Graphical Models Inference and Learning Lecture 2

MVA

2018 – 2019

http://thoth.inrialpes.fr/~alahari/disinflearn

Slides based on material from M. Pawan Kumar

Practical matters

- Paper presentations
  - 18/12: Spatio-temporal video segmentation…
  - 15/01: On Parameter Learning in CRF-based...
  - 12/02: ~~CRFs as RNNs~~ (chosen)

- At most 2 presenters per paper
- Bonus points & you will do better in the quiz

Practical matters

- Course website
  - http://thoth.inrialpes.fr/~alahari/disinflearn
  - (linked from my webpage)

- Questions ?

Recap: Lecture 1

- Graphical Models
  - Making **global** predictions from **local** observations
  - Learning from large quantities of data
- Two types of models studied in the class
  - Bayesian nets
  - Markov nets

Recap: Lecture 1

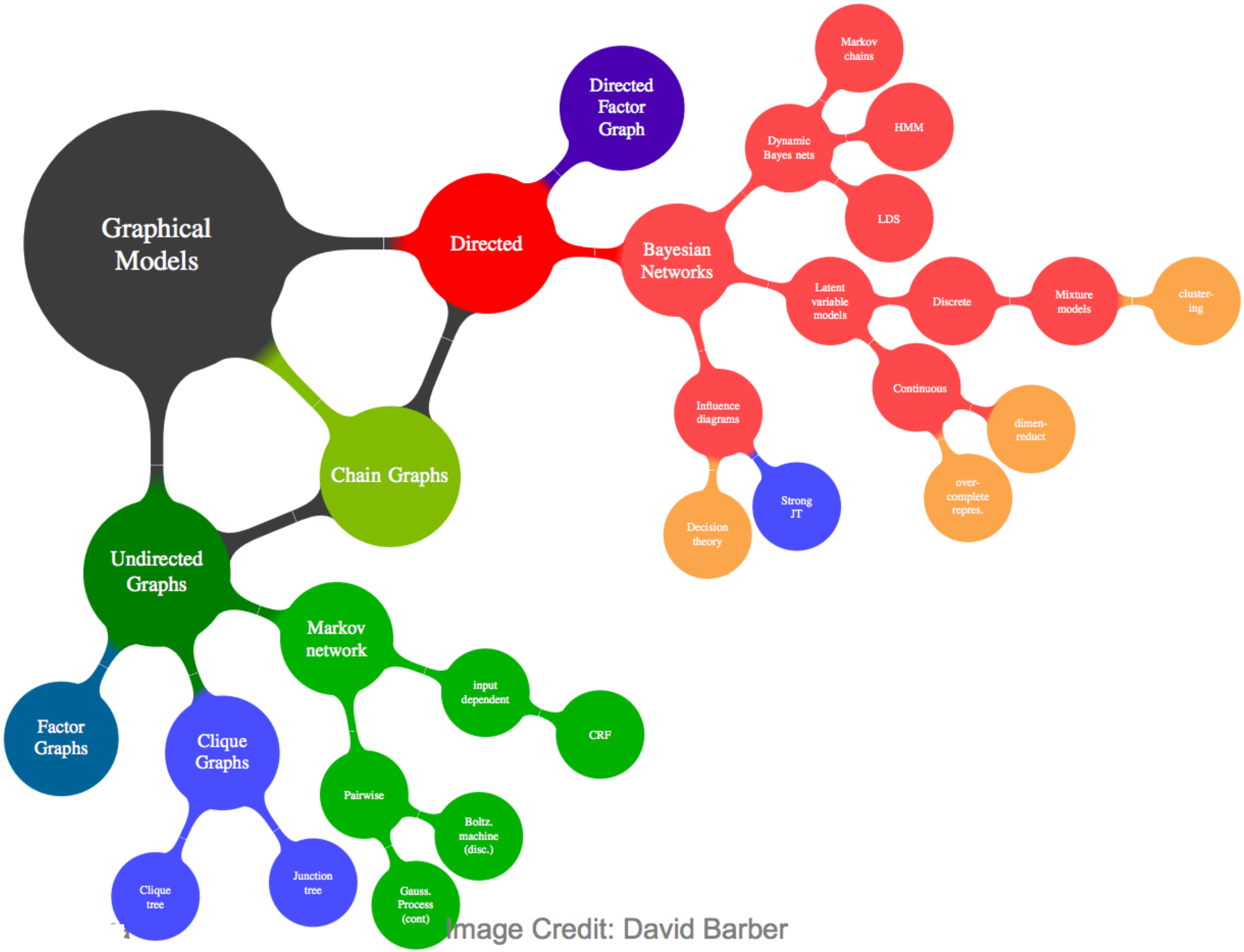
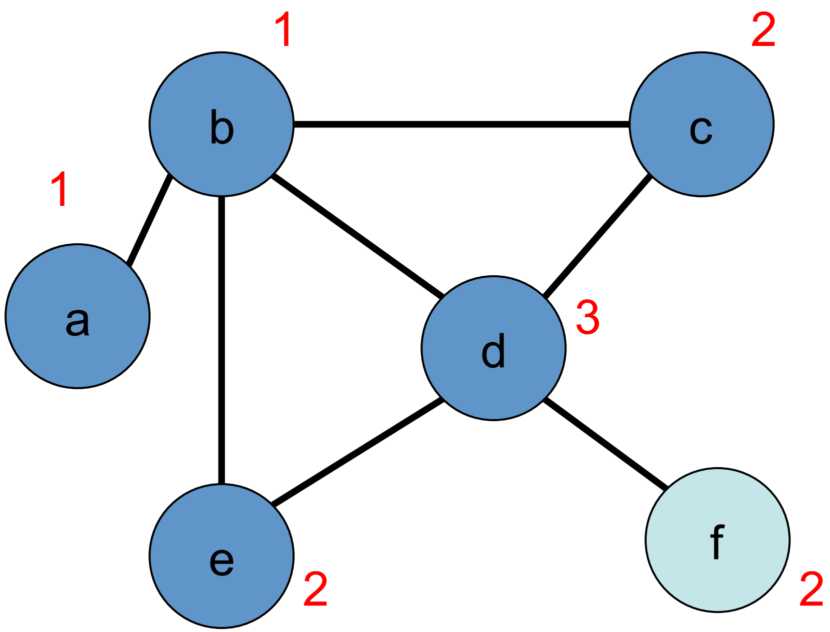


Image Credit: David Barber

Recap: Lecture 1

- Question: What is the core of these models?
- Question: Can you compute probabilities in Markov nets? If yes, how and if no, why?
- Question: What is the difference between Markov and Conditional random fields?

The General Problem



Graph G = (V, E)

Discrete label set L = {1,2,…,h}

Assign a label to each vertex
f: V ➜ L

Cost of a labelling Q(f)

Unary Cost Pairwise Cost

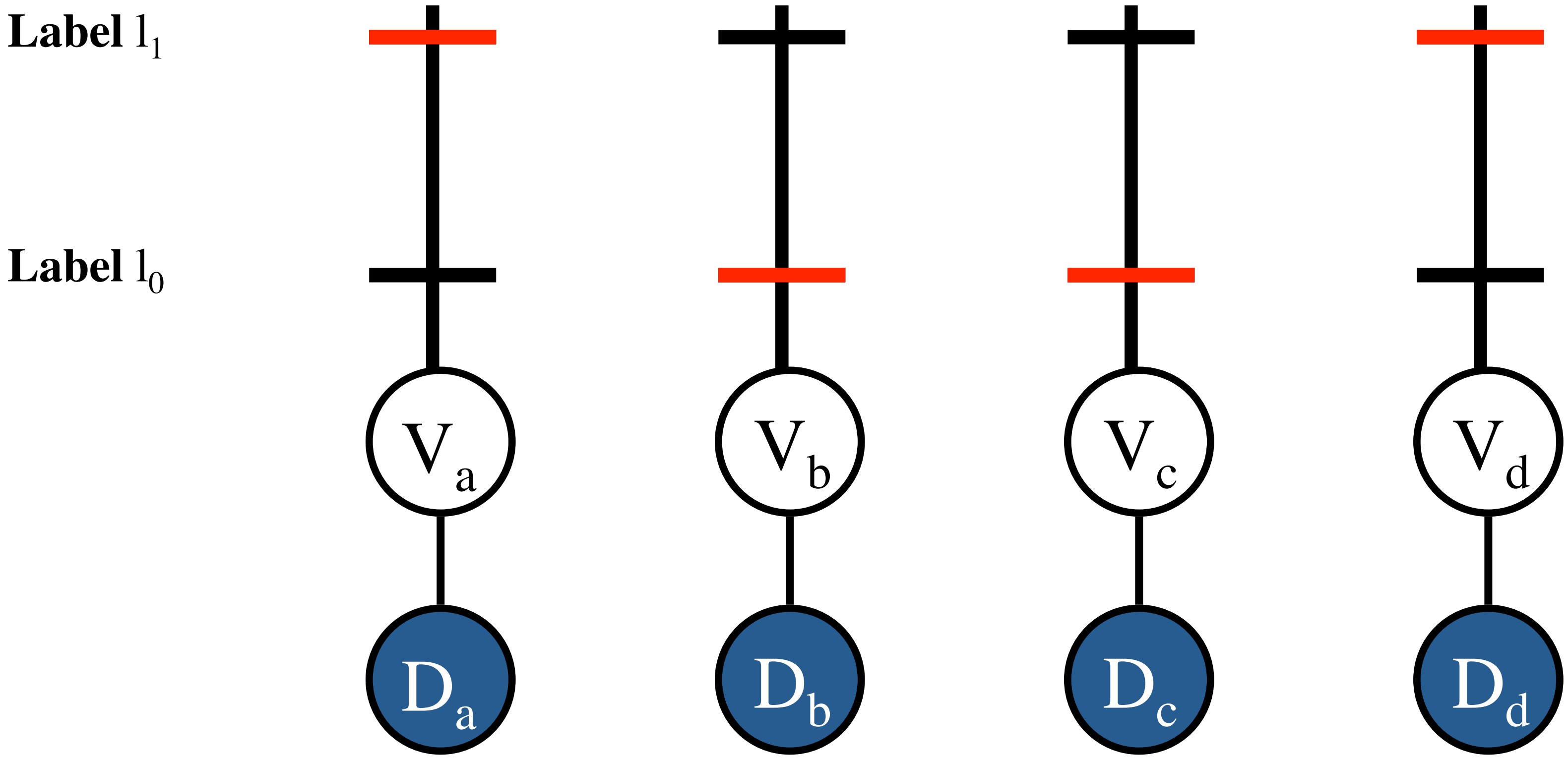
Find f* = arg min Q(f)

Overview

- Basics: problem formulation
  - Energy Function
  - MAP Estimation
  - Computing min-marginals
  - Reparameterization

- Solutions
  - Belief Propagation and related methods [Lecture 2]
  - Graph cuts [Lecture 3]

Energy Function

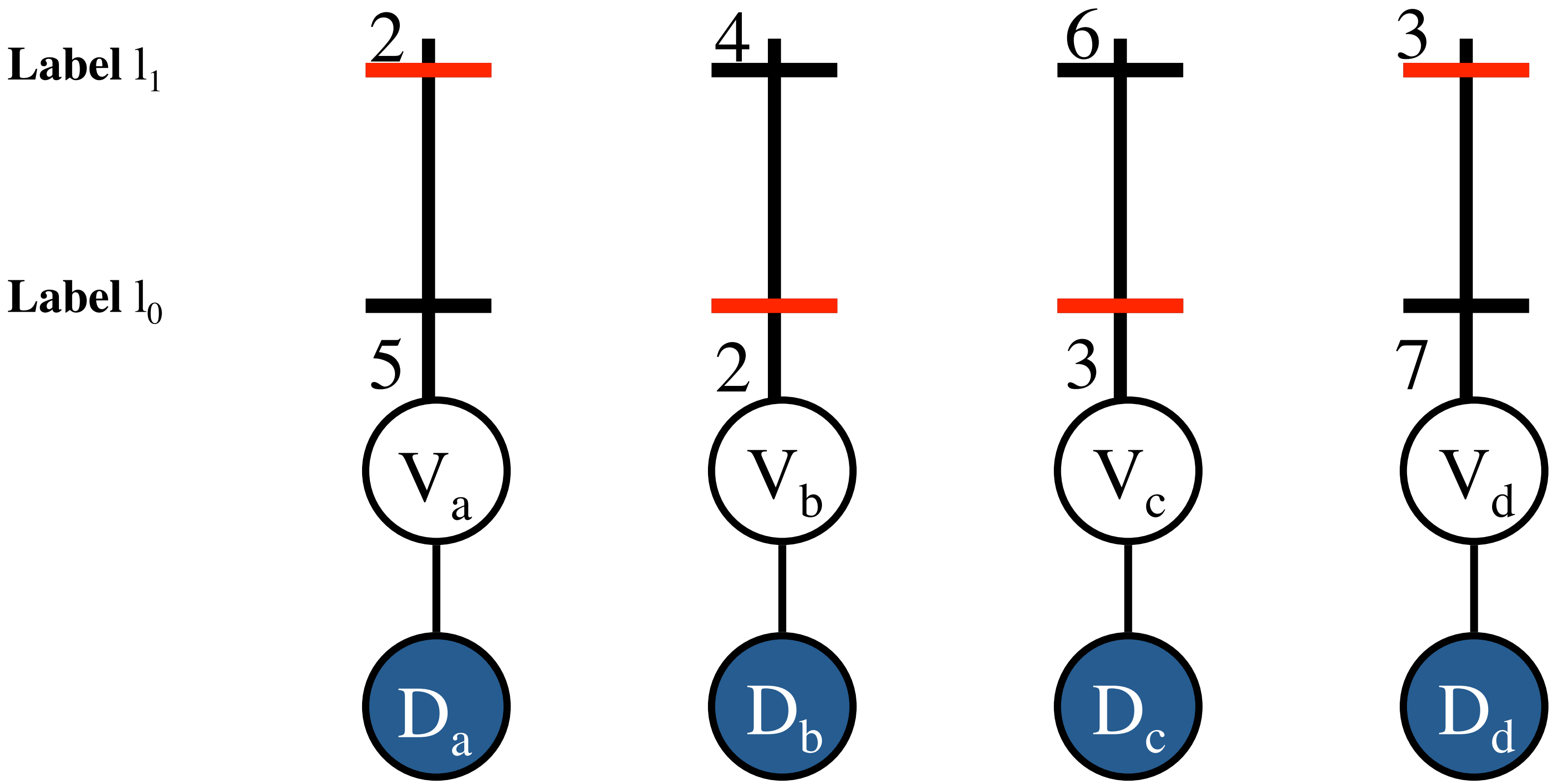


Random Variables $V = \{V_a, V_b, ...\}$

Labels $L = \{l_0, l_1, ...\}$ Data D

Labelling $f: \{a, b, ... \} \rightarrow \{0,1, ...\}$

Energy Function



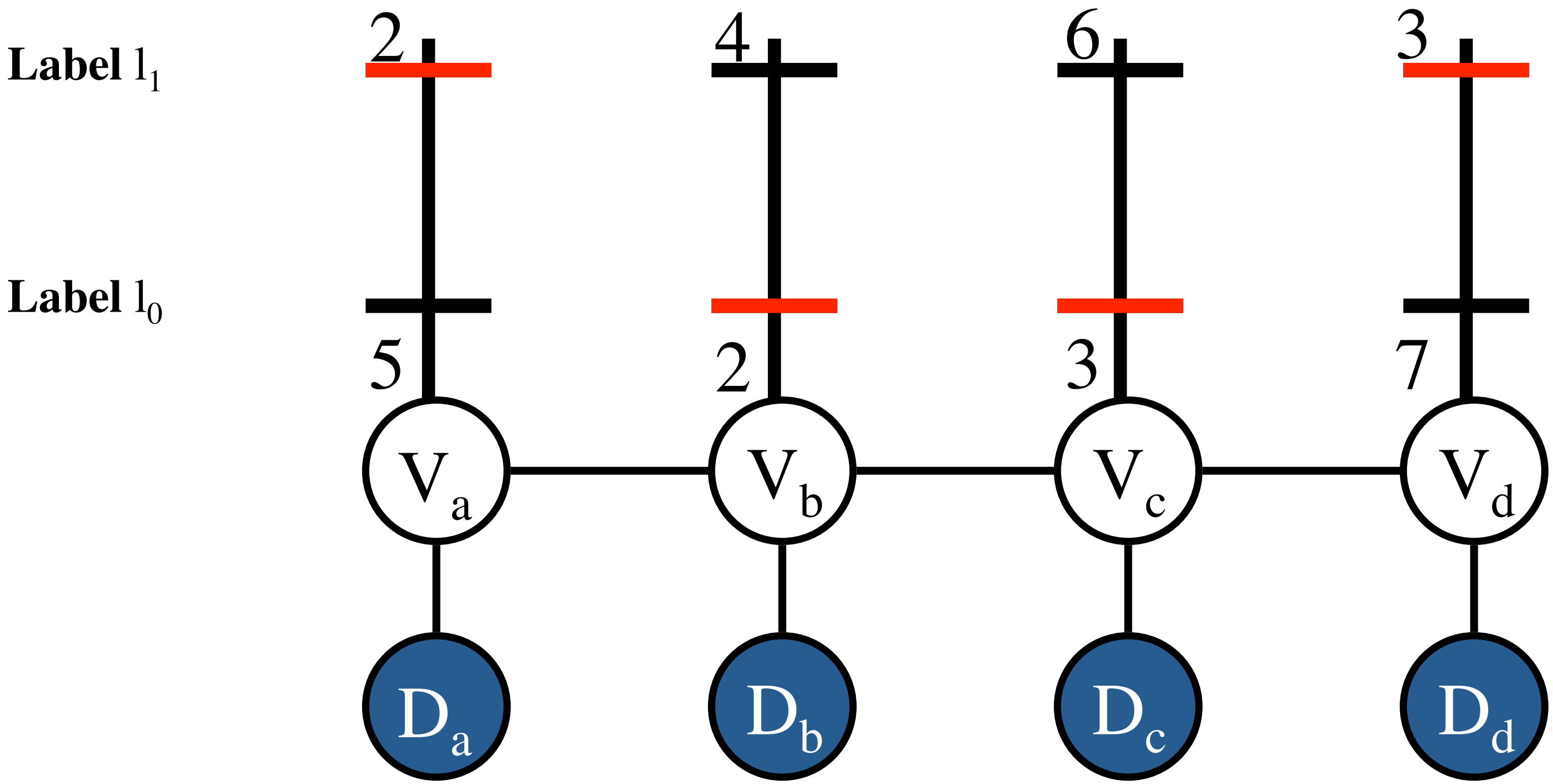
$$Q(f) = \sum_a \theta_{a;f(a)}$$

Unary Potential

Easy to minimize

Neighbourhood

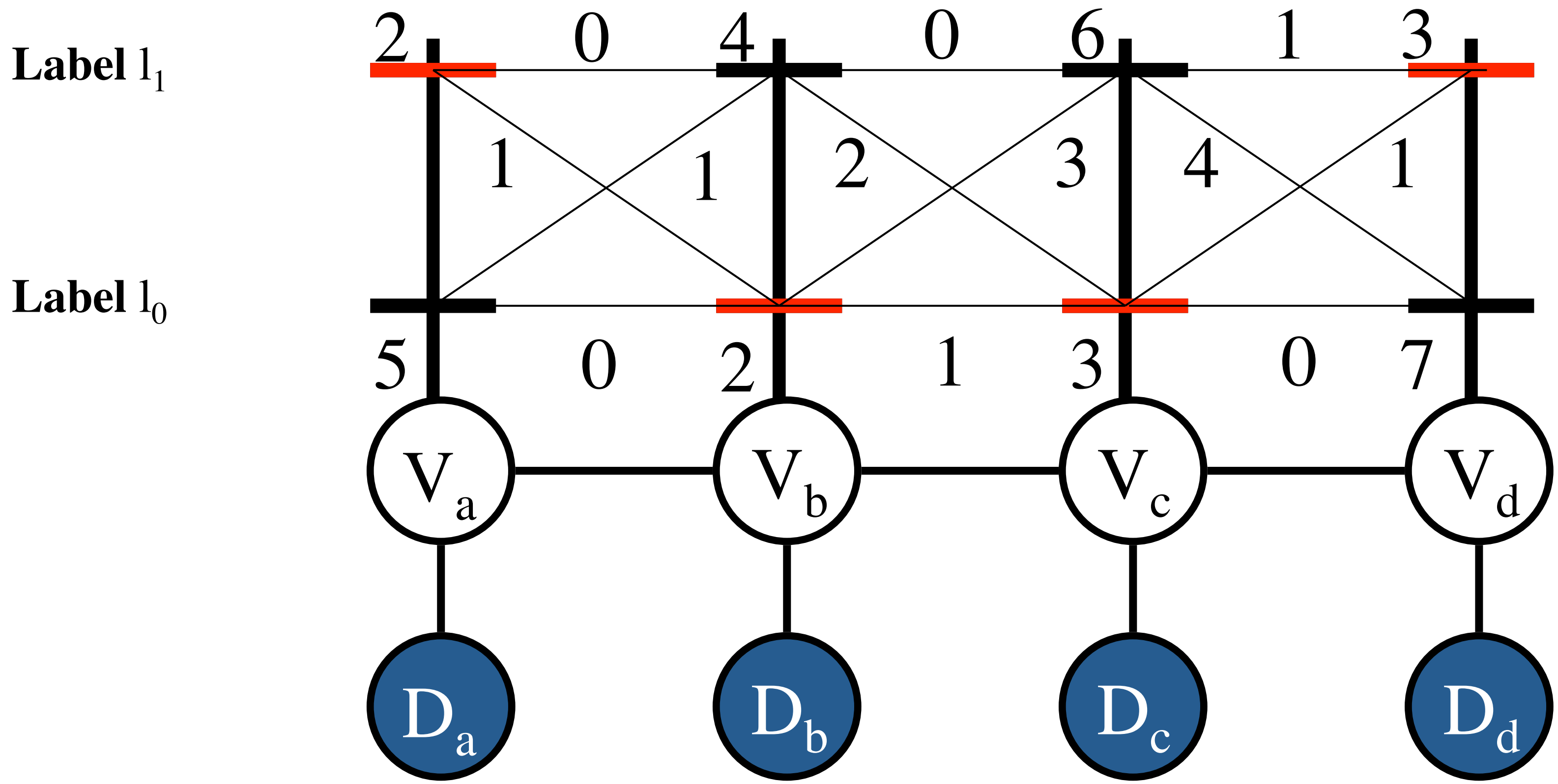
Energy Function



E : $(a,b) \in E$ iff V_a and V_b are neighbours

$E = \{ (a,b) , (b,c) , (c,d) \}$

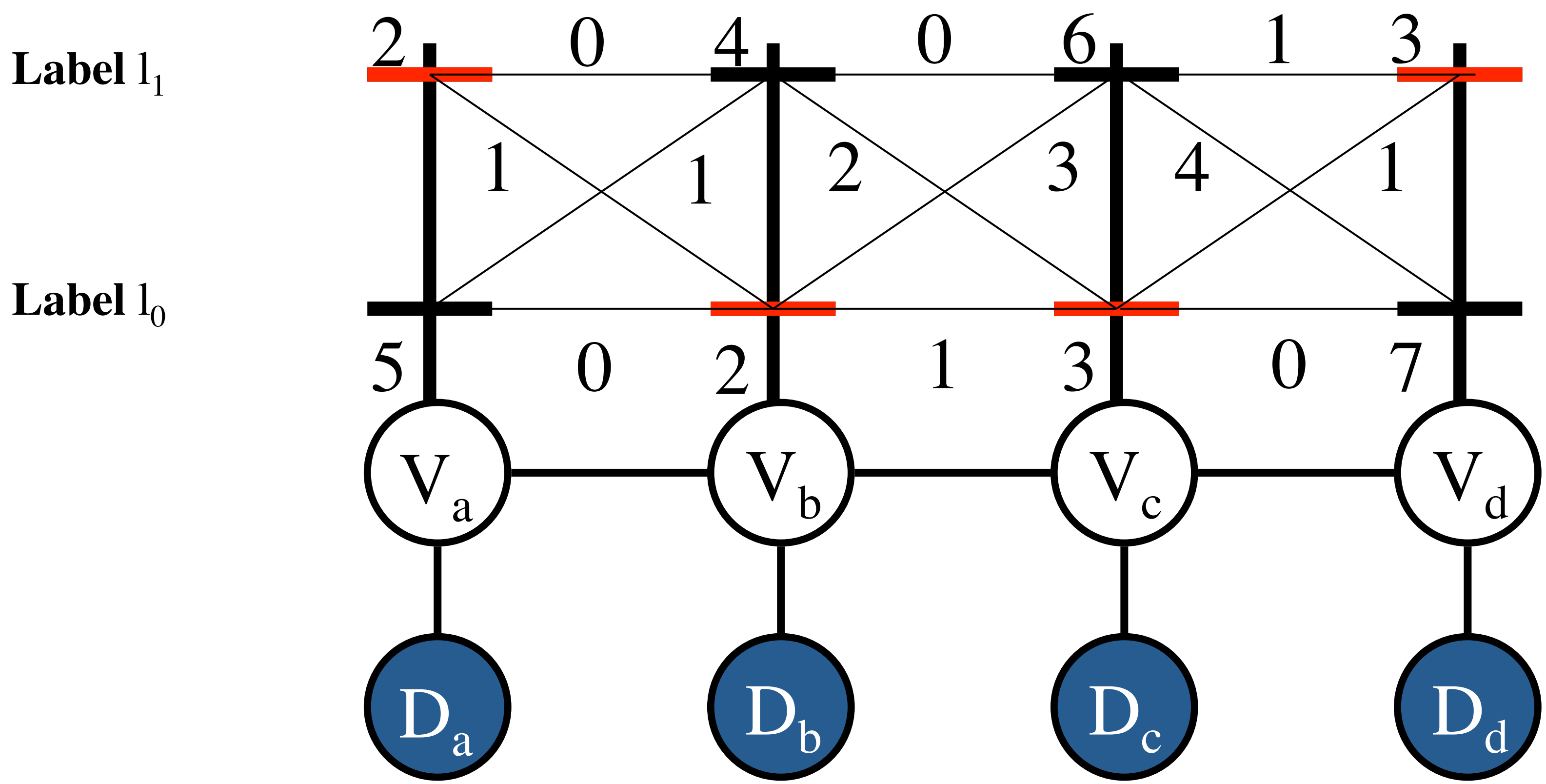
Energy Function



Pairwise Potential

$$Q(f) = \sum_a \theta_{a;f(a)} + \sum_{(a,b)} \theta_{ab;f(a)f(b)}$$

Energy Function



$$Q(f; \theta) = \sum_a \theta_{a;f(a)} + \sum_{(a,b)} \theta_{ab;f(a)f(b)}$$

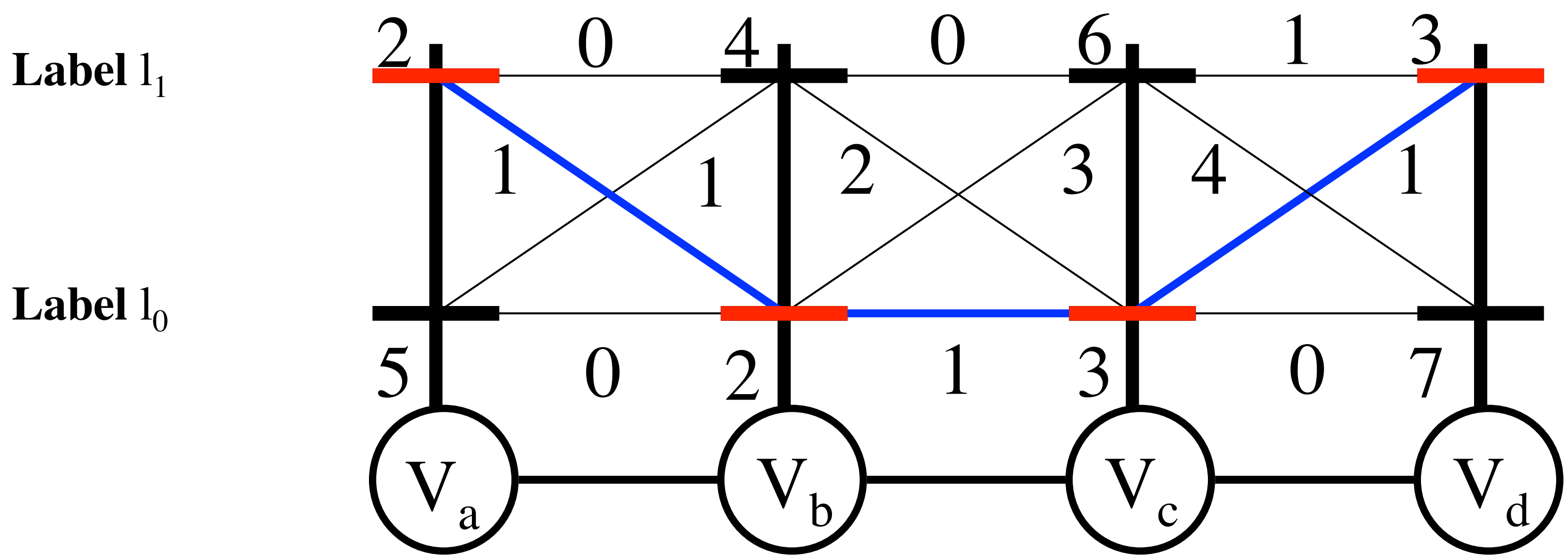
Parameter

Overview

- Basics: problem formulation
  - Energy Function
  - MAP Estimation
  - Computing min-marginals
  - Reparameterization

