Machine Learning

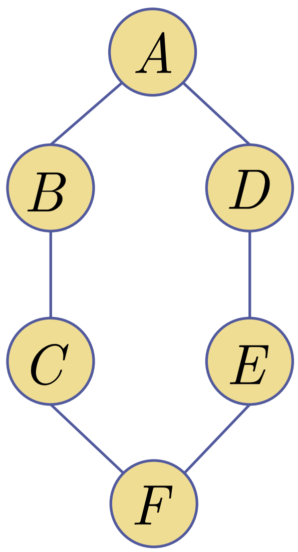
4771

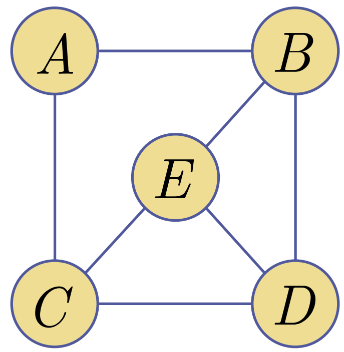
Instructor: Tony Jebara

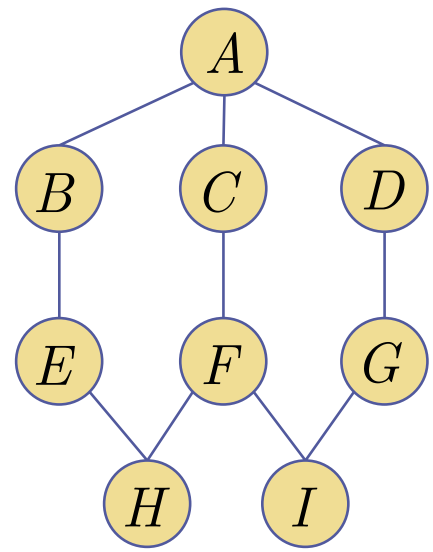
Topic 17

- Triangulation Examples
- Running Intersection Property
- Building a Junction Tree
- The Junction Tree Algorithm

Triangulation Examples

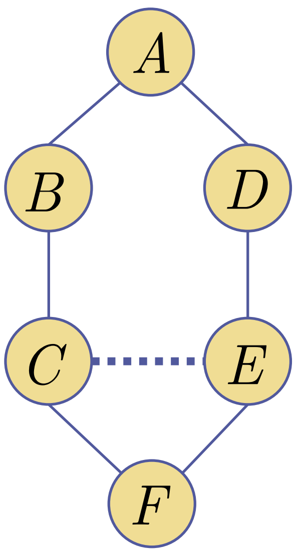


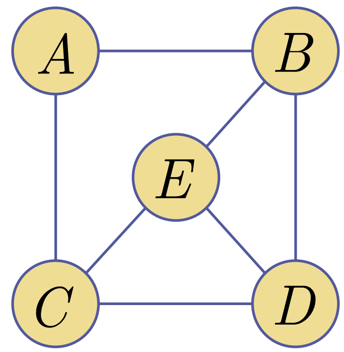


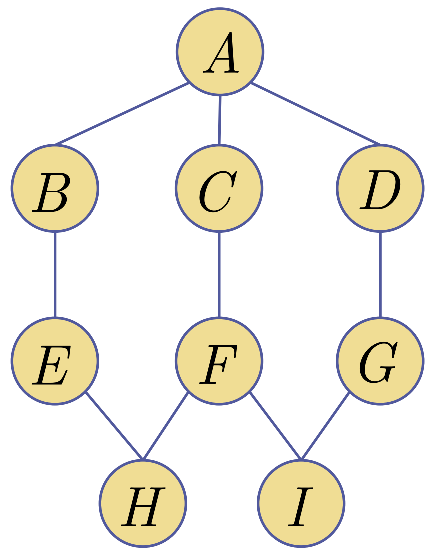


- Cycle: A closed (simple) path, with no repeated vertices other than the starting and ending vertices
- Chordless Cycle: a cycle where no two non-adjacent vertices on the cycle are joined by an edge.
- Triangulated Graph: a graph that contains no chordless cycle of four or more vertices (aka a Chordal Graph).

