Probabilistic Graphical Models

Lectures 16

Max-Sum Variable Elimination and Message Passing
Graph Cuts

Remember: Max Marginal

MAX Marginal

21 I

$$X^*, Y^*, Z^* = \underset{X^*, Y^*, Z^*}{\arg\max}\ P(X,Y,Z)$$

→ May or may not be a probability distribution.

$$g(X) = \max_{Y} \max_{Z} P(X,Y,Z)$$

Assume that we somehow manage to compute $g(X)$.

$$X^{**} = \underset{X}{\arg\max}\ g(X)$$

Assumj that argmax unique solution.

$$X^{**} = X^*$$

Remember: Max Marginal

$$P(X^*, Y^*, Z^*) = \max_X \max_Y \max_Z P(X, Y, Z) = P^*$$

$$g(X^{**}) = \max_X g(X) = \max_X \max_Y \max_Z P(X, Y, Z) = P^*$$

$$P^* = g(X^{**}) = \cancel{\max} \max_Y \max_Z P(X^{**}, Y, Z) = P(X^{**}, Y^{**}, Z^{**}) = P^*$$

$$P(X^{**}, Y^{**}, Z^{**}) = P(X^*, Y^*, Z^*)$$

X^*, Y^*, Z^* unique $\Rightarrow$ $X^{**} = X^*$

$Y^{**} = Y^*$

$Z^{**} = Z^*$

Variable Elimination: Max-Product

$A - B - C - D$ Variable elimination (Max-product)

$$P(A,B,C,D) = \frac{1}{Z}\,\phi_1(A,B)\,\phi_2(B,C)\,\phi_3(C,D)$$

$$\tau(D) = \max_C \max_B \max_A \phi_1(A,B)\,\phi_2(B,C)\,\phi_3(C,D)$$

(Assuming factors are positive $\Rightarrow \max_C \max_B \phi_2(B,C)\,\phi_3(C,D)\,\underbrace{\boxed{\max_A \phi_1(A,B)}}_{\tau_1(B)}$

Product -> Sum

K. N. Toosi
University of Technology

$A - B - C - D$

21 II

$$P(A,B,C,D) = \frac{1}{Z}\,\phi_1(A,B)\,\phi_2(B,C)\,\phi_3(C,D)$$

$$A^*,B^*,C^*,D^* = \underset{A,B,C,D}{\arg\max}\; P(A,B,C,D) = \underset{A,B,C,D}{\arg\max}\; h\big(P(A,B,C,D)\big)$$

any strictly increasing function $\mathbb{R} \to \mathbb{R}$

$$= \underset{A,B,C,D}{\arg\max}\; \log P(A,B,C,D)$$

$$= \underset{A,B,C,D}{\arg\max}\; \log \frac{1}{Z}\,\phi_1\,\phi_2\,\phi_3$$

$$= \underset{A - D}{\arg\max}\; -\log(Z) + \underbrace{\log \phi_1(A,B)}_{\theta_1(A,B)} + \underbrace{\log \phi_2(B,C)}_{\theta_2(B,C)} + \underbrace{\log \phi_3(C,D)}_{\theta_3(C,D)}$$

cost

Product -> Sum

any

function $\mathbb{R} \to \mathbb{R}$

$$= \underset{A,B,C,D}{\operatorname{argmax}} \log P(A,B,C,D)$$

$$= \underset{A,B,C,D}{\operatorname{argmax}} \log \frac{1}{Z} \Phi_1 \Phi_2 \Phi_3$$

$$= \underset{A-D}{\operatorname{argmax}} -\log(Z) + \underbrace{\log \Phi_1(A,B)}_{\theta_1(A,B)} + \underbrace{\log \Phi_2(B,C)}_{\theta_2(B,C)} + \underbrace{\log \Phi_3(C,D)}_{\theta_3(C,D)}$$

cost

$$= \underset{A-D}{\operatorname{argmax}} \theta_1(A,B) + \theta_2(B,C) + \theta_3(C,D)$$

