Factor graphs, cont'd

BP on FGs: Input $G=(V\cup F, E)$, T

Output $\{\hat{P}(x_i)\}_{i\in V}$

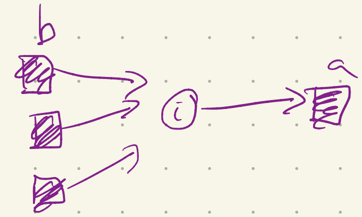
Initialize $\{(m_{i\to a}, m_{a\to i})\}_{(i,a)\in E} = 1$ or random

$\{[\;]\}^{|X|}$

For $t=1 \dots T$

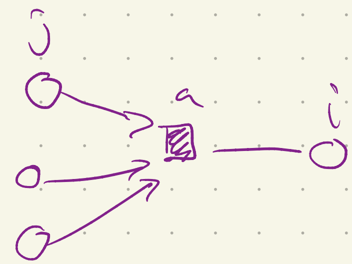
Update $m_{i\to a}$:

$$m_{i\to a}(x_i) = \prod_{b\in\delta i\setminus\{a\}} \tilde{m}_{a\to i}(x_i)$$



Update $\tilde{m}_{a\to i}$:

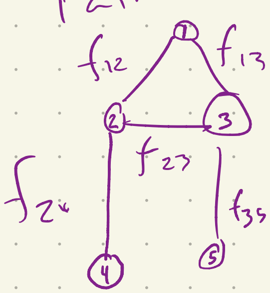
$$\tilde{m}_{a\to i}(x_i) = \sum_{\delta a\setminus\{i\}} f_a(x_{\delta a}) \prod_{j\in\delta a\setminus\{i\}} m_{j\to a}(x_j)$$



Compute Marginal: $m_i(x_i) = \prod_{a\in\delta i} \tilde{m}_{a\to i}(x_i)$

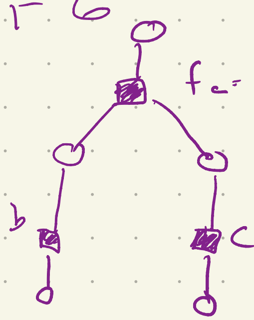
Claim: $\nearrow$ is exact if FG is a tree & $T > TW$

Pairwise MRF $O(|X|^2)$



BP is apx, the loop causes issues

FG



$f_a = f_{12} \cdot f_{13} \cdot f_{23}$

BP is exact

$O(|X|^{\text{max-degree}})$

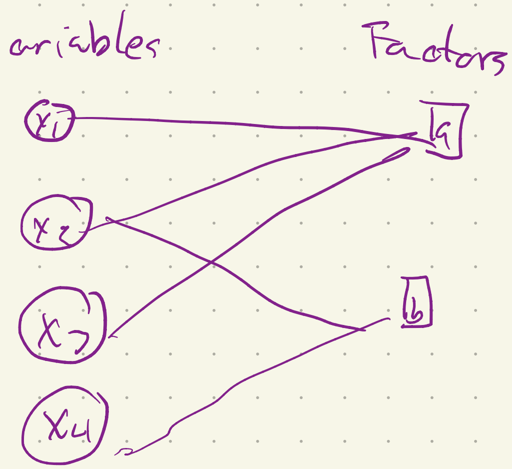
Example: Decoding LDPCs (low density parity check) Codes w BP

Def [LDPC] are a family of codes defined as

$$G = (V \cup F, E)$$

with factors for parity checking

$$f_a(\underbrace{x_{\partial a}}_{= x_a}) = \mathbb{I}(x_1 \overset{\text{XOR}}{\oplus} x_2 \oplus x_3 = 0)$$



Def A Codebook $\{x \in \mathcal{X}^n \mid \text{satisfy all parity checks}\}$ $\to$ $= \{0000,\ 0111,\ 1010,\ 1101\}$

$$|\text{Codebook}| = 2^{|V|-|F|} := 2^K$$

To transmit, one $x \in$ Codebook sent over a noisy channel whose behavior is defined as $P[Y_i \mid X_i]$

	$X_i = 1$	$X_i = 0$
$Y_i = 1$	.7	.3
$Y_i = 0$	.3	.7

& we need to recover X_i from Y_i.

