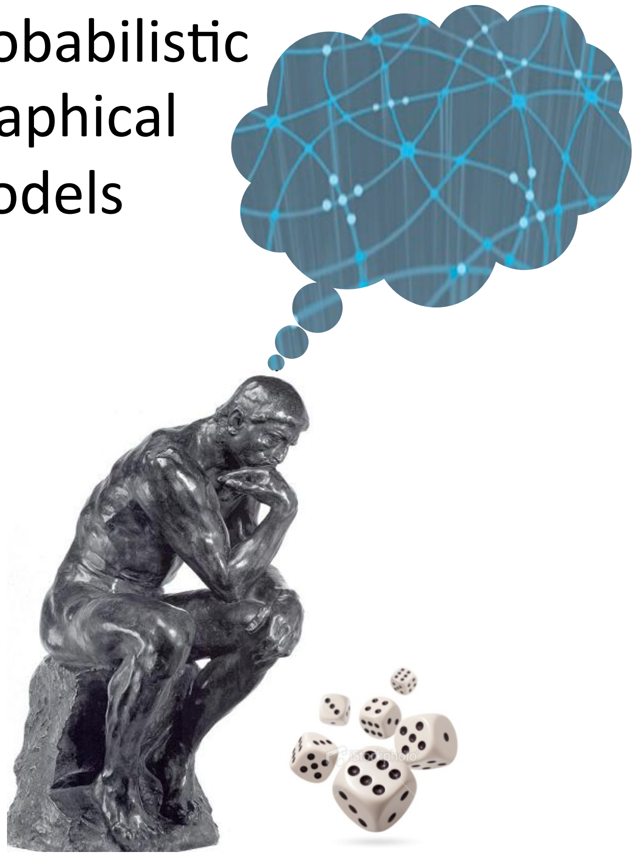
Probabilistic
Graphical
Models



Inference

MAP

Max-Sum Exact Inference

Product ⇨ Summation

$$P_\Phi(\boldsymbol{x}) \propto \prod_k \phi_k(\boldsymbol{D}_k)$$

$$\text{argmax} \prod_k \phi_k(\boldsymbol{D}_k)$$

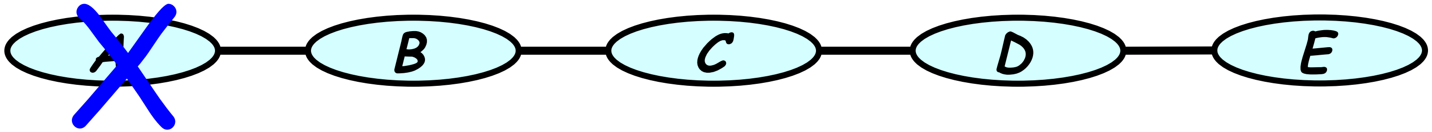
$$\text{argmax} \underbrace{\sum_k \theta_k(\boldsymbol{D}_k)}_{\theta(X_1,\ldots,X_n)}$$

a^1	b^1	8
a^1	b^2	1
a^2	b^1	0.5
a^2	b^2	2

⇩

a^1	b^1	3
a^1	b^2	0
a^2	b^1	-1
a^2	b^2	1

Max-Sum Elimination in Chains



$$\max_D \max_C \max_B \max_A \left(\theta_1(A,B) + \theta_2(B,C) + \theta_3(C,D) + \theta_4(D,E)\right)$$

$$\max_D \max_C \max_B \left(\theta_2(B,C) + \theta_3(C,D) + \theta_4(D,E) + \max_A \theta_1(A,B)\right)$$

$$\max_D \max_C \max_B \left(\theta_2(B,C) + \theta_3(C,D) + \theta_4(D,E) + \lambda_1(B)\right)$$

Factor Summation

a^1	b^1	3
a^1	b^2	0
a^2	b^1	-1
a^2	b^2	1

b^1	c^1	4
b^1	c^2	1.5
b^2	c^1	0.2
b^2	c^2	2

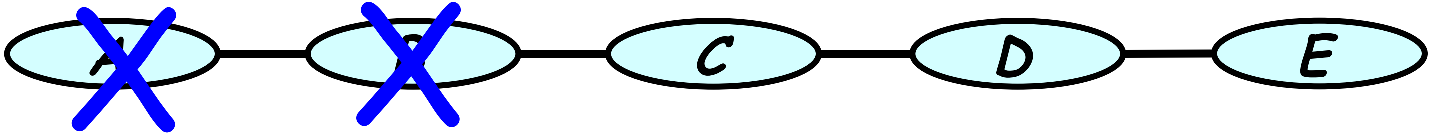
a^1	b^1	c^1	3+4=7
a^1	b^1	c^2	3+1.5=4.5
a^1	b^2	c^1	0+0.2=0.2
a^1	b^2	c^2	0+2=2
a^2	b^1	c^1	-1+4=3
a^2	b^1	c^2	-1+1.5=0.5
a^2	b^2	c^1	1+0.2=1.2
a^2	b^2	c^2	1+2=3