- Solutions
  - Belief Propagation and related methods [Lecture 2]
  - Graph cuts [Lecture 3]

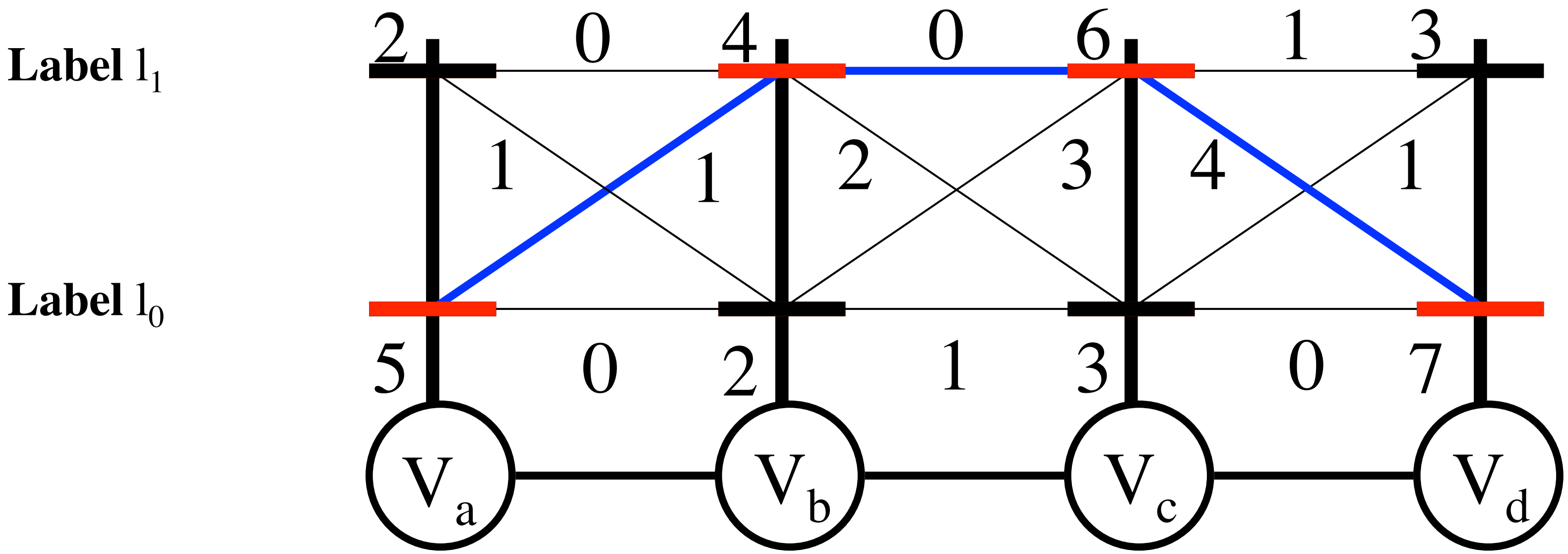
MAP Estimation



$$Q(f;\theta) = \sum_a \theta_{a;f(a)} + \sum_{(a,b)} \theta_{ab;f(a)f(b)}$$

2 + 1 + 2 + 1 + 3 + 1 + 3 = 13

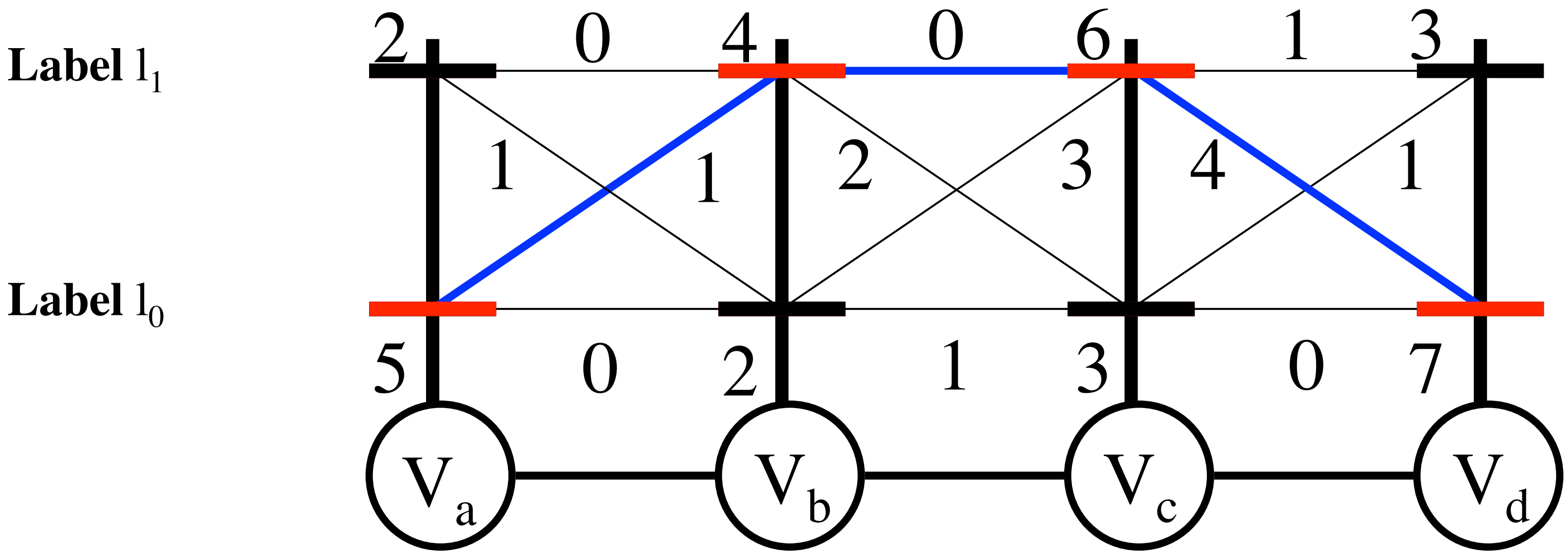
MAP Estimation



$$Q(f;\theta) = \sum_a \theta_{a;f(a)} + \sum_{(a,b)} \theta_{ab;f(a)f(b)}$$

5 + 1 + 4 + 0 + 6 + 4 + 7 = 27

MAP Estimation



$q^* = \min Q(f;\theta) = Q(f^*;\theta)$

$$Q(f;\theta) = \sum_a \theta_{a;f(a)} + \sum_{(a,b)} \theta_{ab;f(a)f(b)}$$

$f^* = \arg\min Q(f;\theta)$

Equivalent to maximizing the associated probability

MAP Estimation

16 possible labellings

$f^* = \{1, 0, 0, 1\}$

$q^* = 13$

f(a)	f(b)	f(c)	f(d)	Q(f; θ)
0	0	0	0	18
0	0	0	1	15
0	0	1	0	27
0	0	1	1	20
0	1	0	0	22
0	1	0	1	19
0	1	1	0	27
0	1	1	1	20

f(a)	f(b)	f(c)	f(d)	Q(f; θ)
1	0	0	0	16
1	0	0	1	13
1	0	1	0	25
1	0	1	1	18
1	1	0	0	18
1	1	0	1	15
1	1	1	0	23
1	1	1	1	16

Computational Complexity

Segmentation

$2^{|V|}$

$|V|$ = number of pixels ≈ 153600

Can we do better than brute-force?

MAP Estimation is NP-hard !!

MAP Inference / Energy Minimization

- Computing the assignment minimizing the energy in NP-hard in general

$$\underset{\mathbf{y}\in\mathcal{Y}}{\operatorname{argmin}}\, E(\mathbf{y};\mathbf{x}) = \underset{\mathbf{y}\in\mathcal{Y}}{\operatorname{argmax}}\, P(\mathbf{y} \mid \mathbf{x})$$

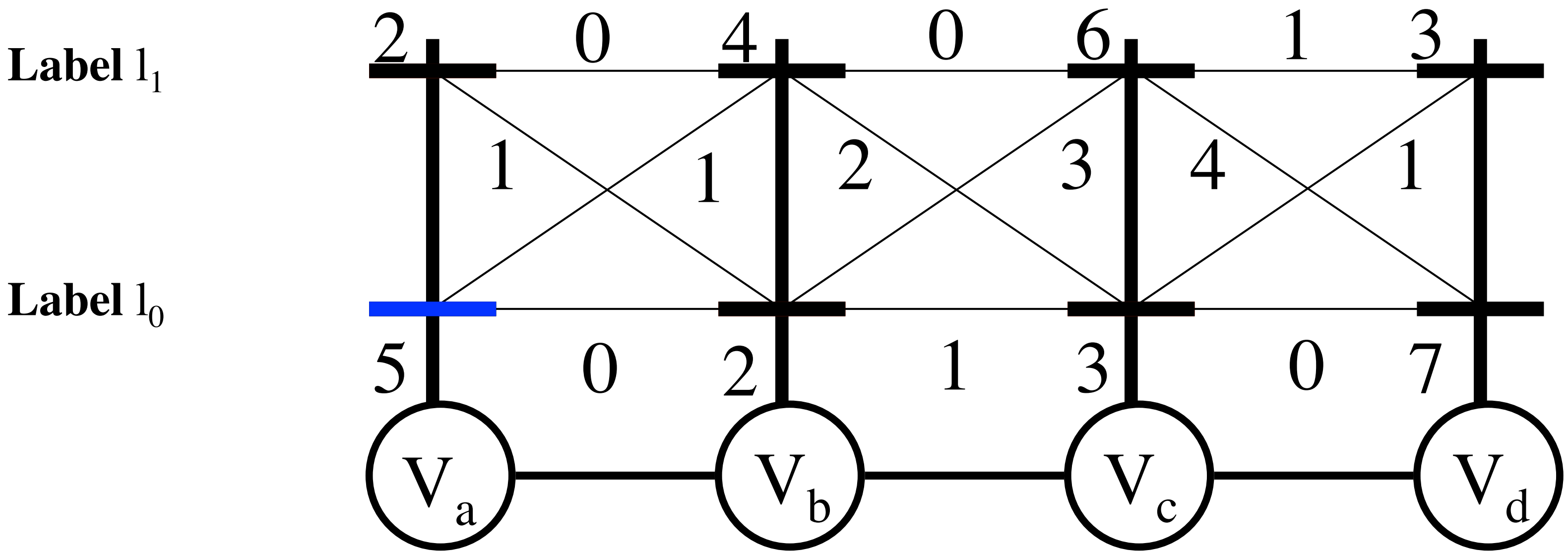
- Exact inference is possible in some cases, e.g.,
  - Low treewidth graphs → message-passing
  - Submodular potentials → graph cuts
- Efficient approximate inference algorithms exist
  - Message passing on general graphs
  - Move-making algorithms
  - Relaxation algorithms

Overview

- Basics: problem formulation
  - Energy Function
  - MAP Estimation
  - Computing min-marginals
  - Reparameterization

- Solutions
  - Belief Propagation and related methods [Lecture 2]
  - Graph cuts [Lecture 3]

Min-Marginals



Not a marginal (no summation)

f* = arg min Q(f; θ) **such that f(a) = i**

Min-marginal $q_{a;i}$

Min-Marginals

16 possible labellings

$q_{a;0} = 15$

f(a)	f(b)	f(c)	f(d)	Q(f; θ)
0	0	0	0	18
0	0	0	1	15
0	0	1	0	27
0	0	1	1	20
0	1	0	0	22
0	1	0	1	19
0	1	1	0	27
0	1	1	1	20

f(a)	f(b)	f(c)	f(d)	Q(f; θ)
1	0	0	0	16
1	0	0	1	13
1	0	1	0	25
1	0	1	1	18
1	1	0	0	18
1	1	0	1	15
1	1	1	0	23
1	1	1	1	16

Min-Marginals

16 possible labellings

$q_{a;1} = 13$

f(a)	f(b)	f(c)	f(d)	Q(f; θ)
0	0	0	0	18
0	0	0	1	15
0	0	1	0	27
0	0	1	1	20
0	1	0	0	22
0	1	0	1	19
0	1	1	0	27
0	1	1	1	20

f(a)	f(b)	f(c)	f(d)	Q(f; θ)
1	0	0	0	16
1	0	0	1	13
1	0	1	0	25
1	0	1	1	18
1	1	0	0	18
1	1	0	1	15
1	1	1	0	23
1	1	1	1	16

Min-Marginals and MAP

- Minimum min-marginal of any variable = energy of MAP labelling

$$\min_i \; q_{a;i}$$

$$\min_i \left(\min_f Q(f; \theta) \quad \text{such that } f(a) = i \right)$$

V_a has to take one label

$$\min_f Q(f; \theta)$$

Summary

Energy Function

$$Q(f; \theta) = \sum_a \theta_{a;f(a)} + \sum_{(a,b)} \theta_{ab;f(a)f(b)}$$

MAP Estimation

$$f^* = \arg\min Q(f; \theta)$$

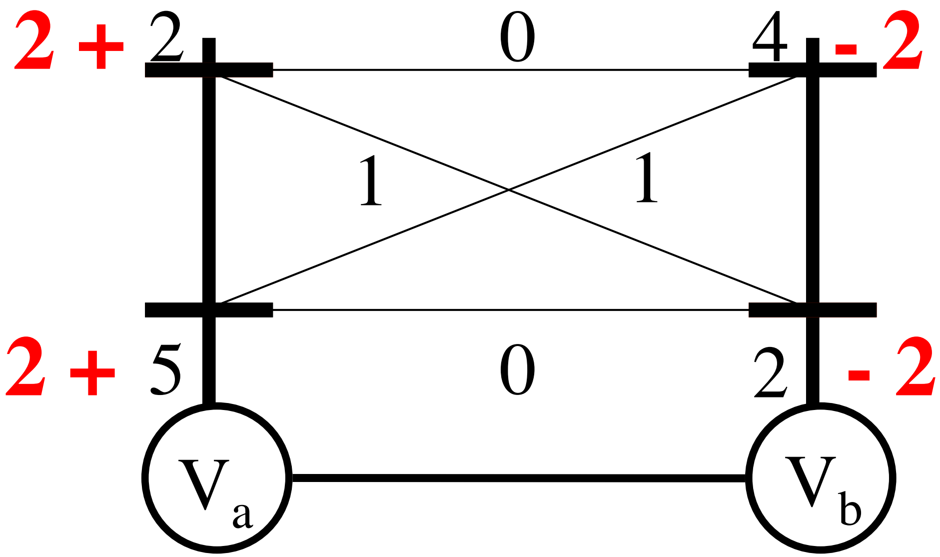
Min-marginals

$$q_{a;i} = \min Q(f; \theta) \quad \text{s.t. } f(a) = i$$

Overview

- Basics: problem formulation
  - Energy Function
  - MAP Estimation
  - Computing min-marginals
  - Reparameterization
- Solutions
  - Belief Propagation and related methods [Lecture 2]
  - Graph cuts [Lecture 3]

Reparameterization



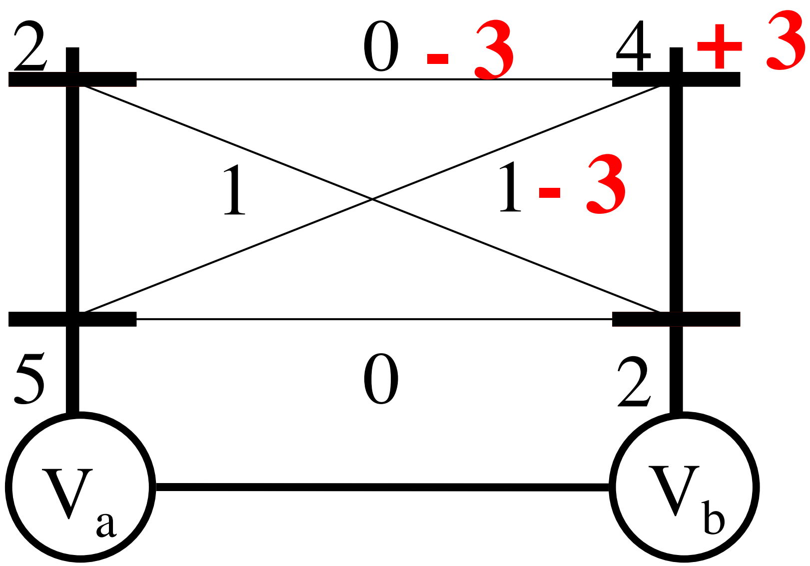
f(a)	f(b)	Q(f; θ)
0	0	7 + 2 - 2
0	1	10 + 2 - 2
1	0	5 + 2 - 2
1	1	6 + 2 - 2

Add a constant to all $\theta_{a;i}$

Subtract that constant from all $\theta_{b;k}$

$Q(f; \theta') = Q(f; \theta)$

Reparameterization



f(a)	f(b)	Q(f; θ)
0	0	7
0	1	10 - 3 + 3
1	0	5
1	1	6 - 3 + 3

Add a constant to one $\theta_{b;k}$

Subtract that constant from $\theta_{ab;ik}$ for all 'i'

$Q(f; \theta') = Q(f; \theta)$

Reparameterization

θ' is a reparameterization of θ, iff

$$Q(f; \theta') = Q(f; \theta), \text{ for all } f \qquad \theta' \equiv \theta$$

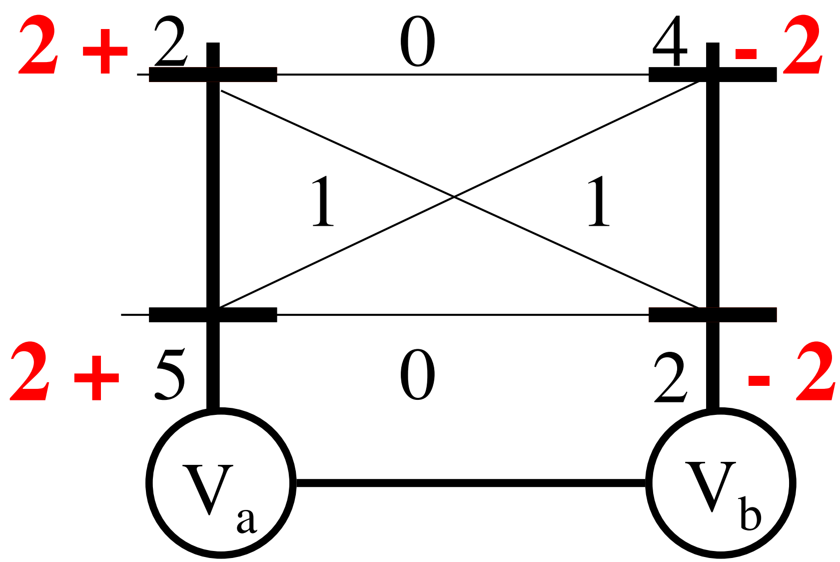
Equivalently

Kolmogorov, PAMI, 2006

$$\theta'_{a;i} = \theta_{a;i} + M_{ba;i}$$

$$\theta'_{b;k} = \theta_{b;k} + M_{ab;k}$$

$$\theta'_{ab;ik} = \theta_{ab;ik} - M_{ab;k} - M_{ba;i}$$



Recap

MAP Estimation

$$f^* = \arg\min Q(f; \theta)$$

$$Q(f; \theta) = \sum_a \theta_{a;f(a)} + \sum_{(a,b)} \theta_{ab;f(a)f(b)}$$

Min-marginals

$$q_{a;i} = \min Q(f; \theta) \quad \text{s.t. } f(a) = i$$

Reparameterization

$$Q(f; \theta') = Q(f; \theta), \text{ for all } f \qquad \theta' \equiv \theta$$

Overview

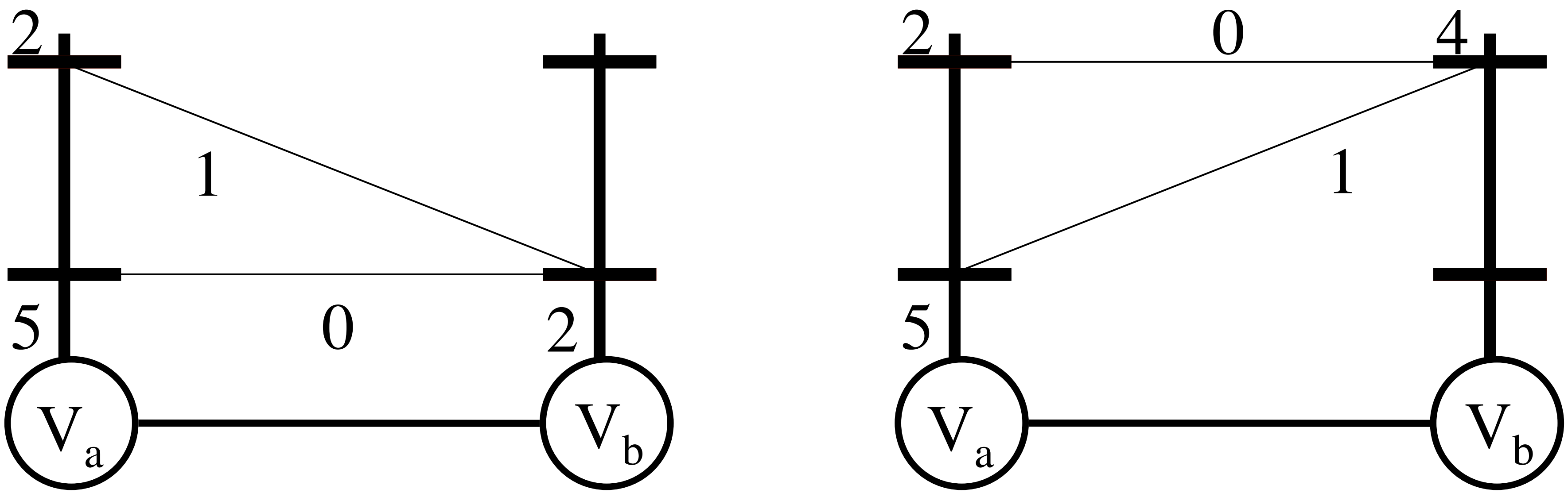
- Basics: problem formulation
  - Energy Function
  - MAP Estimation
  - Computing min-marginals
  - Reparameterization

- Solutions
  - Belief Propagation and related methods [Lecture 2]
  - Graph cuts [Lecture 3]

Belief Propagation

- Remember, some MAP problems are easy
- Belief Propagation gives exact MAP for chains
- Exact MAP for trees
- Clever Reparameterization

Two Variables

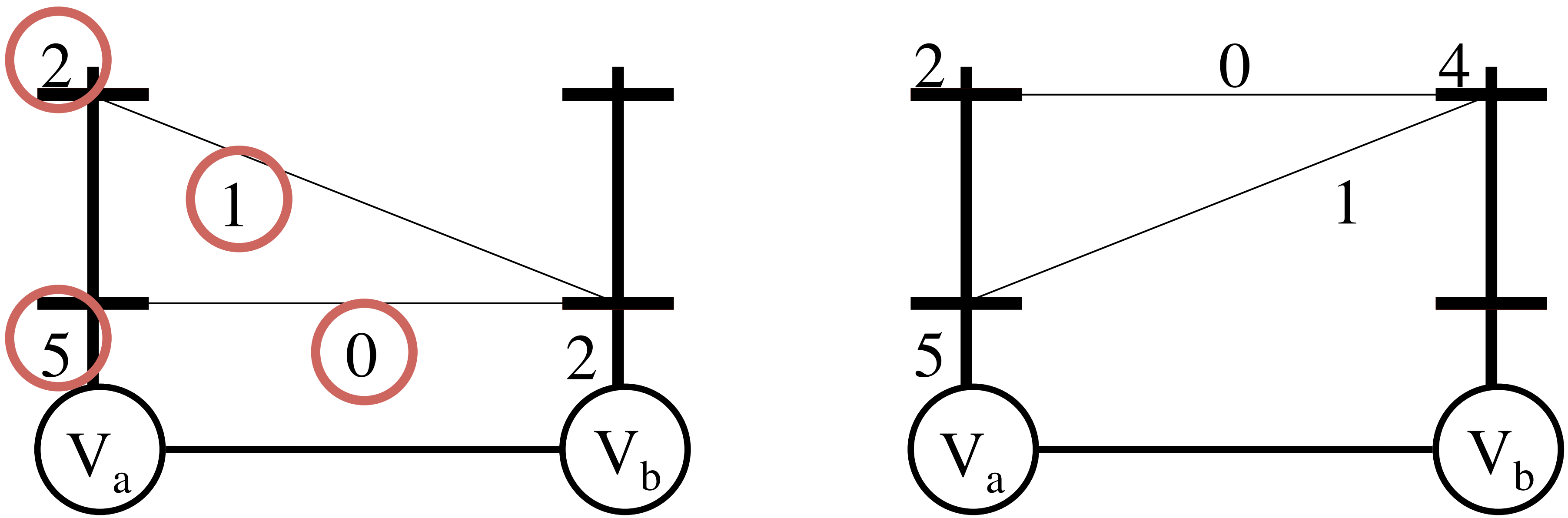


Add a constant to one $\theta_{b;k}$

Subtract that constant from $\theta_{ab;ik}$ for all 'i'