Strategy: Use BP to estimate $P[X_i \mid Y_1, \dots, Y_n]$ $\forall i \in [n]$

& put $$\widehat{X_i} = \left(\text{sign}\left(\log \frac{P[X_i = 1 \mid Y]}{P[X_i = 0 \mid Y]}\right) + 1\right) \cdot \frac{1}{2} \in \{0, 1\}$$

Next: how to convert a BN into a FG easily

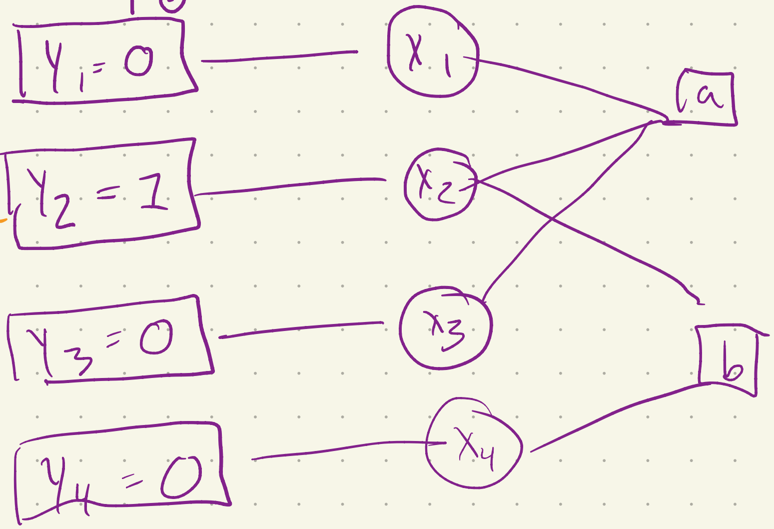
- How to include observed vars in inference + modeling

Example

$$P[X \mid Y] \propto P[X, Y] = \frac{1}{Z} \underbrace{\prod_{a \in F} \mathbb{I}[\oplus x_a = 0]}_{\text{look only for sent code word}} \cdot \underbrace{\prod_{i \in V} P[Y_i \mid X_i]}_{\text{Singleton Factor}}$$