Factor Maximization

a^1	b^1	c^1	7
a^1	b^1	c^2	4.5
a^1	b^2	c^1	0.2
a^1	b^2	c^2	2
a^2	b^1	c^1	3
a^2	b^1	c^2	0.5
a^2	b^2	c^1	1.2
a^2	b^2	c^2	3

a^1	c^1	7
a^1	c^2	4.5
a^2	c^1	3
a^2	c^2	3

Max-Sum Elimination in Chains



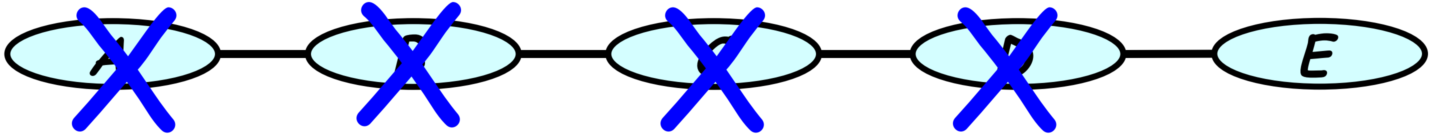
$$\max_D \max_C \max_B \left(\theta_2(B,C) + \theta_3(C,D) + \theta_4(D,E) + \lambda_1(B)\right)$$

$$\max_D \max_C \left(\theta_3(C,D) + \theta_4(D,E) + \max_B \left(\theta_2(B,C) + \lambda_1(B)\right)\right)$$

