Probabilistic Graphical Models

Lecture 7 –Variable Elimination

CS/CNS/EE 155
Andreas Krause

Announcements

- Homework 1 due today in class
- Will get back to you soon with feedback on project proposals.

Key questions

- How do we specify distributions that satisfy particular independence properties?

  → **Representation**

- How can we identify independence properties present in data?

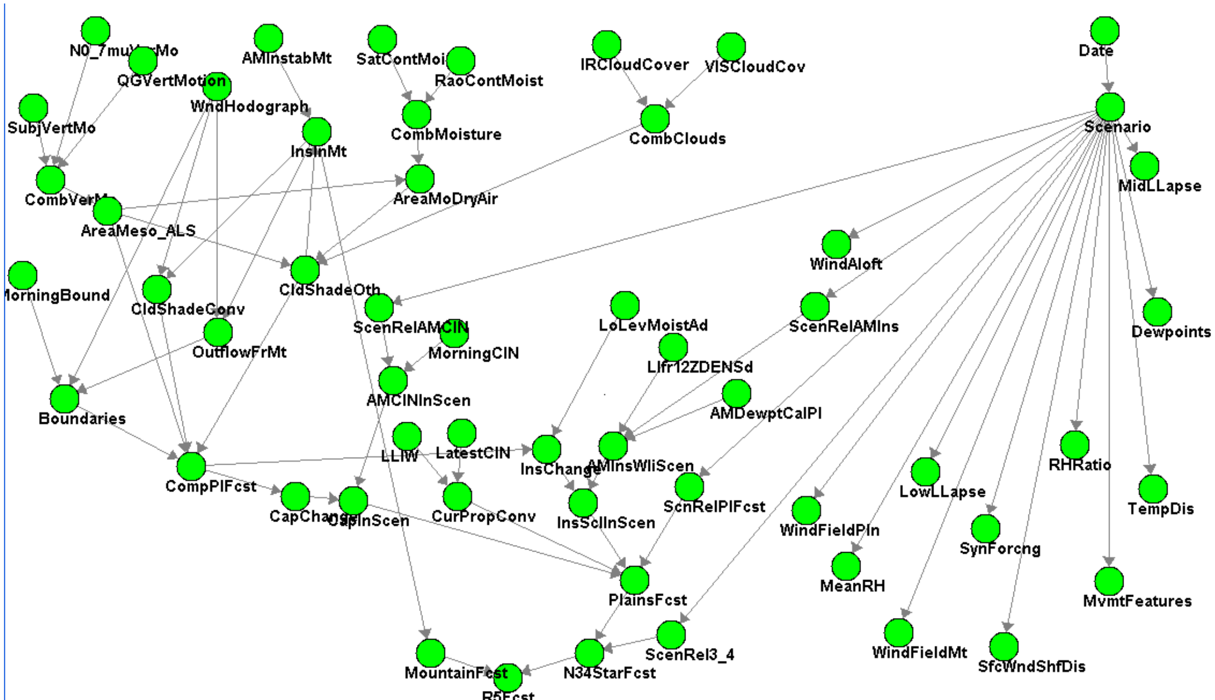
  → **Learning**

- How can we exploit independence properties for efficient computation?

  → **Inference**

Bayesian network inference

- Compact representation of distributions over large number of variables
- (Often) allows efficient **exact inference** (computing marginals, etc.)



JavaBayes applet

HailFinder

56 vars

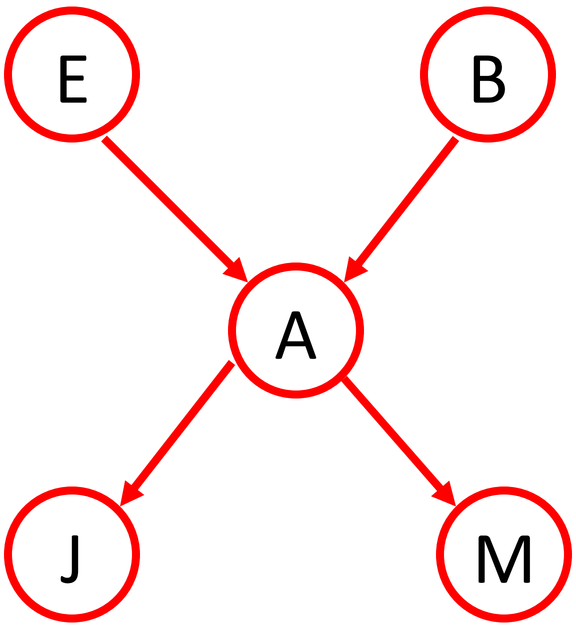
~ 3 states each

➔ ~10^{26} terms

> **10.000 years**

on Top supercomputers

Typical queries: Conditional distribution



- Compute distribution of some variables given values for others

Observe $M=T$

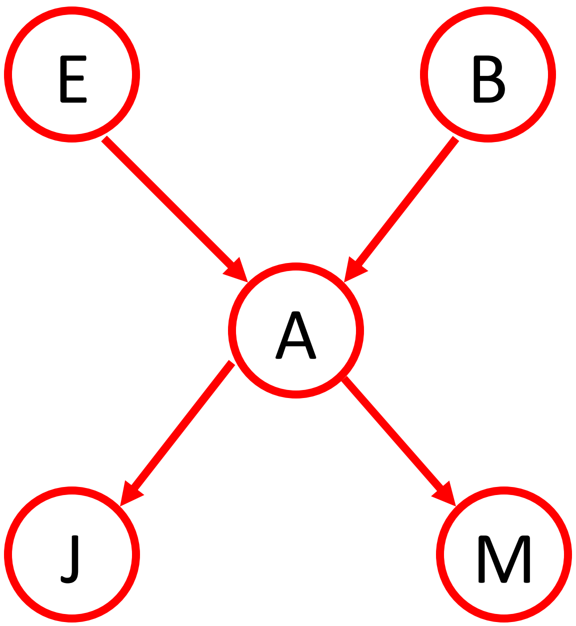
Compute $P(E=T \mid M=T)$

$$P(E=T \mid M=T) = \frac{P(E=T, M=T)}{P(M=T)}$$

$$P(E=T, M=T) = \underbrace{\sum_b \sum_a \sum_j}_{2^3 \text{ terms}} \underbrace{P(E=T, M=T, B=b, A=a, J=j)}_{P(E)P(B)P(A \mid E,B)P(J \mid A)P(M \mid A)}$$

Naive approach exponential in # vars ...

