al, 1995], GES, LAGD,
  Stochastic Greedy Search (SGS^3) [Vandel et al., 2012a],...
- Hybrid approaches
  Max-Min Hill-Climbing (MMHC) [Tsamardinos *et al*, 2006],...

Bayesian network Alarm ($p = 37$, 46 arcs)



Stochastic Greedy Search [Vandel et al., 2012a]

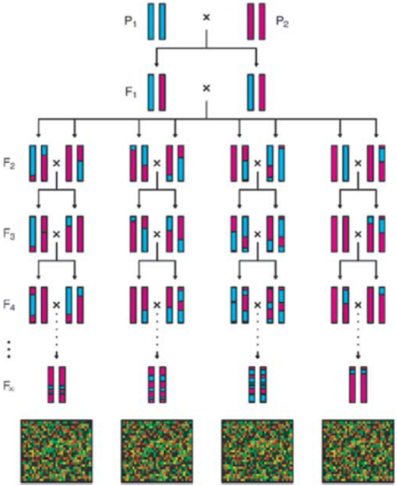
		ALARM			INSURANCE		
Nombre d'observations		50	500	5k	50	500	5k
SGS[1]	FP	44	10	10	18	5	3
	FN	14	4	2	32	20	10
SGS[2]	FP	44	10	10	18	5	3
	FN	14	4	2	32	20	9
SGS[3]	FP	44	**8**	6	19	**4**	**1**
	FN	14	**3**	2	32	**20**	**8**
LAGD	FP	31	11	8	15	**4**	5
	FN	14	4	2	32	**20**	11
GES	FP	**18**	**6**	**4**	**12**	2	3
	FN	**19**	**5**	**2**	**34**	23	12
		HAILFINDER			PIGS		
SGS[1]	FP	18	17	16	558	36	49
	FN	43	24	14	281	0	0
SGS[2]	FP	18	17	16	560	34	49
	FN	43	24	13	284	0	0
SGS[3]	FP	18	17	**16**	587	32	41
	FN	43	24	**13**	281	0	0
LAGD	FP	17	21	20	n/a	n/a	n/a
	FN	42	26	19	n/a	n/a	n/a
GES	FP	**15**	**15**	11	**121**	**2**	**0**
	FN	**43**	**24**	22	**420**	**7**	**0**

Genetical genomics

Gene-expressions vary due to polymorphisms (stationary phenomenon in controlled environment)

Data

- gene-expression levels
- marker genotypes
- marker/gene localisations on the genome



[Jansen and Nap, Trends in Gen., 2001]

Genetical genomics

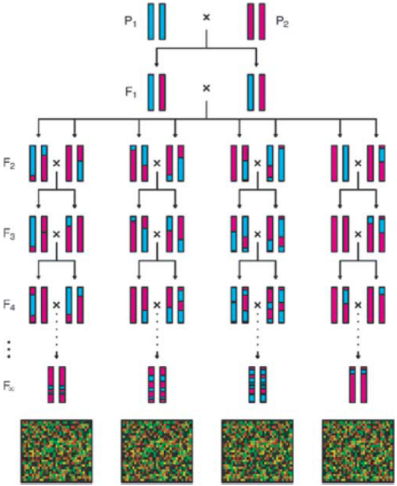
Gene-expressions vary due to polymorphisms (stationary phenomenon in controlled environment)

Data Arabidopsis thaliana

- 34,660 CATMA microarray probes
- 89 SNP marker genotypes
- marker/gene localisations on the genome

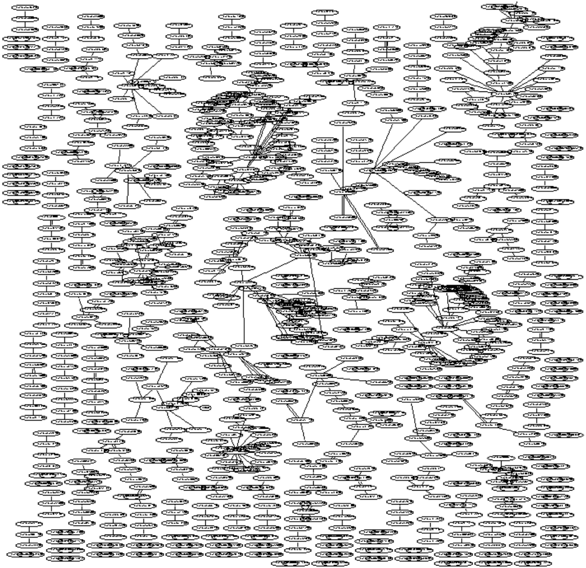
Sample size : $N = 158$ RIL

[Loudet *et al* , Gen. 2008]



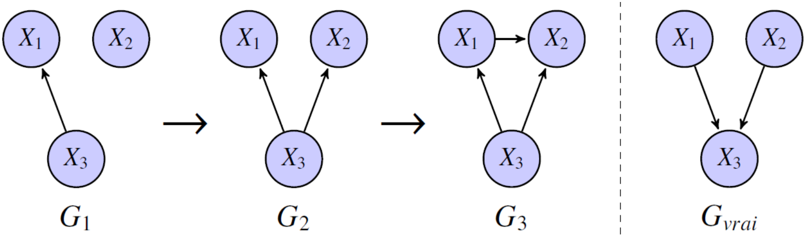
[Jansen and Nap, Trends in Gen., 2001]