X_i \ Y_i:	0	1
0	.7	
1	.3	
0		.3
1		.7
0	.7	
1	.3	
0	.7	
1	.3	

FG

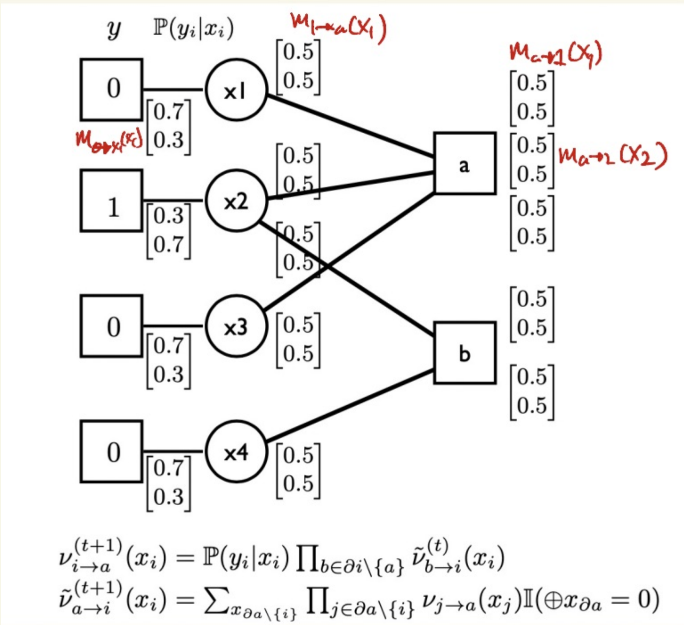


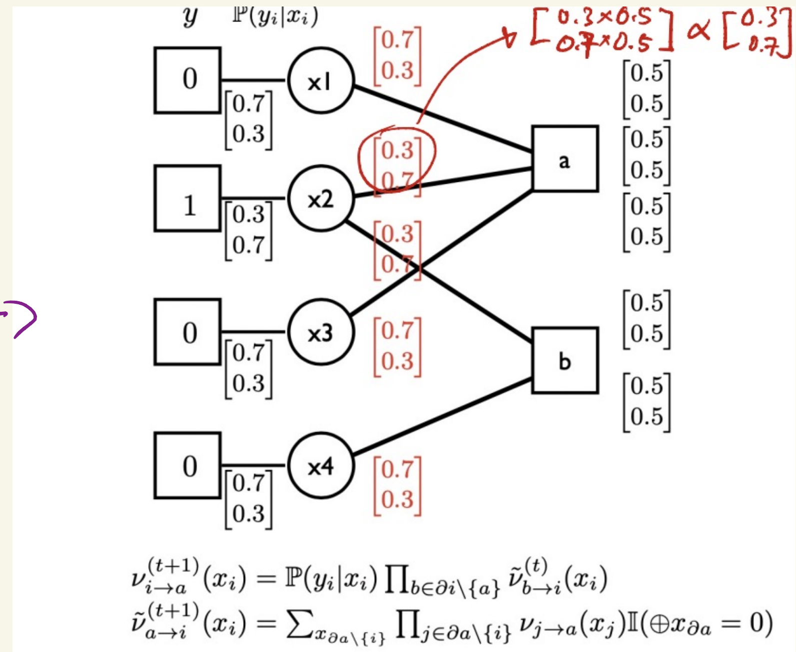
$$f_a = \begin{array}{c|cccc} x_2,x_3 & 00 & 01 & 10 & 11 \\ \hline x_i=0 & & & & \\ x_i=1 & & & & \end{array} \quad \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

$$f_b = \begin{array}{c|cc} x_4 \backslash x_2 & 0 & 1 \\ \hline 0 & & \\ 1 & & \end{array} \quad \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$m_{i \to a}(x_i) = P[Y_i \mid X_i] \cdot \prod_{b \in \partial a \setminus \{i\}} \tilde{m}_{b \to i}(x_i)$$

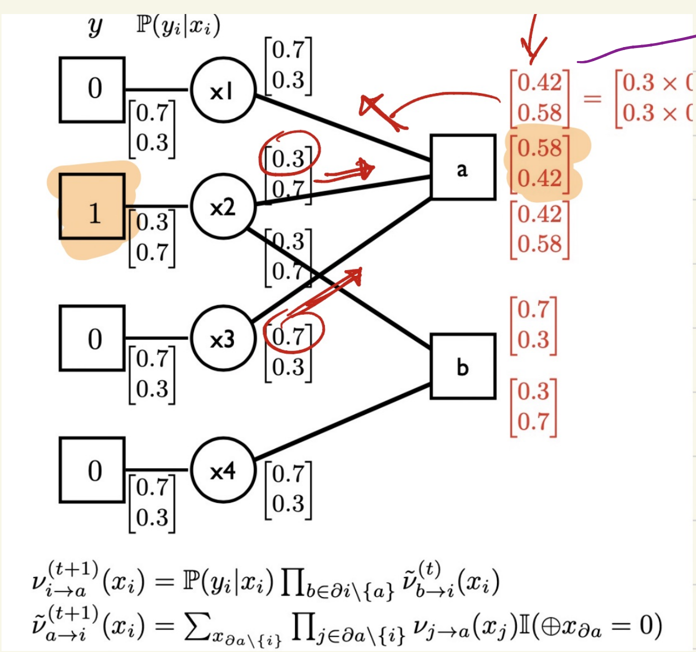
$$m_{a \to i}(x_i) = \sum_{x_{\partial a \setminus \{i\}}} \prod_{j \in \partial a \setminus \{i\}} m_{j \to a}(x_j) \, \mathbb{I}[\oplus x_{\partial a} = 0]$$