Typical queries: Maxizimization



- MPE (Most probable explanation):

  Given values for some vars, compute most likely assignment to all remaining vars

Given J=F, M=T, find

$$(e^*, b^*, a^*) = \arg\max_{e,b,a} P(J=F, M=T, e, b, a)$$

MPE an MAP don't necessarily give same answers...

- MAP (Maximum a posteriori):

  Compute most likely assignment to some variables

$$e^* = \arg\max_e P(J=F, M=T, E=e) =$$

$$= \arg\max \sum_{a,b} P(J=F, M=T, e, a, b)$$

Hardness of inference for general BNs

- Computing conditional distributions:
  - Exact solution: #P-complete
  - Approximate solution: NP-hard:

Absolute approx: Finding $|P(x) - \hat{P}(x)| < \varepsilon$ NP-hard even for $\varepsilon = \frac{1}{2}$

Relative approx: $1-\varepsilon < \frac{\hat{P}(x)}{P(x)} < 1+\varepsilon$ NP hard for $\varepsilon > 0$

- Maximization:
  - MPE: NP-complete
  - MAP: NP^{PP}-complete

$$\max_{x_1 \ldots x_m} \sum_{x_{m+1} \ldots x_{2m}}$$

- Inference in general BNs is really hard ☹
- Is all hope lost?

Inference

- Can exploit structure (conditional independence) to efficiently perform **exact inference** in many practical situations

- For BNs where exact inference is not possible, can use algorithms for **approximate inference** (later this term)

Potential for savings: Variable elimination!

$$X_1 \to X_2 \to X_3 \to X_4 \to X_5$$

$$P(x_5, x_1) = \sum_{x_2}\sum_{x_3}\sum_{x_4} P(x_1)\,P(x_2|x_1)\,P(x_3|x_2)\,P(x_4|x_3)\,P(x_5|x_4)$$

8

distributivity $P(x_1)\sum_{x_2} P(x_2|x_1)\sum_{x_3} P(x_3|x_2)\underbrace{\sum_{x_4} P(x_4|x_3)\,P(x_5|x_4)}_{g_4(x_3,x_5)}$

How many additions: 3

$X_1 \to X_2 \to X_3 \to X_5$ $\quad \underbrace{g_4(x_3,x_5)}_{= P(x_5|x_3)}$ 1

$X_1 \to X_2 \to X_5$ $\quad g_3(x_2,x_5)$ 1

$\underbrace{g_2(x_1,x_5)}_{= P(x_5|x_1)}$ 1

$X_1 \to X_5$

Intermediate solutions are distributions on fewer variables!

Variable elimination algorithm

- Given BN and Query $P(X \mid \mathbf{E}=\mathbf{e})$
- Remove irrelevant variables for $\{X,\mathbf{e}\}$
- Choose an ordering of $X_1,\ldots,X_n$
- Set up initial factors: $f_i = P(X_i \mid \mathbf{Pa_i})$
- For i =1:n, $X_i \notin \{X,\mathbf{E}\}$
  - Collect all factors f that include X_i
  - Generate new factor by marginalizing out X_i

$$g = \sum_{x_i} \prod_j f_j$$

  - Add g to set of factors
- Renormalize $P(x,\mathbf{e})$ to get $P(x \mid \mathbf{e})$

Multiplying factors

$$g = \sum_{x_i} \underbrace{\prod_j}_{} f_j$$

B \ A	T	F
T	.	.
F	.	.

C \ B	T	F
T		
F		

$f_1(A,B)$, $f_2(B,C)$

$f' = f_1 \cdot f_2$

$f'(A,B,C)$

A \ BC	TT	TF	FT	FF
T		.	.	.
F				

$f'(A{=}T, B{=}F, C{=}F) = f_1(A{=}T, B{=}F) \cdot f_2(B{=}F, C{=}F)$

Marginalizing factors

$$g = \sum_{x_i} \underbrace{\prod_j f_j}_{f'}$$

$f'(A,B)$

$A \backslash B$	T	F
T	.	.
F	.	.

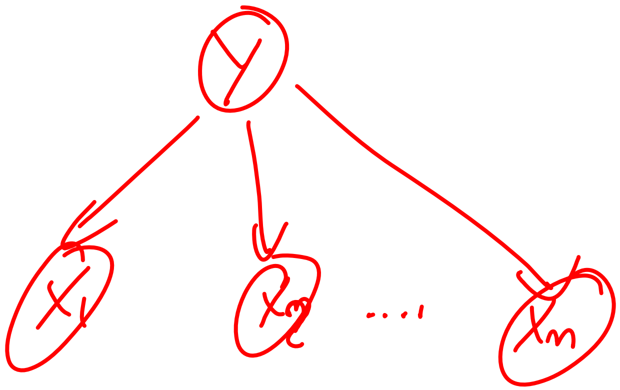
$\Rightarrow g = \sum_A f'(A,B)$

B	$g(B)$
T	
F	

$g(B) = f'(A=T,B) + f'(A=F,B)$

Does the order matter?

Elim X_2 first



$P(X_n | X_1)$

$$g = \sum_{X_2} P(X_2|Y)$$

$$Scope(g) = \{X_2, Y\} = Scope(P(X_2|Y))$$

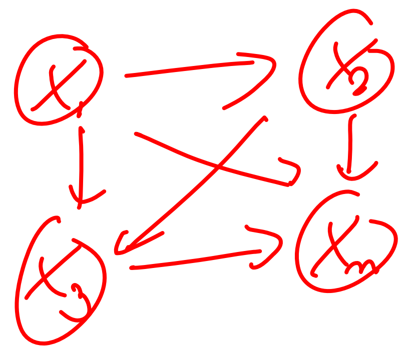
YES!