$\rightarrow$ max-sum

MAX-SUM Variable Elimination

Max-Sum

$$= \arg\max_{A,B,C,D} \log P(A,B,C,D)$$

any ... function $\mathbb{R} \to \mathbb{R}$

$$= \arg\max_{A,B,C,D} \log \frac{1}{Z} \phi_1 \phi_2 \phi_3$$

$$= \arg\max_{A - D} -\log(Z) + \underbrace{\log \phi_1(A,B)}_{\theta_1(A,B)} + \underbrace{\log \phi_2(B,C)}_{\theta_2(B,C)} + \underbrace{\log \phi_3(C,D)}_{\theta_3(C,D)}$$

cost

$$= \arg\max_{A - D} \theta_1(A,B) + \theta_2(B,C) + \theta_3(C,D)$$

$\rightarrow$ max-sum

MAX-SUM Variable Elimination

Max-Sum variable Elimination

K. N. Toosi
University of Technology

$$\sigma_3(D) = \max_C \max_B \max_A \theta_1(A,B)+\theta_2(B,C)+\theta_3(C,D)$$

$$= \max_C \max_B \theta_2(B,C)+\theta_3(C,D)+\boxed{\max_A \theta_1(A,B)}$$

$$= \max_C \max_B \theta_2(B,C)+\theta_3(C,D)+\sigma_1(B)$$

$$= \max_C \theta_3(C,D)+\boxed{\max_B \theta_2(B,C)+\sigma_1(B)}$$

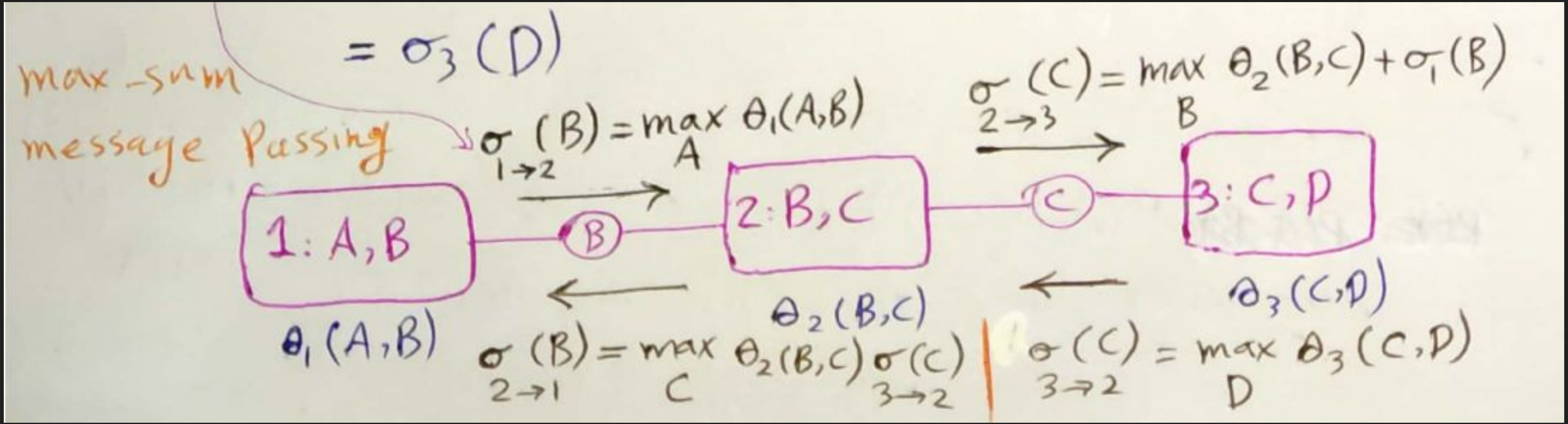
$$= \max_C \theta_3(C,D)+\sigma_2(C)$$

$$= \sigma_3(D)$$

max-sum

$\sigma(C) = \max \theta_2(B,C)+\sigma_1(B)$

Max-Sum message passing



Max-Sum Beliefs

$$\beta_1(A,B) = \theta_1(A,B) + \sigma_{2\to1}(B) = \max_{C,D} \theta(A,B,C,D)$$