Choose the ***right*** constant $\theta'_{b;k} = q_{b;k}$

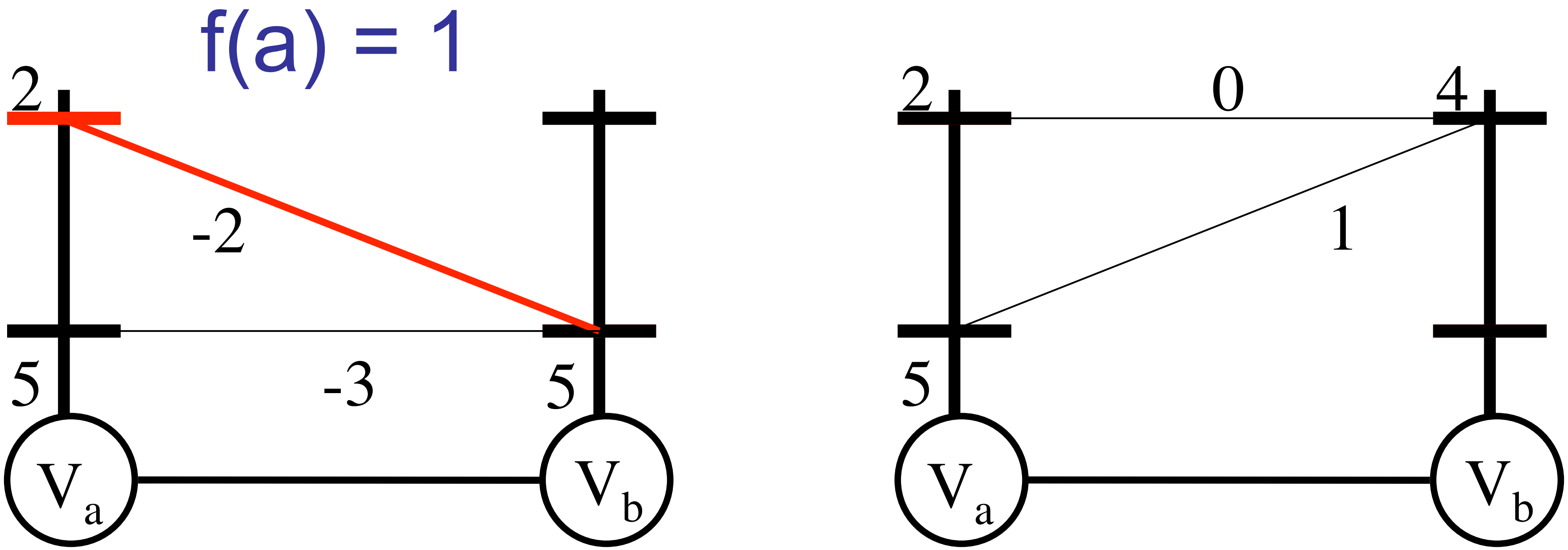
Two Variables



$$M_{ab;0} = \min \begin{cases} \theta_{a;0} + \theta_{ab;00} = 5 + 0 \\ \theta_{a;1} + \theta_{ab;10} = 2 + 1 \end{cases}$$

Choose the ***right*** constant $\theta'_{b;k} = q_{b;k}$

Two Variables

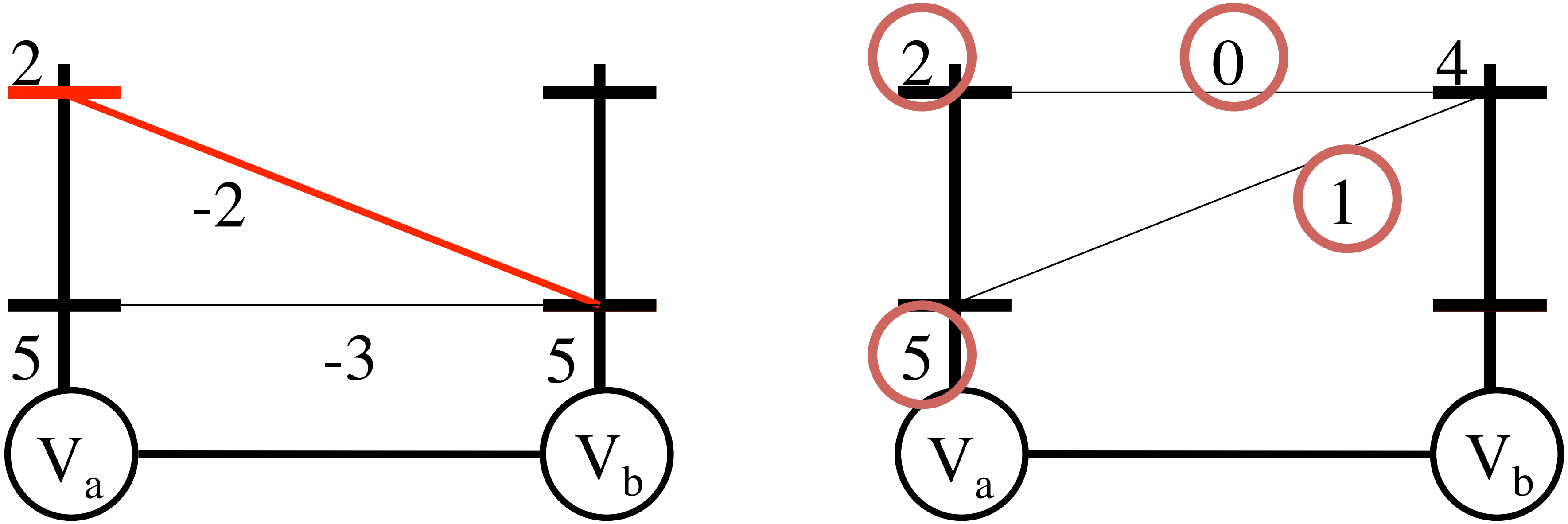


$$\theta'_{b;0} = q_{b;0}$$

Potentials along the red path add up to 0

Choose the ***right*** constant $\theta'_{b;k} = q_{b;k}$

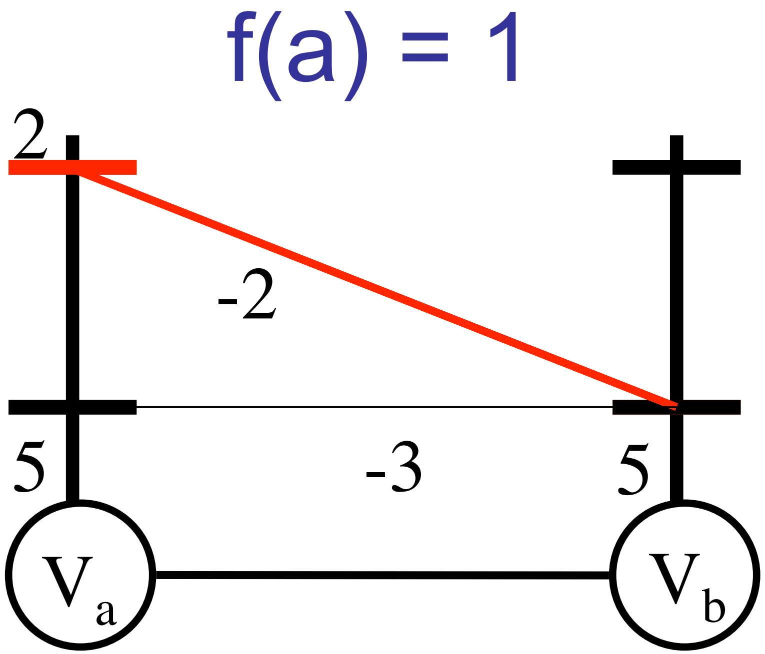
Two Variables



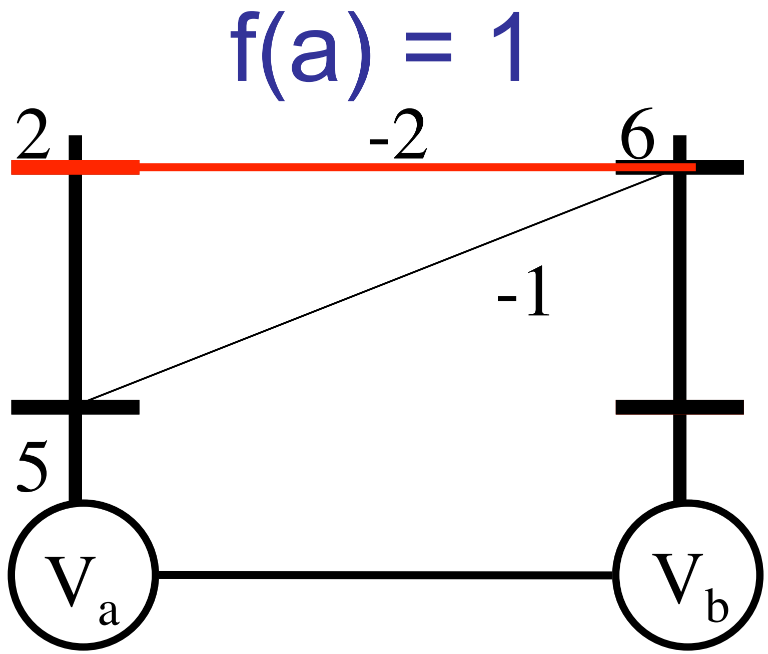
$$M_{ab;1} = \min \begin{matrix} \theta_{a;0} + \theta_{ab;01} = 5 + 1 \\ \theta_{a;1} + \theta_{ab;11} = 2 + 0 \end{matrix}$$

Choose the ***right*** constant $\theta'_{b;k} = q_{b;k}$

Two Variables



$\theta'_{b;0} = q_{b;0}$

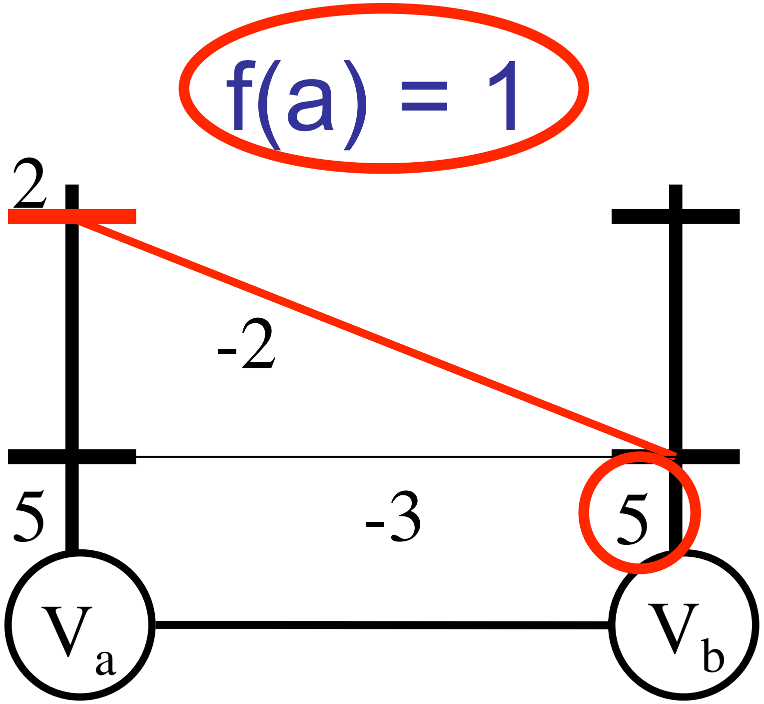


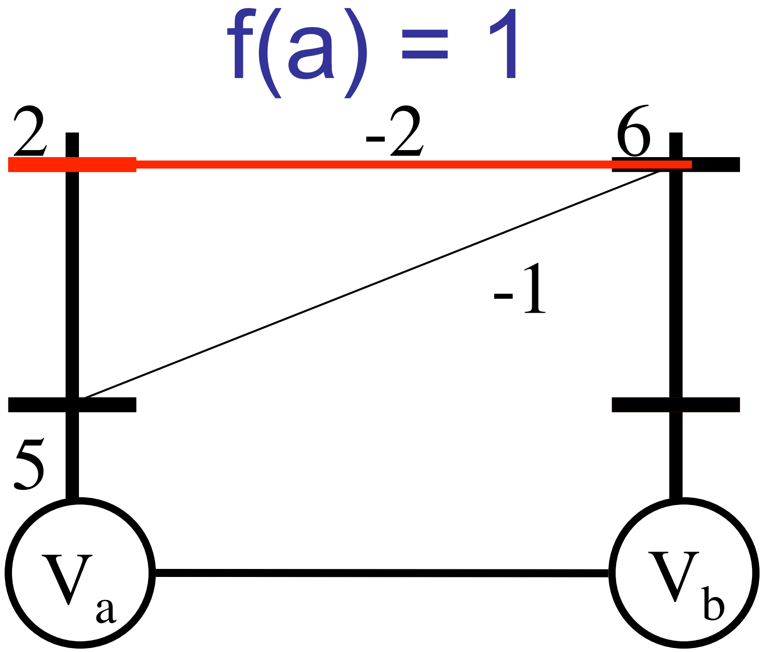
$\theta'_{b;1} = q_{b;1}$

Minimum of min-marginals = MAP estimate

Choose the ***right*** constant $\theta'_{b;k} = q_{b;k}$

Two Variables





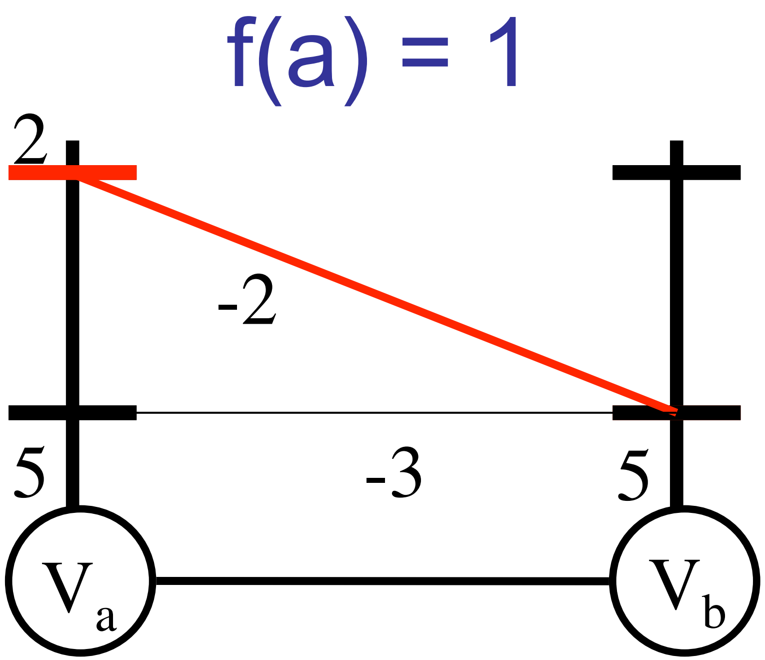
$\theta'_{b;0} = q_{b;0}$

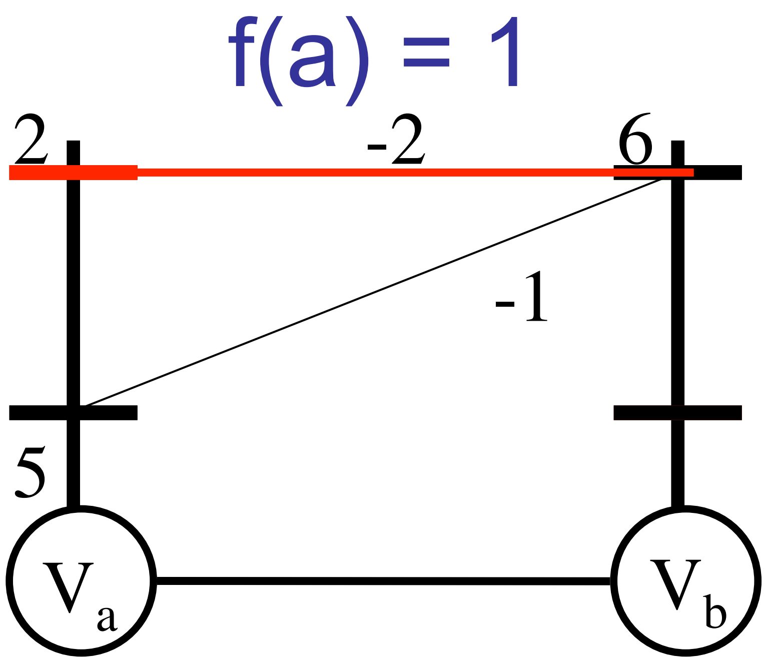
$\theta'_{b;1} = q_{b;1}$

f*(b) = 0 f*(a) = 1

Choose the ***right*** constant $\theta'_{b;k} = q_{b;k}$

Two Variables





$\theta'_{b;0} = q_{b;0}$ $\theta'_{b;1} = q_{b;1}$

We get all the min-marginals of V_b

Choose the ***right*** constant $\theta'_{b;k} = q_{b;k}$

Recap

We only need to know two sets of equations

General form of Reparameterization

$$\theta'_{a;i} = \theta_{a;i} + M_{ba;i} \qquad \theta'_{b;k} = \theta_{b;k} + M_{ab;k}$$

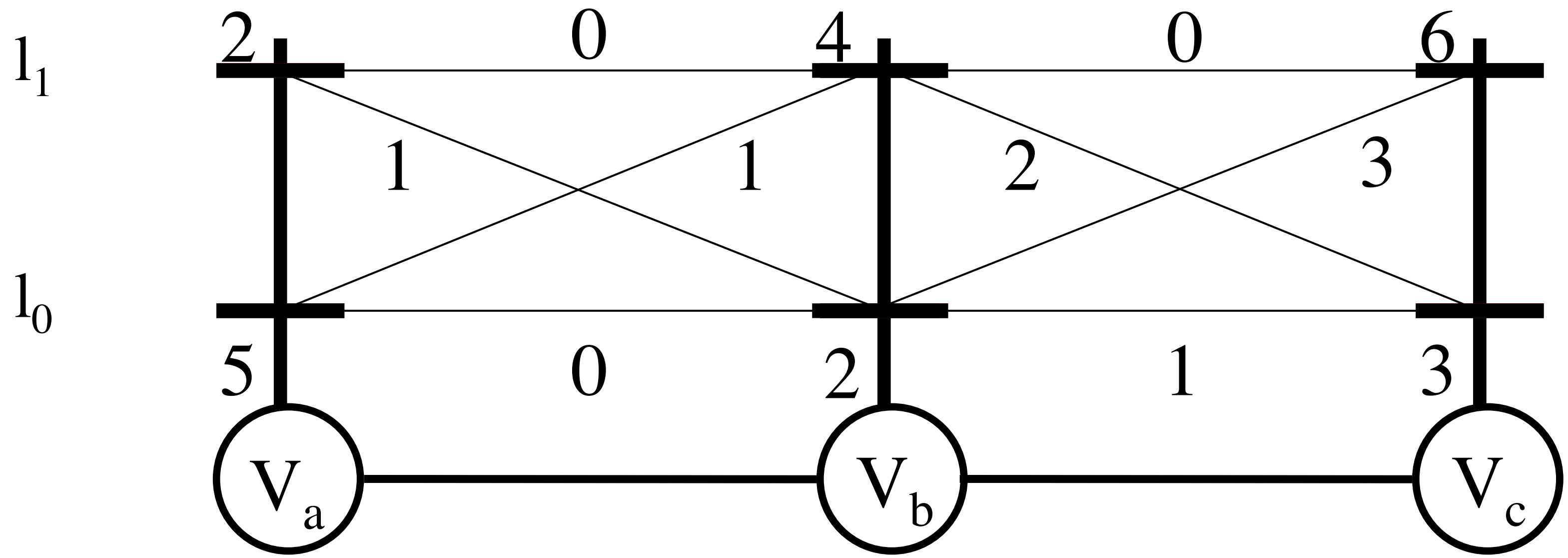
$$\theta'_{ab;ik} = \theta_{ab;ik} - M_{ab;k} - M_{ba;i}$$