$$\max_D \max_C \left(\theta_3(C,D) + \theta_4(D,E) + \lambda_2(C)\right)$$

Daphne Koller

Max-Sum Elimination in Chains

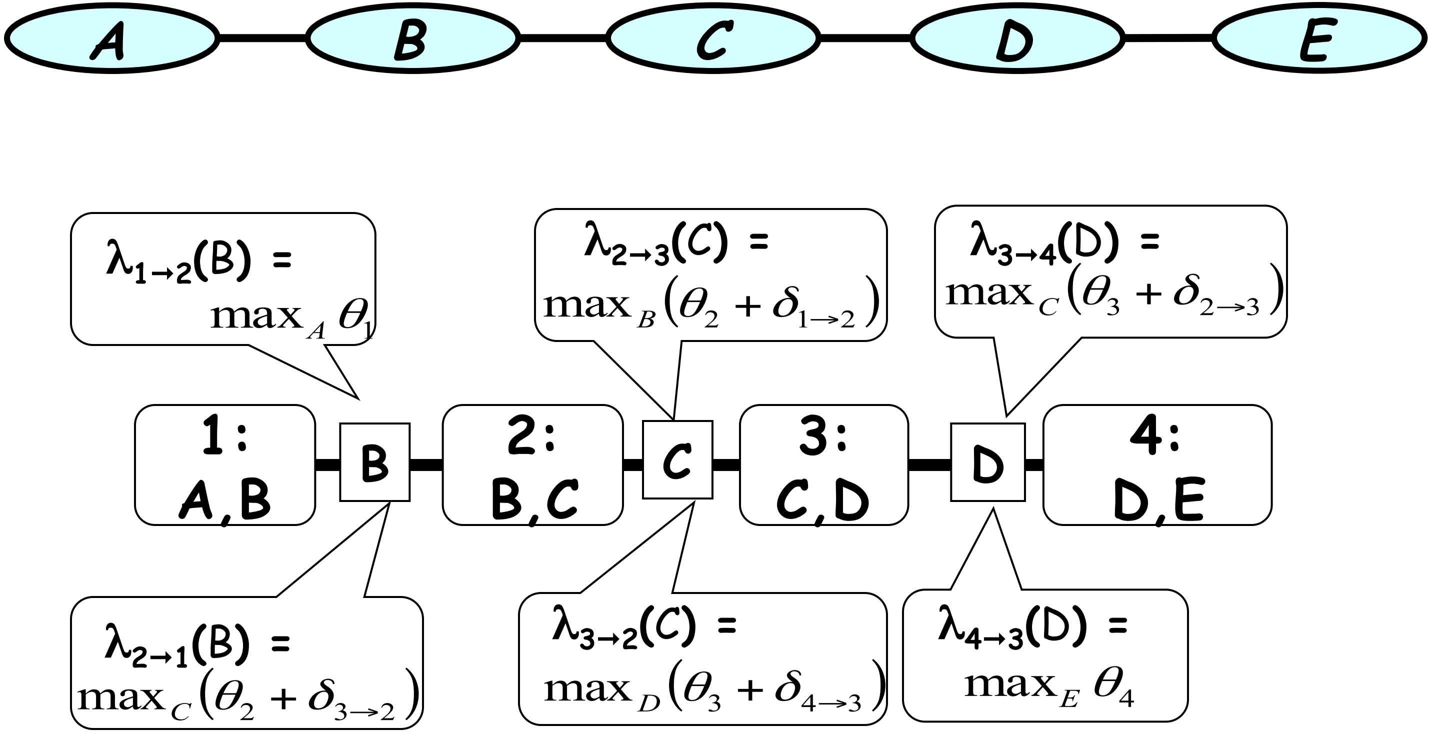


$$\max_D \max_C \left(\theta_3(C,D) + \theta_4(D,E) + \lambda_2(C)\right)$$

$$\max_D \left(\theta_4(D,E) + \lambda_3(D)\right)$$

$$\lambda_4(E) \qquad \lambda_4(e)$$

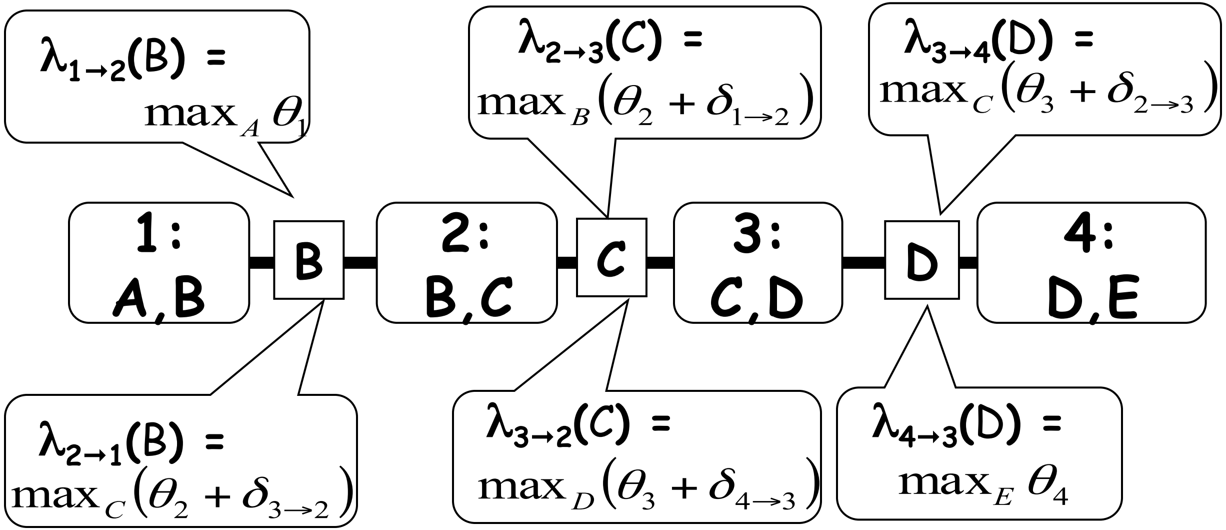
Max-Sum in Clique Trees



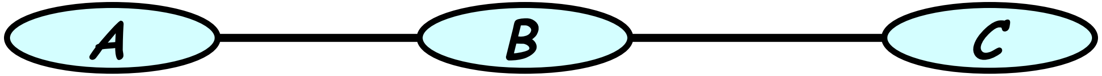
Daphne Koller

Convergence of Message Passing

- Once C_i receives a final message from all neighbors except C_j, then $\lambda_{i \to j}$ is also final (will never change)
- Messages from leaves are immediately final



