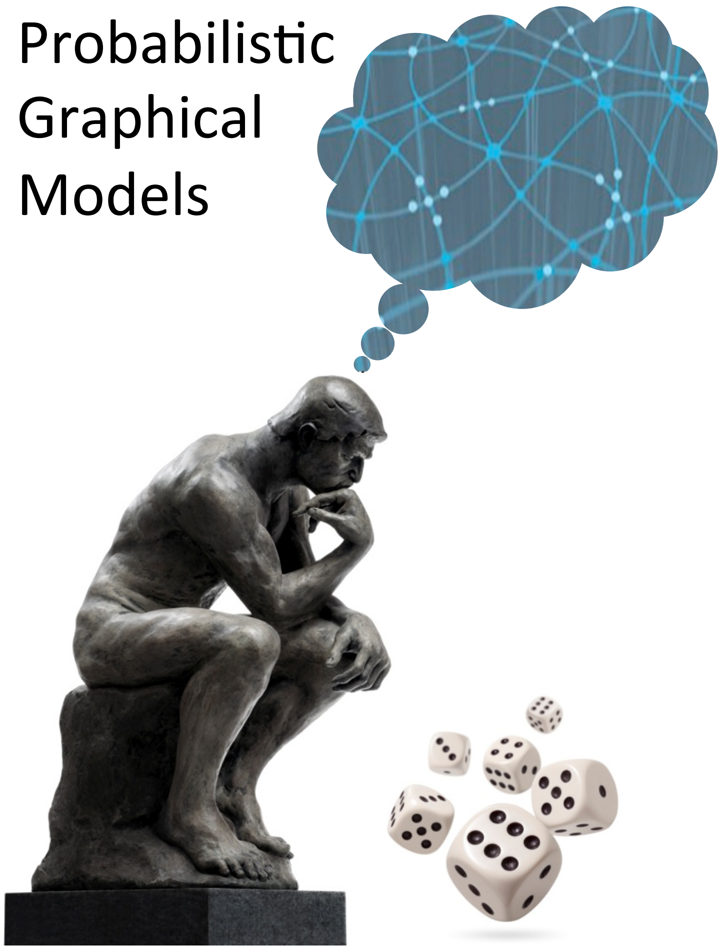
Probabilistic
Graphical
Models



Inference

MAP

Max-Sum
Exact Inference

Product ⇨ Summation

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

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

$\log \phi_k(D_k)$

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

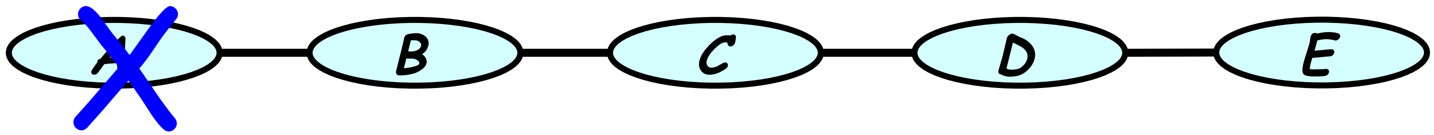
a^1	b^1	8
a^1	b^2	1
a^2	b^1	0.5
a^2	b^2	2