Reparameterization of (a,b) in Belief Propagation

$$M_{ab;k} = \min_i \{ \theta_{a;i} + \theta_{ab;ik} \}$$

$$M_{ba;i} = 0$$

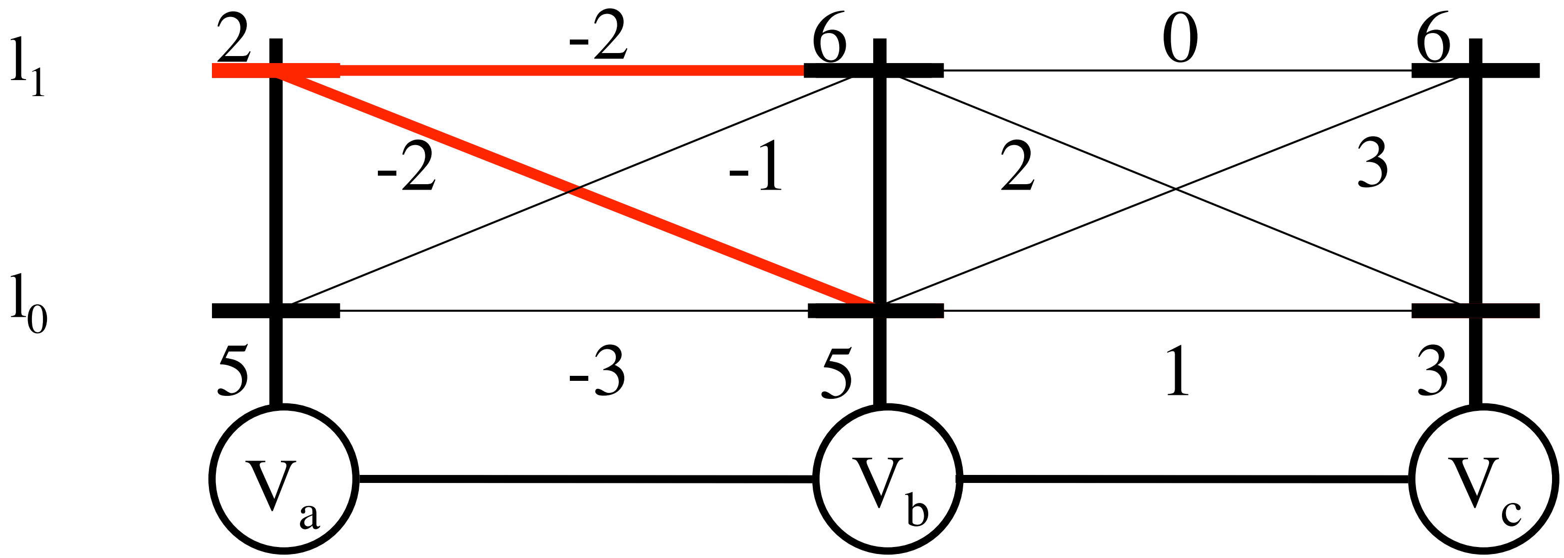
Three Variables



Reparameterize the edge (a,b) as before

Three Variables

f(a) = 1



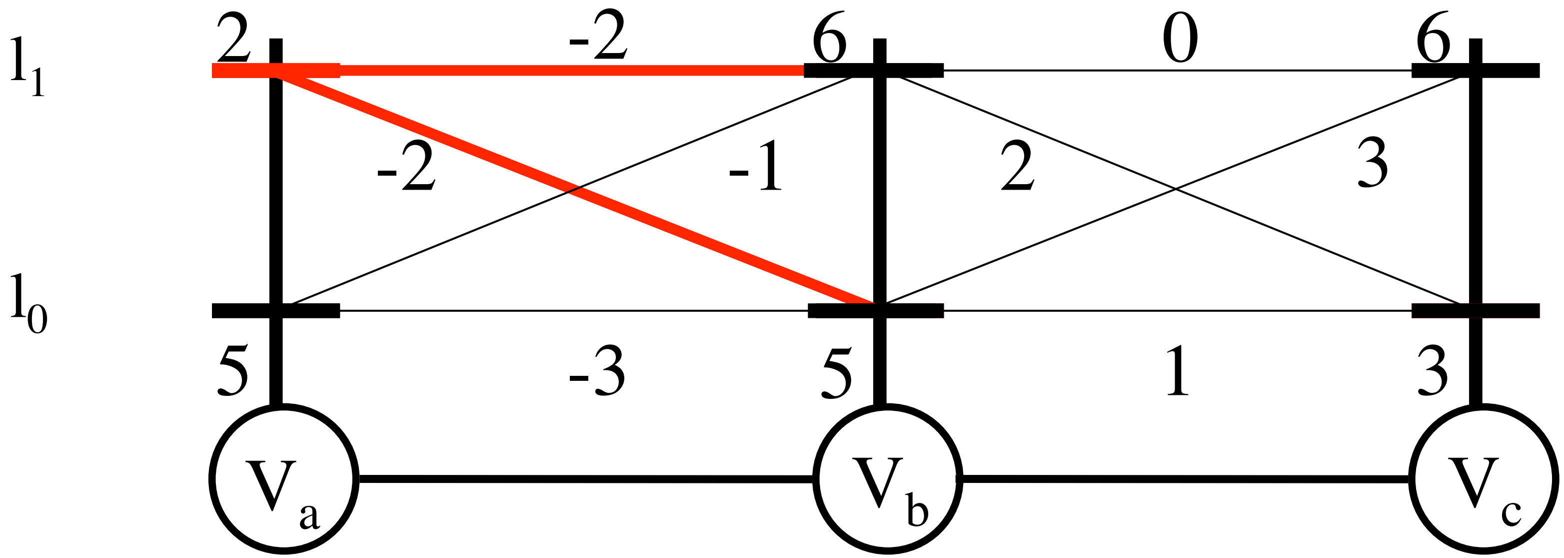
f(a) = 1

Reparameterize the edge (a,b) as before

Potentials along the red path add up to 0

Three Variables

f(a) = 1

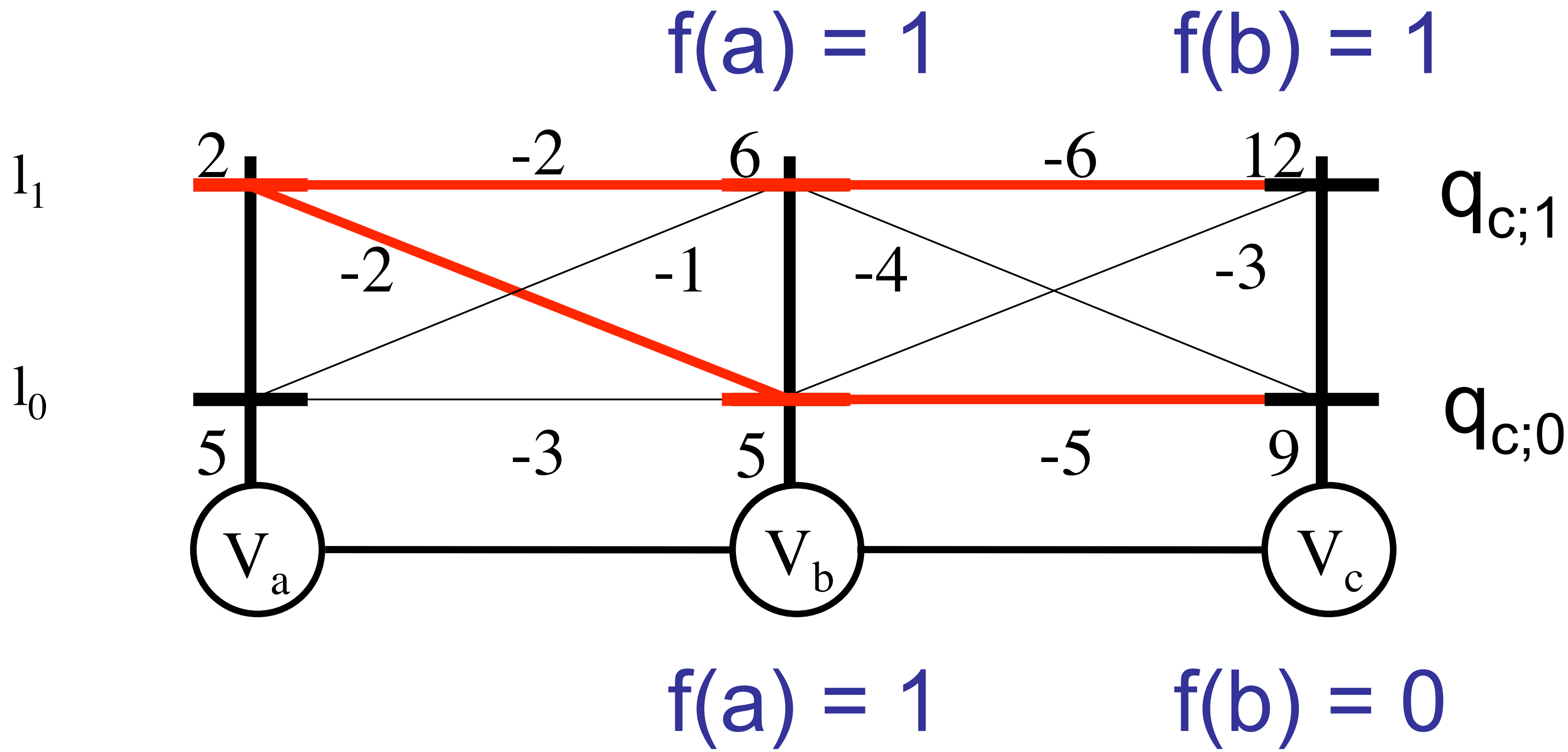


f(a) = 1

Reparameterize the edge (b,c) as before

Potentials along the red path add up to 0

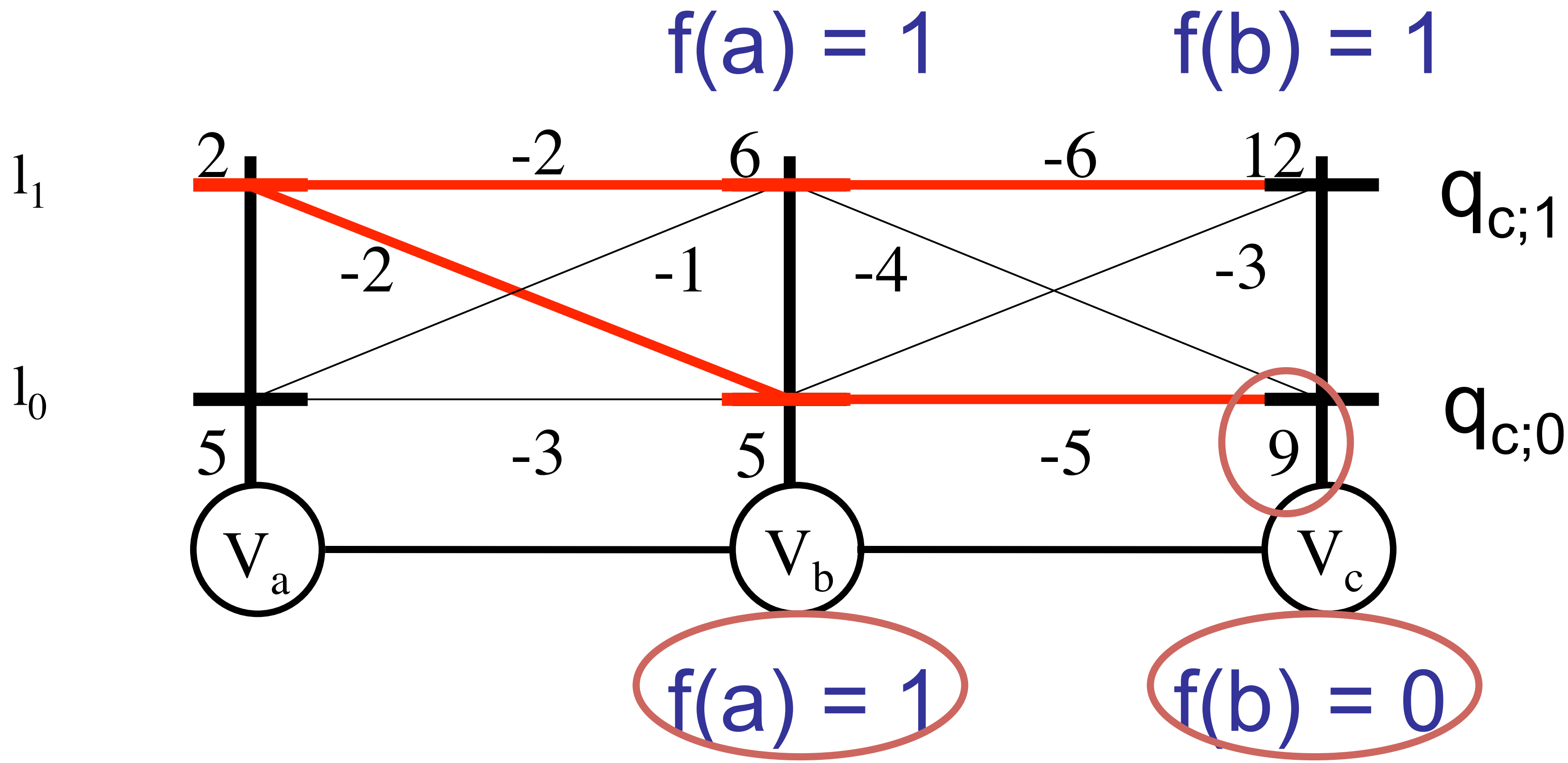
Three Variables



Reparameterize the edge (b,c) as before

Potentials along the red path add up to 0

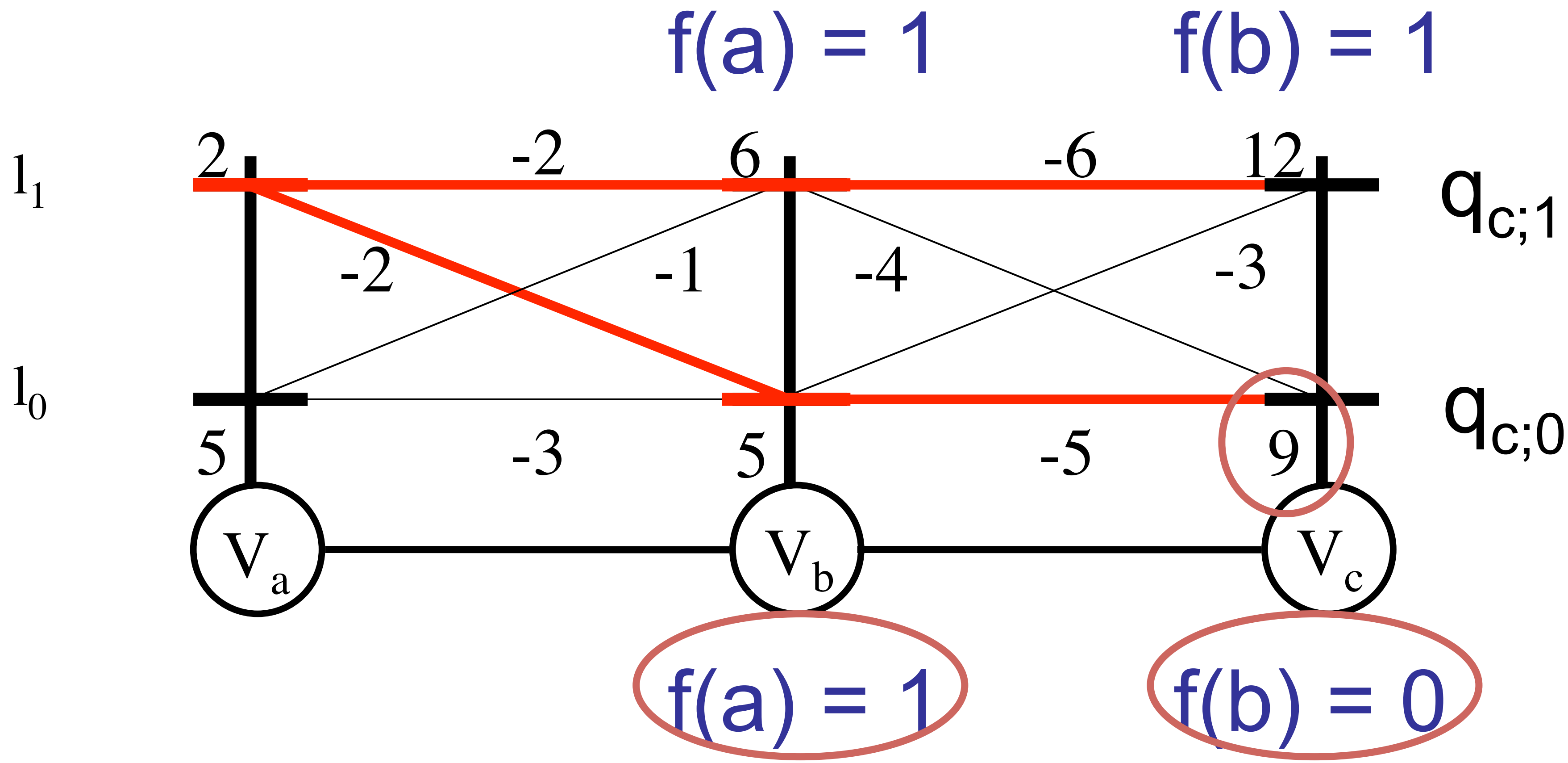
Three Variables



f*(c) = 0 f*(b) = 0 f*(a) = 1

Generalizes to any length chain

Three Variables



f*(c) = 0 f*(b) = 0 f*(a) = 1

Only Dynamic Programming

Why Dynamic Programming?

3 variables ≡ 2 variables + book-keeping

n variables ≡ (n-1) variables + book-keeping

Start from left, go to right

Reparameterize current edge (a,b)

$$M_{ab;k} = \min_i \{ \theta_{a;i} + \theta_{ab;ik} \}$$

$$\theta'_{b;k} = \theta_{b;k} + M_{ab;k} \quad \theta'_{ab;ik} = \theta_{ab;ik} - M_{ab;k}$$

Repeat

Why Dynamic Programming?

Messages Message Passing

Why stop at dynamic programming?

Start from left, go to right

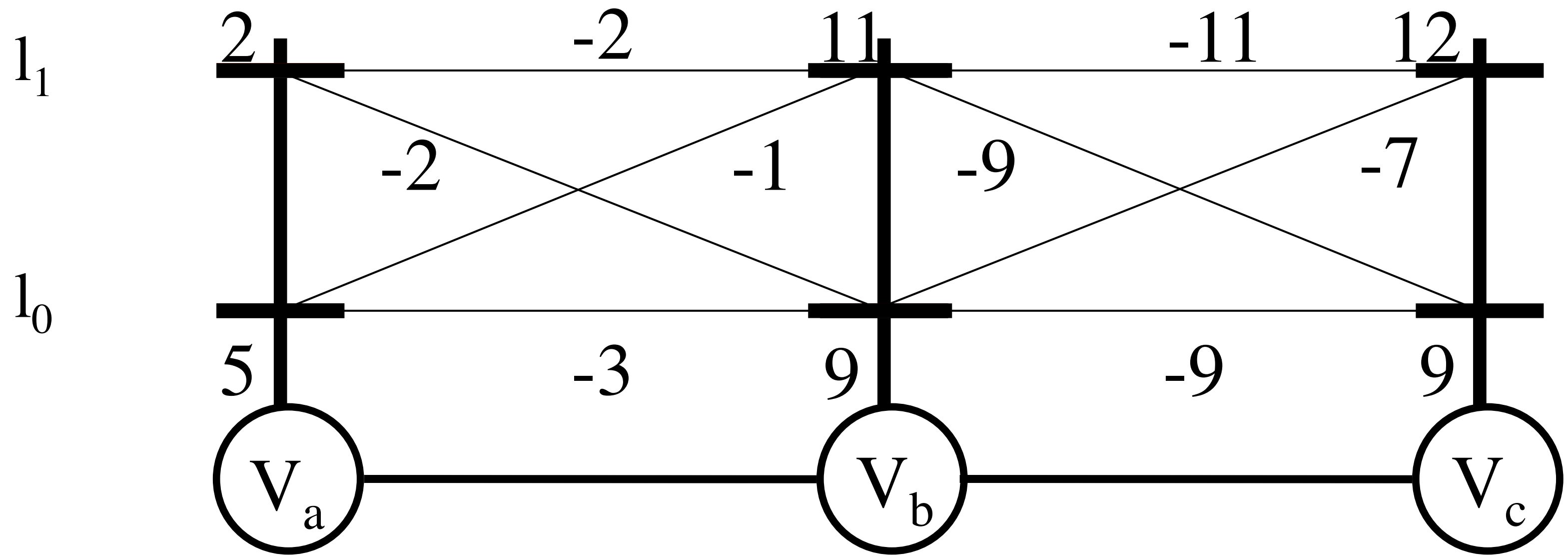
Reparameterize current edge (a,b)

$$M_{ab;k} = \min_i \{ \theta_{a;i} + \theta_{ab;ik} \}$$

$$\theta'_{b;k} = \theta_{b;k} + M_{ab;k} \quad \theta'_{ab;ik} = \theta_{ab;ik} - M_{ab;k}$$

Repeat

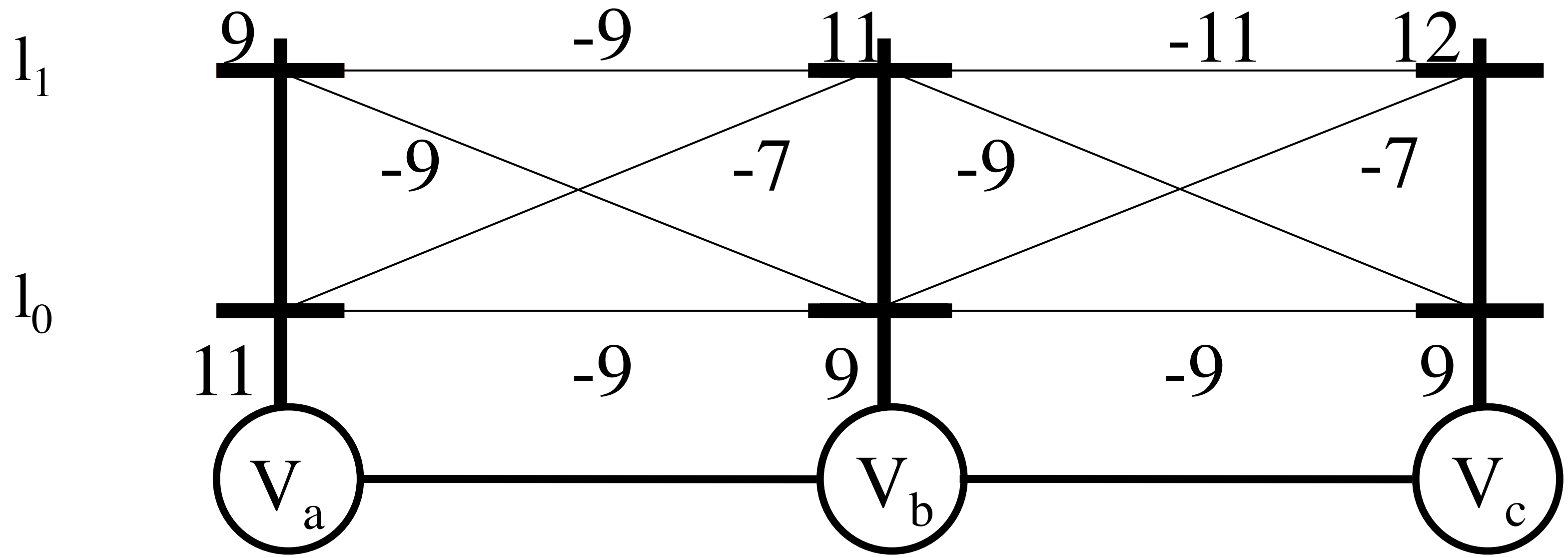
Three Variables



Reparameterize the edge (c,b) as before

$$\theta'_{b;i} = q_{b;i}$$

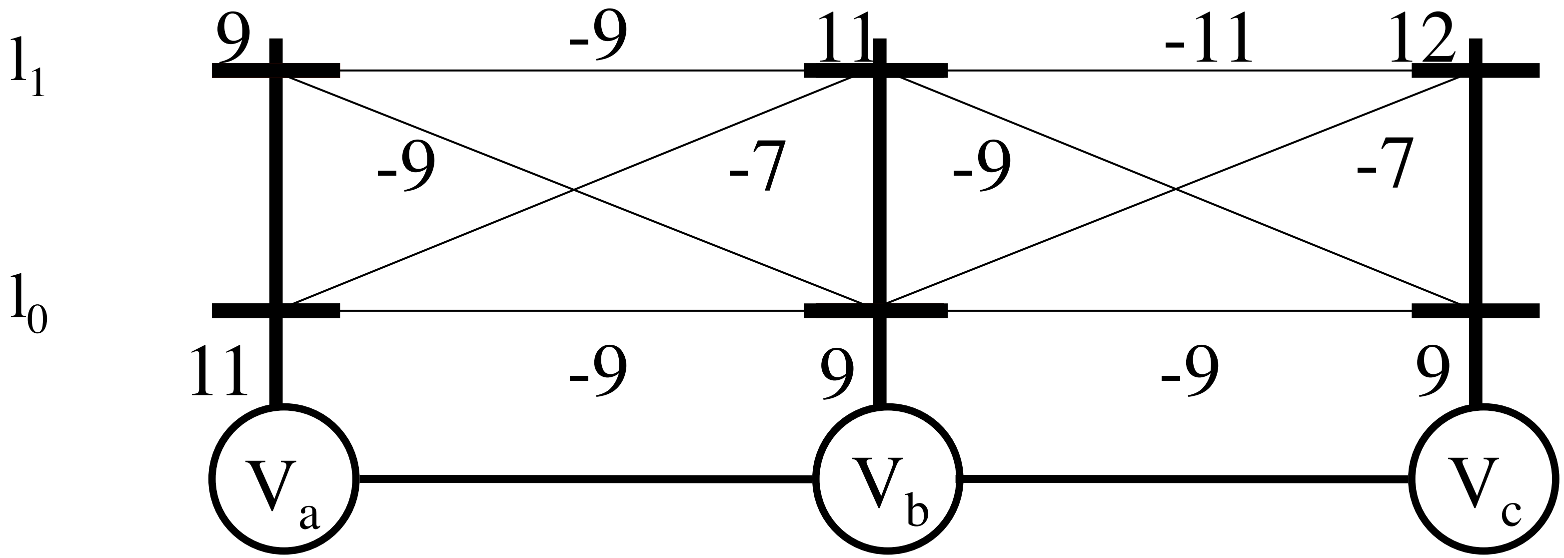
Three Variables



Reparameterize the edge (b,a) as before

$$\theta'_{a;i} = q_{a;i}$$

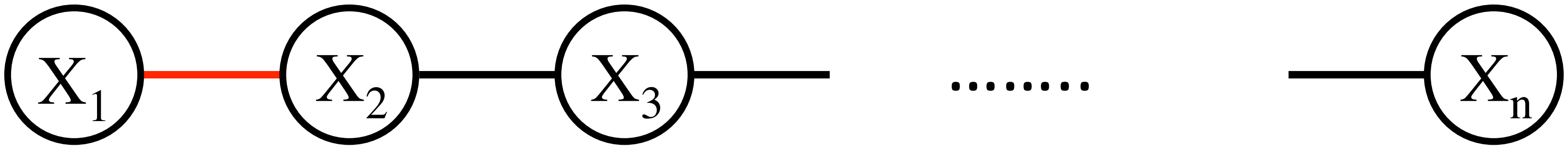
Three Variables



Forward Pass → ← Backward Pass

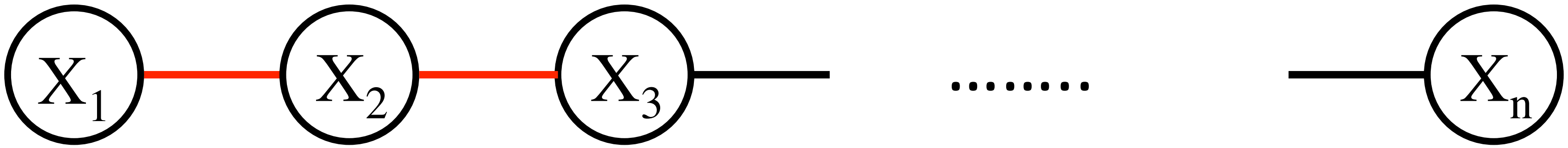
All min-marginals are computed

Chains



Reparameterize the edge (1,2)

Chains



Reparameterize the edge (2,3)

Chains

Reparameterize the edge (n-1,n)

Min-marginals $e_n(i)$ for all labels

Belief Propagation on Chains

Start from left, go to right

Reparameterize current edge (a,b)

$$M_{ab;k} = \min_i \{ \theta_{a;i} + \theta_{ab;ik} \}$$

$$\theta'_{b;k} = \theta_{b;k} + M_{ab;k} \quad \theta'_{ab;ik} = \theta_{ab;ik} - M_{ab;k}$$

Repeat till the end of the chain

Start from right, go to left

Repeat till the end of the chain

Belief Propagation on Chains

- Generalizes to chains of any length

- A way of computing reparam constants

- Forward Pass - Start to End
  - MAP estimate
  - Min-marginals of final variable

- Backward Pass - End to start
  - All other min-marginals

Computational Complexity

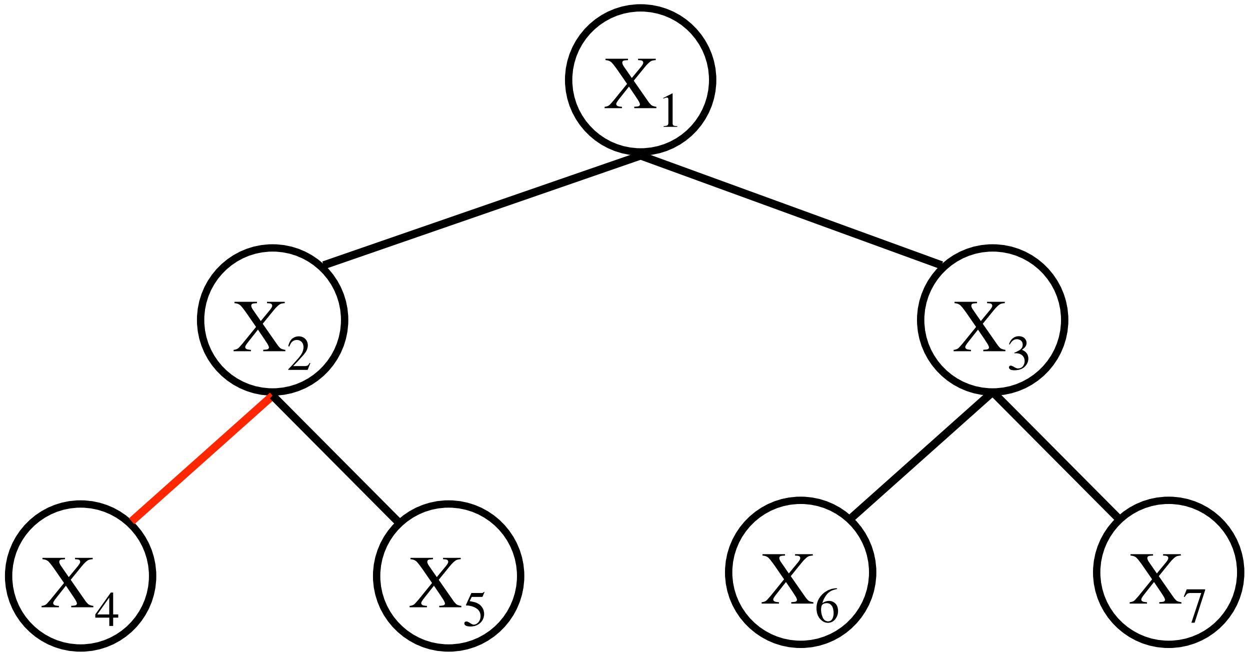
Number of reparameterization constants = (n-1)h

Complexity for each constant = O(h)