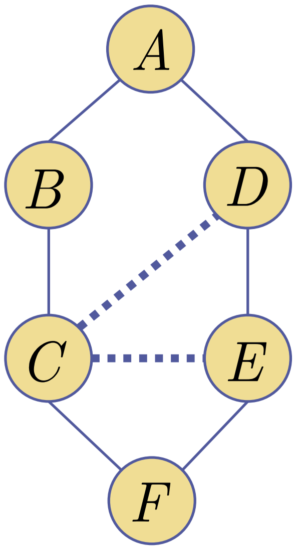
Triangulation Examples

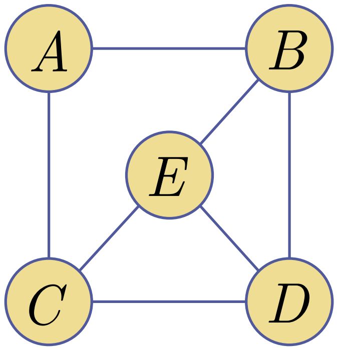


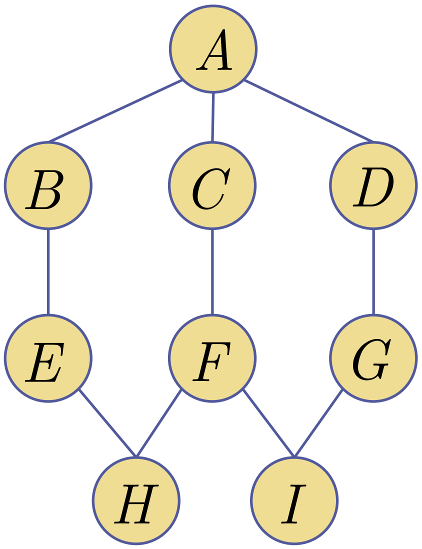




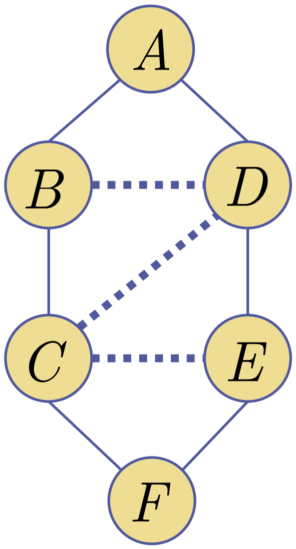
Triangulation Examples

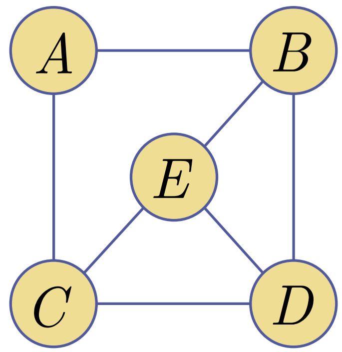


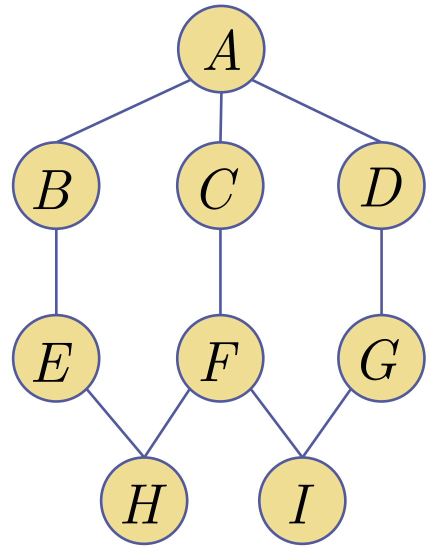