$$\beta_2(B,C) = \theta_2(B,C) + \sigma_{1\to2}(B) + \sigma_{3\to2}(C) = \max_{A,D} \theta(A,B,C,D)$$

$$\beta_3(C,D) = \theta_3(C,D) + \sigma_{2\to3}(C) = \max_{A,B} \theta(A,B,C,D)$$

$$\theta(A,B,C,D) = \theta_1(A,B) + \theta_2(B,C) + \theta_3(C,D)$$

$$\boxed{\text{Beliefs} = \text{max-marginals}}$$

Max-Sum Sepset Beliefs

Sepset beliefs:

$$\beta_{1,2}(B) = \sigma_{1\to 2}(B) + \sigma_{2\to 1}(B) = \max_{A,C,D} \theta(A,B,C,D)$$

$$\beta_{2,3}(C) = \sigma_{2\to 3}(C) + \sigma_{3\to 2}(C) = \max_{A,B,D} \theta(A,B,C,D)$$

How to use max-marginals

K. N. Toosi
University of Technology

How to make use of max-marginals?

$$A^*, B^*, C^*, D^* = \underset{A,B,C,D}{\arg\max}\ \theta(A,B,C,D) \qquad A, \dots D \in \{0,1\}$$

but we do not want to try out all 2^4 combinations

How to use max-marginals

1- Assume that (A^*, B^*, C^*, D^*) is unique

As we saw before $\quad A^*, B^* = \underset{A,B}{\arg\max}\ \beta_1(A,B)$

$\quad \| $

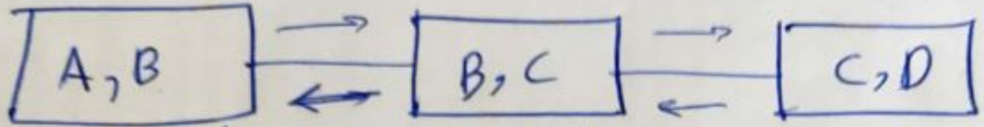
$\quad B^*, C^* = \underset{B,C}{\arg\max}\ \beta_2(B,C)$

Exact Solution

↳ Cluster beliefs will agree upon sepset beliefs.

$$\max_{A} \beta_1(A,B) = \max_{C} \beta_2(B,C) = \beta_{12}(B) = \sigma_{1\to 2}(B) + \sigma_{2\to 1}(B)$$

How to use max-marginals



$\theta_1(A,B)$ $\theta_2(B,C)$ $\theta_3(C,D)$

$\beta_1(A,B)$ $\beta_2(B,C)$ $\beta_3(C,D)$

$$\theta(A,B,C,D) = \theta_1(A,B)\,\theta_2(B,C)\,\theta_3(C,D)$$

$$\boxed{0 \quad 1 \quad 0 \quad 1}$$

$$A^*, B^*, C^*, D^* \in \operatorname{argmax}\ \theta(A,B,C,D)$$

$$A^{**}, B^{**}, C^{**}, D^{**} \in \operatorname{argmax}\ \theta(A,B,C,D)$$

$$\boxed{1 \quad 0 \quad 1 \quad 0}$$

$$B^* \neq B^{**}$$

How to use max-marginals

A

2- what if argmax is not unique?

We can choose any of the solutions as long as solutions on different clusters are consistent.