Genetical genomics on Arabidopsis thaliana



Consensus BN 2,775 transcripts with high eQTL (LOD$\geq$ 3, bootstrap$\geq$ 0.3) [Vandel et al., 2012b]

Wrong arc orientation issue



Stochastic Greedy Search with post-processing [Vandel et al., 2012a]

		ALARM			INSURANCE		
Nombre d'observations		50	500	5k	50	500	5k
SGS[1]	FP	39	8	5	17	5	3
	FN	15	4	2	32	21	10
SGS[2]	FP	39	7	5	17	4	3
	FN	15	4	2	32	20	9
SGS[3]	FP	40	**6**	**2**	17	**4**	**1**
	FN	14	**3**	**2**	32	**20**	**8**
GES	FP	**18**	6	4	**12**	2	3
	FN	**19**	5	2	**34**	23	12
		HAILFINDER			PIGS		
SGS[1]	FP	15	16	13	546	3	7
	FN	43	24	14	281	0	0
SGS[2]	FP	16	16	13	548	3	7
	FN	43	24	14	284	0	0
SGS[3]	FP	**15**	16	**13**	574	**1**	2
	FN	**43**	24	**13**	281	**0**	0
GES	FP	**15**	**15**	11	**121**	2	**0**
	FN	**43**	**24**	22	**420**	7	**0**

Search and score

- Superexponential number of DAGs (e.g., $> 10^9$ for $p = 7$)
- NP-hard problem with two parents or more [Chickering, 1996]
- Various scoring functions (BIC/MDL, BDeu,...)
- Greedy search, tabu search, genetic algorithms,...
- Exact methods
  - integer linear programming [Bartlett and Cussens, 2015]
  - dynamic programming [Yuan and Malone, 2013]
  - *constraint programming [van Beek and Hoffmann, 2015]*

Integer linear programming

Definition of a 0/1 Linear Program (01LP)

$$
\begin{array}{rll}
\min\limits_{x_1,x_2,\ldots,x_n} & \sum_{i=1}^{n} & c_i x_i \\
\text{such that} & \sum_{i=1}^{n} & a_{i,1} x_i \leq b_1 \\
 & \sum_{i=1}^{n} & a_{i,2} x_i \leq b_2 \\
 & & \ldots \\
 & \sum_{i=1}^{n} & a_{i,m} x_i \leq b_m \\
 & x_i \in \{0,1\} & \forall i \in \{1,\ldots,n\}
\end{array}
$$

State-of-the-art solvers : SCIP, CPLEX, Gurobi, Xpress-MP,...

- Linear relaxation x^* using continuous domains
- Cutting planes (linear inequalities implied by the problem)
- Branching

01LP formulation

Family 0/1 variables

Create $x_{v \leftarrow \mathbf{W}}$ for each node $v \in \mathbf{V}$ and candidate parent set $\mathbf{W} \subseteq \mathbf{V}$.
$x_{v \leftarrow \mathbf{W}} = 1$ iff $\mathbf{W}$ is the parent set for v.

e.g. $x_{A \leftarrow \emptyset} = 1$, $x_{B \leftarrow \emptyset} = 1$, $x_{C \leftarrow \{A,B\}} = 1$

At most $p2^{p-1}$ candidate parent sets.

In practice, we limit the number of parents per node (typically to 3 or 4).

BIC score allows at most $l = \log(\frac{2N}{\log N})$ parents for sample size N, ie.
$l = 4$ for $N = 450$.

01LP formulation

BDeu score

We have $P(Graph|Data) \propto P(Graph)P(Data|Graph)$.
$P(Data|Graph)$ is called *marginal likelihood*.

BDeu score maximizes $P(Data|Graph)$ using Dirichlet priors on the CPT parameters.

$$\begin{aligned} \max \quad \log P(Data|Graph) &= \\ \max \quad & \sum_{v \in \mathbf{V}} \sum_{\mathbf{W} \subseteq \mathbf{V}} c_{v \leftarrow \mathbf{W}} x_{v \leftarrow \mathbf{W}} \\ \text{such that} \quad & \ldots \end{aligned}$$

Pruning can be applied on the list of candidate parent sets [de Campos and Ji, 2010] :

$$\text{remove } x_{v \leftarrow \mathbf{W}} \text{ if } \exists \mathbf{U} \subseteq \mathbf{V}, c_{v \leftarrow \mathbf{U}} \geq c_{v \leftarrow \mathbf{W}}$$

01LP formulation

01LP model

$$
\begin{array}{rccr}
\max & \sum_{v \in \mathbf{V}} \sum_{\mathbf{W} \subseteq \mathbf{V}} c_{v \leftarrow \mathbf{W}} x_{v \leftarrow \mathbf{W}} & & \\
\text{such that} & \sum_{\mathbf{W} \subseteq \mathbf{V}} x_{v \leftarrow \mathbf{W}} = 1 & \forall v \in \mathbf{V} & \\
 & \sum_{v \in \mathbf{C}} \sum_{\mathbf{W} : \mathbf{W} \cap \mathbf{C} = \emptyset} x_{v \leftarrow \mathbf{W}} \geq 1 & \forall \mathbf{C} \subseteq \mathbf{V} & (1) \\
 & x_{v \leftarrow \mathbf{W}} \in \{0, 1\} & \forall v \in \mathbf{V}, \mathbf{W} \subseteq \mathbf{V} & (2)
\end{array}
$$

At most $2^{p-1} - 1$ *cluster constraints* (1),
Add them on-the-fly (removing one cycle/two circuits at a time)

Solve using a *branch-and-cut* method.