Total complexity = $O(nh^2)$

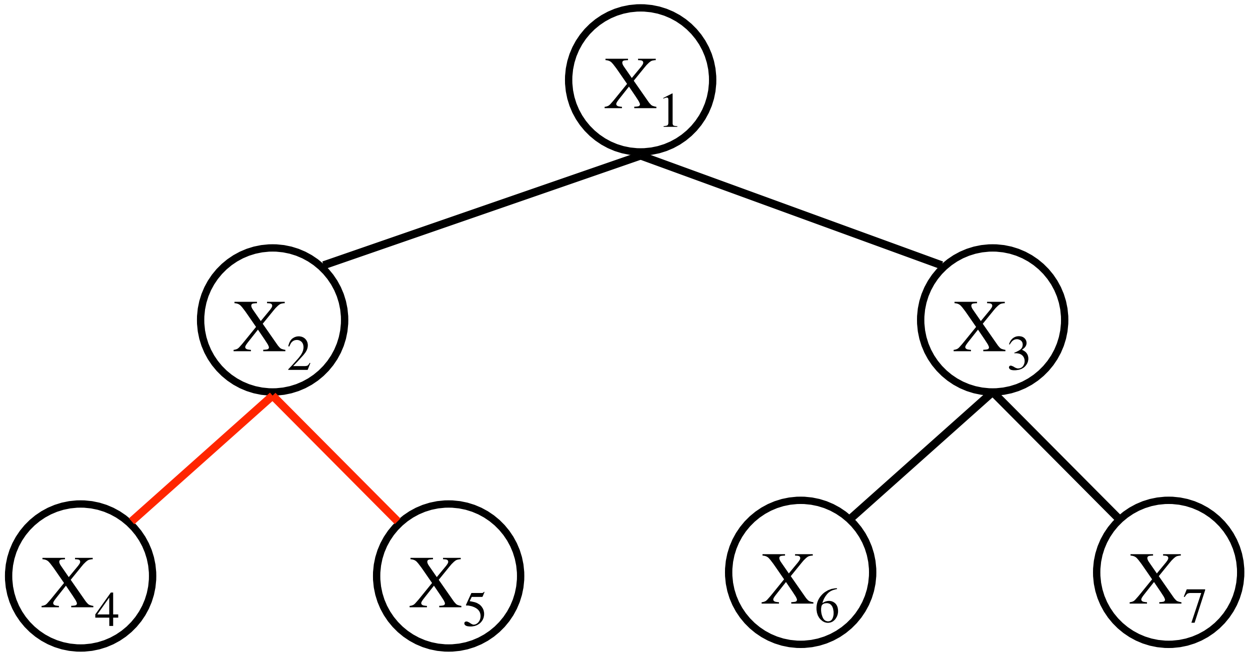
Better than brute-force $O(h^n)$

Trees



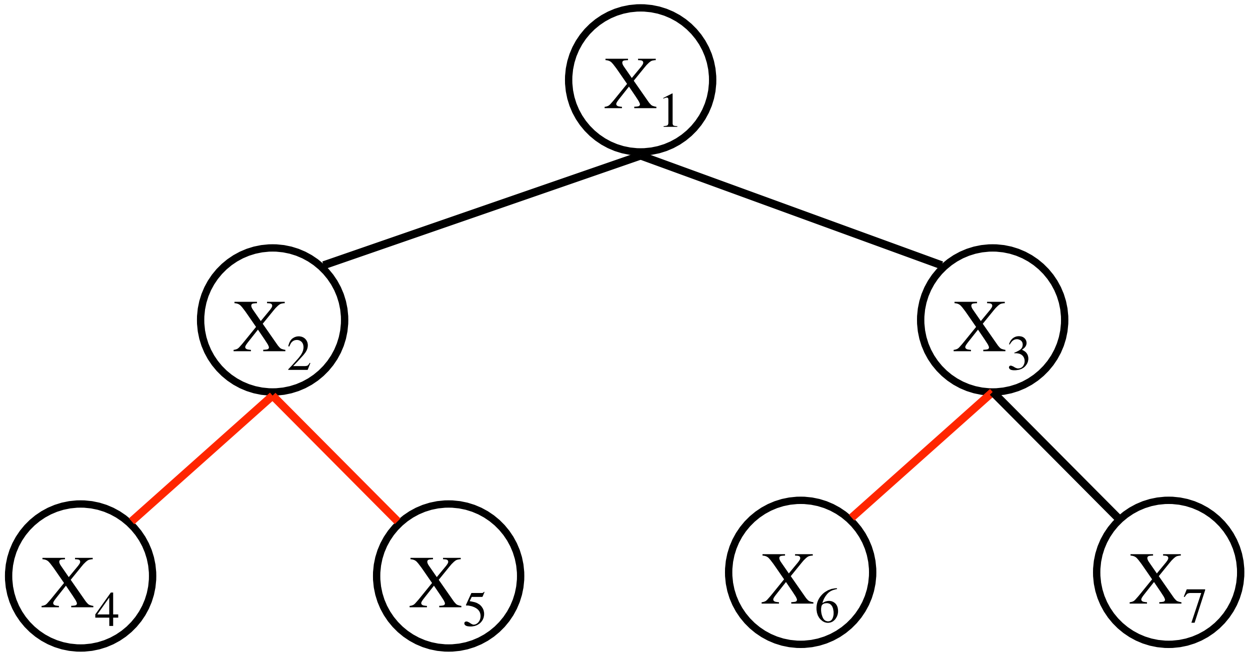
Reparameterize the edge (4,2)

Trees



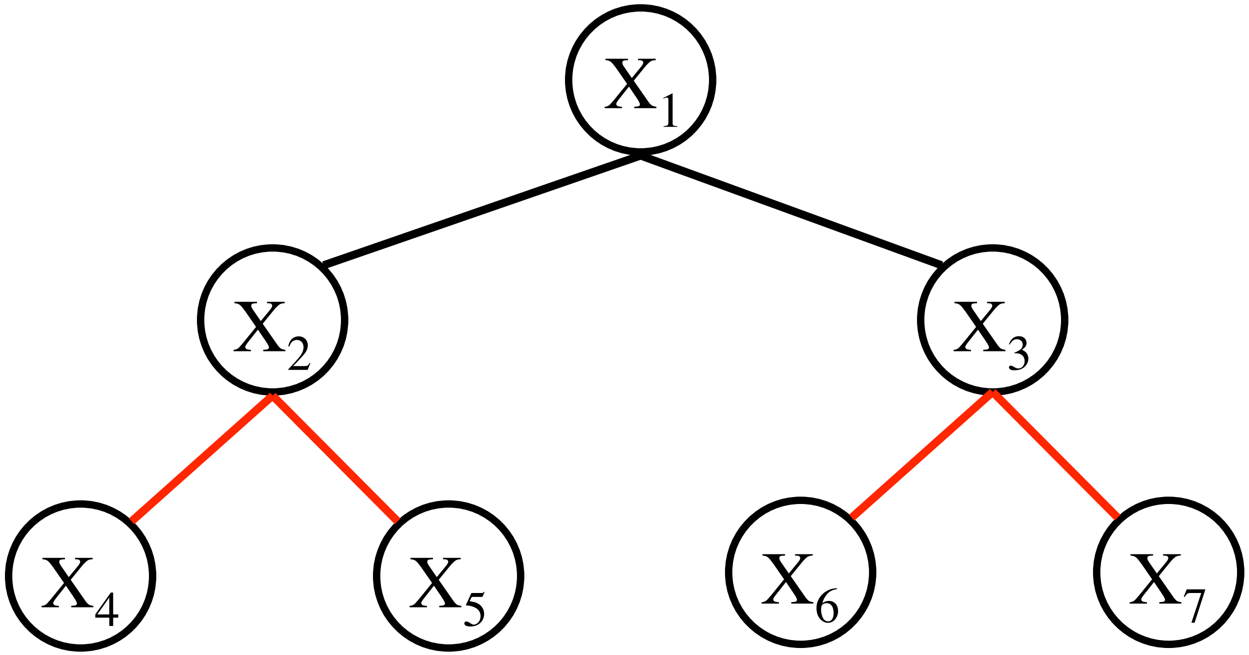
Reparameterize the edge (5,2)

Trees



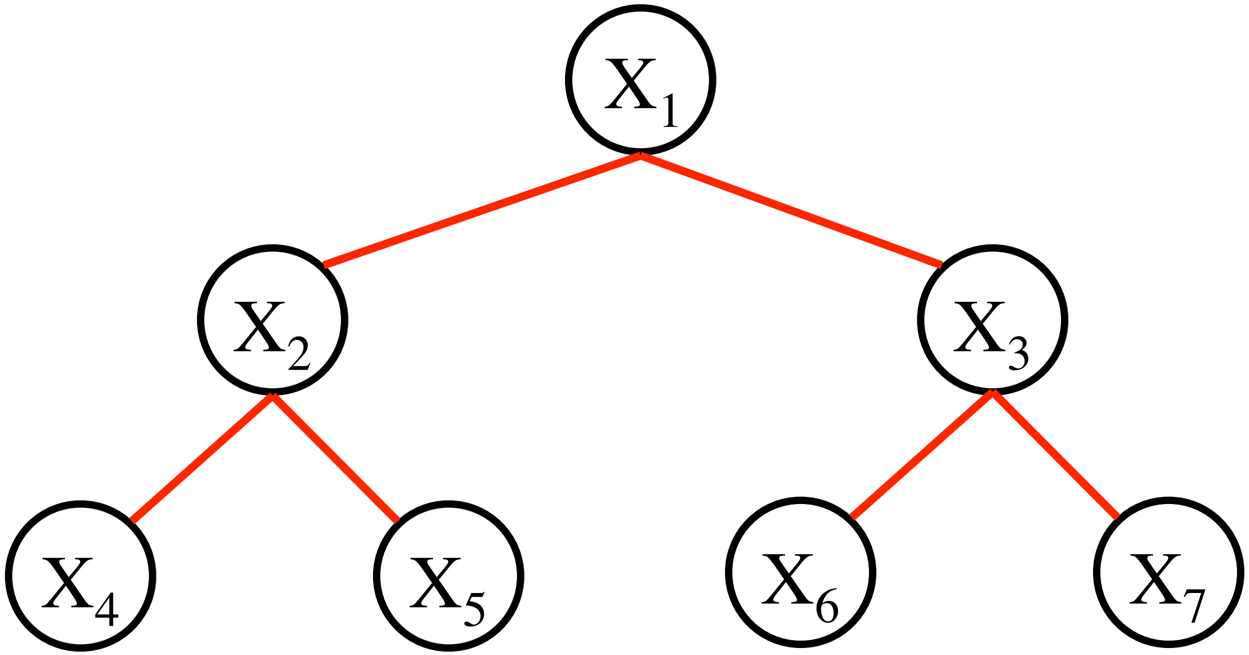
Reparameterize the edge (6,3)

Trees



Reparameterize the edge (7,3)

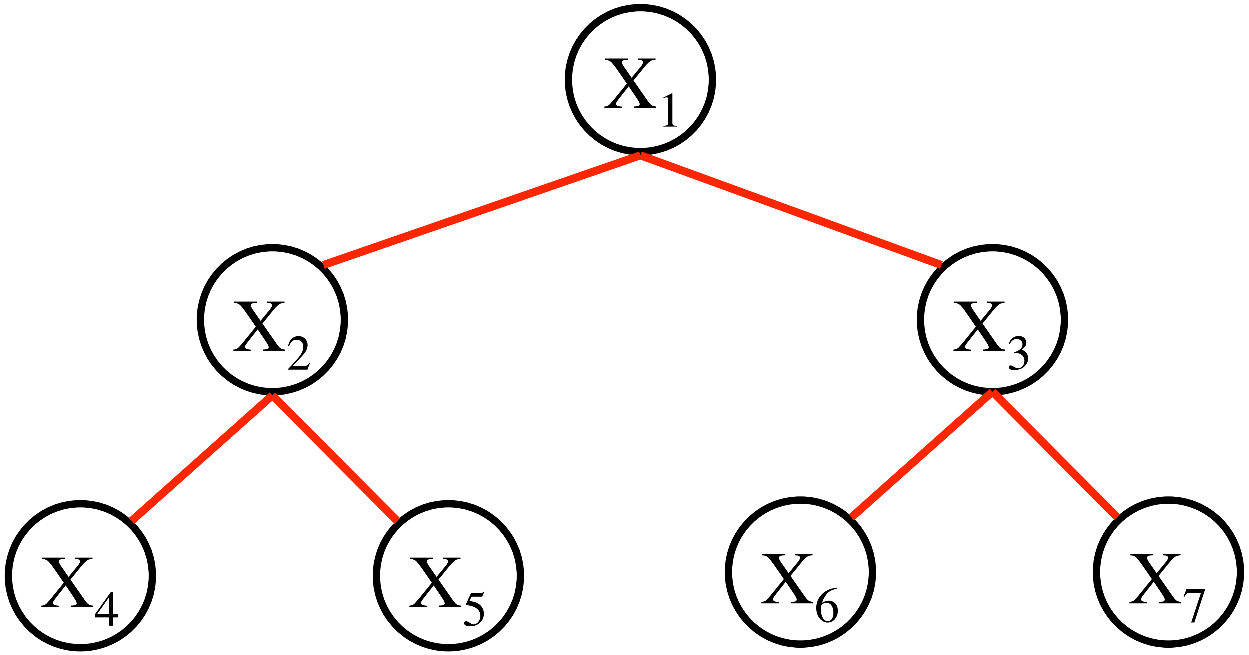
Trees



Reparameterize the edge (3,1)

Min-marginals $e_1(i)$ for all labels

Trees

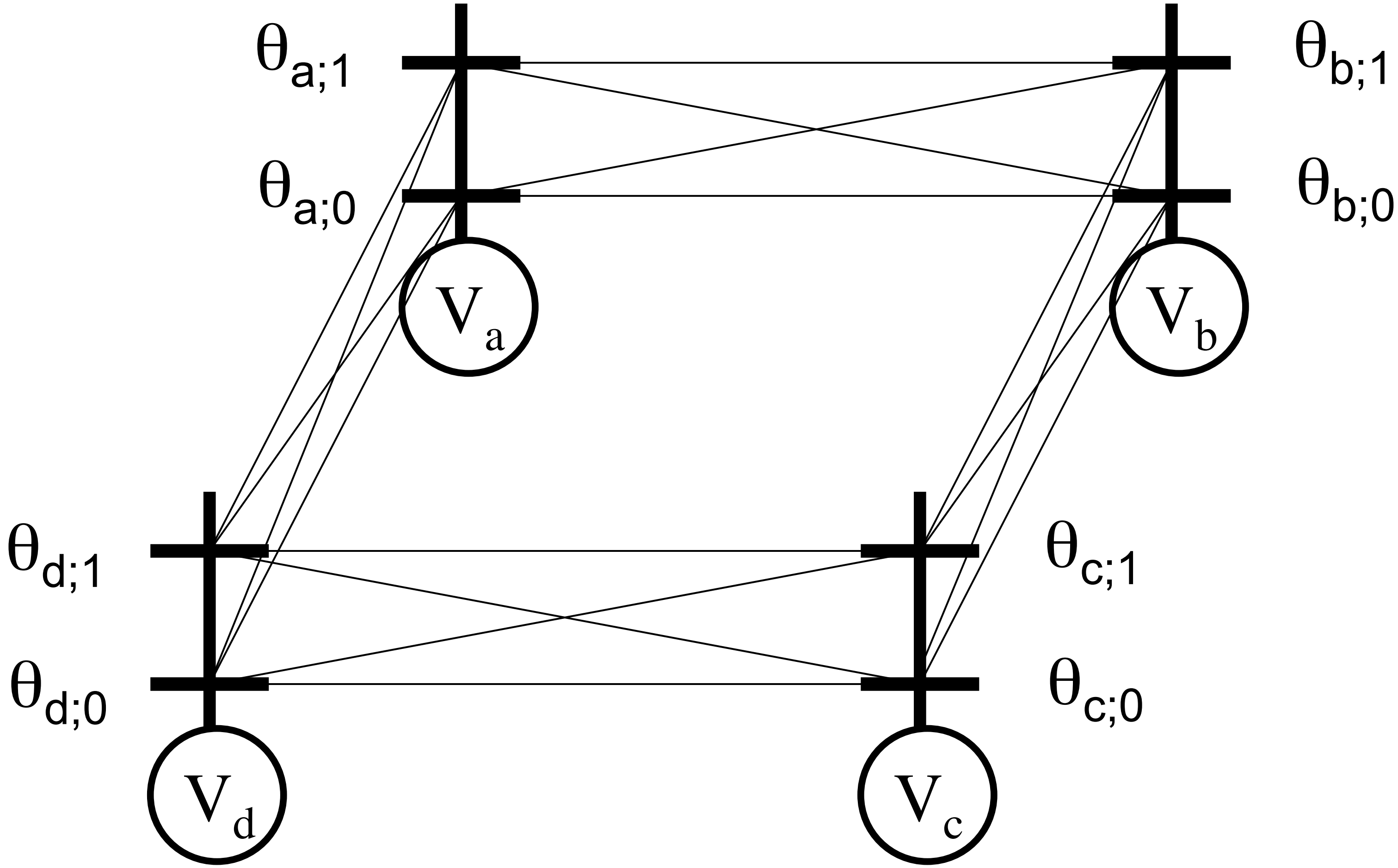


Start from leaves and move towards root

Pick the minimum of min-marginals

Backtrack to find the best labeling **x**

Belief Propagation on Cycles

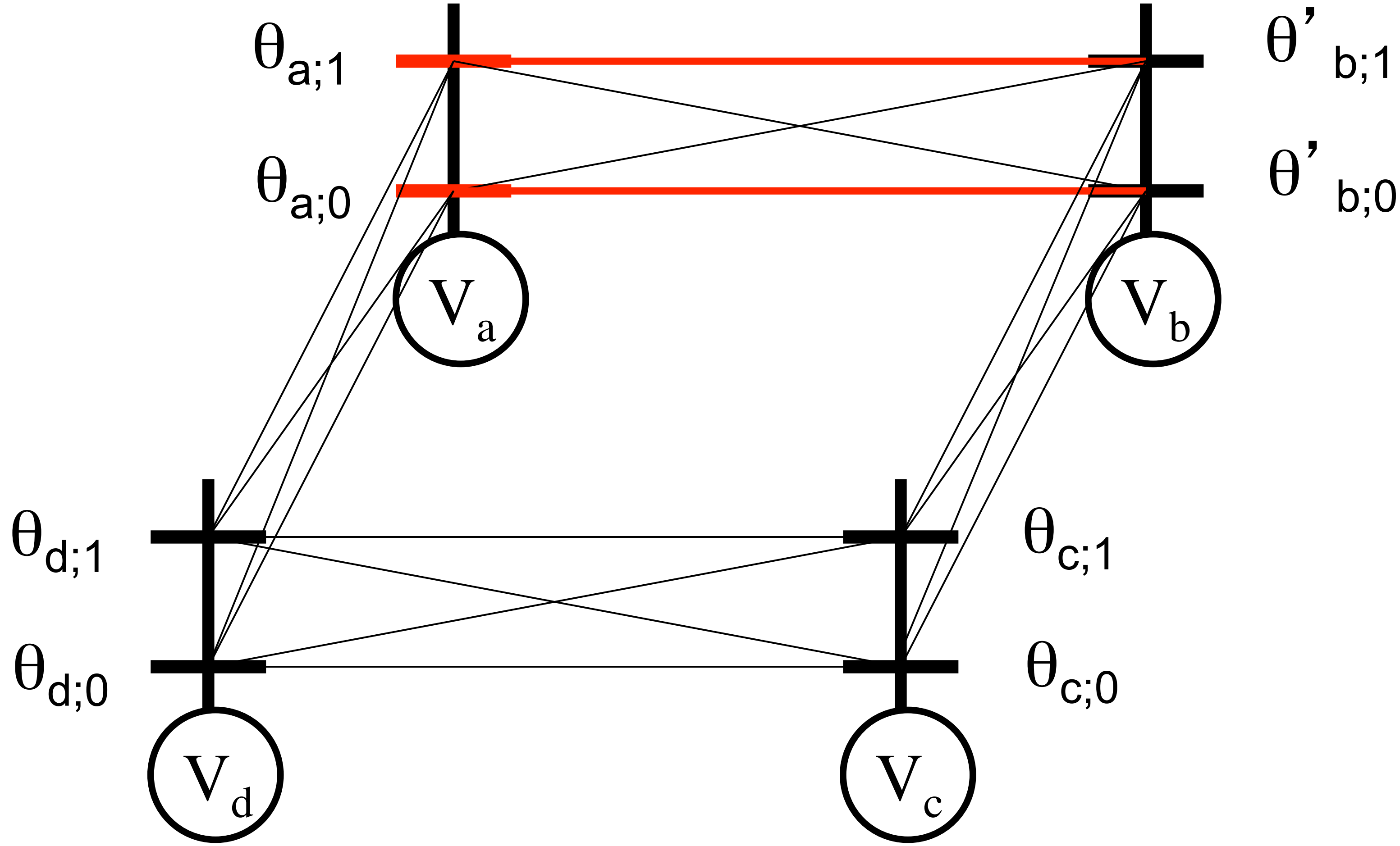


Where do we start?

Arbitrarily

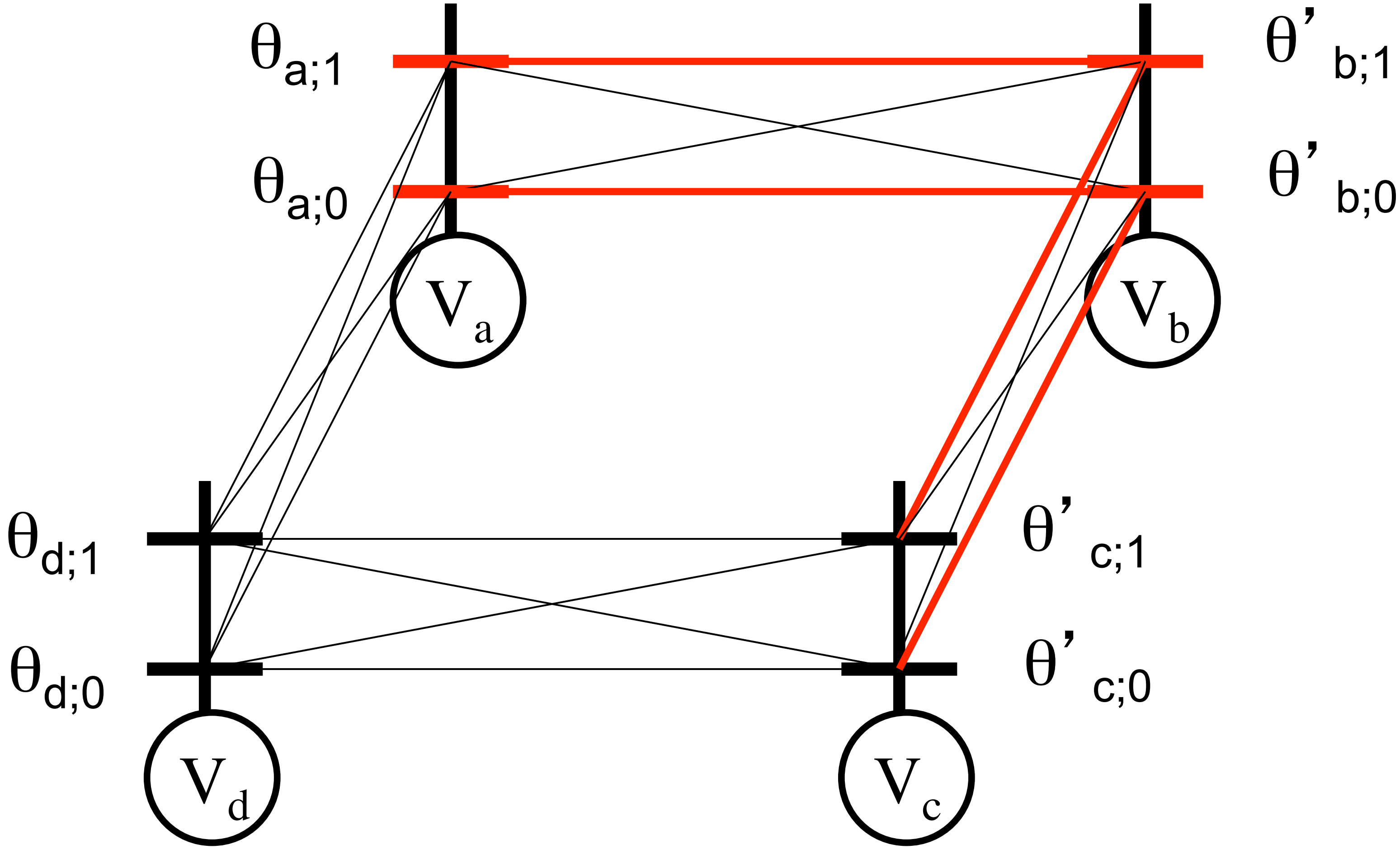
Reparameterize (a,b)

Belief Propagation on Cycles



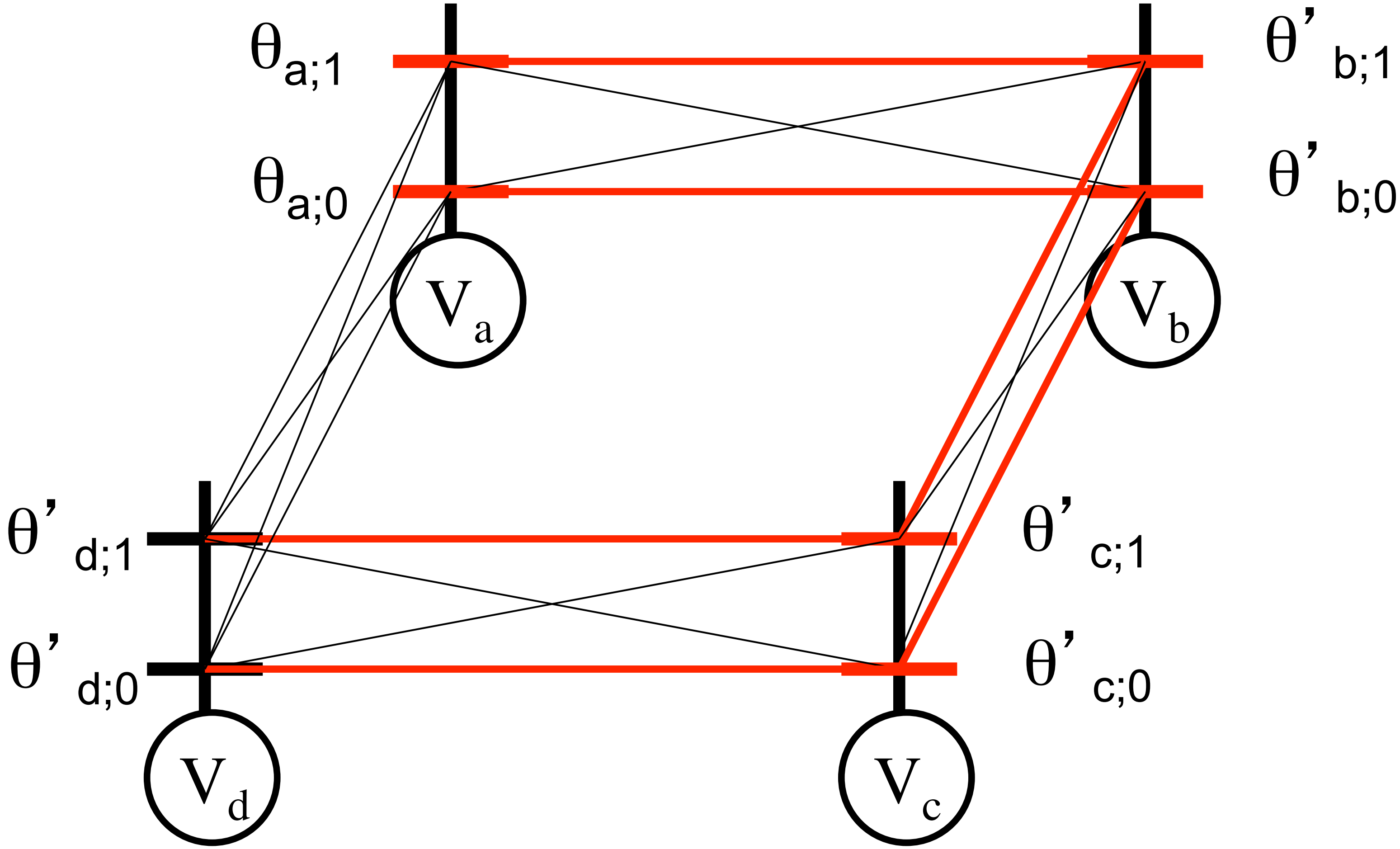
Potentials along the red path add up to 0