⇩ $\log_2$

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

Daphne Koller

Max-Sum Elimination in Chains



$\theta(A,B,C,D,E)$

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

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

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

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

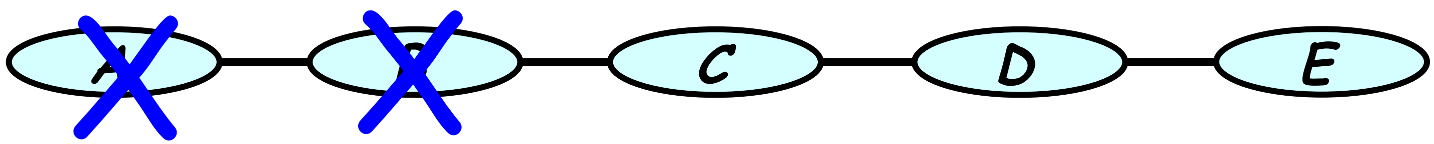
Factor Maximization

max - marginalization

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

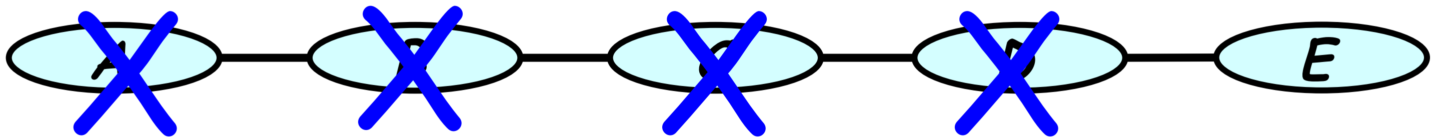


$$\max_D \max_C \max_B \left(\theta_2(B,C) + \theta_3(C,D) + \theta_4(D,E) + \lambda_1(B)\right)$$

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

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

Max-Sum Elimination in Chains



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

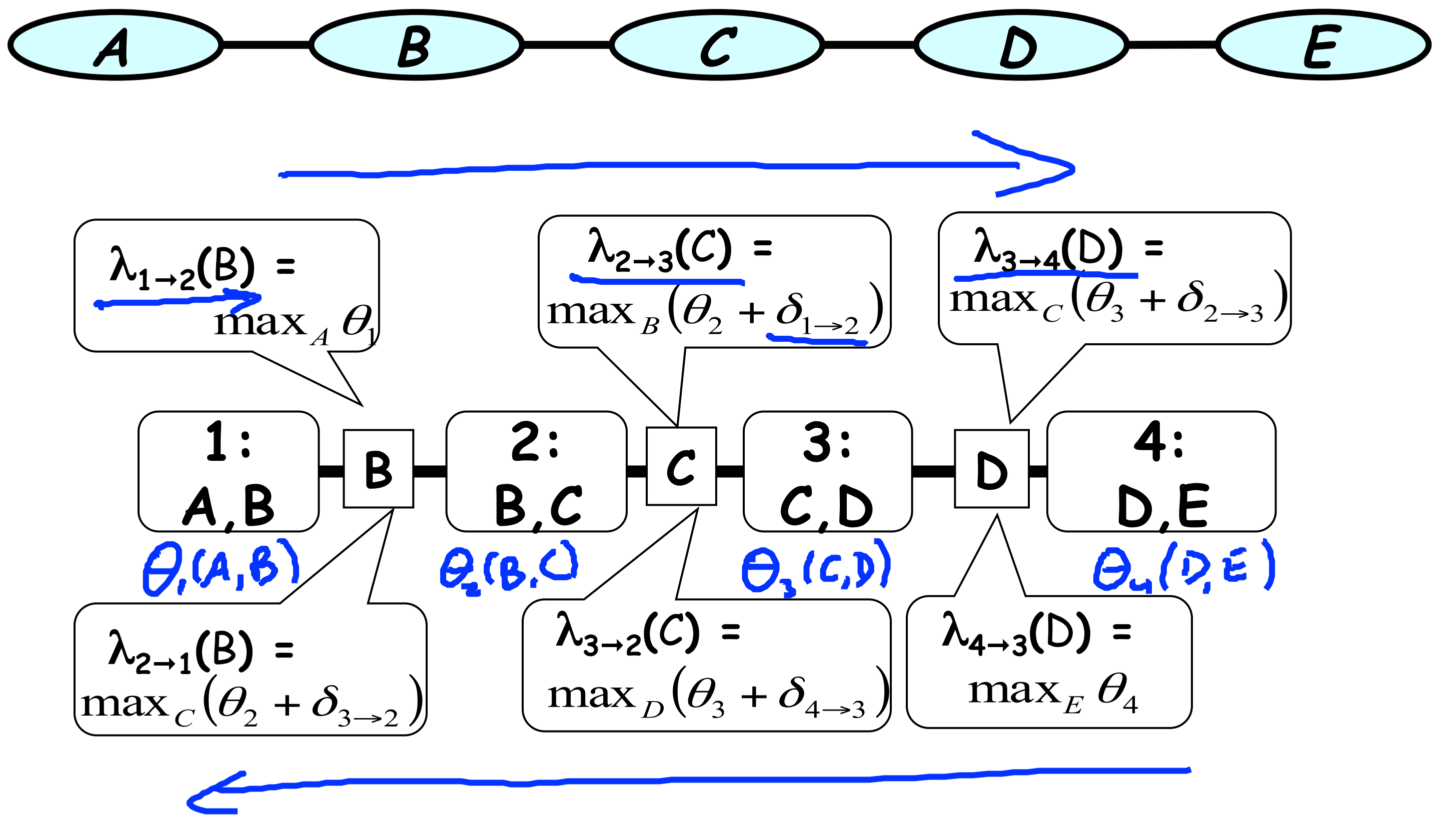
$$\max_D \left(\theta_4(D,E) + \lambda_3(D)\right)$$