Simple Example



a^1	b^1	3
a^1	b^2	0
a^2	b^1	-1
a^2	b^2	1

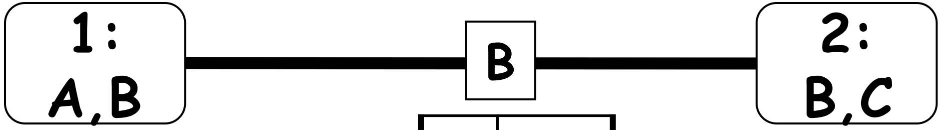
b^1	c^1	4
b^1	c^2	1.5
b^2	c^1	0.2
b^2	c^2	2

a^1	b^1	c^1	3+4=7
a^1	b^1	c^2	3+1.5=4.5
a^1	b^2	c^1	0+0.2=0.2
a^1	b^2	c^2	0+2=2
a^2	b^1	c^1	-1+4=3
a^2	b^1	c^2	-1+1.5=0.5
a^2	b^2	c^1	1+0.2=1.2
a^2	b^2	c^2	1+2=3

Simple Example

a^1	b^1	3
a^1	b^2	0
a^2	b^1	-1
a^2	b^2	1

b^1	c^1	4
b^1	c^2	1.5
b^2	c^1	0.2
b^2	c^2	2



b^1	3
b^2	1

b^1	4
b^2	2

a^1	b^1	3+4=7
a^1	b^2	0+2=2
a^2	b^1	-1+4=3
a^2	b^2	1+2=3

b^1	c^1	4+3=7
b^1	c^2	1.5+3=4.5
b^2	c^1	0.2+1=1.2
b^2	c^2	2+1=3

Max-Sum BP at Convergence

- Each clique contains max-marginal

$$\beta_i(\boldsymbol{C}_i) = \max_{\boldsymbol{W}_i} \theta(\boldsymbol{C}_i, \boldsymbol{W}_i) \qquad \boldsymbol{W}_i = \{X_1, \ldots, X_n\} - \boldsymbol{C}_i$$

- Cliques are necessarily calibrated
  – Agree on sepsets $\max_{\boldsymbol{C}_i - \boldsymbol{S}_{i,j}} \beta_i(\boldsymbol{C}_i) = \max_{\boldsymbol{C}_j - \boldsymbol{S}_{i,j}} \beta_j(\boldsymbol{C}_j)$

a^1	b^1	3+4=7
a^1	b^2	0+2=2
a^2	b^1	-1+4=3
a^2	b^2	1+2=3

b^1	c^1	4+3=7
b^1	c^2	1.5+3=4.5
b^2	c^1	0.2+1=1.2
b^2	c^2	2+1=3

Summary

- The same clique tree algorithm used for sum-product can be used for max-sum
- As in sum-product, convergence is achieved after a single up-down pass
- Result is a max-marginal at each clique $\boldsymbol{C}$:
  - For each assignment $\mathbf{c}$ to $\boldsymbol{C}$, what is the score of the best completion to $\mathbf{c}$

Probabilistic
Graphical
Models

Inference

MAP

Finding a MAP Assignment

Decoding a MAP Assignment

- Easy if MAP assignment is unique
  - Single maximizing assignment at each clique
  - Whose value is the θ value of the MAP assignment
  - Due to calibration, choices at all cliques must agree

a^1	b^1	c^1	7
a^1	b^1	c^2	4.5
a^1	b^2	c^1	0.2
a^1	b^2	c^2	2
a^2	b^1	c^1	3
a^2	b^1	c^2	0.5
a^2	b^2	c^1	1.2
a^2	b^2	c^2	3

a^1	b^1	3+4=7
a^1	b^2	0+2=2
a^2	b^1	-1+4=3
a^2	b^2	1+2=3

b^1	c^1	4+3=7
b^1	c^2	1.5+3=4.5
b^2	c^1	0.2+1=1.2
b^2	c^2	2+1=3

Decoding a MAP assignment

- If MAP assignment is not unique, we may have multiple choices at some cliques
- Arbitrary tie-breaking may not produce a MAP assignment

a^1	b^1	2
a^1	b^2	1
a^2	b^1	1
a^2	b^2	2

b^1	c^1	2
b^1	c^2	1
b^2	c^1	1
b^2	c^2	2

Decoding a MAP assignment

- If MAP assignment is not unique, we may have multiple choices at some cliques
- Arbitrary tie-breaking may not produce a MAP assignment
- Two options:
  - Slightly perturb parameters to make MAP unique
  - Use traceback procedure that incrementally builds a MAP assignment, one variable at a time