Belief Propagation on Cycles



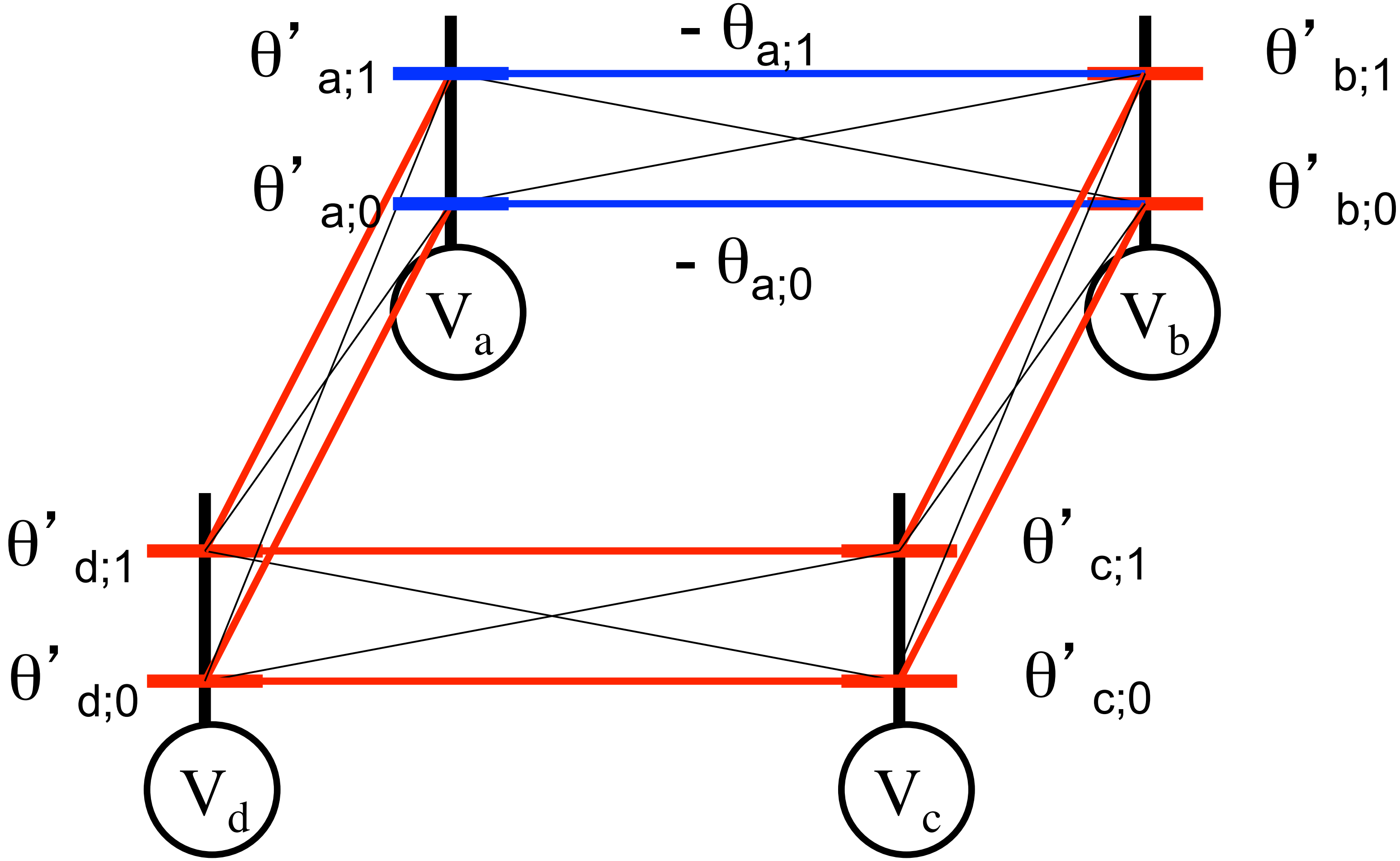
Potentials along the red path add up to 0

Belief Propagation on Cycles



Potentials along the red path add up to 0

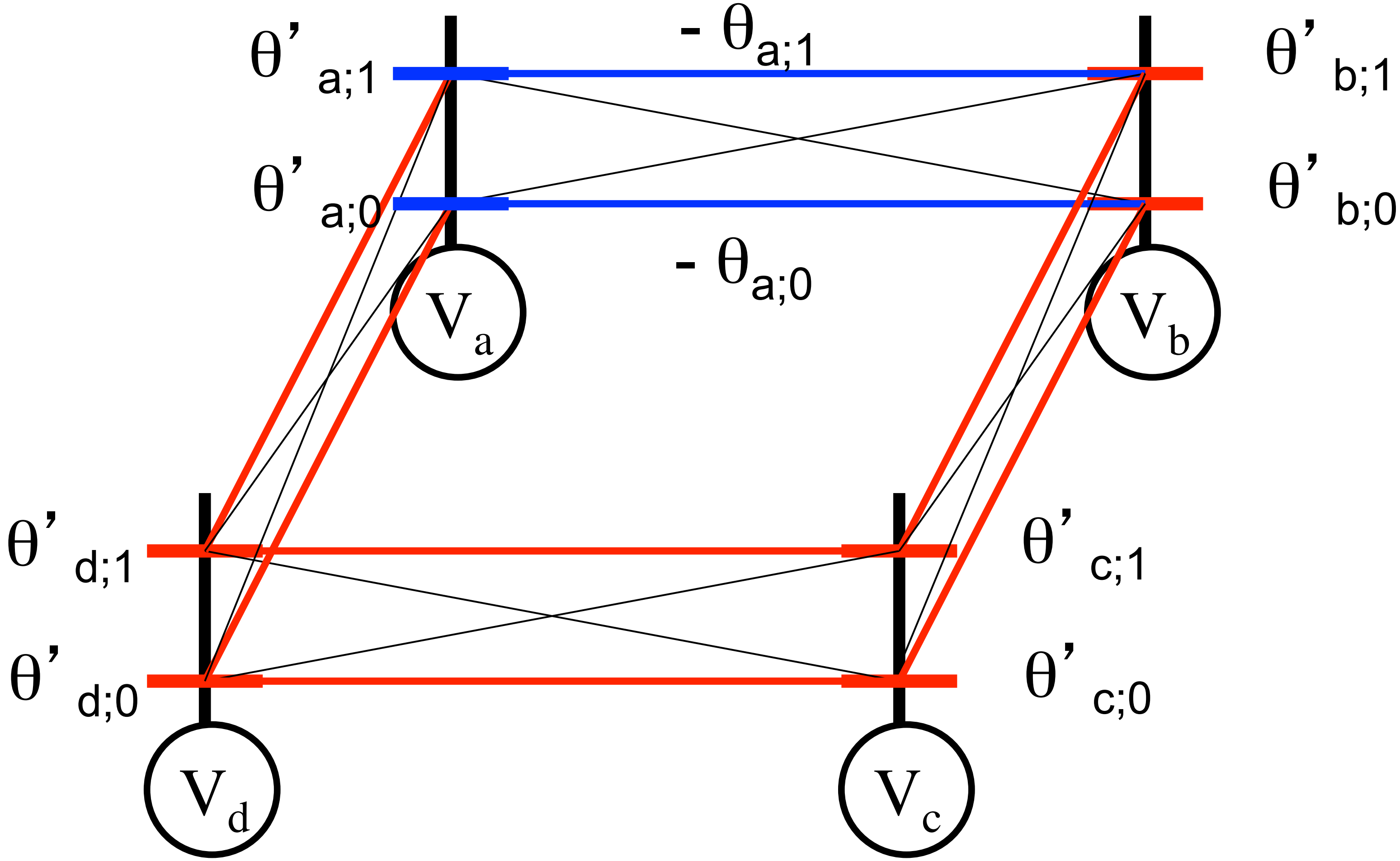
Belief Propagation on Cycles



$\theta'_{a;0} - \theta_{a;0} = q_{a;0}$ $\theta'_{a;1} - \theta_{a;1} = q_{a;1}$

Potentials along the red path add up to 0

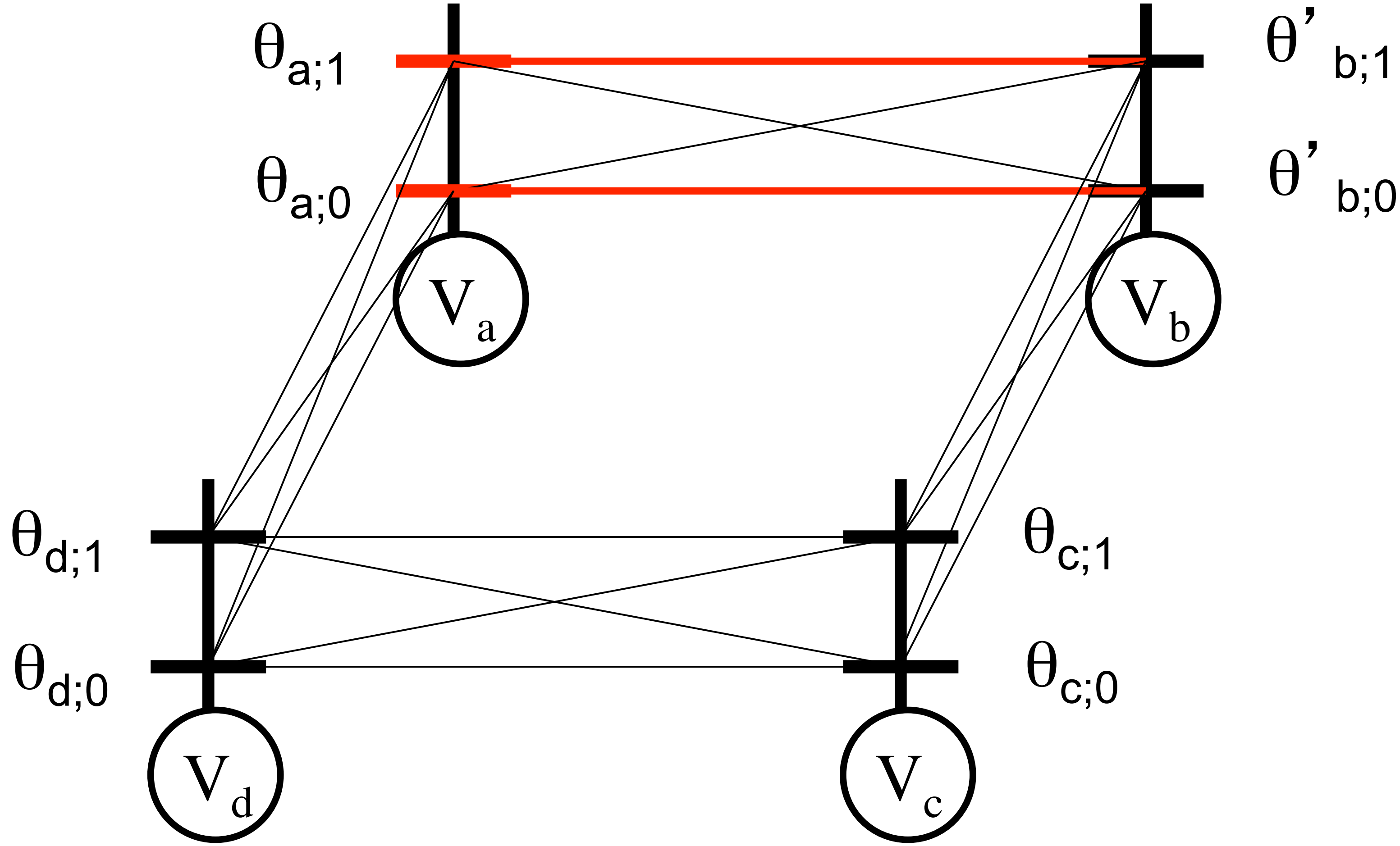
Belief Propagation on Cycles



$\theta'_{a;0} - \theta_{a;0} = q_{a;0}$ $\theta'_{a;1} - \theta_{a;1} = q_{a;1}$

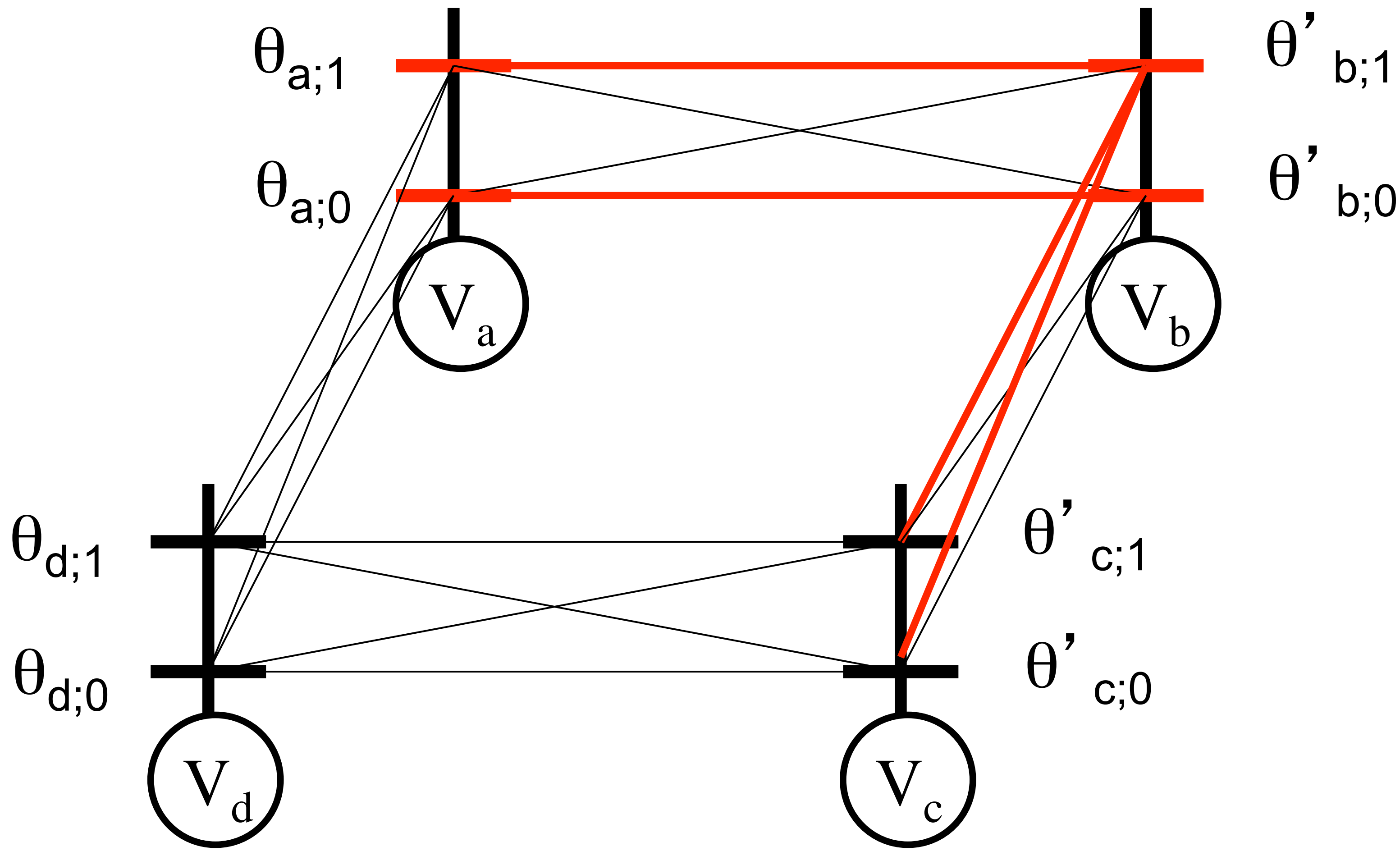
Pick minimum min-marginal. Follow red path.

Belief Propagation on Cycles



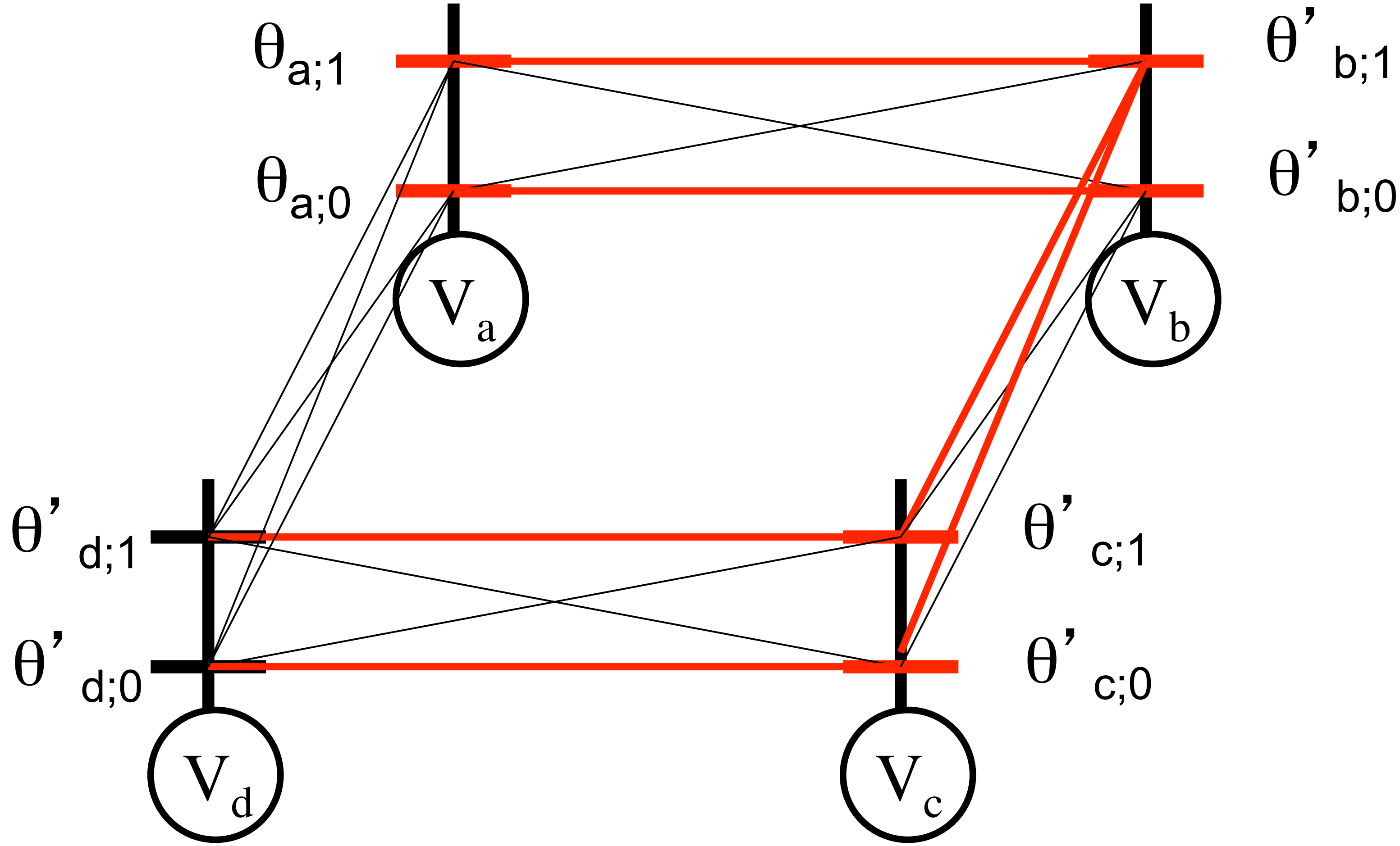
Potentials along the red path add up to 0

Belief Propagation on Cycles



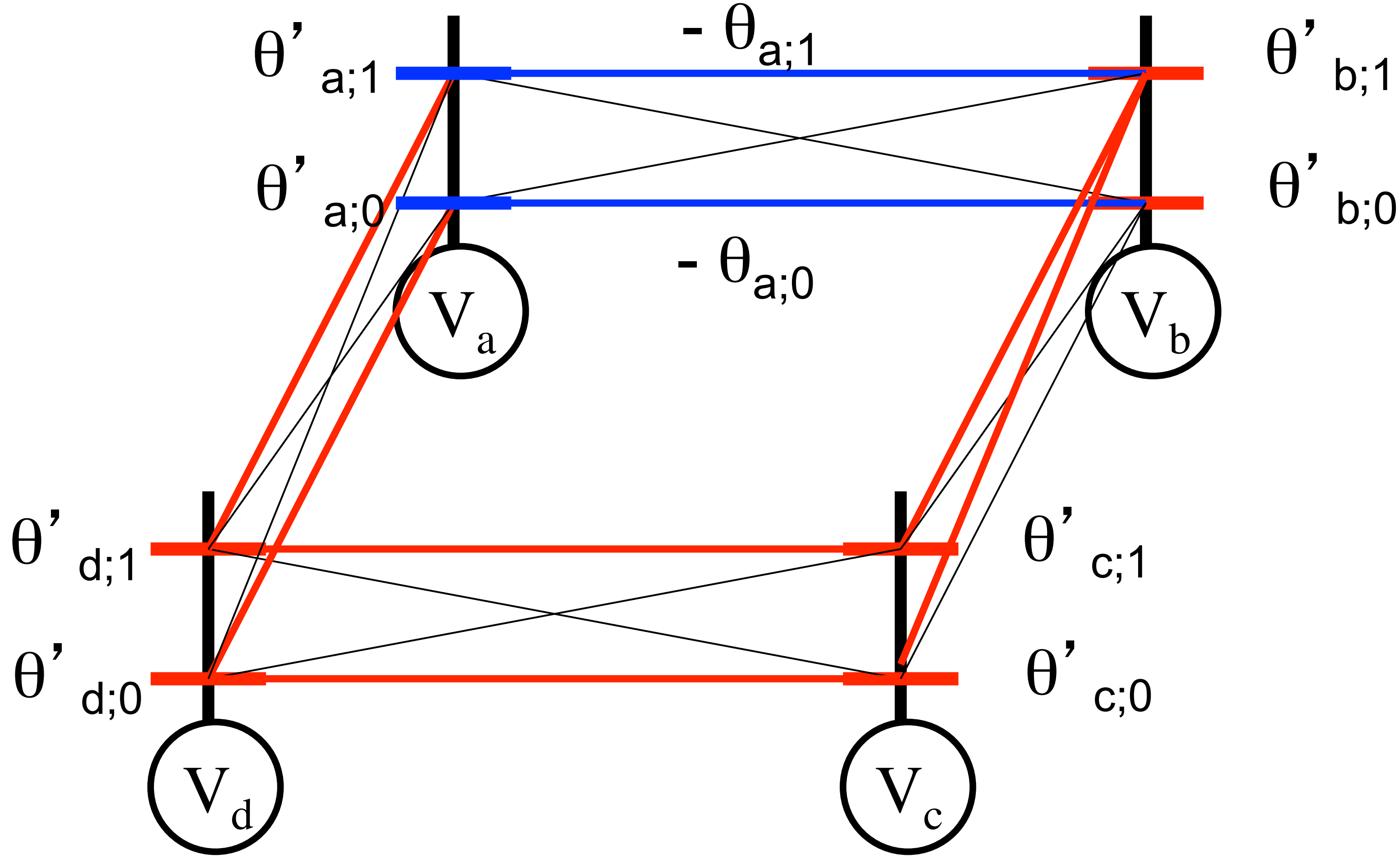
Potentials along the red path add up to 0

Belief Propagation on Cycles



Potentials along the red path add up to 0

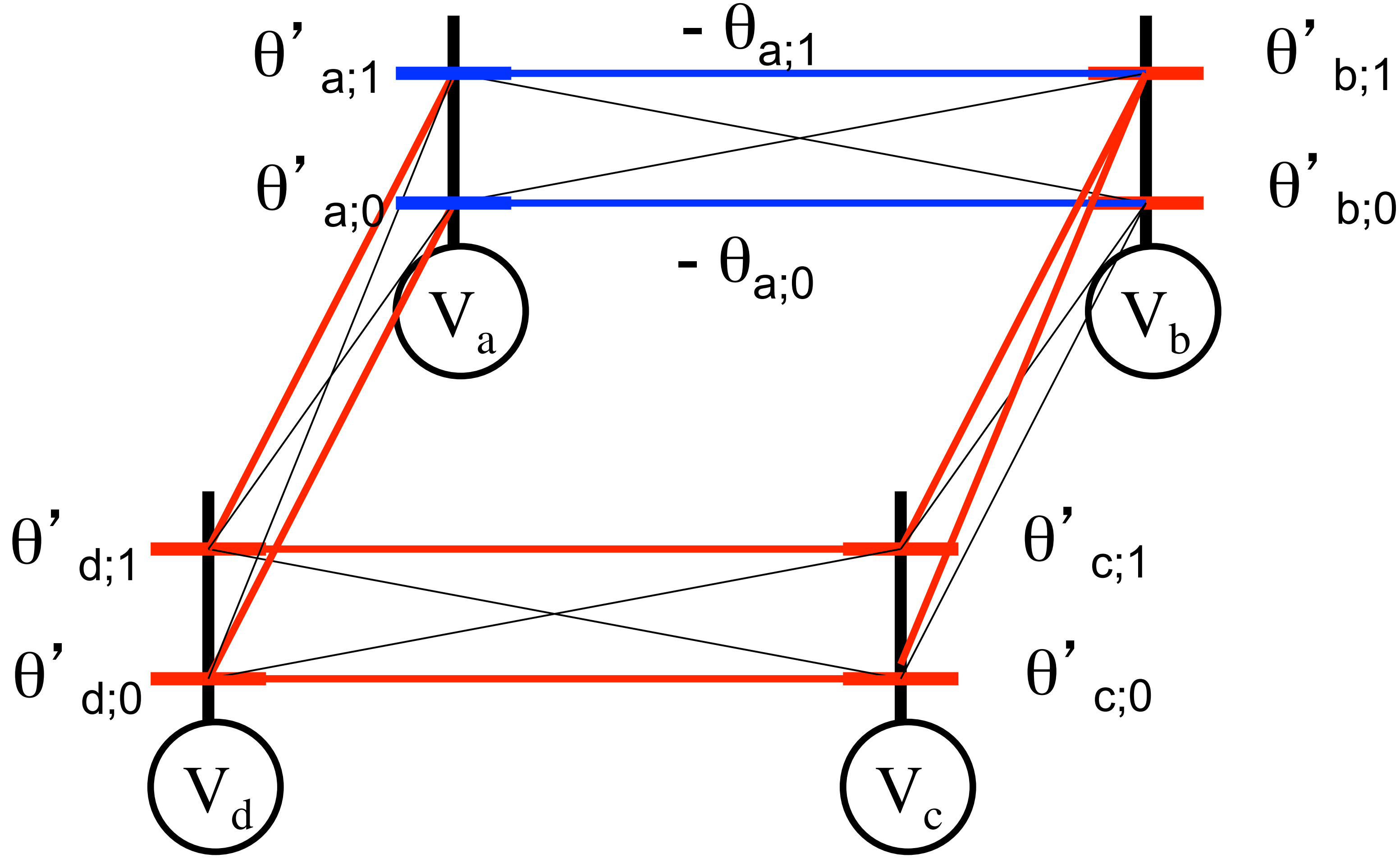
Belief Propagation on Cycles



$\theta'_{a;1} - \theta_{a;1} = q_{a;1} \leq \theta'_{a;0} - \theta_{a;0} = q_{a;0}$