→ start from a clusters and choose consistent solutions

→ Add a very small random number to each function $\theta(X_i, X_j)$

MAP inference using junction tree

21 IV

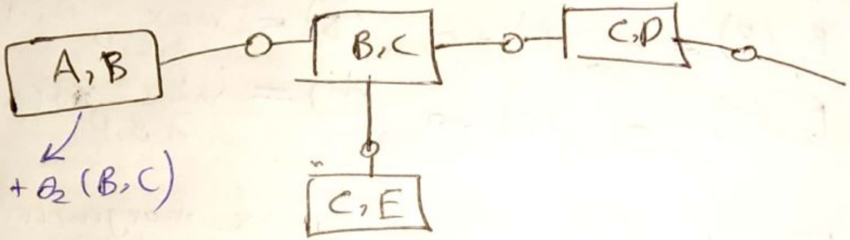
→ try to build a cluster tree from our PGM (Junction tree)

→ Compute messages (max-sum)

→ Compute Beliefs (max-sum)

→ find solution $X_1^*, X_2^*, \ldots, X_n^* = \operatorname{argmax} P(X_1 \ldots X_n)$ as argmax over beliefs

$$\theta(A, B, \ldots) = \boxed{\theta_1(A) + \theta_2(B, C)} + \theta_3(C, D) + \theta_4(C, E) + \ldots$$