If elim. Y first

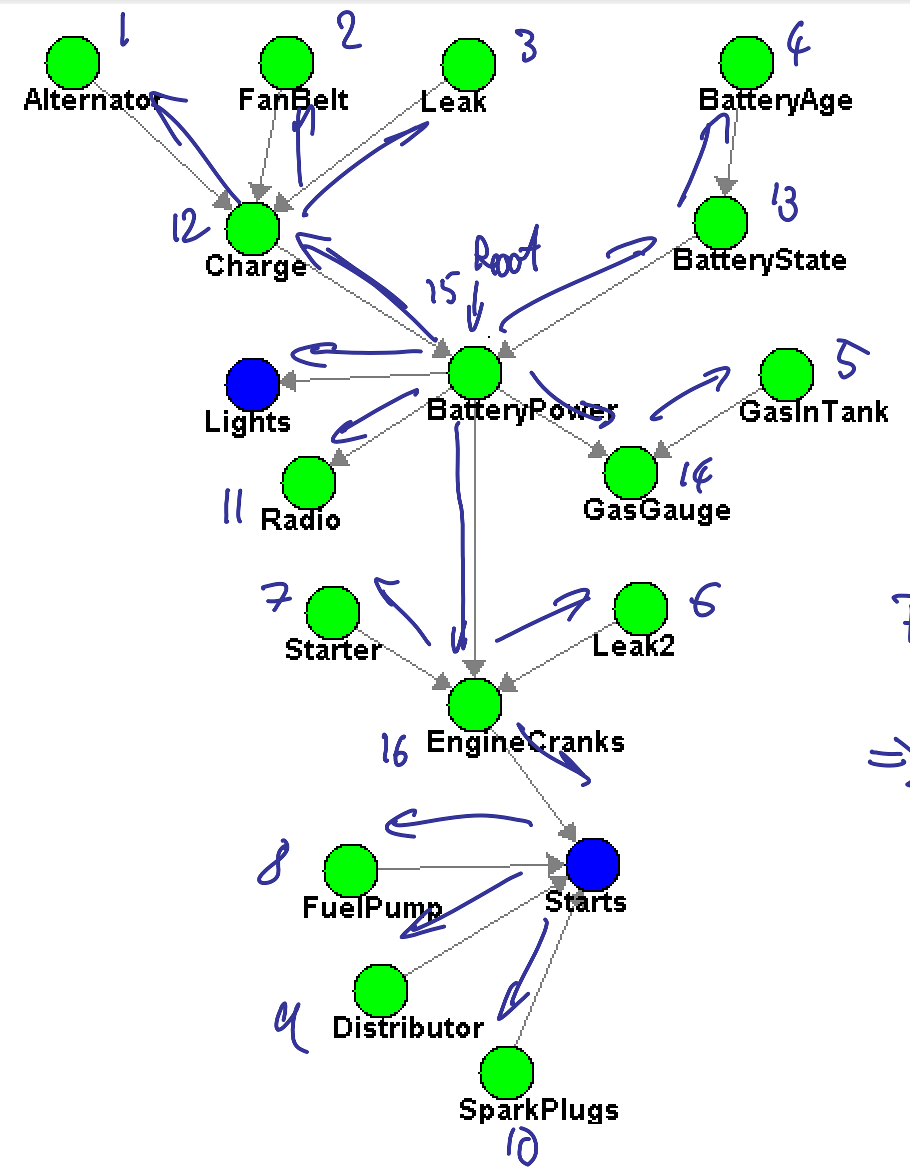
$$g = \sum_{Y} \left(\prod_j P(X_j|Y) \right) P(Y)$$

$$Scope(g) = \{X_1 \ldots X_n\}$$

exponential in size



Variable elimination for polytrees



BN G Polytree if skeleton(G) is tree

Drop edge directions
Pick some root in skeleton
Direct edges outward from root

Eliminate vars (not in query) in inverse topological order of resulting directed tree

Factor sizes do <u>not</u> increase

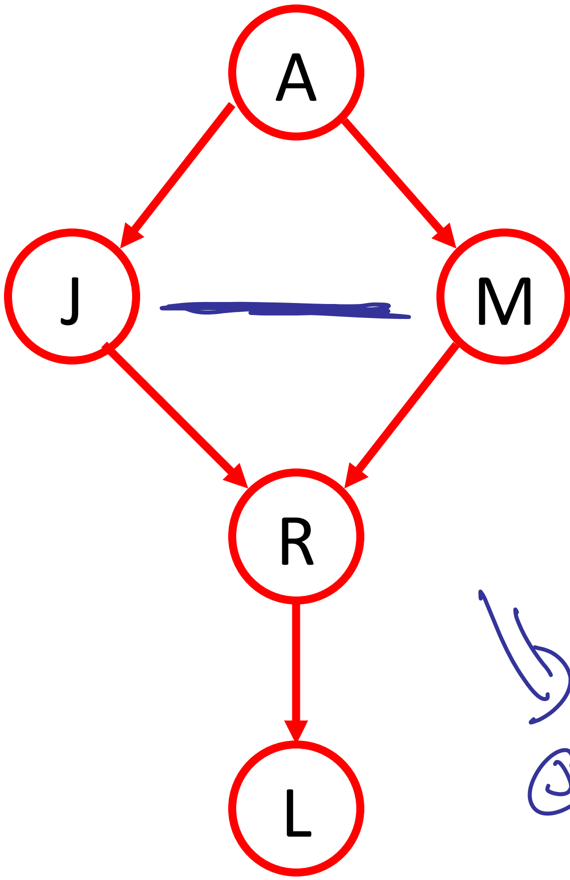
$\Rightarrow$ Inference possible in $O(n)$ steps

Complexity of variable elimination

- Tree graphical models
  - Using correct elimination order, factor sizes do not increase!
  - Inference in linear time!!