Branch-and-cut

Cluster constraints

- Finding cluster **C** such that x^* violates acyclicity the most by solving a 01LP subproblem (still NP-hard [Cussens et al., 2016])

$$\max \sum_{v \in \mathbf{V}} \sum_{\mathbf{W} \subseteq \mathbf{V}} x^*_{v \leftarrow \mathbf{W}} y_{v \leftarrow \mathbf{W}} - \sum_{v \in \mathbf{V}} z_v$$

$$\text{s.t.} \quad y_{v \leftarrow \mathbf{W}} \implies z_v, \quad y_{v \leftarrow \mathbf{W}} \implies \bigvee_{w \in \mathbf{W}} z_w$$

$$\sum_{v \in \mathbf{V}} \sum_{\mathbf{W} \subseteq \mathbf{V}} x^*_{v \leftarrow \mathbf{W}} y_{v \leftarrow \mathbf{W}} - \sum_{v \in \mathbf{V}} z_v > -1$$

$$y_{v \leftarrow \mathbf{W}}, z_v \in \{0, 1\}$$

with $y_{v \leftarrow \mathbf{W}} \in \{0, 1\}$ for each variable $x^*_{v \leftarrow \mathbf{W}} > 0$ and $z_v = 1$ iff v belongs to cluster **C**.

We want to add more cuts...

Branch-and-cut

Cluster constraints

- Finding clusters **C** corresponding to elementary circuits in a *consensus* directed graph $G^{and} = (\mathbf{V}, \mathbf{E})$ such that

$$(u \to v) \in \mathbf{E} \text{ iff } \sum_{\mathbf{W} \subseteq \mathbf{V} : u \in \mathbf{W}} x^*_{v \leftarrow \mathbf{W}} = 1$$

e.g., $x^*_{C \leftarrow \{A,B\}} = 1, x^*_{B \leftarrow \{A,C\}} = 1, x^*_{A \leftarrow \{B,C\}} = 1$:
6 circuits of size 2 ($A \to B \to A$, $B \to A \to B$, $A \to C \to A$, $C \to A \to C$, $B \to C \to B$, $C \to B \to C$)
2 circuits of size 3 ($A \to B \to C \to A$ and $A \to C \to B \to A$)
$\Longrightarrow$ 4 cluster constraints :

- $\mathbf{C_1} = \{A, B\} : x_{A \leftarrow \emptyset} + x_{A \leftarrow \{C\}} + x_{B \leftarrow \emptyset} + x_{B \leftarrow \{C\}} \geq 1$
- ...
- $\mathbf{C_4} = \{A, B, C\} : x_{A \leftarrow \emptyset} + x_{B \leftarrow \emptyset} + x_{C \leftarrow \emptyset} \geq 1$

In practice, we limit the size of the circuits (up to 6).

Branch-and-cut

Extra features

- *Value Propagation* : removes simple unfeasible assignments.
  e.g., if $x_{B \leftarrow \{A,\ldots\}} = 1$ and $x_{C \leftarrow \{B,\ldots\}} = 1$ then $x_{A \leftarrow \{C,\ldots\}} = 0$
- *Sink-finding Primal Heuristic* (see later on dynamic programming)
- Generic cutting planes (Gomory, Strong Chvátal-Gomory, Zero-Half [Nemhauser and Wolsey, 1988])
- *Set Packing Constraint* $\sum_{v \in \mathbf{C}} \sum_{\mathbf{W}:\mathbf{C}\setminus\{v\}\subseteq \mathbf{W}} x_{v \leftarrow \mathbf{W}} \leq 1 \quad \forall \mathbf{C} \subseteq \mathbf{V}$

e.g., $x_{A \leftarrow \{B,C\}} + x_{B \leftarrow \{A,C\}} + x_{C \leftarrow \{A,B\}} \leq 1$

In practice, we limit to SPCs with $|\mathbf{C}| \leq 4$.

BN datasets characteristics [Bartlett and Cussens, 2015]

Name	Equivalent Sample Size	Number of Variables	Parent Set Limit	Number of Parent Sets
car	1	7	6	35
asia	10	8	2	127
insurance	1	27	6	341
mildew	1	35	3	3520
tic-tac-toe	10	10	3	112
flag	10	30	5	24 892
dermatology	10	35	3	5059
hailfinder	1	56	4	4330
kr-vs-kp	10	37	2	12 877
soybean-large	2	36	2	10 351
sponge	1	46	4	11 042
zoo	10	18	4	6461
alarm	1	37	4	8445
diabetes	1	413	2	4441
carpo	1	60	3	16 391
lung-cancer	10	57	2	8294

Practical impact of B&C features [Bartlett and Cussens, 2015]