$\theta_1(A) + \theta_2(B, C)$

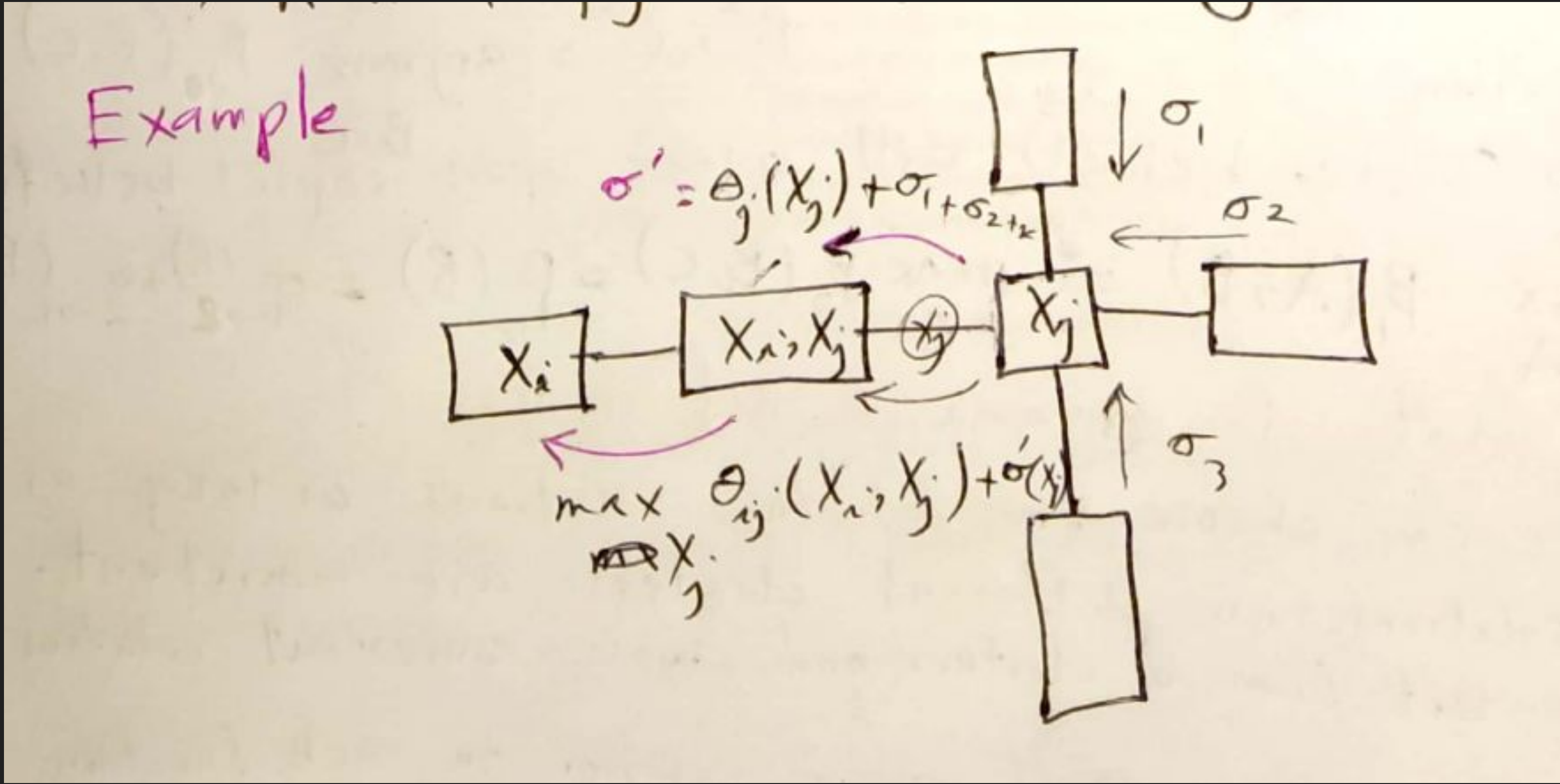
MAP inference using cluster graphs

What if we cannot build a cluster tree with few variables in each cluster?

$\Rightarrow$ Build a cluster graph (with loops) possessing FPP & RIP.

$\Rightarrow$ Run loopy message passing.

MAP inference using cluster graphs



Energy-based systems

$P(X) = 1/Z \exp(\theta(X))$ where $\theta(X) = \Sigma_c \ \theta_c(X_c)$

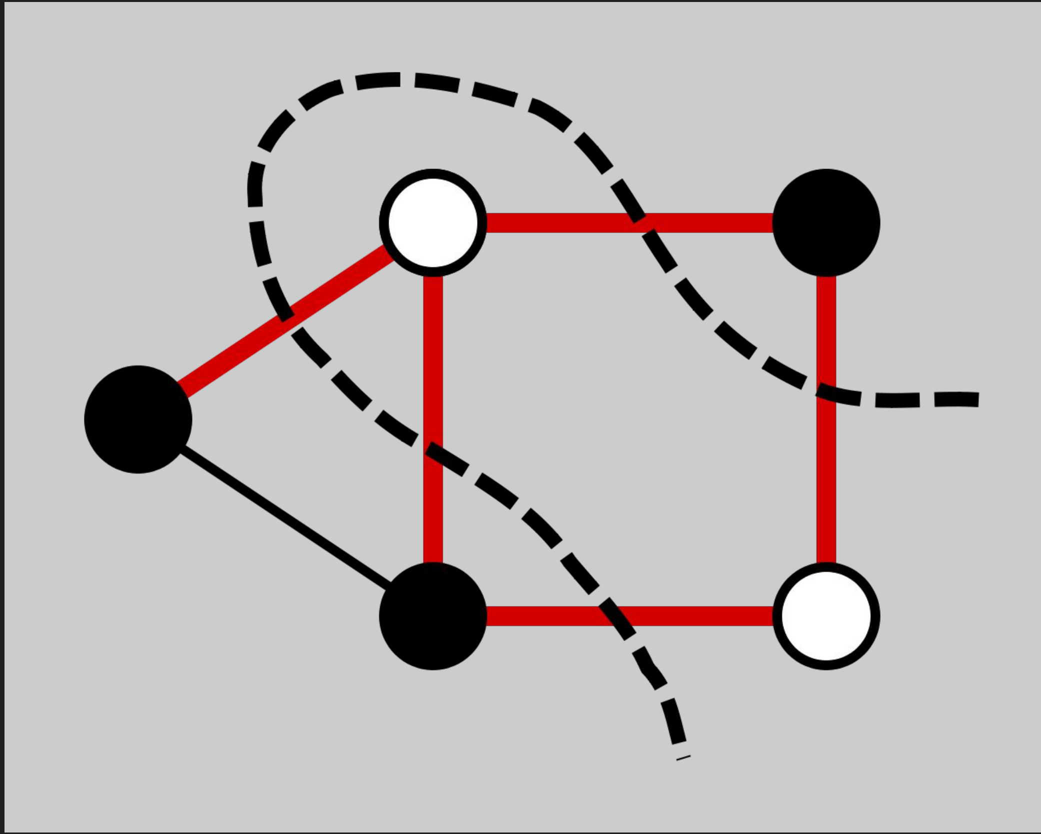
$P(X) = 1/Z \exp(-E(X))$ where $E(X) = \Sigma_c \ E_c(X_c)$