- General graphical models
  - Ultimately NP-hard..
  - Need to understand what happens if there are loops

Variable elimination with loops



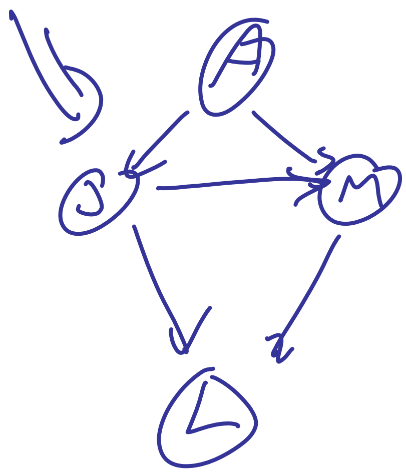
$P(A,L)$

Elim. order R, J, M

$$P(A,L) = \sum_{r\,j\,m} P(A)\, P(J|A)\, P(M|A)\, P(R|J,M)\, P(L|R)$$

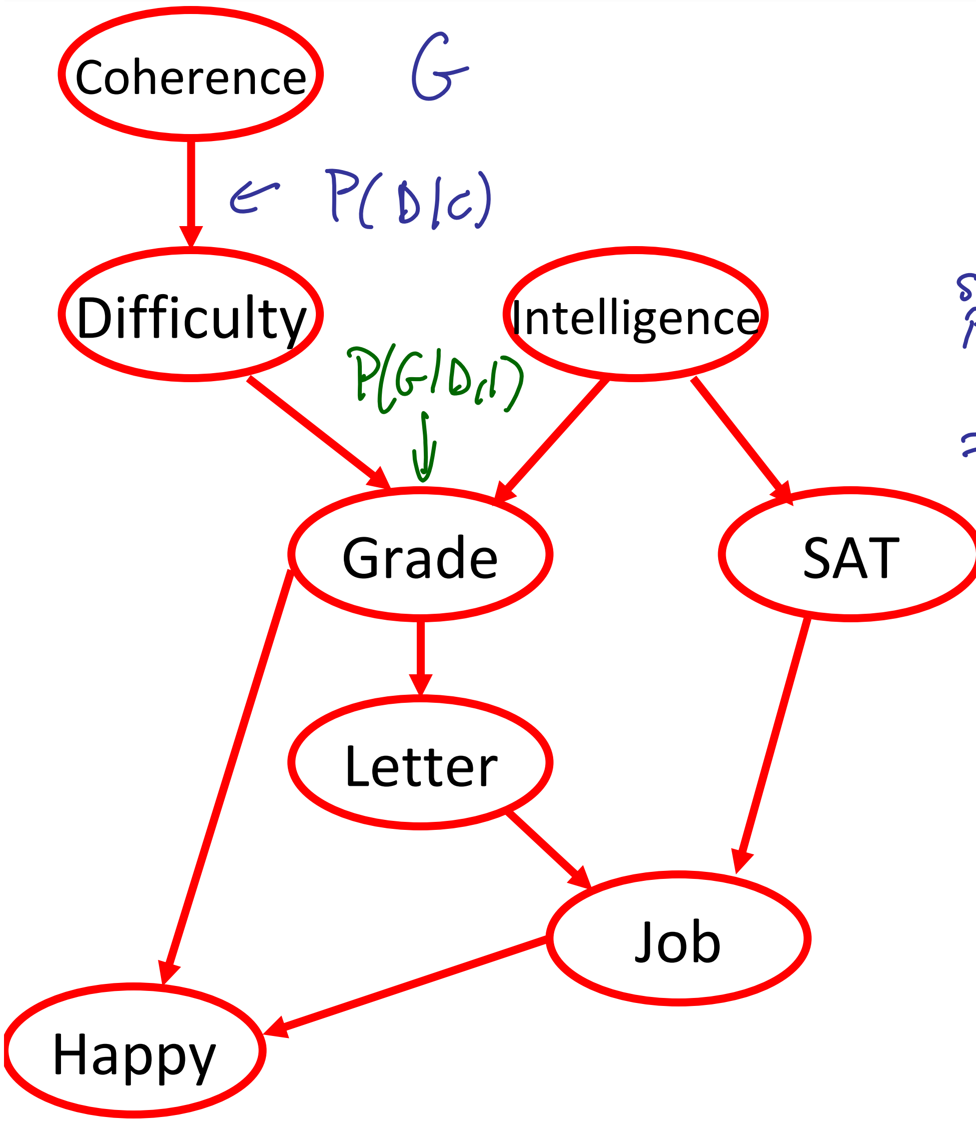
$$= \sum_{j\,m} P(A)\, P(J|A)\, P(M|A) \underbrace{\sum_r P(r|j,m)\, P(L|r)}_{g}$$