Network	Baseline	No Cuts of Type			Without Solver Feature		
		G	SCG	ZH	SPH	SPC	VP
car	0.26 s	0.01 s	0.01 s	0.01 s	0.01 s	0.15 s	0.01 s
asia	0.36 s	0.35 s	**0.97 s**	0.34 s	0.34 s	**0.42 s**	0.36 s
insurance	0.83 s	0.76 s	**1.00 s**	0.74 s	0.81 s	**1.47 s**	0.81 s
Mildew	1.20 s	1.16 s	1.16 s	1.12 s	**1.21 s**	**2.03 s**	**1.21 s**
tic-tac-toe	9.40 s	5.23 s	9.18 s	2.55 s	9.26 s	8.41 s	9.30 s
flag	39.99 s	36.79 s	17.77 s	19.14 s	36.09 s	**70.85 s**	35.95 s
dermatology	32.17 s	31.19 s	21.16 s	27.49 s	31.63 s	28.84 s	29.74 s
hailfinder	112.56 s	79.23 s	61.90 s	87.56 s	**118.97 s**	**226.23 s**	111.93 s
kr-vs-kp	124.37 s	80.64 s	75.48 s	71.91 s	**125.81 s**	96.75 s	122.47 s
soybean-large	98.41 s	92.83 s	**130.14 s**	82.02 s	89.90 s	**110.38 s**	97.14 s
alarm	200.64 s	**280.12 s**	112.78 s	**244.59 s**	**201.50 s**	108.71 s	**227.45 s**
Diabetes	–	–	–	–	–	–	–
sponge	195.03 s	**225.24 s**	**300.02 s**	**231.95 s**	**230.15 s**	**214.46 s**	191.36 s
zoo	264.79 s	**290.40 s**	191.36 s	166.54 s	**266.65 s**	174.74 s	214.03 s
carpo	483.69 s	**528.10 s**	**587.44 s**	**513.89 s**	**577.18 s**	**589.40 s**	**510.62 s**
lung-cancer	670.76 s	646.50 s	651.57 s	583.45 s	670.66 s	642.77 s	627.88 s

Structure Learning with dynamic programming

Key Idea

Any Bayesian network has at least one *sink* node.
$\Longrightarrow$ choose its best parent set without introducing any directed cycle.

Recursive definition

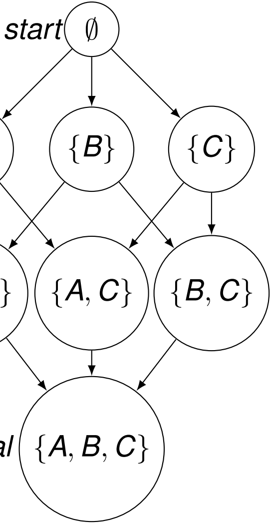
$$\log P(Data|Graph) \equiv L(\mathbf{V}) = \max_{s \in \mathbf{V}} (L(\mathbf{V} \setminus \{s\}) + \max_{\mathbf{W} \subseteq \mathbf{V} \setminus \{s\}} c_{s \leftarrow \mathbf{W}})$$

First, find optimal structures for single variables.
Then, build optimal subnetworks for increasing larger variable sets until **V**.

Time and space complexity in $O(2^p)$ [Silander and Myllymäki, 2006].

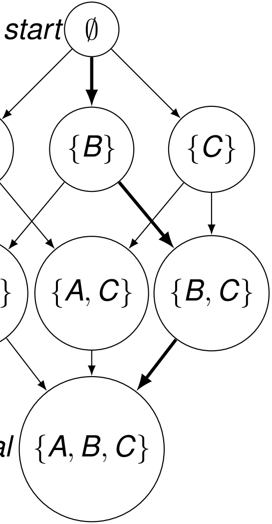
In practice, memory limits $p \leq 30$.

A simple order graph on three variables



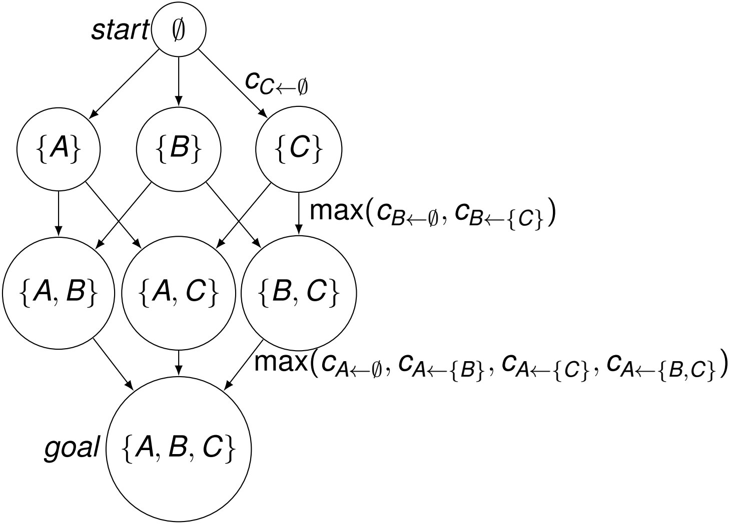
Each path from *start* to *goal* corresponds to a variable ordering