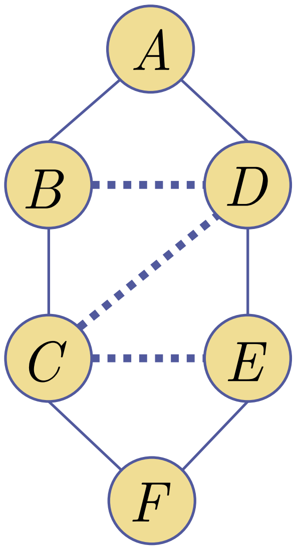
Triangulation Examples

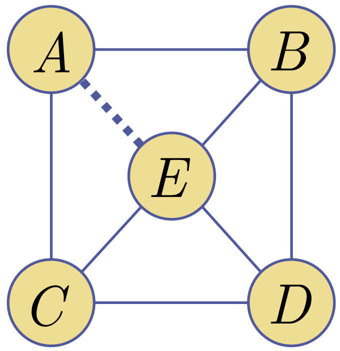


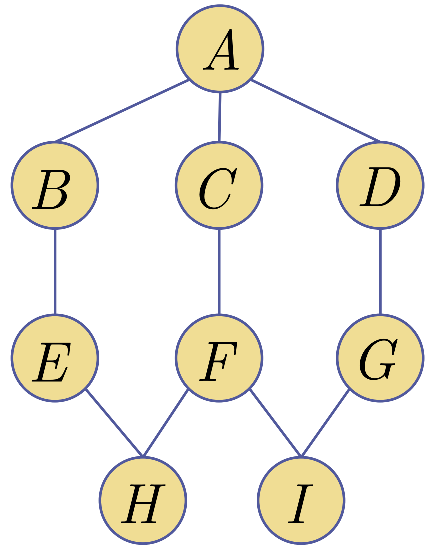




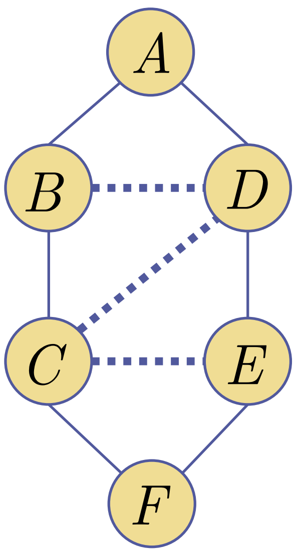
Triangulation Examples

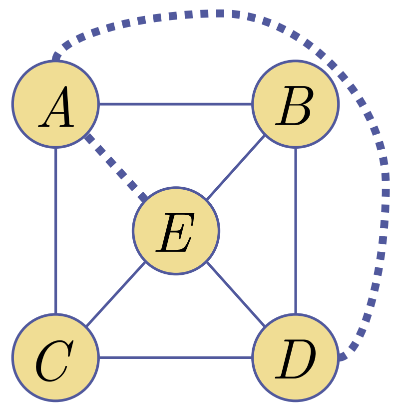