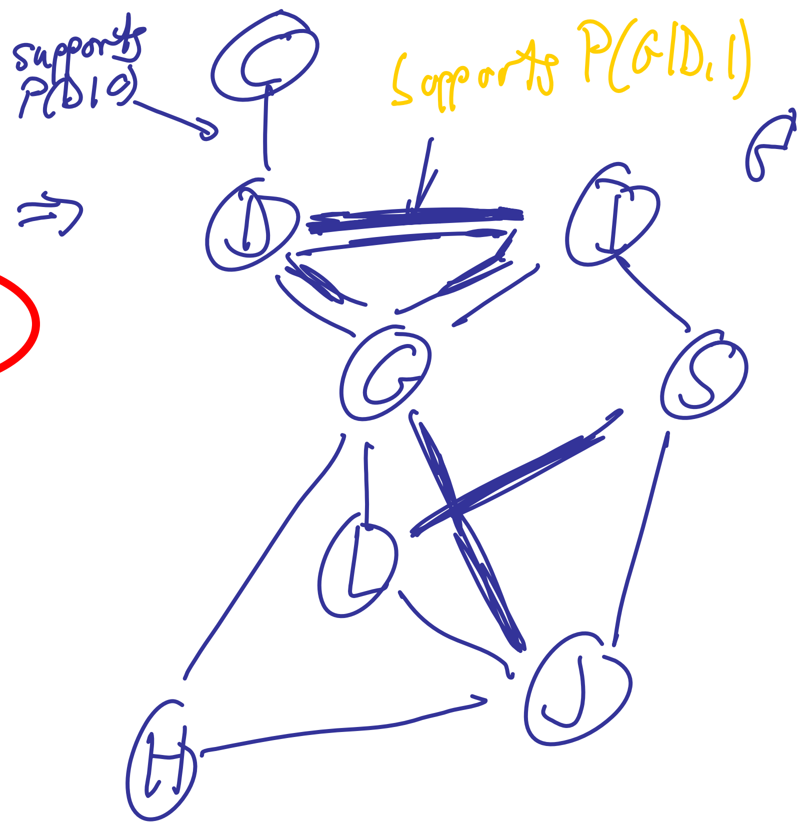
$\text{Scope}(g) = \{L, J, M\}$



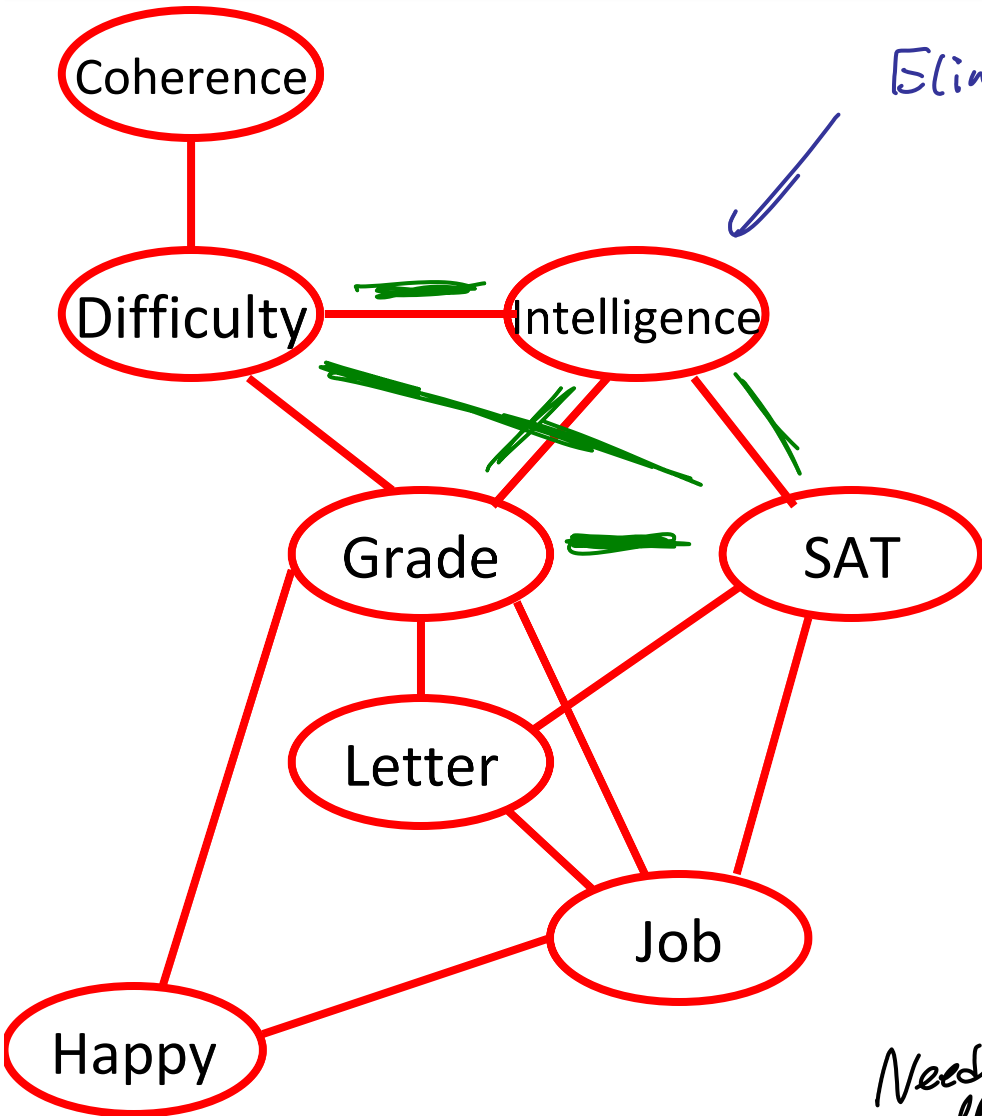
Elimination as graph transformation: Moralization

Want that all factors of G are supported on cliques of G'





Elimination: Filling edges



Eliminate

$\Rightarrow$ Introduce new factor

$$g = \sum_{I} \prod_{\substack{\text{Cliques } C \\ \text{incident} \\ \text{with } I}} f_C$$

$$\text{Scope}(g) = \bigcup_{\substack{\text{Cliques } C \\ \text{inc. w } I}} \text{Scope}(f_C) \setminus \{I\}$$

Want to make sure that resulting graph supports g

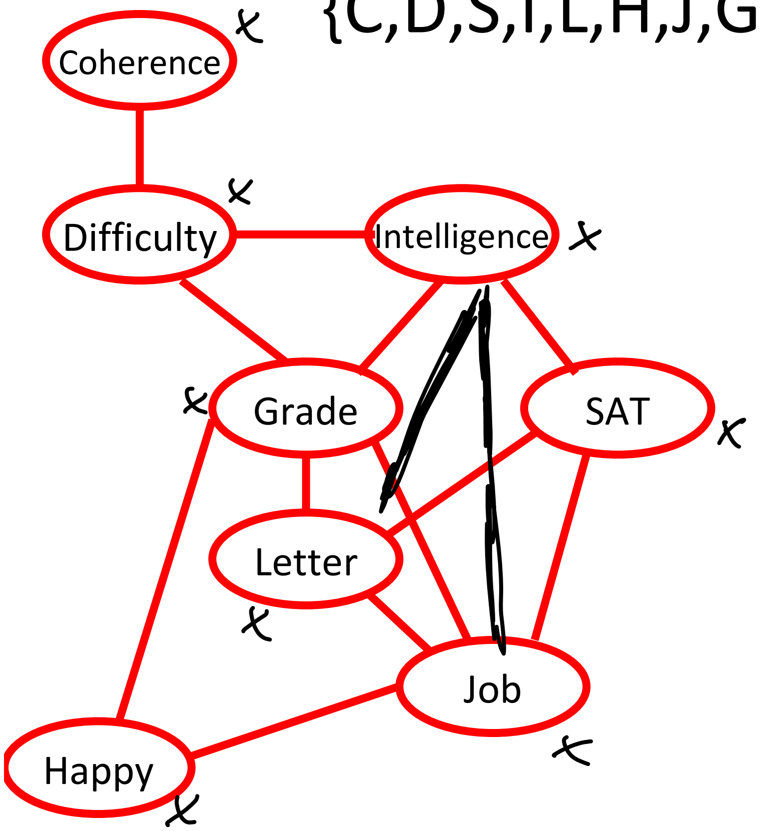
Here: $\text{Scope}(g) = \{D, G, S\}$

Not supported!