A simple order graph on three variables



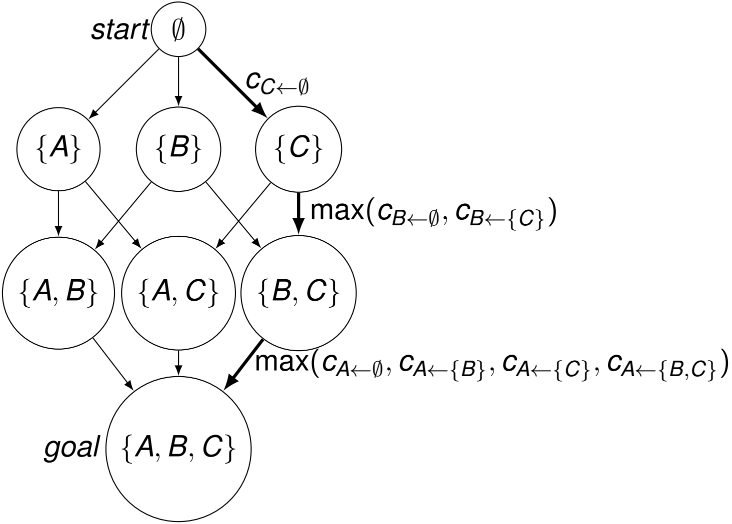
E.g., $\emptyset \rightarrow \{B\} \rightarrow \{B, C\} \rightarrow \{A, B, C\}$ corresponds to order (B, C, A)

A simple order graph on three variables



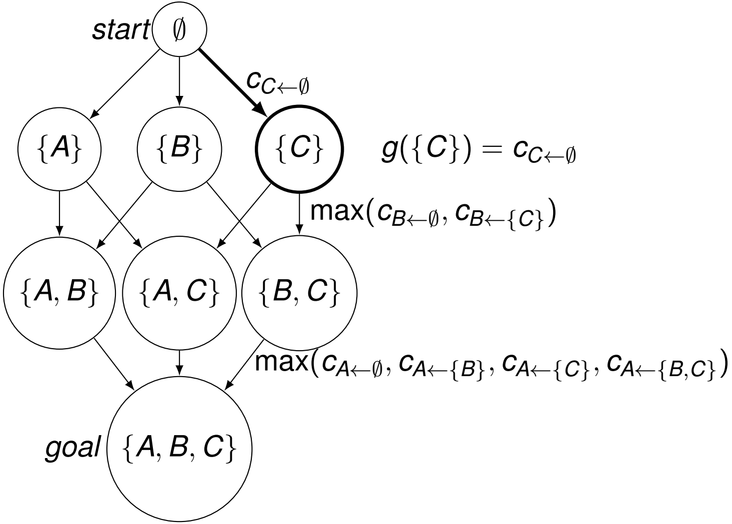
Each arc $\mathbf{S_1} \rightarrow \mathbf{S_2}$ has a score $\max_{\mathbf{W}\subseteq\mathbf{S_1}\setminus\{s\}} c_{s\leftarrow\mathbf{W}}$ with $s = \mathbf{S_1} \cap \mathbf{S_2}$

A simple order graph on three variables



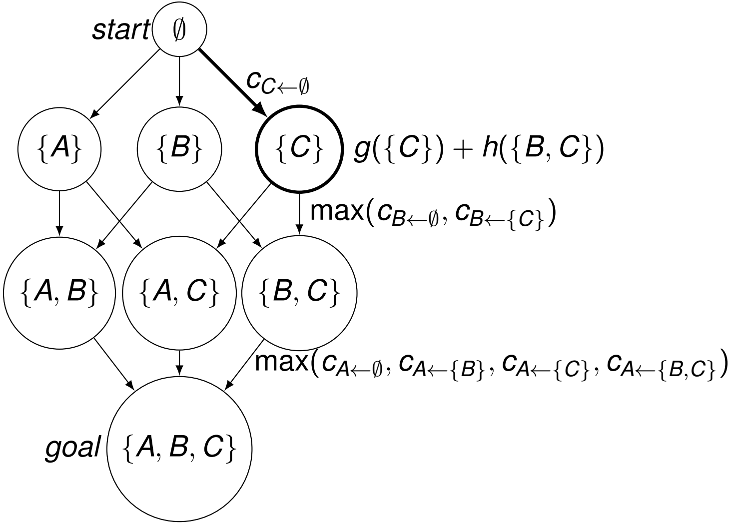
A^* find an optimal path in the order graph [Yuan and Malone, 2013]

A simple order graph on three variables



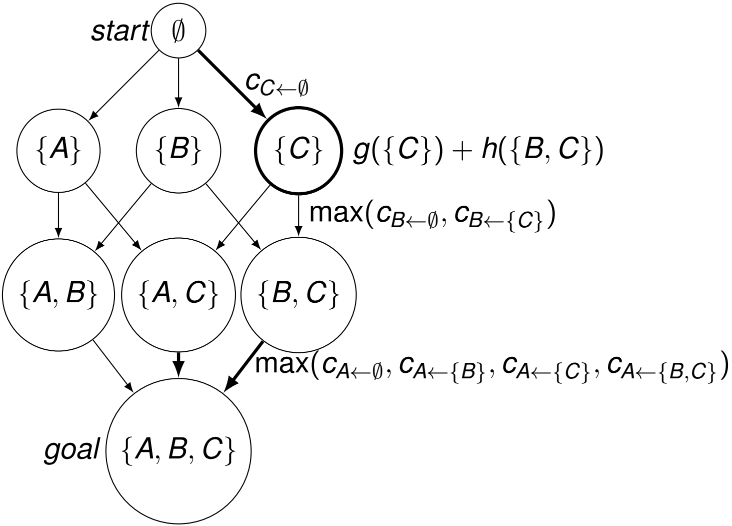
A^* explores the nodes in the order graph using a best-first traversal