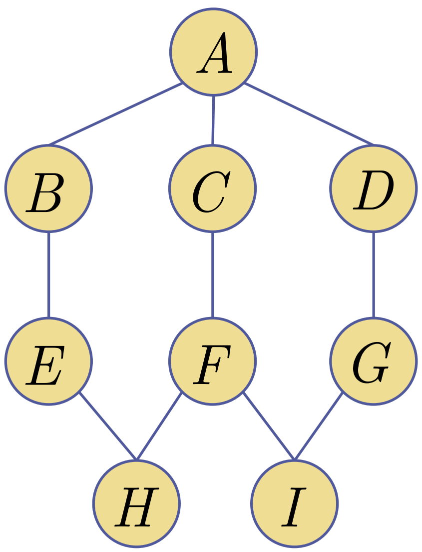




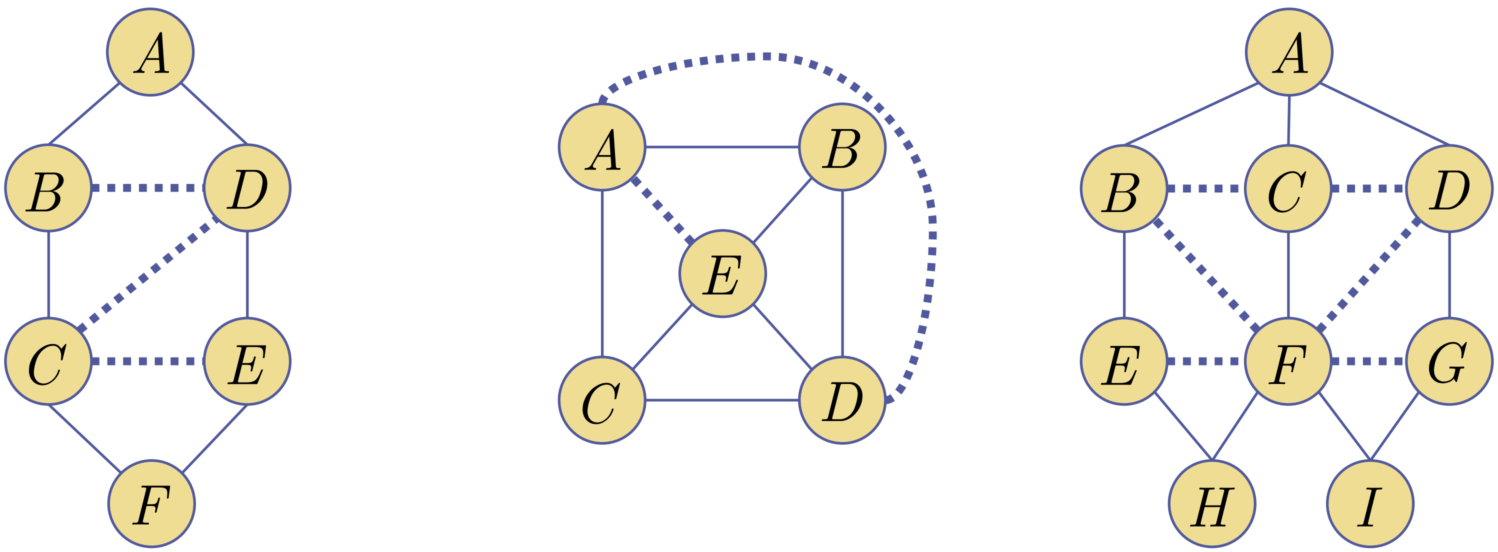
Triangulation Examples



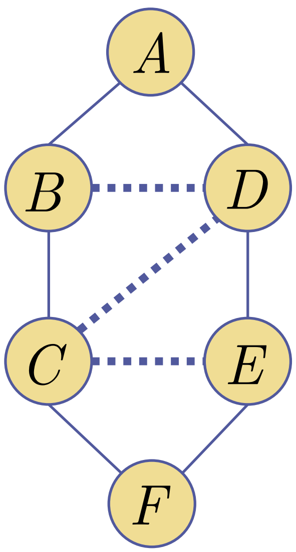


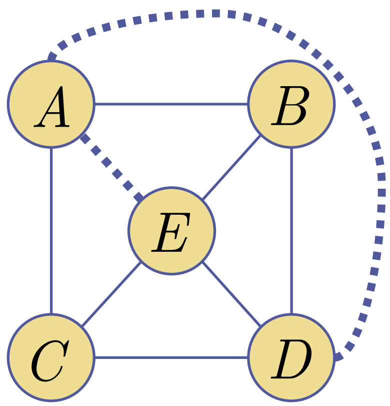


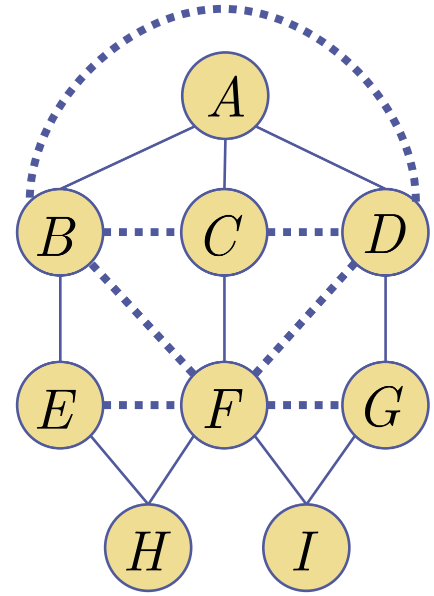
Triangulation Examples



Triangulation Examples

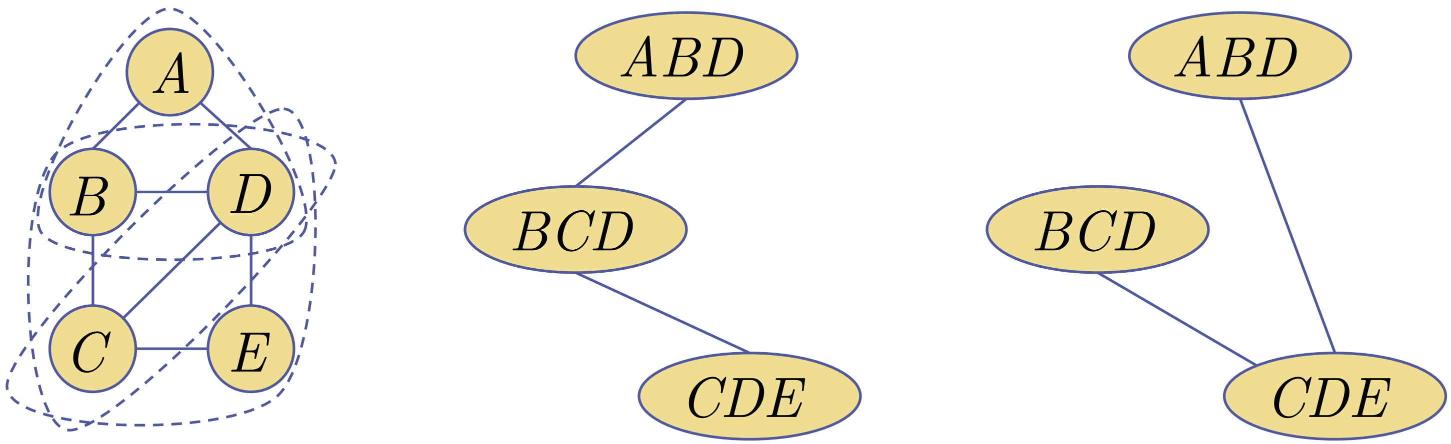