Need to fill in edges by connecting all neighbors of I into clique

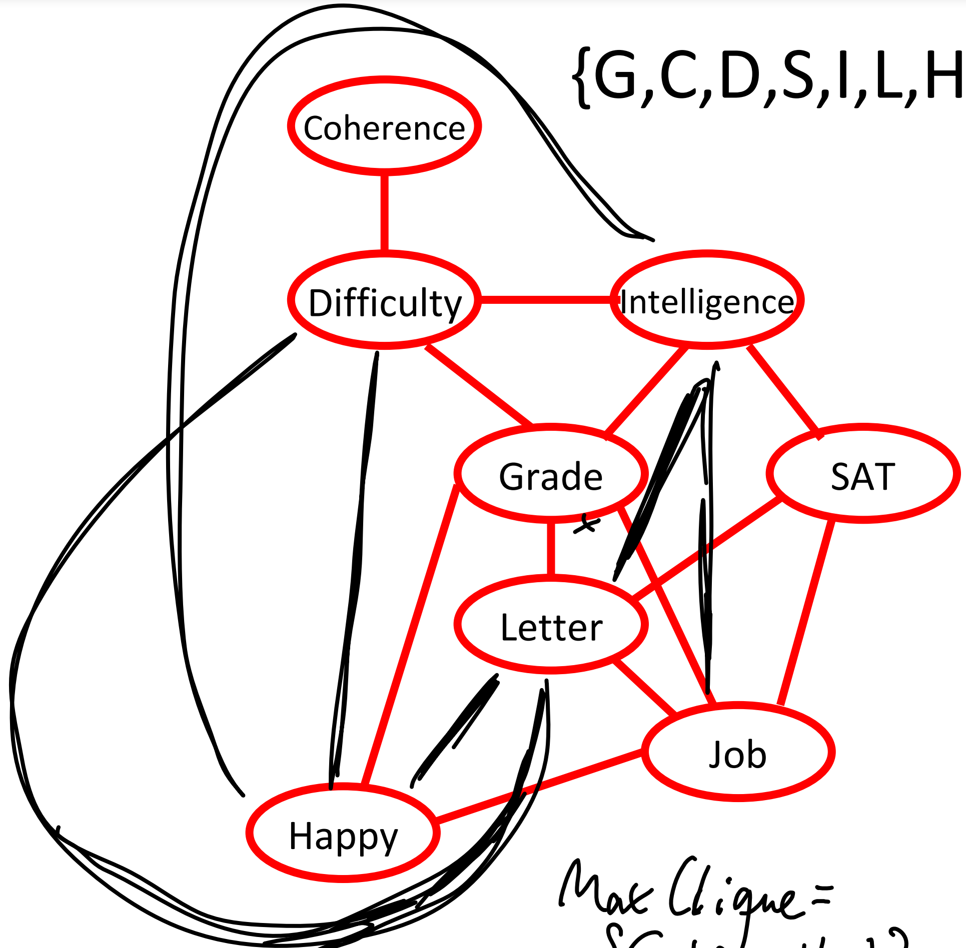
Impact of elimination order

{C,D,S,I,L,H,J,G}



Max Clique: {I,G,L,J}
=> Treewidth 3

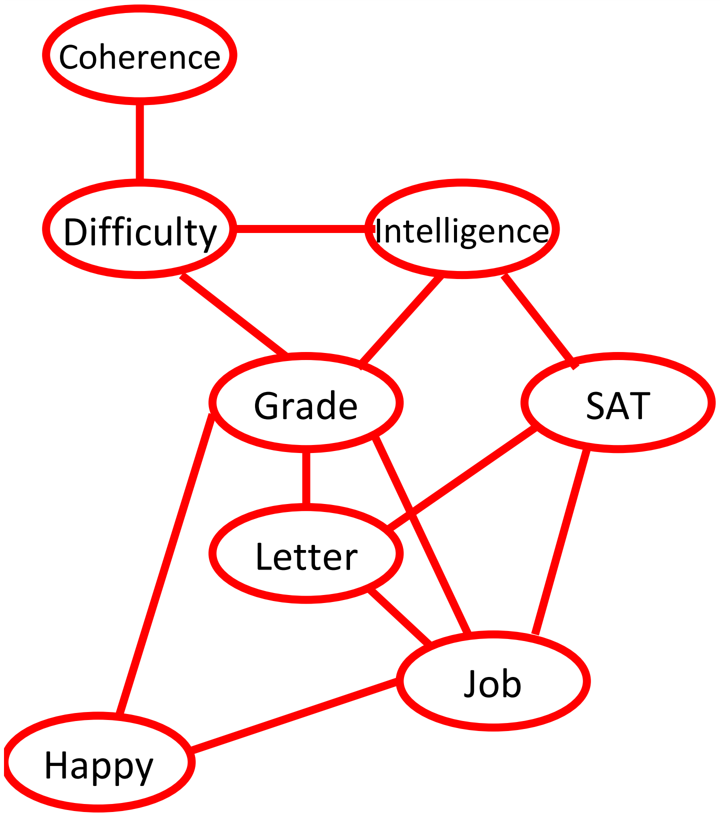
{G,C,D,S,I,L,H,J}



Max Clique =
{G,I,D,L,H,J}

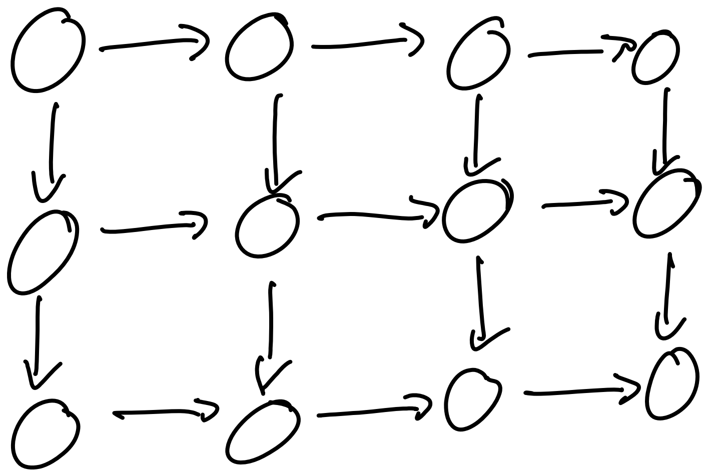
- Different elim. order induce different graphs! Treewidth 5

Induced graph and VE complexity



- **Theorem**:
  - All factors arising in VE are defined over cliques (fully connected subgraphs) of the induced graph
  - All maximal cliques of induced graph arise as factors in VE
- Treewidth for ordering = Size of largest clique of induced graph -1
- Treewidth of a graph = minimal treewidth under optimal ordering
- VE exponential in treewidth!

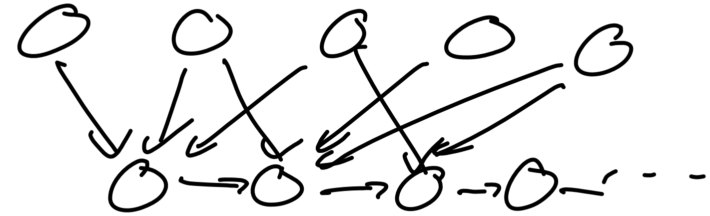
Compact representation ➔ small treewidth?