Potentials along the red path add up to 0

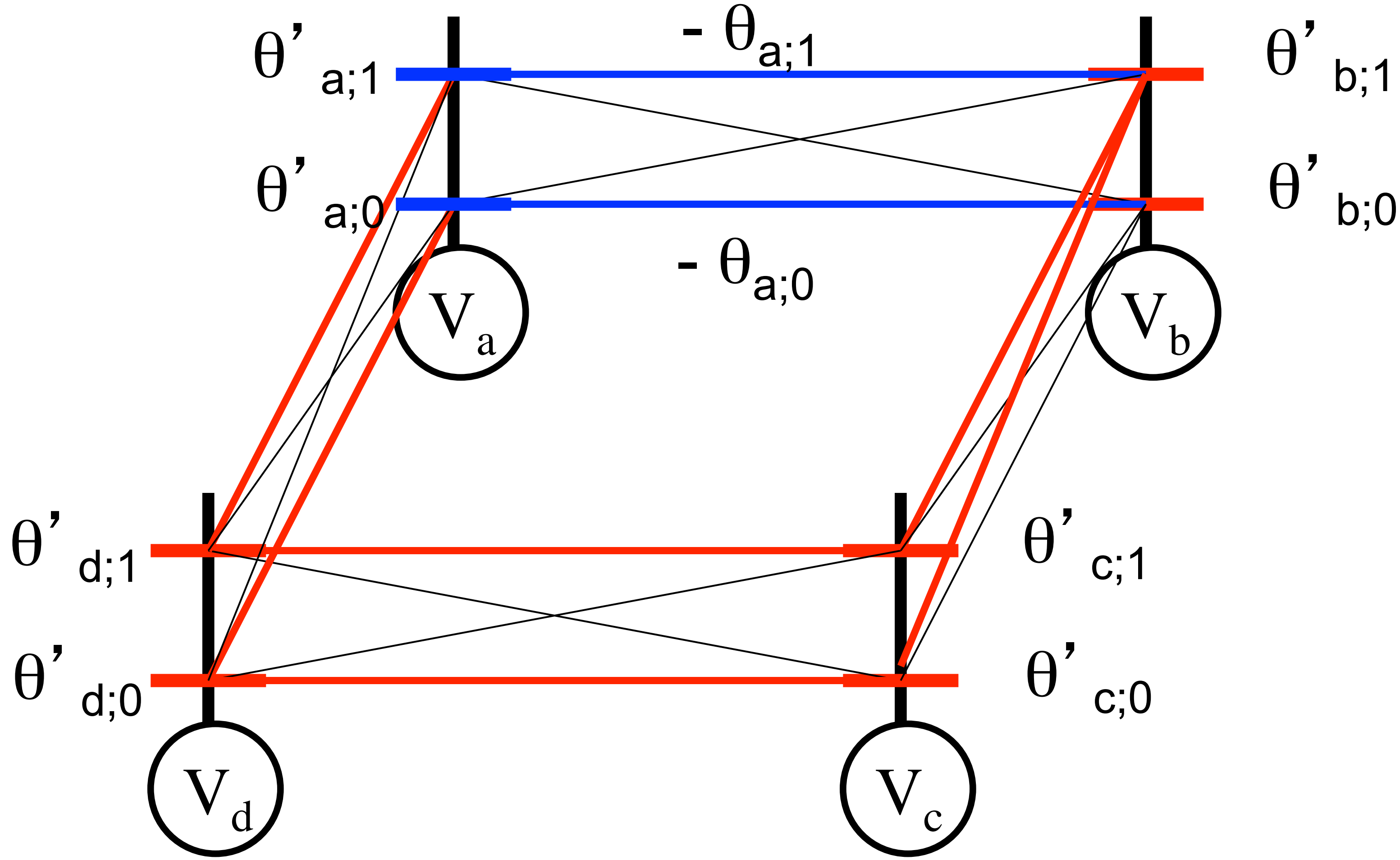
Belief Propagation on Cycles



$$\theta'_{a;1} - \theta_{a;1} = q_{a;1} \leq \theta'_{a;0} - \theta_{a;0} = q_{a;0}$$

Problem Solved

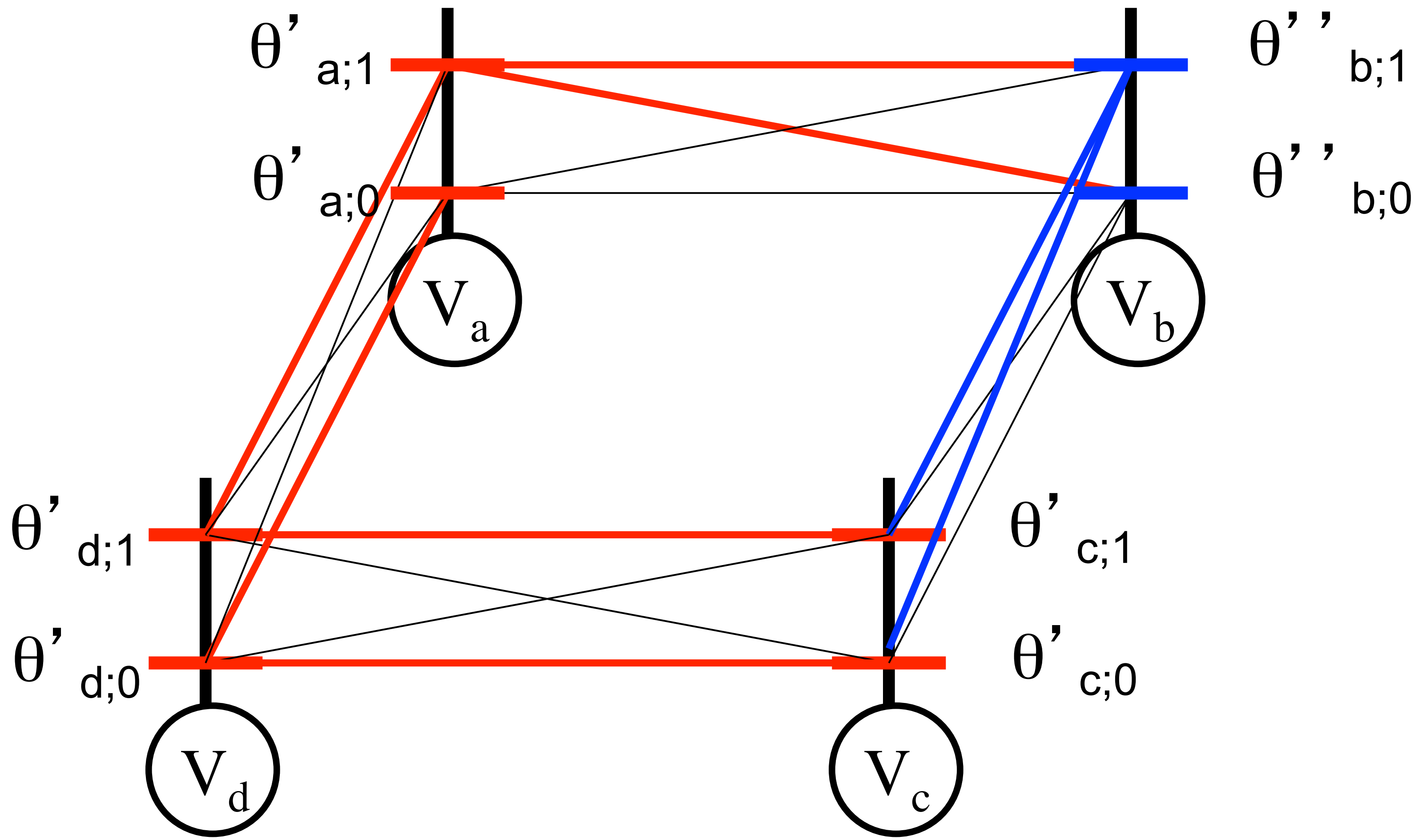
Belief Propagation on Cycles



$\theta'_{a;1} - \theta_{a;1} = q_{a;1} \geq \theta'_{a;0} - \theta_{a;0} = q_{a;0}$

Problem Not Solved

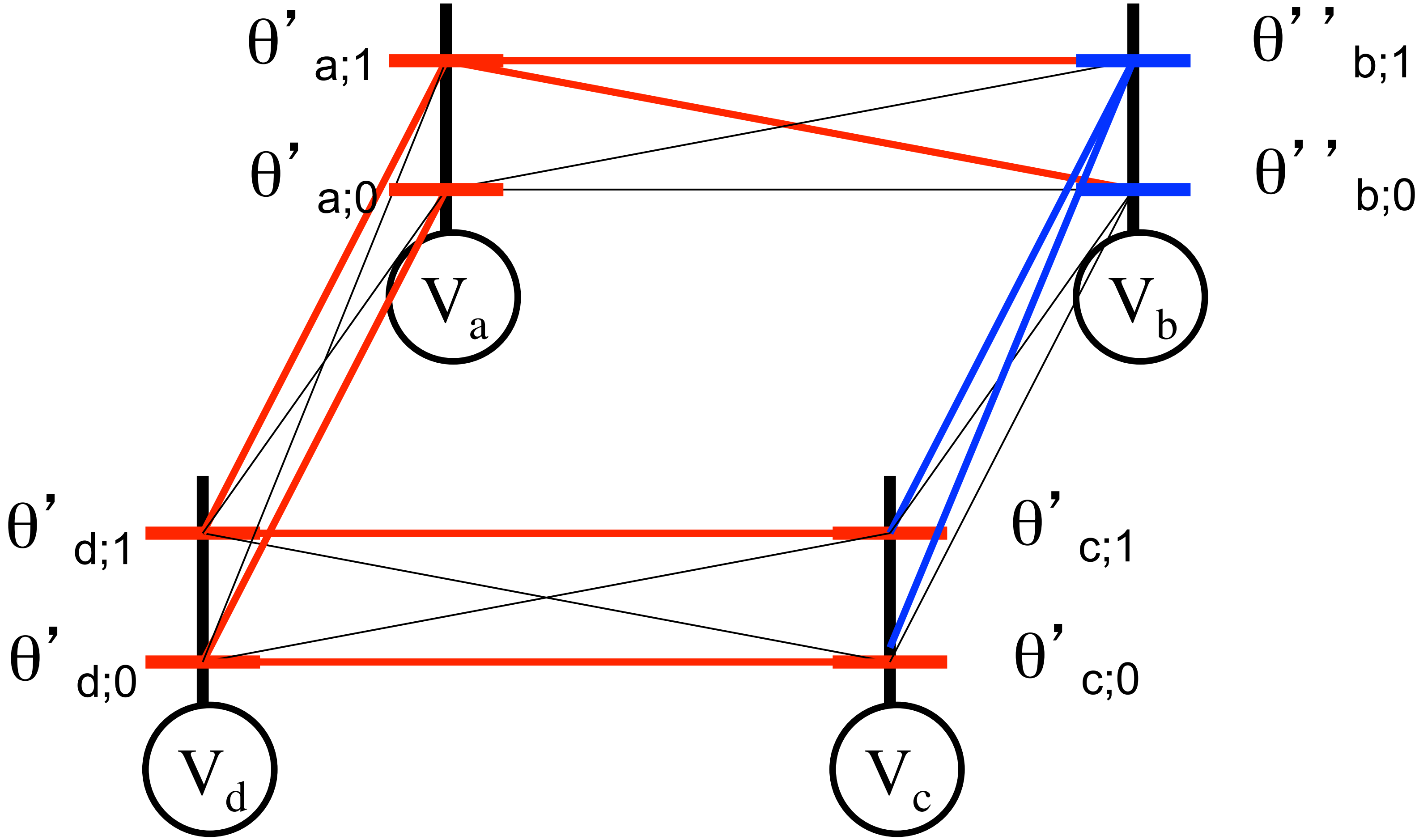
Belief Propagation on Cycles



Reparameterize (a,b) again

But doesn't this overcount some potentials?

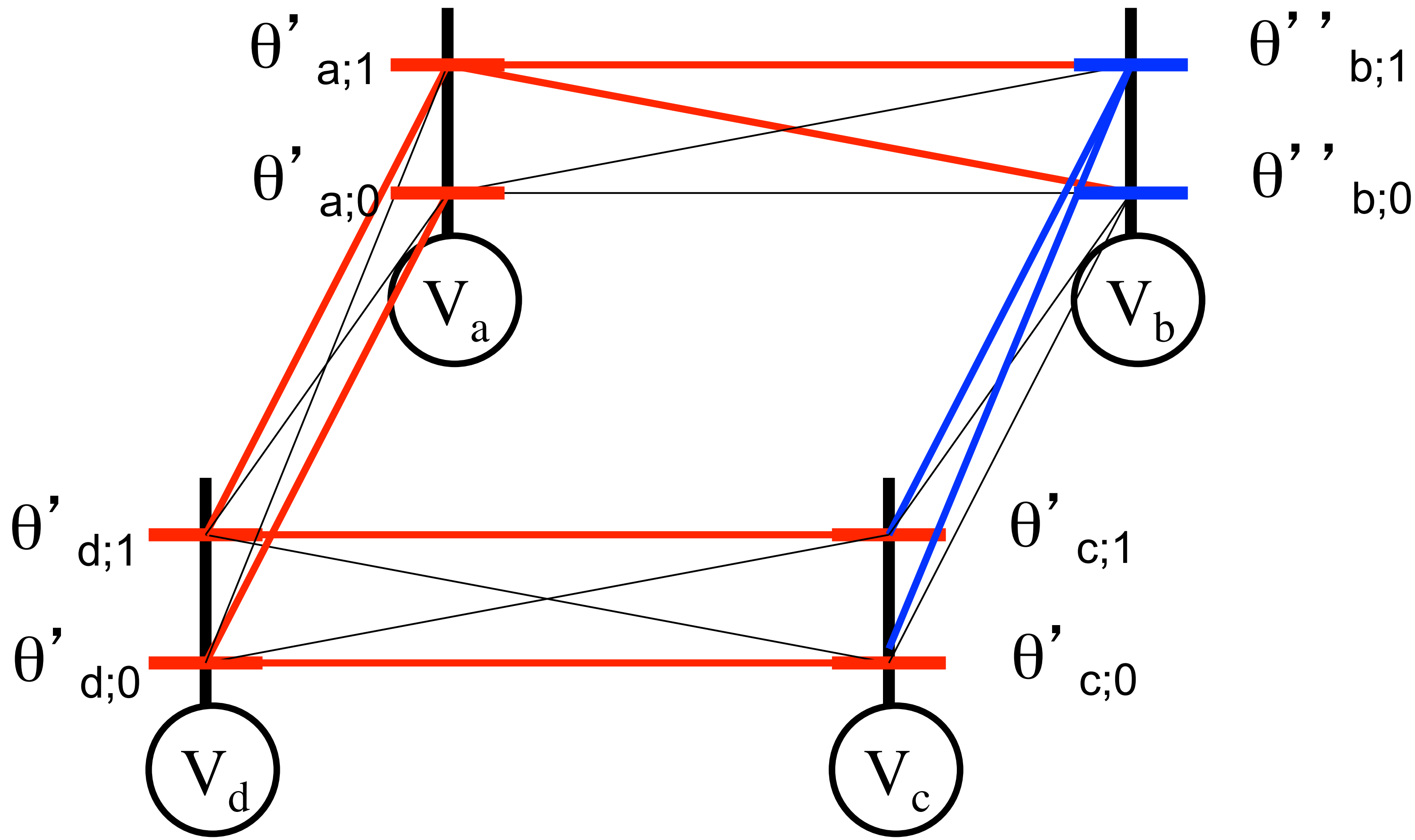
Belief Propagation on Cycles



Reparameterize (a,b) again

Yes. But we will do it anyway

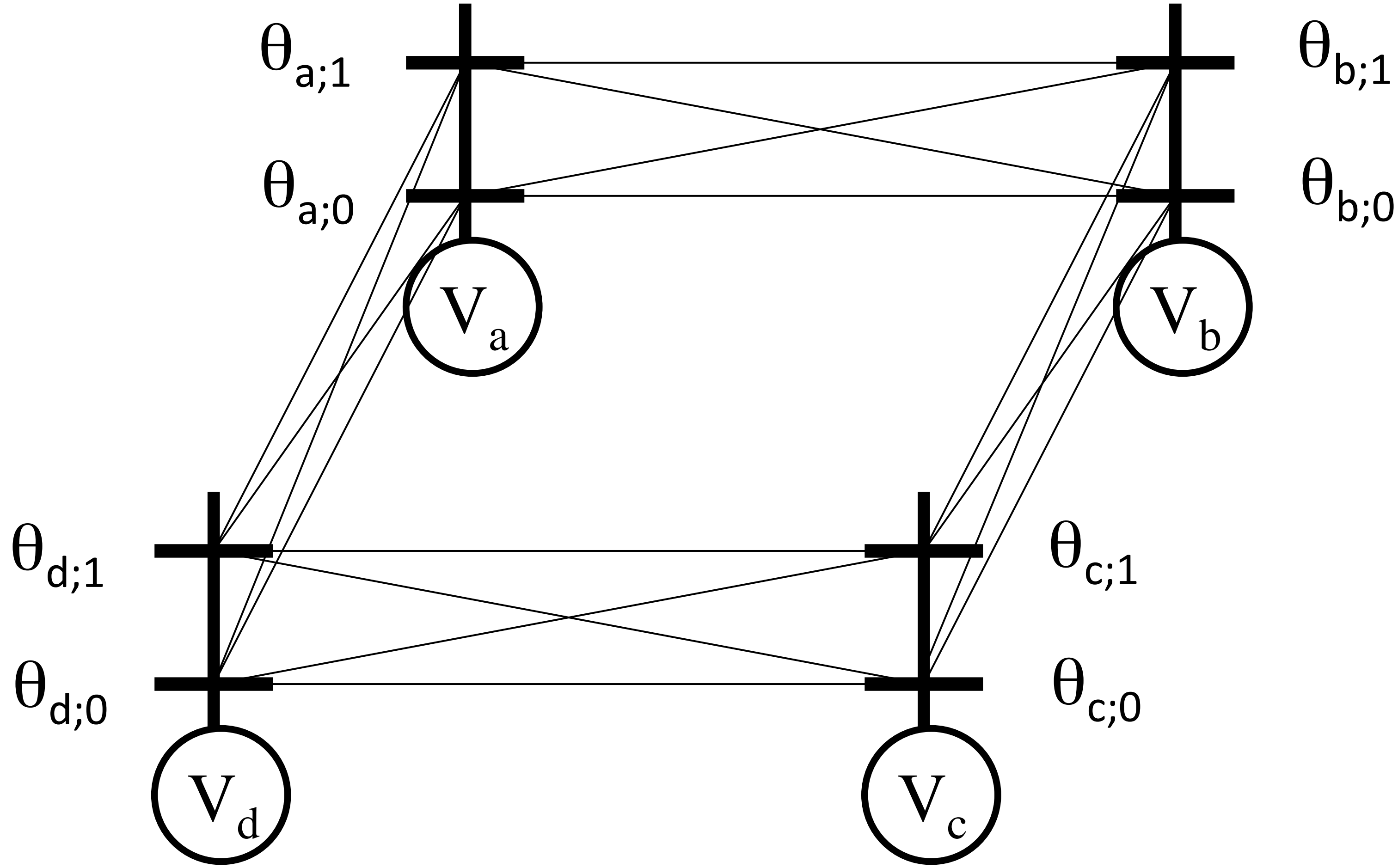
Belief Propagation on Cycles



Keep reparameterizing edges in some order

Hope for convergence and a good solution

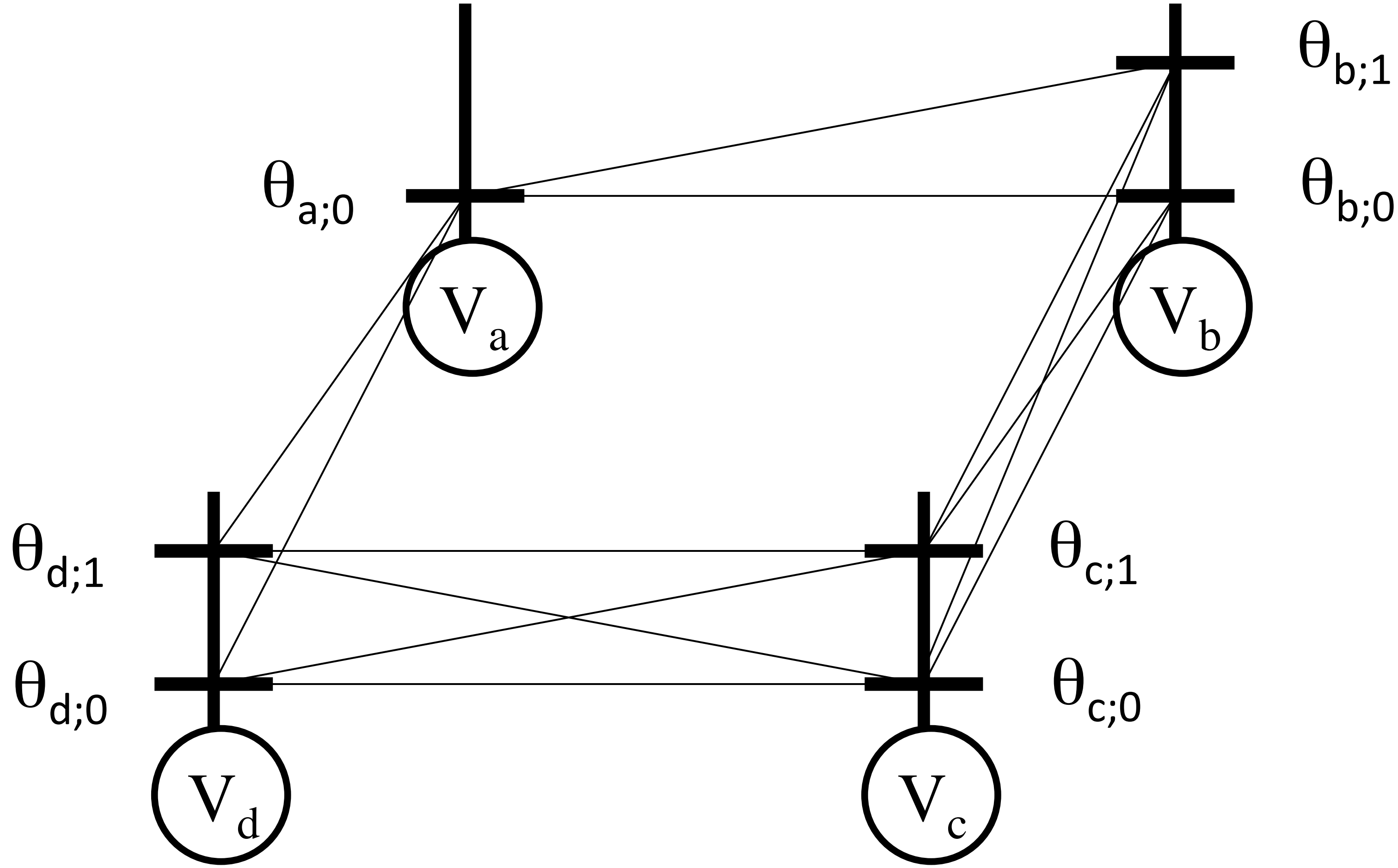
Belief Propagation on Cycles



Any suggestions?

Fix V_a to label l_0

Belief Propagation on Cycles

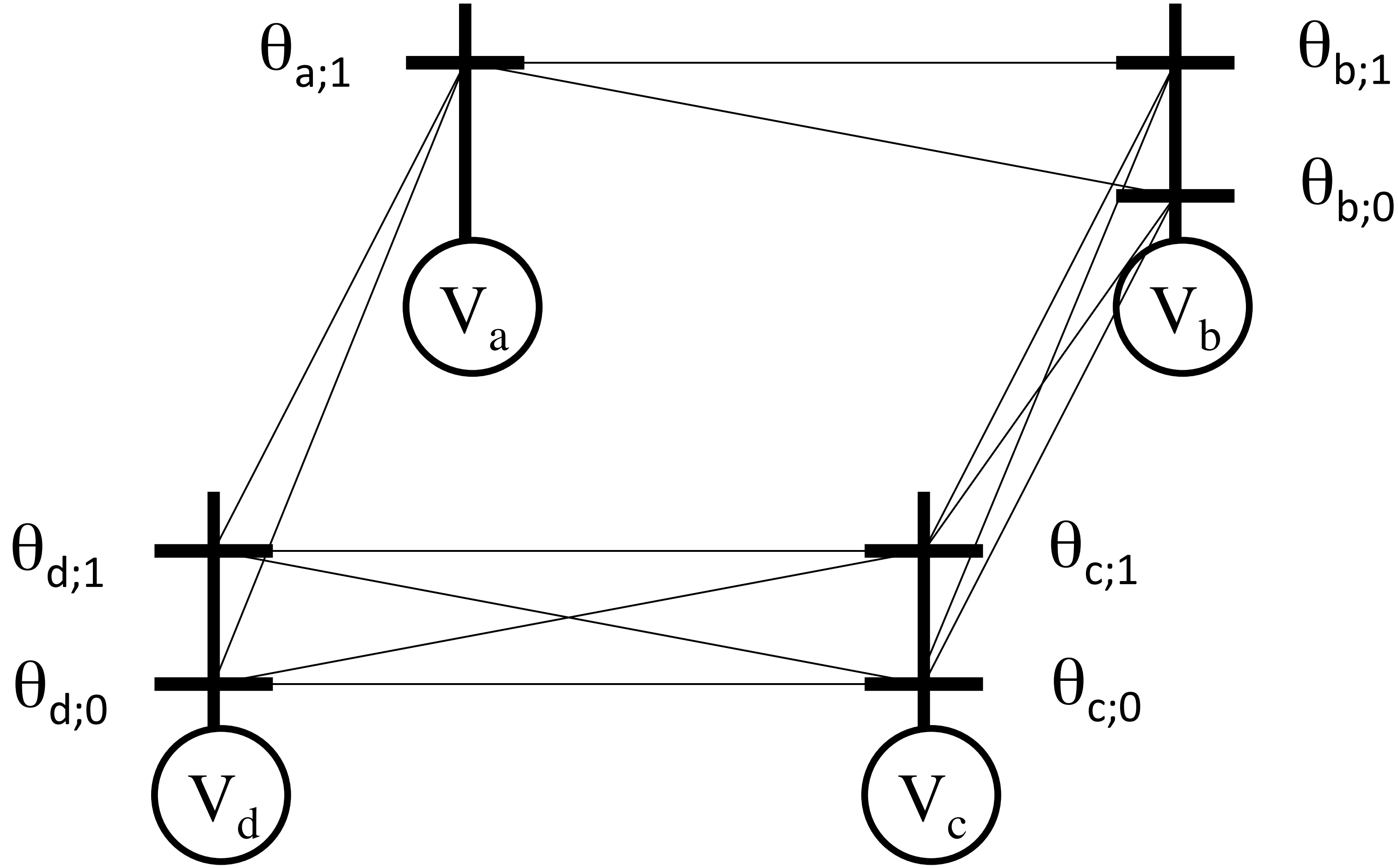


Any suggestions?