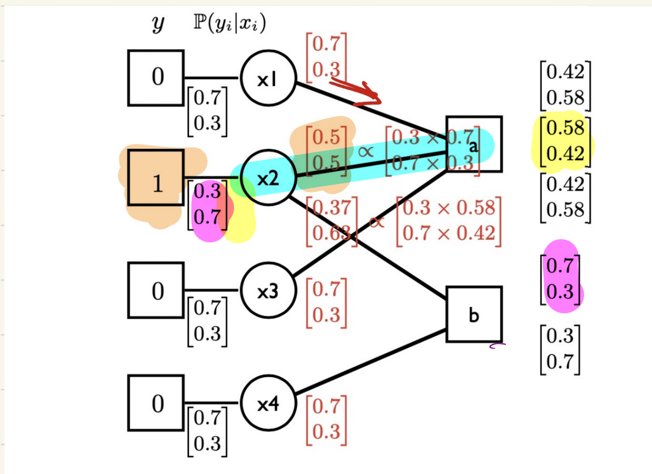
$m_{i\to a}$ update

- Normalization within message doesn't matter
- Num. stability then better if normalize after each update



$$\tilde m_{a\to 1}(x_1) = \sum_{\underbrace{x_{\partial a\setminus\{1\}}}_{x_{2,3}}} \prod_{j\in\partial a\setminus\{1\}} m_{j\to a}(x_j)\,\mathbb{1}[\oplus x_{\partial a}=0]$$

$$= [x_2=x_3=0]\ .3\cdot .7\cdot \mathbb{1}[x_2\oplus x_3=0]$$
$$[x_2=x_3=1] + \ .7\cdot .3\cdot \mathbb{1}[x_2\oplus x_3=0]$$
$$[x_2=1, x_3=0] + \ .7\cdot .7\cdot \cancel{\mathbb{1}[x_2\oplus x_3=0]}$$
$$[x_3=1, x_2=0] + \ .3\cdot .3\ \cancel{\mathbb{1}[x_2\oplus x_3=0]}$$
$$= .3\cdot .7\cdot 2 = .42$$

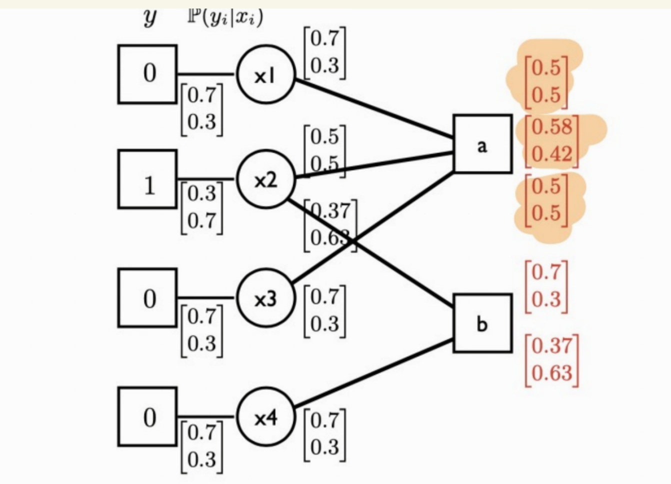


$$\nu_{i\to a}^{(t+1)}(x_i) = \mathbb{P}(y_i|x_i) \prod_{b\in\partial i\setminus\{a\}} \tilde{\nu}_{b\to i}^{(t)}(x_i)$$
$$\tilde{\nu}_{a\to i}^{(t+1)}(x_i) = \sum_{x_{\partial a\setminus\{i\}}} \prod_{j\in\partial a\setminus\{i\}} \nu_{j\to a}(x_j)\mathbb{I}(\oplus x_{\partial a} = 0)$$

$$m_{2\to a}(x_2) = P[y_2|x_2] \prod_{b\in\partial 2\setminus\{a\}} \tilde{m}_{b\to 2}(x_2)$$