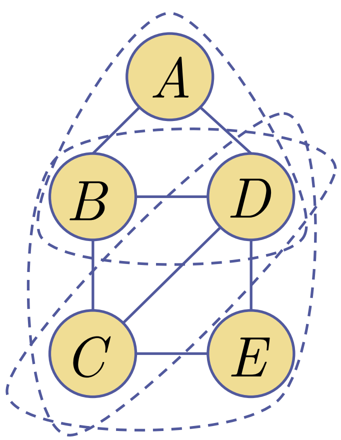
Running Intersection Property

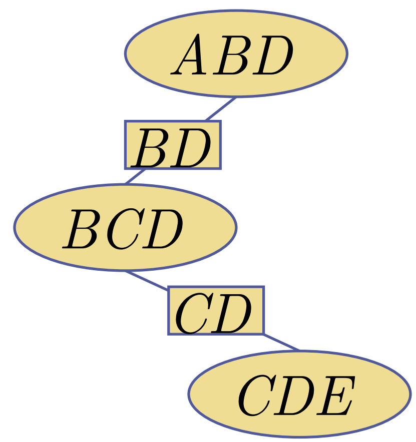
- Junction Tree must satisfy Running Intersection Property
- RIP: On unique path connecting clique V to clique W, all other cliques share nodes in $V \cap W$