replace Max by Min

Other MAP inference methods

- Graph Cuts
  - Roof Duality/QPBO
  - Move-based algorithms (Alpha expansion, Alpha-beta swap, fusion moves, etc.)
- Dual Decomposition
- Integer programming

Graph Cuts



Graph Cuts - Simple Case

K. N. Toosi
University of Technology

$$P(X) = P(X_1, X_2, \dots, X_n) = \frac{1}{Z} e^{-E(X)}$$

$$E(X) = E(X_1, X_2, \dots, X_n) = \sum_{i=1}^{n} E_i(X_i) + \sum_{(i,j)\in E} E_{ij}(X_i, X_j)$$

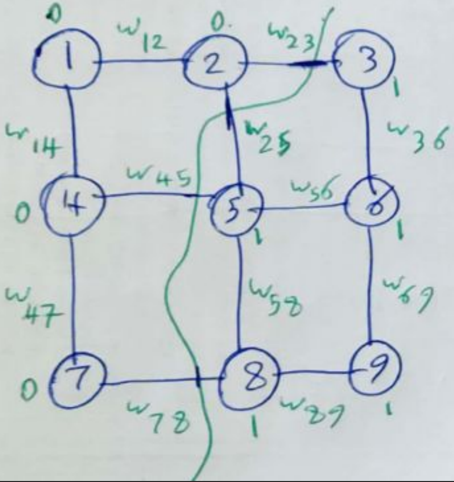
$$X_i \in \{0, 1\} \qquad E_{ij}(X_i, X_j) = w_{ij}\, 1(X_i \neq X_j)$$

What is a cut? (simple case)

$$E(X) = \sum_{i=1}^{n} E_i(X_i) + \sum_{(i,j)\in\mathcal{E}} E_{ij}(X_i, X_j)$$

$$= \sum_{i=1}^{n} E_i(X_i) + \sum_{(i,j)\in\mathcal{E}} w_{ij}\, 1(X_i \neq X_j)$$

$$X_i \in \{0,1\}$$



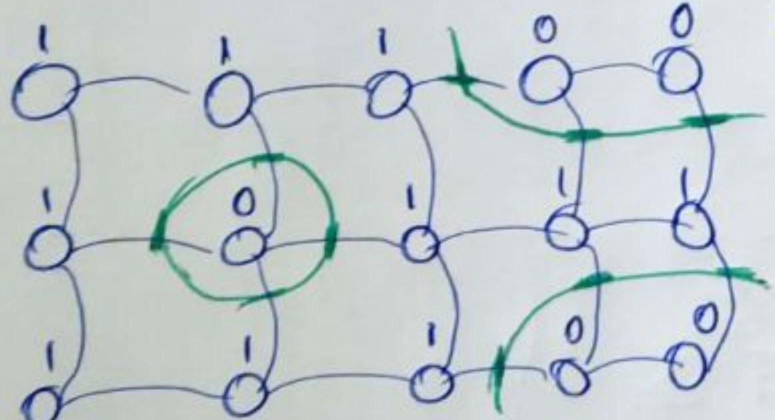
$cut = w_{23} + w_{25} +$

$\overset{X_0}{} \; \overset{X_1}{} \; \overset{X_2}{} \; \cdots \; \overset{X_9}{}$

$cut(0,0,1,0,1,1,0,1,1)$

$= w_{23} + w_{25} + w_{45} + w_{78}$

What is a cut? (simple case)