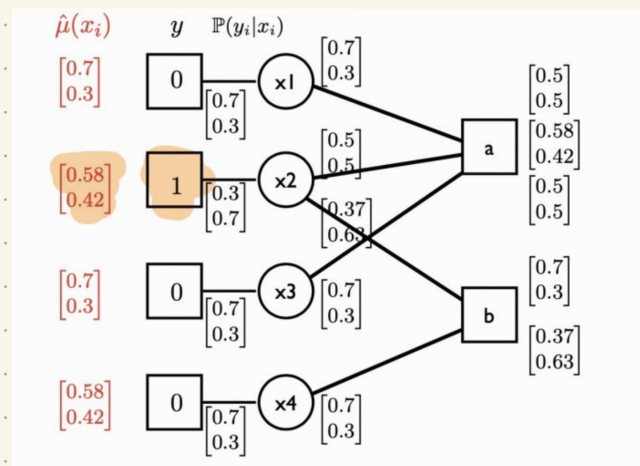
$$m_{2\to b}(x_2) = P[y_2|x_2] \prod_{a\in\partial 2\setminus\{b\}} \tilde{m}_{a\to 2}(x_2)$$

$\Downarrow$



$$\nu_{i\to a}^{(t+1)}(x_i) = \mathbb{P}(y_i|x_i) \prod_{b\in\partial i\setminus\{a\}} \tilde{\nu}_{b\to i}^{(t)}(x_i)$$
$$\tilde{\nu}_{a\to i}^{(t+1)}(x_i) = \sum_{x_{\partial a\setminus\{i\}}} \prod_{j\in\partial a\setminus\{i\}} \nu_{j\to a}(x_j)\mathbb{I}(\oplus x_{\partial a} = 0)$$

$\Rightarrow$



$$\nu_{i\to a}^{(t+1)}(x_i) = \mathbb{P}(y_i|x_i) \prod_{b\in\partial i\setminus\{a\}} \tilde{\nu}_{b\to i}^{(t)}(x_i)$$
$$\tilde{\nu}_{a\to i}^{(t+1)}(x_i) = \sum_{x_{\partial a\setminus\{i\}}} \prod_{j\in\partial a\setminus\{i\}} \nu_{j\to a}(x_j)\mathbb{I}(\oplus x_{\partial a} = 0)$$

Codebook {0000, 0111, 1010, 1101} $\longleftrightarrow$ min hamming distance

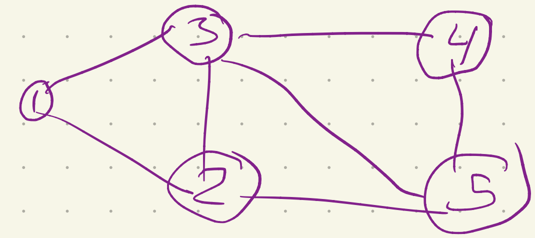
$\boxed{\begin{array}{l}\text{Output } 0000 \\ \text{Received } 0100\end{array}}$

Junction Tree Alg : Elim on A junction $\boxed{\text{tree}}$ → exact

- Fixing the tree, inference is easy using most methods
- Similar to elimination alg [exact, ordering]
- the data structure is meant to support efficient elimination

- MRF ⟶ Clique tree
  [non-unique]