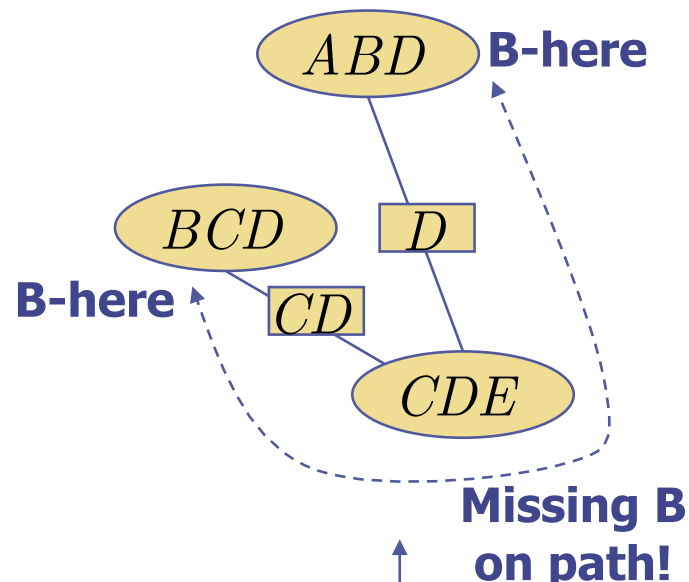


Running Intersection Property

- Junction Tree must satisfy Running Intersection Property
- RIP: On unique path connecting clique V to clique W, all other cliques share nodes in $V \cap W$







HINT: Junction Tree has largest total separator cardinality

$$\left|\Phi\right| = \left|\phi\left(B,D\right)\right| + \left|\phi\left(C,D\right)\right| = 2+2$$

$$\left|\Phi\right| = \left|\phi\left(C,D\right)\right| + \left|\phi\left(D\right)\right| = 2+1$$

Forming the Junction Tree

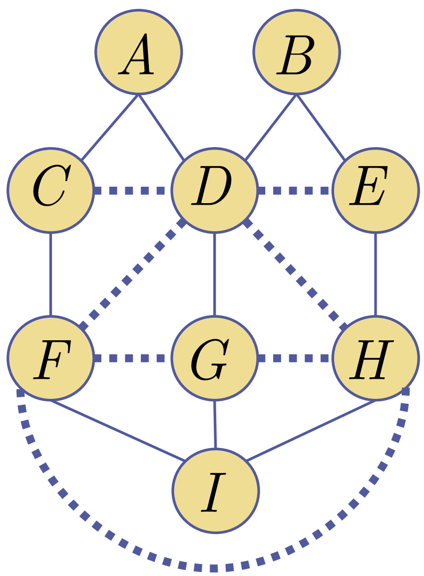
- Goal: connect k cliques into a tree… k^{k-2} possibilities!
- For each, check Running Intersection Property, too slow…
- Theorem: a valid (RIP) Junction Tree connection is one that maximizes the cardinality of the separators

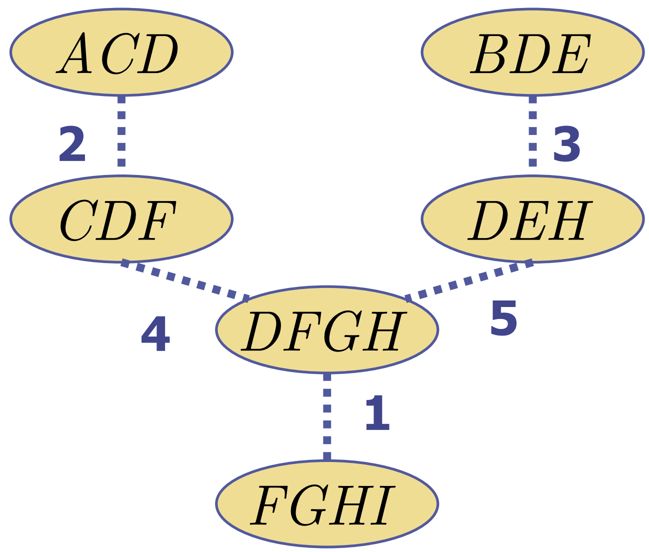
$$JT^* = \arg\max_{TREE\,STRUCTURES} \left|\Phi\right|$$
$$= \arg\max_{TREE\,STRUCTURES} \sum_S \left|\phi\left(X_S\right)\right|$$