Fix V_a to label l_0

Equivalent to a tree-structured problem

Belief Propagation on Cycles

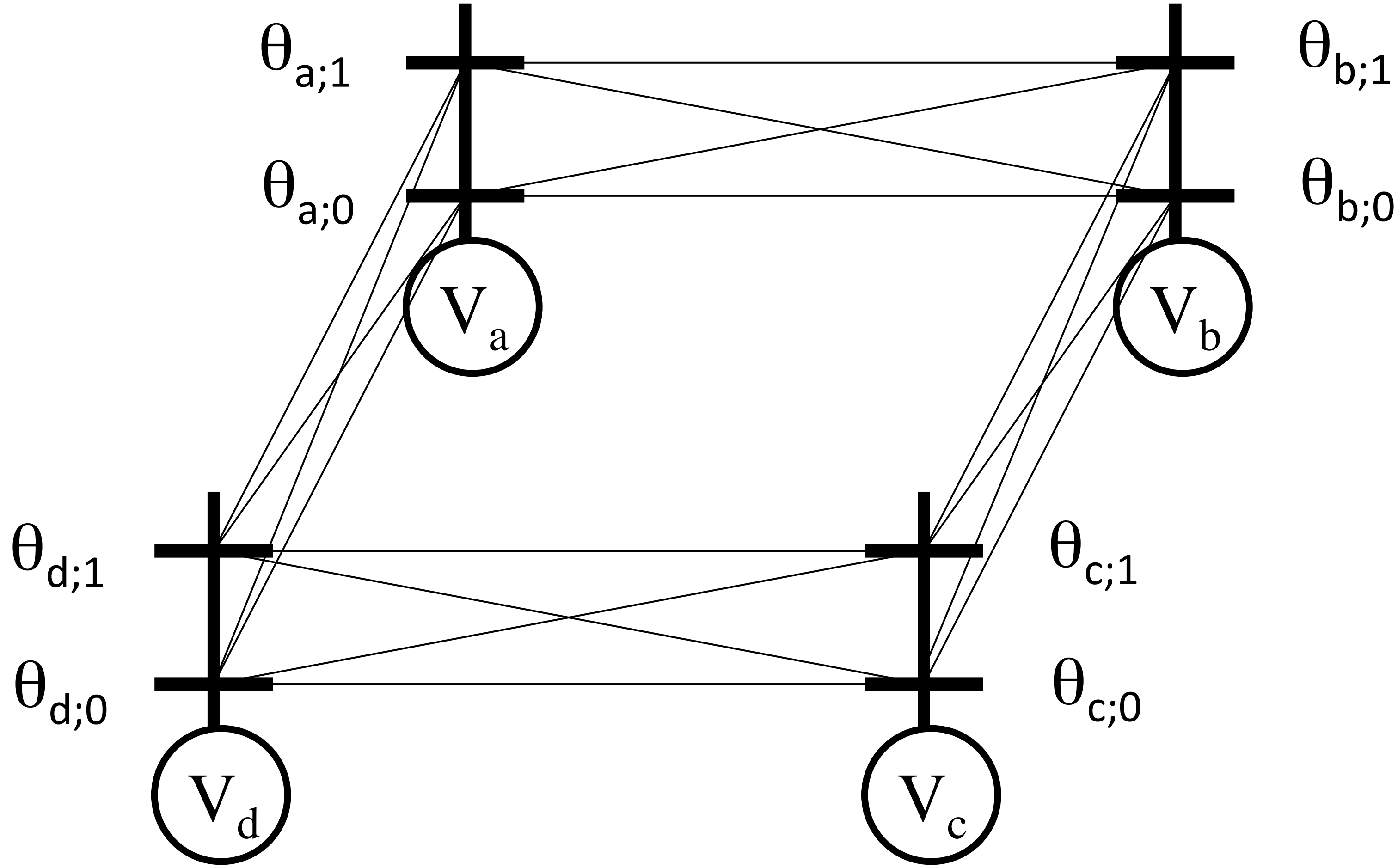


Any suggestions?

Fix V_a to label l_1

Equivalent to a tree-structured problem

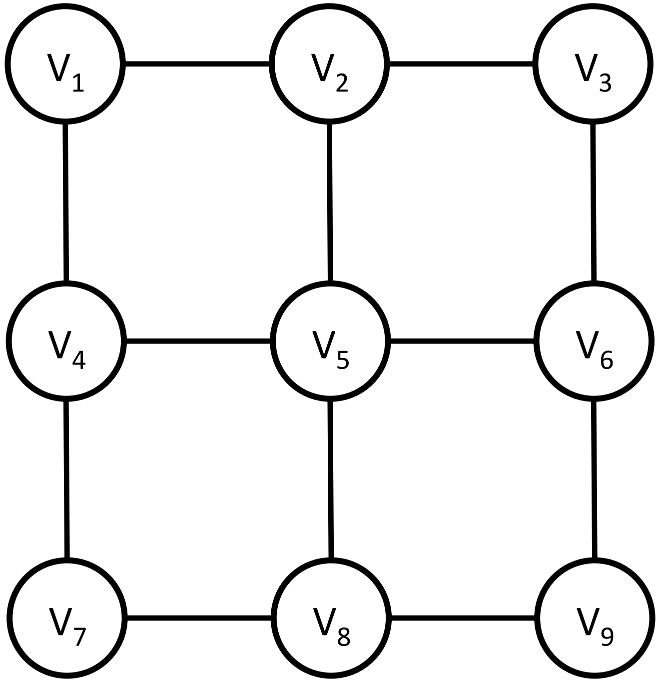
Belief Propagation on Cycles



This approach quickly becomes infeasible

Choose the minimum energy solution

Loopy Belief Propagation



Keep reparameterizing edges in some order

Hope for convergence and a good solution

Belief Propagation

- Generalizes to any arbitrary random field

- Complexity per iteration ?

$$\mathbf{O(nh^2)}$$

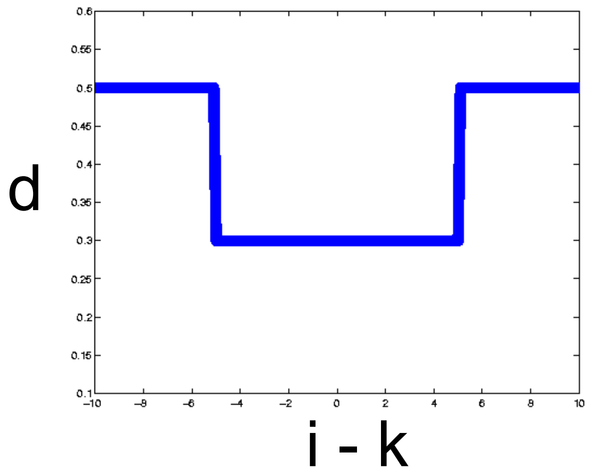
- Memory required ?

$$\mathbf{O(nh)}$$

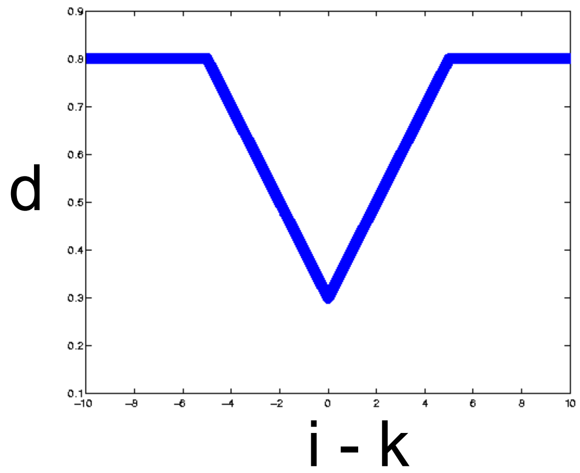
Computational Issues of BP

Complexity per iteration $O(nh^2)$

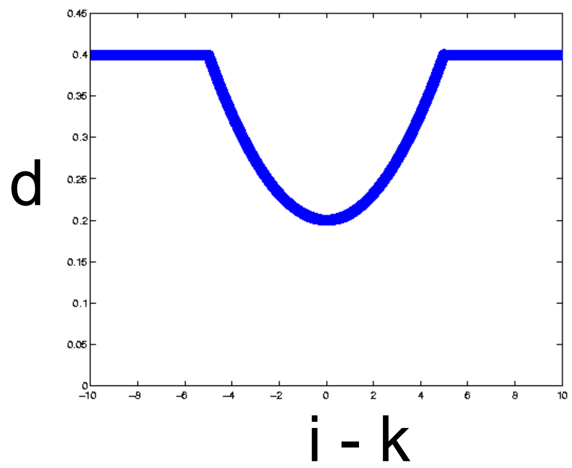
Special Pairwise Potentials $\theta_{ab;ik} = w_{ab}d(|i\text{-}k|)$



Potts



Truncated Linear



Truncated Quadratic

$O(nh)$

Felzenszwalb & Huttenlocher, 2004

Summary of BP

Exact for chains

Exact for trees

Approximate MAP for general cases

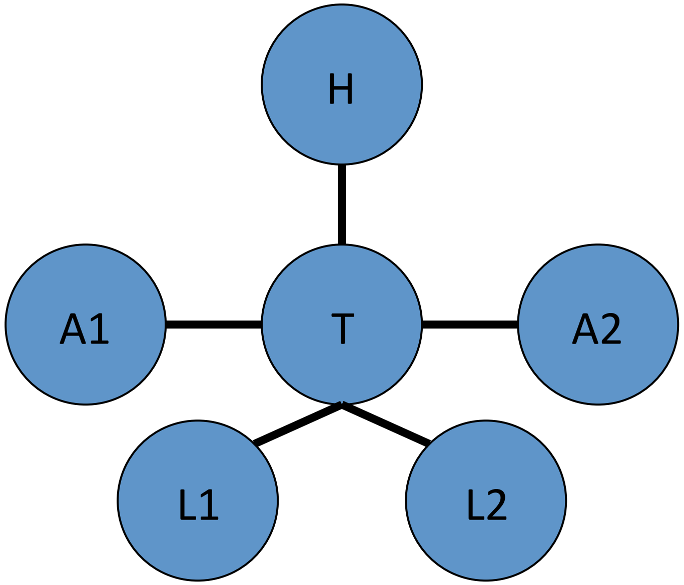
Not even convergence guaranteed

So can we do something better?

Results

Object Detection

Felzenszwalb and Huttenlocher, 2004



Labels - Poses of parts

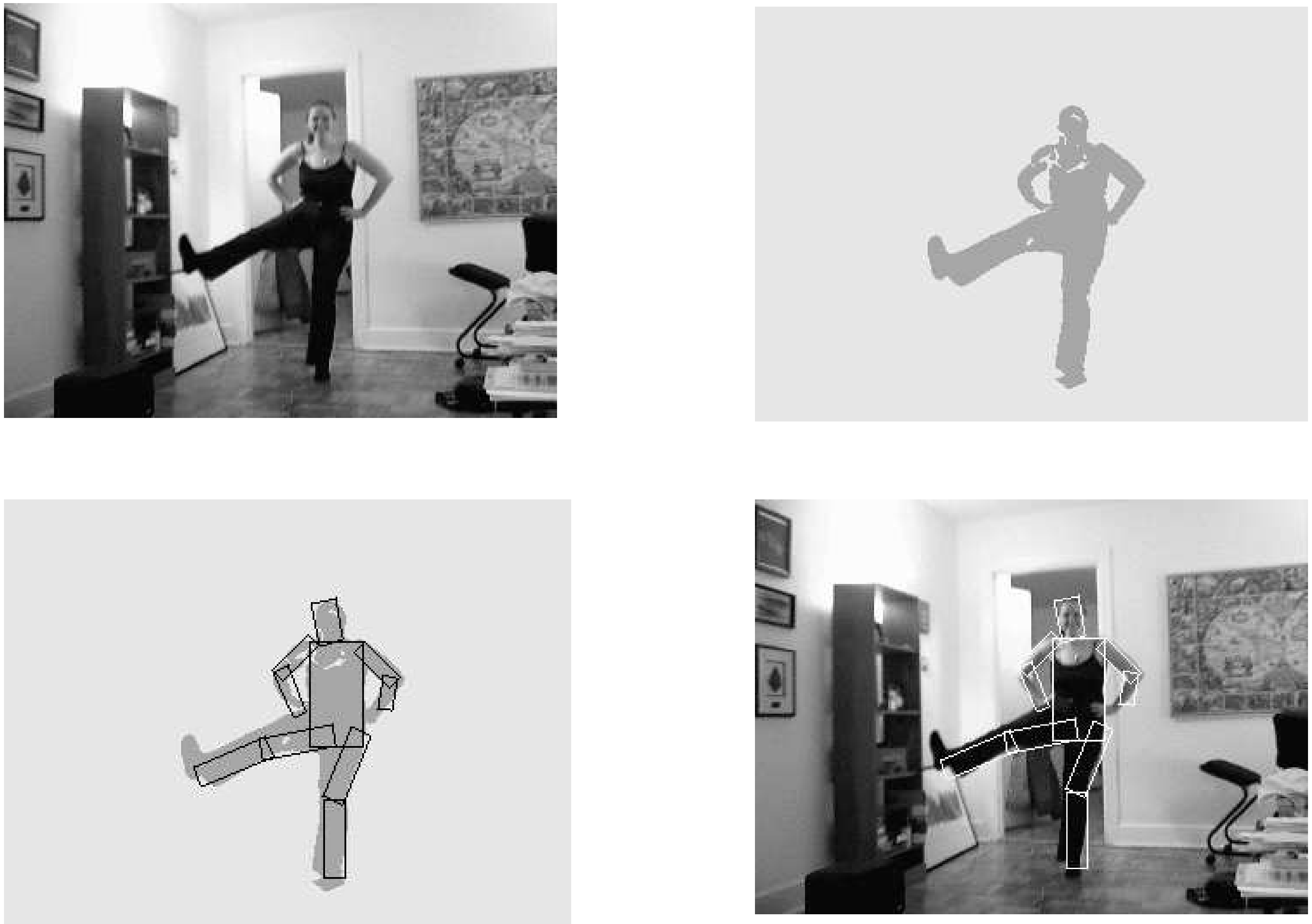
Unary Potentials:
Fraction of foreground pixels

Pairwise Potentials:
Favour Valid Configurations

Results

Object Detection

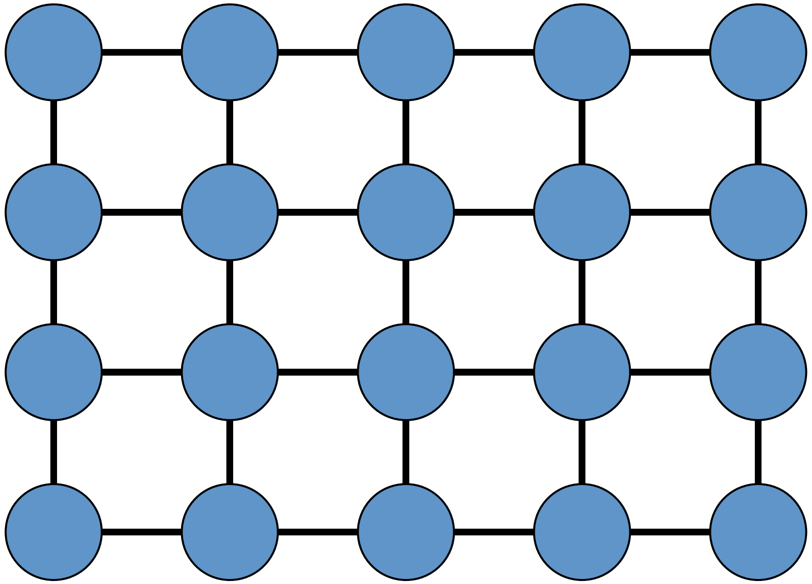
Felzenszwalb and Huttenlocher, 2004



Results

Binary Segmentation

Szeliski et al. , 2008



Labels - {foreground, background}

Unary Potentials: -log(likelihood) using learnt fg/bg models

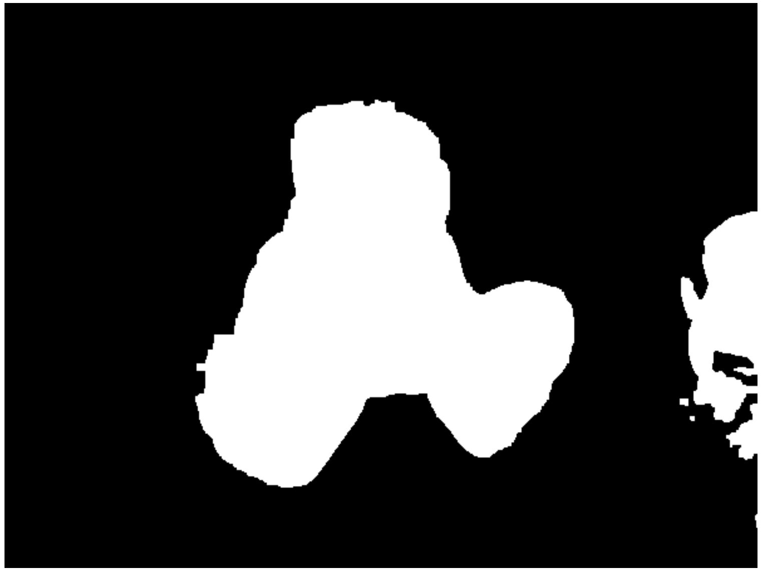
Pairwise Potentials: 0, if same labels

$1 - \lambda \exp(|D_a - D_b|)$, if different labels

Results

Binary Segmentation

Szeliski et al. , 2008



Belief Propagation

Labels - {foreground, background}

Unary Potentials: -log(likelihood) using learnt fg/bg models

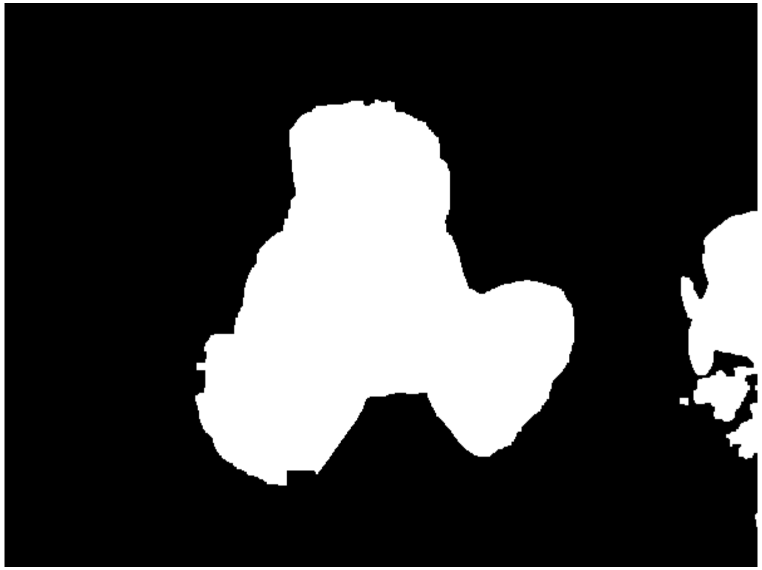
Pairwise Potentials: 0, if same labels

$1 - \lambda \exp(|D_a - D_b|)$, if different labels

Results

Binary Segmentation

Szeliski et al. , 2008



Global optimum

Labels - {foreground, background}

Unary Potentials: -log(likelihood) using learnt fg/bg models

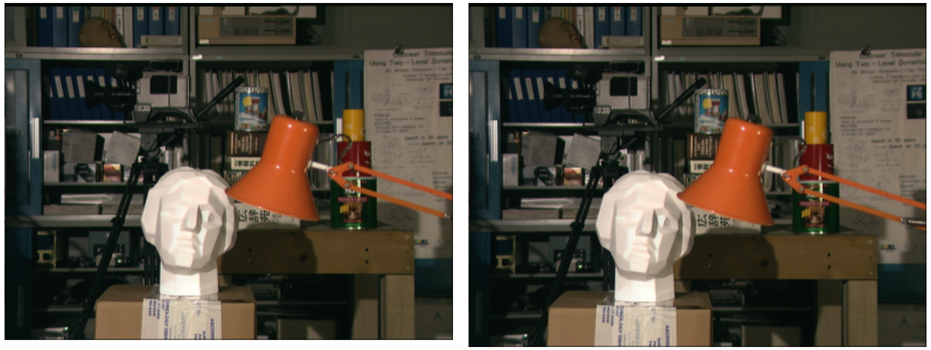
Pairwise Potentials: 0, if same labels

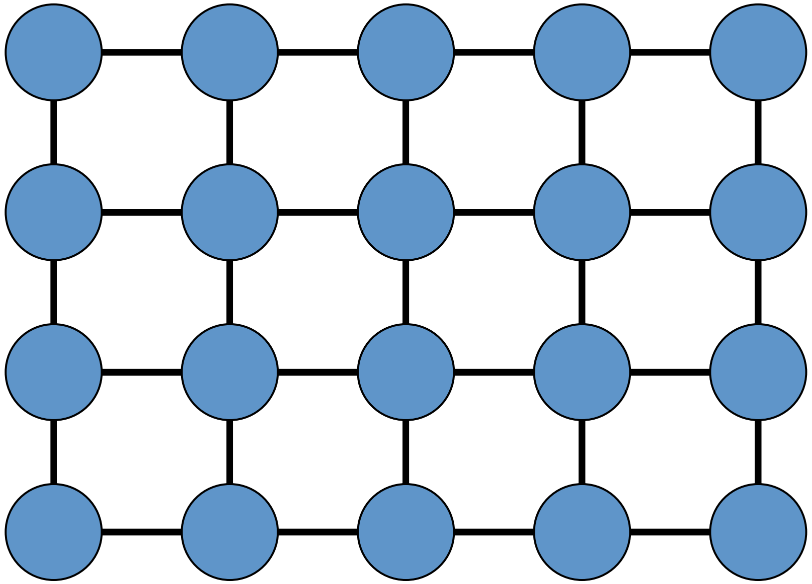
$1 - \lambda \exp(|D_a - D_b|)$, if different labels

Results

Stereo Correspondence

Szeliski et al. , 2008





Labels - {disparities}

Unary Potentials: Similarity of pixel colours

Pairwise Potentials: 0, if same labels

$1 - \lambda \exp(|D_a - D_b|)$, if different labels

Results

Stereo Correspondence


Szeliski et al. , 2008

Belief Propagation

Labels - {disparities}

Unary Potentials: Similarity of pixel colours

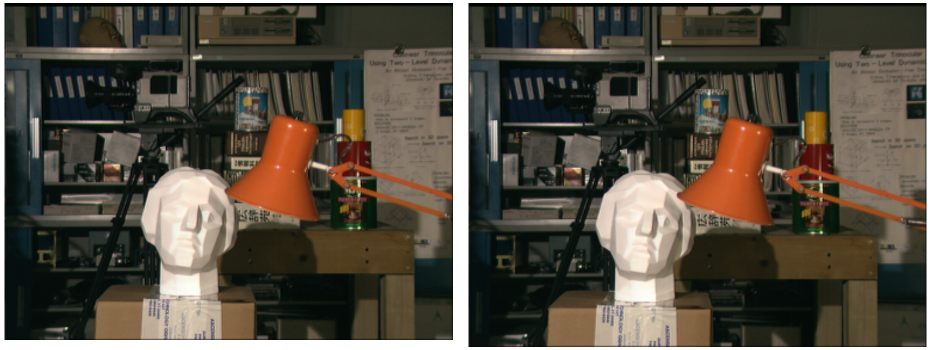
Pairwise Potentials: 0, if same labels

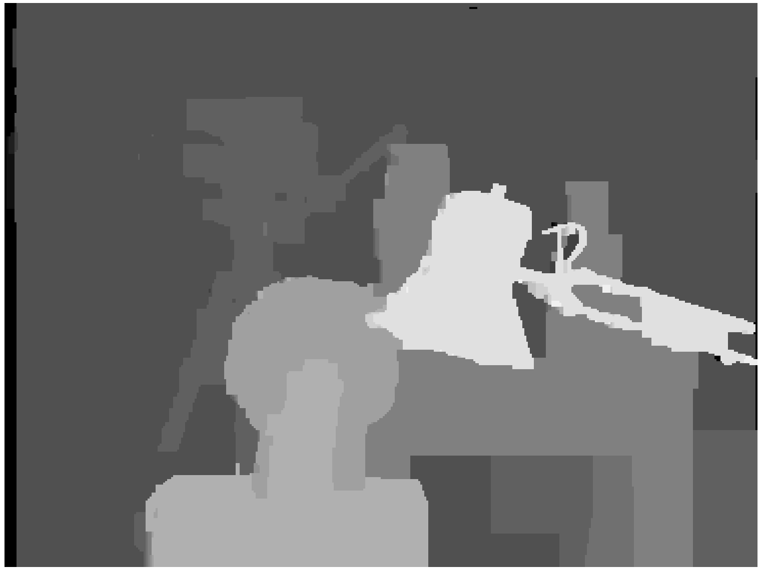
$1 - \lambda \exp(|D_a - D_b|)$, if different labels

Results

Stereo Correspondence

Szeliski et al. , 2008





Global optimum

Labels - {disparities}

Unary Potentials: Similarity of pixel colours

Pairwise Potentials: 0, if same labels

$1 - \lambda \exp(|D_a - D_b|)$, if different labels

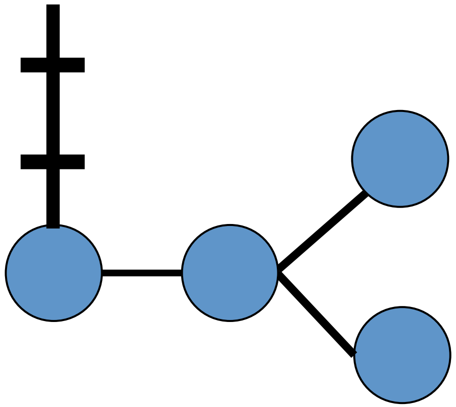
Other alternatives

- Integer linear programming and relaxation

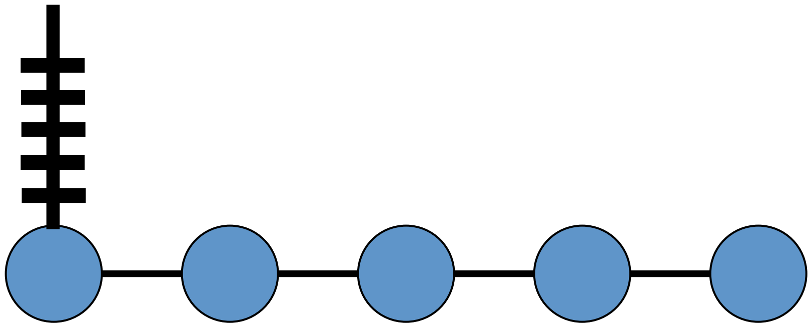
- TRW, Dual decomposition methods

- Extensively studied
  - Schlesinger, 1976
  - Koster et al., 1998, Chekuri et al., ’01, Archer et al., ’04
  - Wainwright et al., 2001, Kolmogorov, 2006
  - Globerson and Jaakkola, 2007, Komodakis et al., 2007
  - Kumar et al., 2007, Sontag et al., 2008, Werner, 2008
  - Batra et al., 2011, Werner, 2011, Zivny et al., 2014

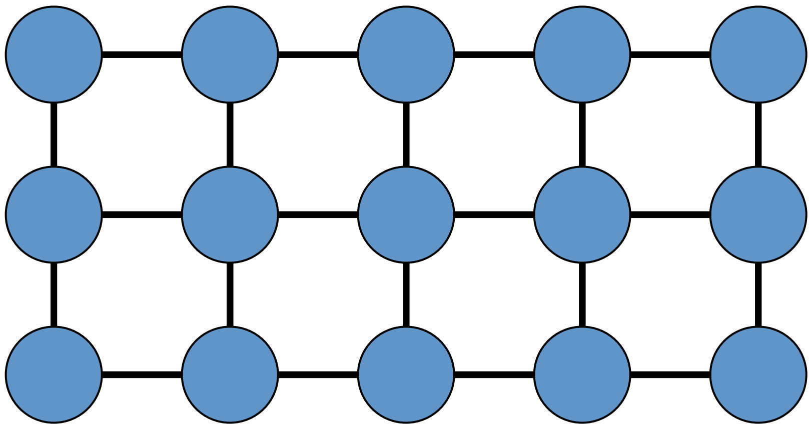
Where do we stand ?



Chain/Tree, 2-label: Use BP



Chain/Tree, multi-label: Use BP



Grid graph: Use TRW,
dual decomposition,
relaxation