$\boxed{M}$



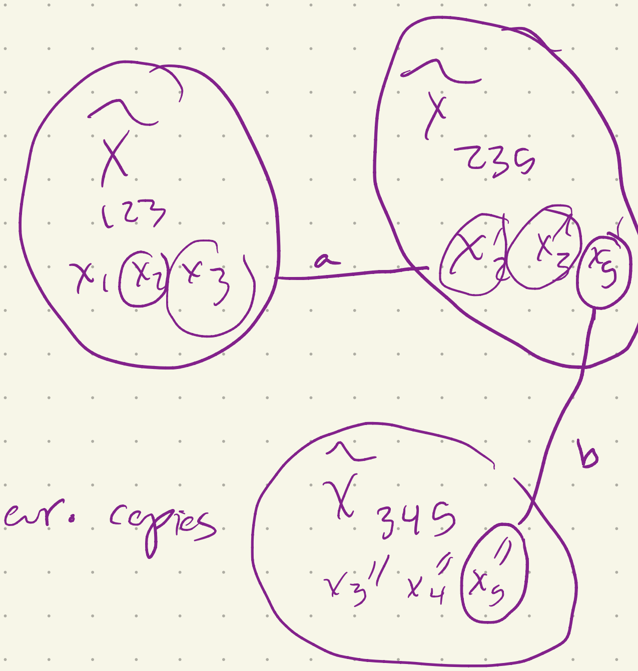
One example clique tree for M

$$P(X) \propto f_{123}\, f_{235}\, f_{345}$$

– Create a joint node for each clique
$\tilde{x}_C \in \mathcal{X}^{|C|}$

- each node has a local copy of its vars

$$\tilde{x}_{123} = (x_1, x_2, x_3)$$
$$\tilde{x}_{235} = (x_2', x_3', x_5')$$
$$\tilde{x}_{345} = (x_3'', x_4'', x_5'')$$



- Assign edges to form a tree
  & ensure consistency across var. copies

$$\tilde{P}(\tilde{x}_{123}, \tilde{x}_{235}, \tilde{x}_{345}) = \frac{1}{Z} f_{123}(\tilde{x}_{123}) \cdot f_{235}(\tilde{x}_{235}) f_{345}(\tilde{x}_{345}) \cdot$$

$$\mathbb{I}([\tilde{x}_{123}]_2 = [\tilde{x}_{235}]_1) \cdot \mathbb{I}([\tilde{x}_{123}]_3 = [\tilde{x}_{235}]_2) \cdot \mathbb{I}([\tilde{x}_{235}]_3 = [\tilde{x}_{345}]_3)$$

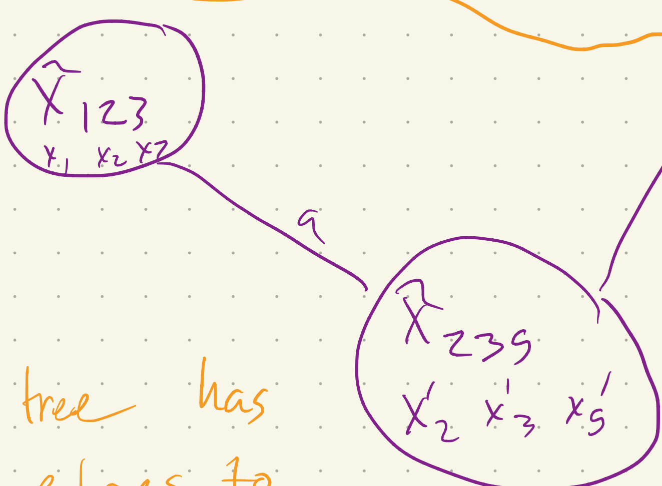
$$\mathbb{I}([\tilde{x}_{345}]_1 = [x_{235}]_2)$$