- Use very fast Kruskal algorithm:
  1) Init Tree with all cliques unconnected (no edges)
  2) Compute size of separators between all pairs
  3) Connect the two cliques with the biggest separator cardinality which doesn't create a loop in current Tree (maintains Tree structure)
  4) Stop when all nodes are connected, else goto 3

Kruskal Example

- Start with unconnected cliques (after triangulation)





	ACD	BDE	CDF	DEH	DFGH	FGHI
ACD	-	1	2	1	1	0
BDE		-	1	2	1	0
CDF			-	1	2	1
DEH				-	2	1
DFGH					-	3
FGHI						-

Junction Tree Probabilities

- We now have a valid Junction Tree!
- What does that mean?
- Recall probability for undirected graphs:

$$p(X) = p(x_1, \ldots, x_M) = \tfrac{1}{Z} \prod_C \psi(X_C)$$

- Can write junction tree as potentials of its cliques:

$$p(X) = \tfrac{1}{Z} \prod_C \tilde{\psi}(X_C)$$

- Alternatively: clique potentials over separator potentials:

$$p(X) = \frac{1}{Z} \frac{\prod_C \psi(X_C)}{\prod_S \phi(X_S)}$$

- This doesn't change/do anything! Just less compact…
- Like *de-absorbing* smaller cliques from maximal cliques:

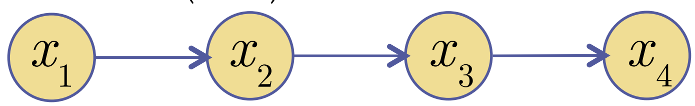
$$\tilde{\psi}(A,B,D) = \frac{\psi(A,B,D)}{\phi(B,D)}$$ ← **…gives back original formula if** $\phi(B,D) \triangleq 1$

Junction Tree Probabilities

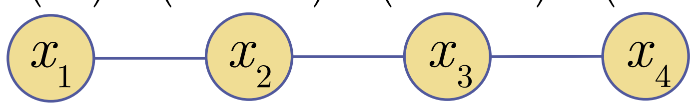
- Can quickly converted directed graph into this form:

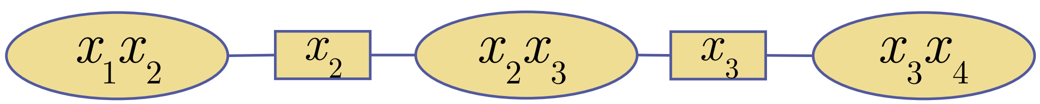
$$p(X) = \frac{1}{Z}\frac{\prod_C \psi(X_C)}{\prod_S \phi(X_S)}$$

- Example:



$$p(X) = p(x_1)\,p(x_2 \mid x_1)\,p(x_3 \mid x_2)\,p(x_4 \mid x_3)$$





$$p(X) = \frac{1}{1}\frac{p(x_1,x_2)\,p(x_3 \mid x_2)\,p(x_4 \mid x_3)}{1 \times 1}$$

$$= \frac{1}{Z}\frac{\psi(x_1,x_2)\,\psi(x_2,x_3)\,\psi(x_3,x_4)}{\phi(x_2)\,\phi(x_3)}$$

By inspection, can just cut & paste CPTs as clique and separator potential functions

Junction Tree Algorithm

- Running the JTA converts clique potentials & separator potentials into marginals over their variables … and does not change p(X)

$$\psi(A,B,D) \to p(A,B,D)$$
$$\phi(B,D) \to p(B,D)$$
$$\psi(B,C,D) \to p(B,C,D)$$

- Don't want just normalization!