$n \times n$ grid

treewidth n

$\frac{\# \text{parents}}{\text{node}} = 2$



compact

$\# \text{parents} \leq 4$

treewidth ??

Finding the optimal elimination order

- **Theorem**: Deciding whether there exists an elimination order with induced with at most K is NP-hard
  - Proof by reduction from MAX-CLIQUE

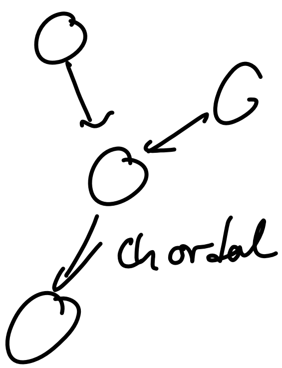
- In fact, can find elimination order in time exponential in treewidth
- Finding optimal ordering as hard as inference…

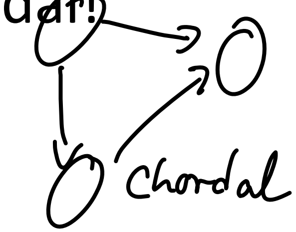
- For which graphs can we find optimal elimination order?

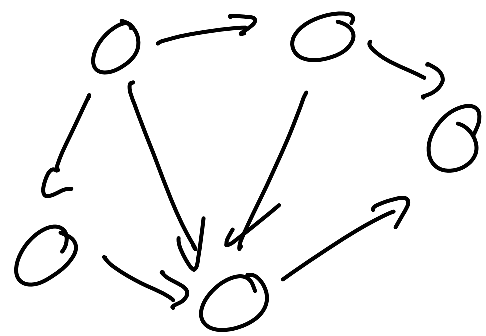
Finding optimal elimination order

- For poly trees can find optimal ordering (saw before)

- A graph is called chordal if every cycle of length ~~\geq~~ $\geq$ 4 has a chord (an edge between some pair of non-consecutive nodes)
  - Every tree is chordal!