A simple order graph on three variables



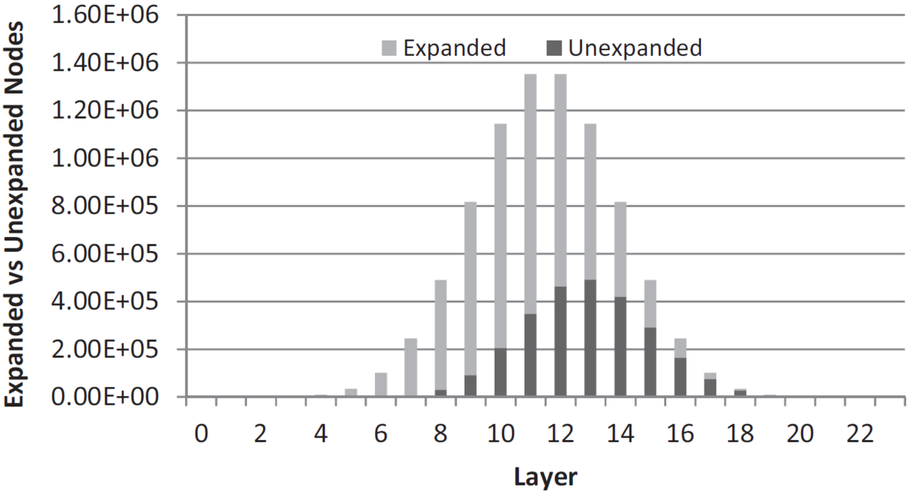
A^* exploits an upper bound h on the remaining distance

A simple order graph on three variables



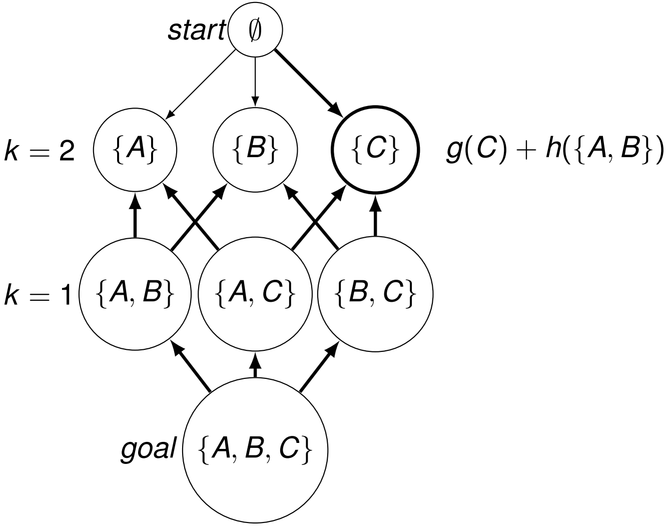
$h_1(\mathbf{U}) = \sum_{v \in \mathbf{V} \setminus \mathbf{U}} \max_{\mathbf{W} \subseteq \mathbf{V} \setminus \{v\}} c_{v \leftarrow \mathbf{W}}$

Experiments with A^* [Yuan and Malone, 2013]



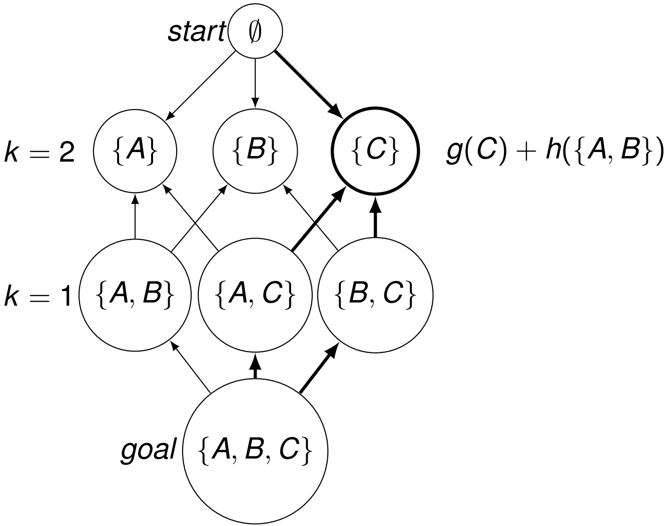
Parkinson ($p = 23, N = 195$), A^* using h_1 takes $\approx$ 100 seconds to solve.

A stronger heuristic h_k for A^*



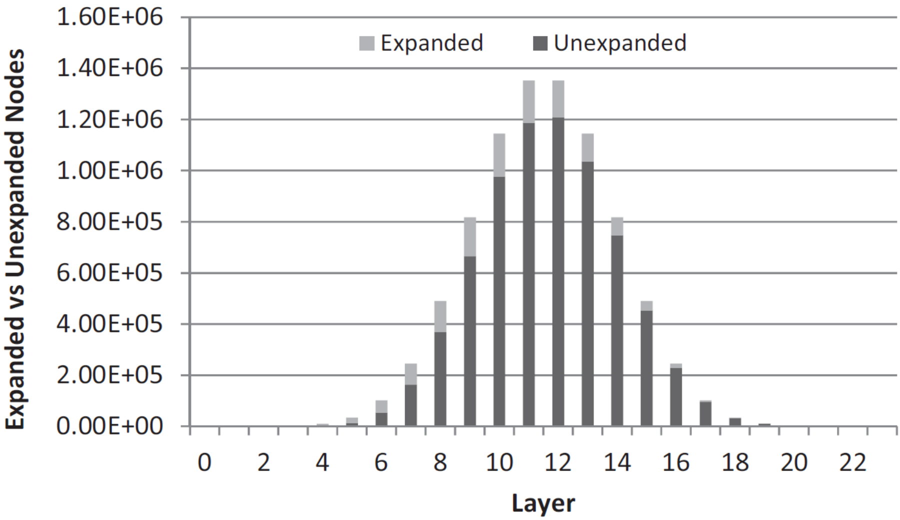
Explore from the goal up to k layers, complexity in $O(\frac{n!}{(n-k)!})$