$$\frac{\psi(A,B,D)}{\sum_{A,B,D} \psi(A,B,D)} \neq p(A,B,D)$$

- These marginals should all agree & be consistent

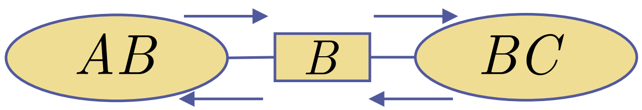
$$\psi(A,B,D) \to p(A,B,D) \qquad \to \sum_A p(A,B,D) = \tilde{p}(B,D)$$
$$\phi(B,D) \to p(B,D) \qquad \to p(B,D)$$
$$\psi(B,C,D) \to p(B,C,D) \qquad \to \sum_C p(B,C,D) = \tilde{\tilde{p}}(B,D)$$

ALL EQUAL

- Consistency: all distributions agree on submarginals
- JTA sends messages between cliques & separators dividing each by the others marginals until consistency…

Junction Tree Algorithm

- Send message from each clique *to* its separators of what it thinks the submarginal on the separator is.
- Normalize each clique by incoming message *from* its separators so it agrees with them



$$V = \{A, B\} \quad S = \{B\} \quad W = \{B, C\}$$

If agree: $\sum_{V \setminus S} \psi_V = \phi_S = p(S) = \phi_S = \sum_{W \setminus S} \psi_W$ **...Done!**

Else:

Send message From V to W...

$$\phi_S^* = \sum_{V \setminus S} \psi_V$$

$$\psi_W^* = \frac{\phi_S^*}{\phi_S} \psi_W$$

$$\psi_V^* = \psi_V$$

Send message From W to V...

$$\phi_S^{**} = \sum_{W \setminus S} \psi_W^*$$

$$\psi_V^{**} = \frac{\phi_S^{**}}{\phi_S^*} \psi_V^*$$

$$\psi_W^{**} = \psi_W^*$$

Now they Agree...Done!

$$\sum_{V \setminus S} \psi_V^{**} = \sum_{V \setminus S} \frac{\phi_S^{**}}{\phi_S^*} \psi_V^*$$

$$= \frac{\phi_S^{**}}{\phi_S^*} \sum_{V \setminus S} \psi_V^*$$

$$= \phi_S^{**} = \sum_{W \setminus S} \psi_W^{**}$$

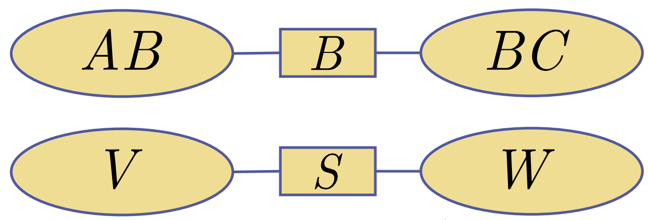
Junction Tree Algorithm

- When “Done”, all clique potentials are marginals and all separator potentials are submarginals!
- Note that p(X) is unchanged by message passing step:

$$\phi_S^* = \sum\nolimits_{V \setminus S} \psi_V$$

$$\psi_W^* = \frac{\phi_S^*}{\phi_S} \psi_W$$

$$\psi_V^* = \psi_V$$



$$p\left(X\right) = \frac{1}{Z} \frac{\psi_V^* \psi_W^*}{\phi_S^*} = \frac{1}{Z} \frac{\psi_V \frac{\phi_S^*}{\phi_S} \psi_W}{\phi_S^*} = \frac{1}{Z} \frac{\psi_V \psi_W}{\phi_S}$$

- Potentials set to conditionals (or slices) become marginals!

$$\psi_{AB} = p\left(B \mid A\right) p\left(A\right) = p\left(A,B\right) \longrightarrow \phi_B^* = \sum\nolimits_A \psi_{AB} = \sum\nolimits_A p\left(A,B\right) = p\left(B\right)$$

$$\psi_{BC} = p\left(C \mid B\right) \longrightarrow \psi_{BC}^* = \frac{\phi_S^*}{\phi_S} \psi_{BC} = \frac{p\left(B\right)}{1} p\left(C \mid B\right) = p\left(B,C\right)$$

$$\phi_B = 1$$