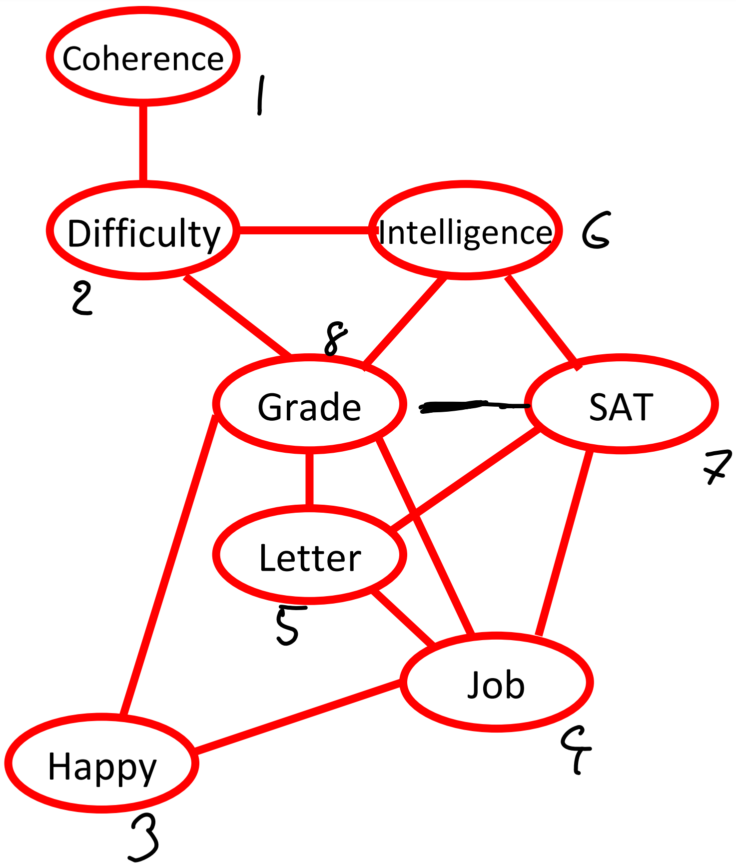






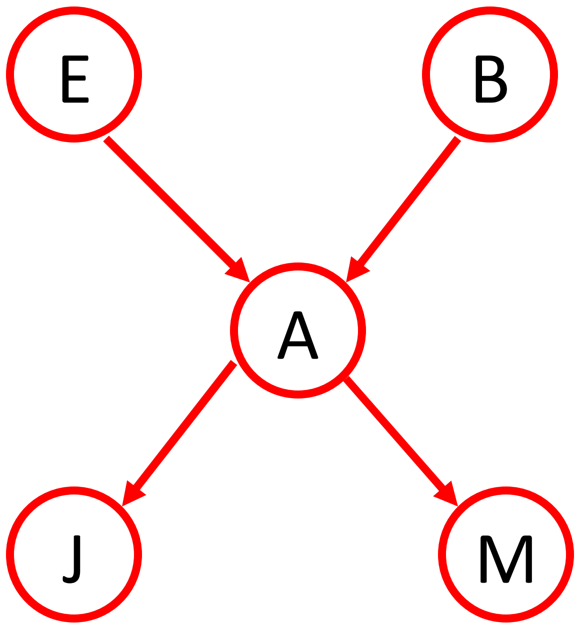
- Can find optimal elimination ordering for chordal graphs

Minimal fill heuristic



- Unmark all nodes
- For i = 1:n
  - Find unmarked node X such that adding X adds fewest additional edges
  - Set X as i-th var. in ordering
  - Fill in edges added by eliminating X
  - Mark X
- Often very effective!
- In fact, finds optimal order for chordal graphs!
- May want to weigh by factor size

Maxizimization queries



- MPE (Most probable explanation):

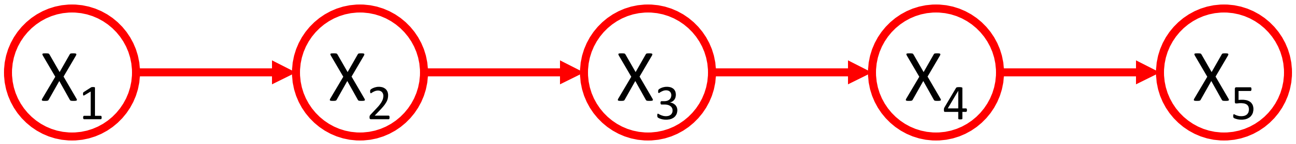
  Given values for some vars, compute most likely assignment to all remaining vars

Given J=F, M=T, find

$$(e^*, b^*, a^*) = \underset{e,b,a}{\operatorname{argmax}}\ P(J=F, M=T, e, b, a)$$

$$\underset{X_{\{1\dots n\}\setminus E}}{\operatorname{argmax}} P(x_1 \dots x_n \mid e) = \underset{x}{\operatorname{argmax}}\ P(x_1 \dots x_n, e)$$

Variable elimination for MPE

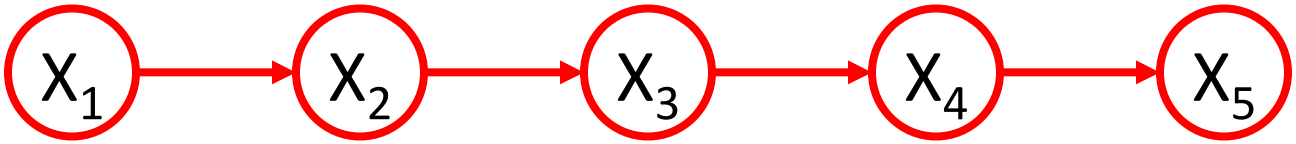


$$\max_{x_2 x_3 x_4} P(X_1)\, P(X_2|X_1)\, P(X_3|X_2)\, P(X_4|X_3)\, P(X_5|X_4)$$

$$= P(X_1) \underbrace{\max_{x_2} P(X_2|X_1) \underbrace{\max_{x_3} P(X_3|X_2) \underbrace{\max_{x_4} P(X_4|X_3)\, P(X_5|X_4)}_{g_4(X_3, X_5)}}_{g_3(X_2, X_5)}}_{g_2(X_1, X_5)}$$

How can we retrieve the argmax ??

Recovering the MPE



$$x_1^* = \arg\max_{x_1} P(x_1)\, g_2(x_1, x_5)$$

$$x_2^* = \arg\max_{x_2} \cancel{P(x_1^*)}\, P(x_2 \mid x_1^*)\, g_3(x_2, x_5)$$

$$x_3^* = \arg\max_{x_3} P(x_3 \mid x_2^*)\, g_4(x_3, x_5)$$

$$\vdots$$

Variable elimination for MPE: Forward pass

- Given BN and MLE query max P(x_1,…,x_n | **E**=**e**)
- Choose an ordering of $X_1,\ldots,X_n$
- Set up initial factors: $f_i = P(X_i \mid \mathbf{Pa_i})$
- For i =1:n, $X_i \notin \{X, \mathbf{E}\}$
  - Collect all factors f that include X_i
  - Generate new factor by maximizing over X_i

$$g = \max_{x_i} \prod_j f_j$$

  - Add g to set of factors

Variable elimination for MPE: Backward pass

- Variables $x_1^*,\ldots,x_n^*$ will contain MPE
- For $i = \underline{n:\overset{-1:}{1}}$, $X_i \notin \{X,\mathbf{E}\}$
  - Take factors $f_1,..,f_m$ used when eliminating X_i
  - Plug in values $x_{i+1}^*,\ldots,x_n^*$ into these factors

$$x_i^* = \underset{x_i}{\text{argmax}} \prod_j f_j(x_i, x_{i+1}^*, \ldots, x_n^*)$$

can only depend on x_i

Maximizing factors

$$g = \max_{x_i} \underbrace{\prod_j f_j}_{f'}$$

f'

A\B	T	F
T	.	.
F	.	.

$$g = \max_A f'$$

$$g(B) = \max_A f'(A, B)$$

Summary so far

- Variable elimination complexity exponential in induced width for elimination ordering
- Finding optimal ordering is NP-hard
- Many good heuristics
  - Exact for trees, chordal graphs
- Ultimately, inference is NP-hard
- Only difference between cond. prob. queries and MPE is $\sum$ vs. max

- Variable elimination building block for many exact and approximate inference techniques

Tasks

- Homework 2 due in class Wednesday Nov 4