A stronger heuristic h_k for A^*



Given a partition $\{\mathbf{S_1}, \ldots, \mathbf{S_p}\}$ of $\mathbf{V} \setminus \mathbf{U}$ with $|\mathbf{S_i}| \leq k$, $h_k(U) = \sum_{\mathbf{S_i}} h(\mathbf{S_i})$

Experiments with A^* [Yuan and Malone, 2013]



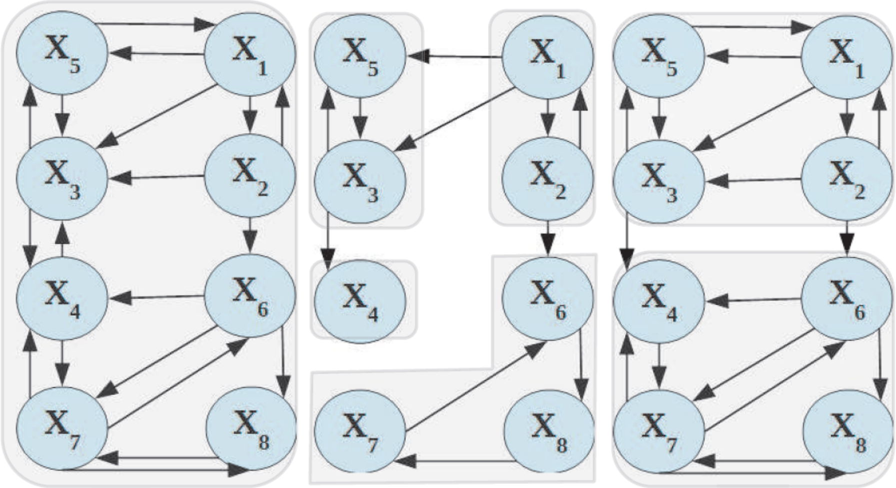
Parkinson ($p = 23, N = 195$), A^* using *static* $h_{k=12}$ takes $\approx$ 10 seconds

How to partition **V** in *quasi-independent* subproblems?

var.	**POPS**			
X_1	$\{X_2\}$	$\{X_5\}$		
X_2	$\{X_1\}$			
X_3	$\{X_1, X_5\}$	$\{X_1, X_2\}$	$\{X_2, X_4\}$	$\{X_1\}$
X_4	$\{X_3\}$	$\{X_6\}$	$\{X_7\}$	
X_5	$\{X_1, X_3\}$	$\{X_3\}$		
X_6	$\{X_2, X_7\}$	$\{X_7\}$		
X_7	$\{X_8\}$	$\{X_6, X_4\}$		
X_8	$\{X_6\}$	$\{X_7\}$		

List of non-dominated candidate parent sets [Fan and Yuan, 2015]

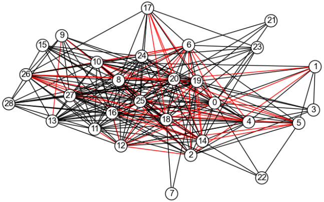
How to partition **V** in *quasi-independent* subproblems?



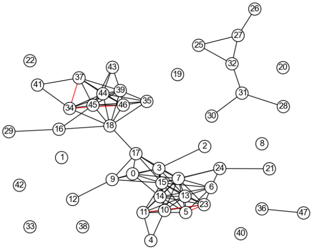
All / Top-1 / Top-2

parent relation graphs divided into SCCs [Fan and Yuan, 2015]

Examples of parent relation graphs (BIC score)

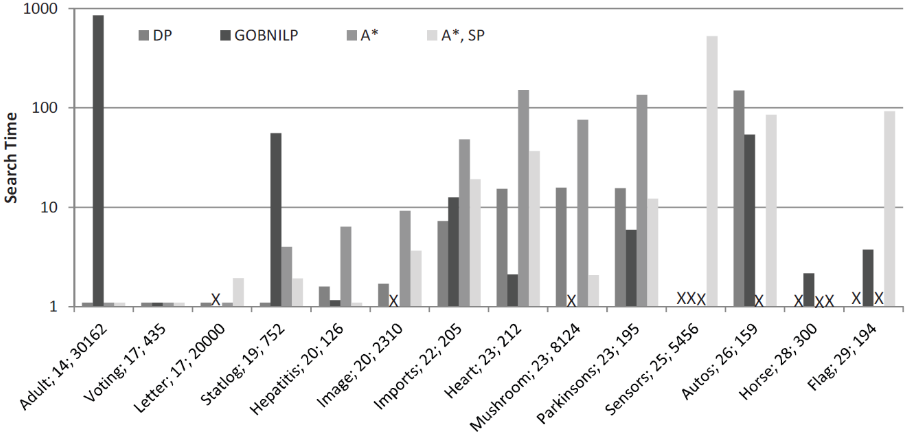


Flag ($p = 29, N = 194$)