$$:= f_a(\tilde{x}_{123}, \tilde{x}_{235}) \cdot f_b(\tilde{x}_{345})$$

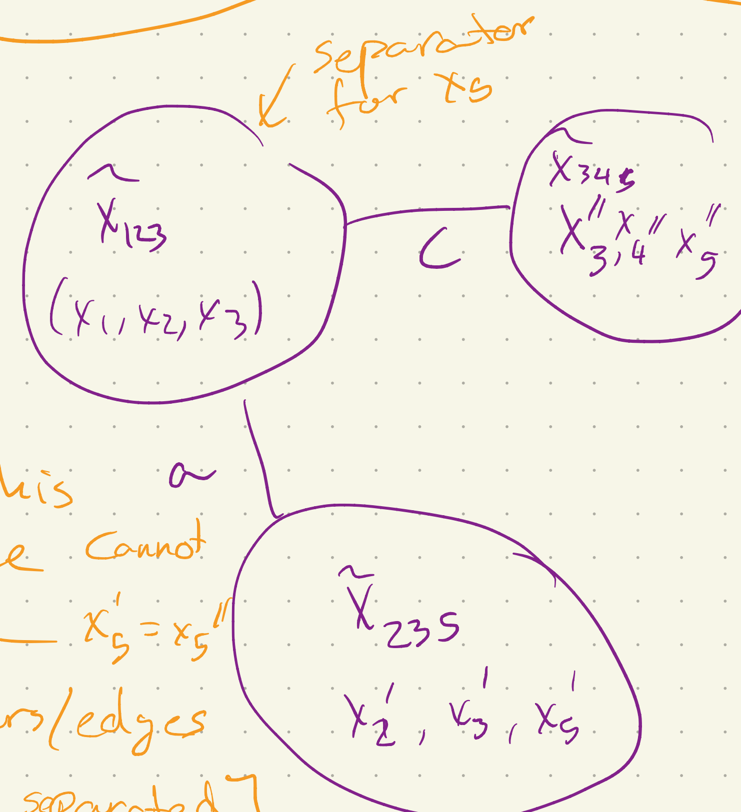
Consistency checks across edges

* If all local vars can be made consistent, we have global consistency

$$P(x_1) = \sum_{x_2 x_3} \tilde{P}(\tilde{x}_{123})$$



This tree has suff edges to guarantee global consistency



This tree cannot ensure $x_5' = x_5''$

[the factors/edges a+c are separated]

Q: When will a tree be globally consistent?

Def $\boxed{\text{Junction tree property}}$

For a MRF $G=(V,E)$ with $\mathcal{C}$ the set of maximal cliques

a tree T over $\mathcal{C}$, node $i \in V$ satisfies JTP if

all cliques containing i form a connected subtree $\in T$.

$\boxed{\text{Def}}$ T is a junction tree if $\forall i \in V$ it has JTP.

Q: When will G have a JT? [existence]

A: when G is chordal (any ≥ 4 loop has a chord)

Q How to find a JT for G? ① [construction]

Q If G isn't chordal, is all hope lost? ②

① Algo sketch to find JT given $G=(V,E)$, $\mathcal{C}$ the set of max cliques

- Consider K_n on $\mathcal{C}$
- Assign weights on edges base on # shared vars $\Leftrightarrow$ 2 cliques
- Find max-wt spanning tree [using, eg, Kruskal's alg]

$\boxed{\text{Claim}}$ A clique tree T is a junction tree $\Longleftrightarrow$ it's a max wt spanning tree

$$w(T) = \sum_{(C_e, C_k) \in T} |C_e \cap C_k|$$

$\forall$ trees T

$$= \sum_{(C_e, C_k) \in T} \sum_{i \in V} \mathbb{I}[i \in C_e \cap i \in C_k]$$

$$= \sum_{i \in V} \sum_{(C_\ell, C_k) \in T} \mathbb{I}[i \in C_\ell \wedge i \in C_k]$$

$$\le \sum_{i \in V} \sum_{(C_\ell, C_k) \in T} m_i - 1$$

m_i: # of cliques containing node i

Since cliques containing i is a subtree / forest on m_i nodes

$$\le \sum_{i \in V} (m_i - 1)$$

$=$ only when subgraph on i's cliques is connected

$\Rightarrow$ T is a junction tree $\Longleftarrow\!\Rightarrow$ T achieves max wt

D

② If G isn't chordal, what can you do?

Algo Sketch

- Choose an elim ordering
- Find reconstituted graph [chordal] $\Rightarrow$ Junction tree
- Construct complete clique graph
  Find max-wt spanning tree

We can run any exact algo on this JT and get exact inference.

Diff elim orderings $\rightarrow$ different [max clique sizes] / (complexities)

@ best $O(|X|^{\text{tree width}})$

How good is BP for $|X| < \infty$?

- When G is a tree, it's exact
- $\exists$ exactly 1 loop, it converges but might be wrong [Weiss 2000]
- Max-wt matching prob $\Rightarrow$ converges + exact [Bayati, Shah, Sharma 2005]
- In the limit of a large graph density estimation provides an asymptotic perf est.

Useful for

- Analysis of Compressed sensing

$$\min_{x \in \mathbb{R}^d} \|Ax - b\|^2 + \lambda \|x\|_1$$

$\hookrightarrow$ Underdetermined system $\qquad \rightarrow$ sparse

- Analysis of LDPC decoders, first provable capacity-achieving codes [Luby '98]
- Analysis of near-opt crowdsourcing [Karger Oh Shah 2011]