best value that I can get if we mandate $E=e$

$$\lambda_4(E) \qquad \lambda_4(e) = \max_{a,b,c,d} \Theta(a,b,c,d,e)$$

max-marginal

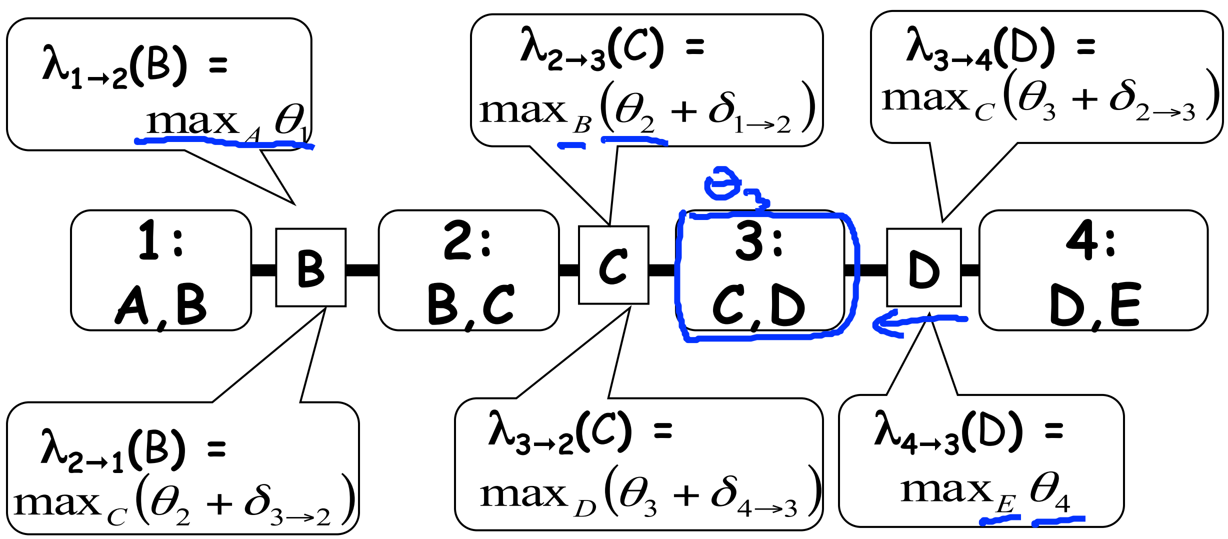
Max-Sum in Clique Trees



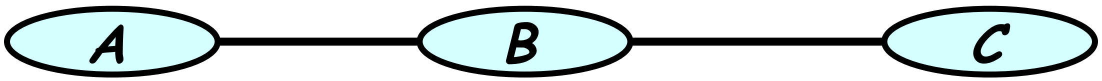
Daphne Koller

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



$\theta_1(A,B)$

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

$\theta_2(B,C)$

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

$\theta = \theta_1 + \theta_2$

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

Daphne Koller

Simple Example

θ_1

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

θ_2

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

max-marg$_A$

1: A,B — **B** — **2: B,C**

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

Daphne Koller

Max-Sum BP at Convergence

- Beliefs at each clique are max-marginals

$$\beta_i(C_i) = \theta_i(C_i) + \underbrace{\sum_k \lambda_{k\to i}}$$

incoming msgs

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

- Calibration: cliques agree on shared variables

$$\max_{\boldsymbol{C}_i - \boldsymbol{S}_{i,j}} \beta_i(\boldsymbol{C}_i) = \max_{\boldsymbol{C}_j - \boldsymbol{S}_{i,j}} \beta_j(\boldsymbol{C}_j)$$

$\beta_1 = \theta_1 + \lambda_{2\to 1}$

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 7
b^2 3

b_1 7
b_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

$\beta_2 = \theta_2 + \lambda_{1\to 2}$

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 ***C***:
  - For each assignment **c** to ***C***, what is the score of the best completion to **c**