Barley ($p = 48, N = 1000$)

Comparative Results [Yuan and Malone, 2013]



using Minimum Description Length (≈BIC) score instead of BDeu

Constraint programming [van Beek and Hoffmann, 2015]

- dual model for acyclicity : DAG encoding and topological order
- symmetry breaking constraints
- same lower bound as dynamic programming A_k^*

Comparative Results [van Beek and Hoffmann, 2015]

Benchmark	n	N	BDeu				BIC			
			d	GOBN. v1.4.1	A* v2015	CPBayes v1.0	d	GOBN. v1.4.1	A* v2015	CPBayes v1.0
shuttle	10	58,000	812	58.5	0.0	0.0	264	2.8	0.1	0.0
adult	15	32,561	768	1.4	0.1	0.0	547	0.7	0.1	0.0
letter	17	20,000	18,841	5,060.8	1.3	1.4	4,443	72.5	0.6	0.2
voting	17	435	1,940	16.8	0.3	0.1	1,848	11.6	0.4	0.1
zoo	17	101	2,855	177.7	0.5	0.2	554	0.9	0.4	0.1
tumour	18	339	274	1.5	0.9	0.2	219	0.4	0.9	0.2
lympho	19	148	345	1.7	2.1	0.5	143	0.5	1.0	0.2
vehicle	19	846	3,121	90.4	2.4	0.7	763	4.4	2.1	0.5
hepatitis	20	155	501	2.1	4.9	1.1	266	1.7	4.8	1.0
segment	20	2,310	6,491	2,486.5	3.3	1.3	1,053	13.2	2.4	0.5
mushroom	23	8,124	438,185	OT	255.5	561.8	13,025	82,736.2	34.4	7.7
autos	26	159	25,238	OT	918.3	464.2	2,391	108.0	316.3	50.8
insurance	27	1,000	792	2.8	583.9	107.0	506	2.1	824.3	103.7
horse colic	28	300	490	2.7	15.0	3.4	490	3.2	6.8	1.2
steel	28	1,941	113,118	OT	902.9	21,547.0	93,026	OT	550.8	4,447.6
flag	29	194	1,324	28.0	49.4	39.9	741	7.7	12.1	2.6
wdbc	31	569	13,473	2,055.6	OM	11,031.6	14,613	1,773.7	1,330.8	1,460.5
water	32	1,000					159	0.3	1.6	0.6
mildew	35	1,000	166	0.3	7.6	1.5	126	0.2	3.6	0.6
soybean	36	266					5,926	789.5	1,114.1	147.8
alarm	37	1,000					672	1.8	43.2	8.4
bands	39	277					892	15.2	4.5	2.0
spectf	45	267					610	8.4	401.7	11.2
sponge	45	76					618	4.1	793.5	13.2
barley	48	1,000					244	0.4	1.5	3.4
hailfinder	56	100					167	0.1	9.9	1.5
hailfinder	56	500					418	0.5	OM	9.3
lung cancer	57	32					292	2.0	OM	10.5
carpo	60	100					423	1.6	OM	253.6
carpo	60	500					847	6.9	OM	OT

Softwares

- GOBNILP `www.cs.york.ac.uk/aig/sw/gobnilp`
- A^* `www.urlearning.org`
- CPBayes `cs.uwaterloo.ca/~vanbeek`

Conclusions and Perspectives

- A^* is limited by the number of variables (≈ 60)
- GOBNILP by the number of candidate parent sets ($\approx 20,000$)

- Hybrid approaches combining approximate & exact methods
- Add prior knowledge ($P(Graph|Data) \propto P(Graph)P(Data|Graph)$)
- Easy to add user constraints in generic 01LP or CP frameworks

Bartlett, M. and Cussens, J. (2015).
Integer linear programming for the bayesian network structure learning problem.
Artificial Intelligence.

Cussens, J., Järvisalo, M., Korhonen, J. H., and Bartlett, M. (2016).
Bayesian network structure learning with integer programming: Polytopes, facets, and complexity.
Journal of Artificial Intelligence Research.

de Campos, C. P. and Ji, Q. (2010).
Properties of bayesian dirichlet scores to learn bayesian network structures.
In *Proc. of AAAI-00*, Atlanta, Georgia, USA.

Fan, X. and Yuan, C. (2015).
An improved lower bound for bayesian network structure learning.
In *Proc. of AAAI-15*, Austin, Texas.

Silander, T. and Myllymäki, P. (2006).
A simple approach for finding the globally optimal bayesian network structure.
In *Proc. of UAI'06*, Cambridge, MA, USA.

van Beek, P. and Hoffmann, H.-F. (2015).
Machine learning of bayesian networks using constraint programming.
In *Proc. of CP'15*, Cork, Ireland.

Vandel, J., Mangin, B., and de Givry, S. (2012a).
New Local Move Operators for Bayesian Network Structure Learning.
In *Probabilistic Graphical Models (PGM2012)*.

Vandel, J., Mangin, B., Vignes, M., Leroux, D., Loudet, O., Martin-Magniette, M.-L., and de Givry, S. (2012b).
Inférence de réseaux de régulation de gènes au travers de scores étendus dans les réseaux bayésiens.
Revue d'Intelligence Artificielle, 26(6):679–708.

Yuan, C. and Malone, B. (2013).
Learning optimal bayesian networks: A shortest path perspective.
Journal of Artificial Intelligence Research.