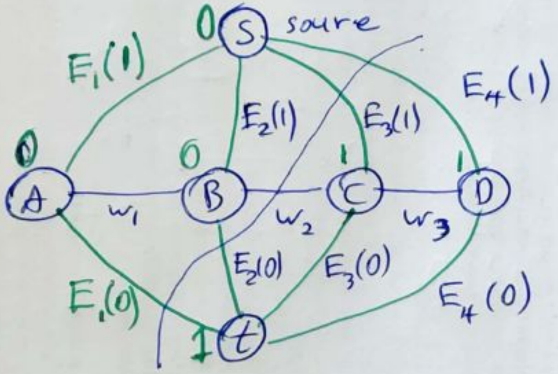


$$E(X) = \sum_{(i,j)\in E} w_{ij} \cdot \mathbb{1}(X_i \neq X_j)$$

Energy function as cut values

$E(X_1, X_2, X_3, X_4)$

$$E(A,B,C,D) = E_1(A) + E_2(B) + E_3(C) + E_4(D)$$
$$+ w_1\, 1(A \neq B) + w_2\, 1(B \neq C) + w_3\, 1(C \neq D)$$



$$\mathrm{Cut}(A,B,C,D) = E(A,B,C,D)$$

General Case and Submodularity

Special Case $E(X_i, X_j) = w_{ij} \mathbb{1}(X_i \neq X_j)$

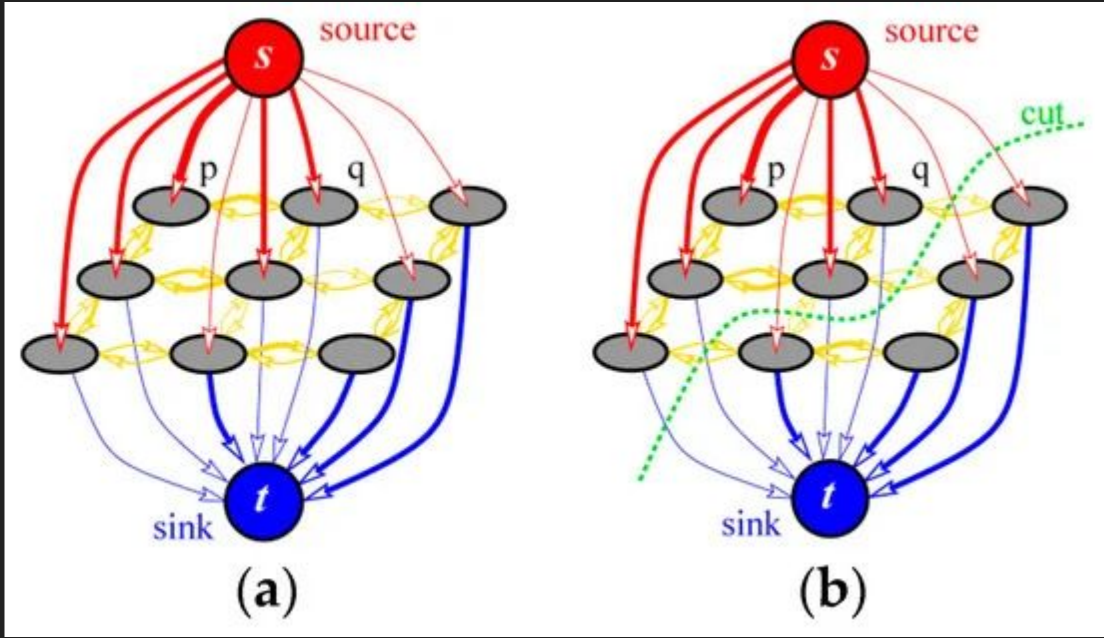
$\underline{E_{ij}(X_i, X_j)}$

$$E_{ij}(0,0) + E_{ij}(1,1) \leq E_{ij}(1,0) + E_{ij}(1,1)$$

for all $(i,j) \in \mathcal{E}$

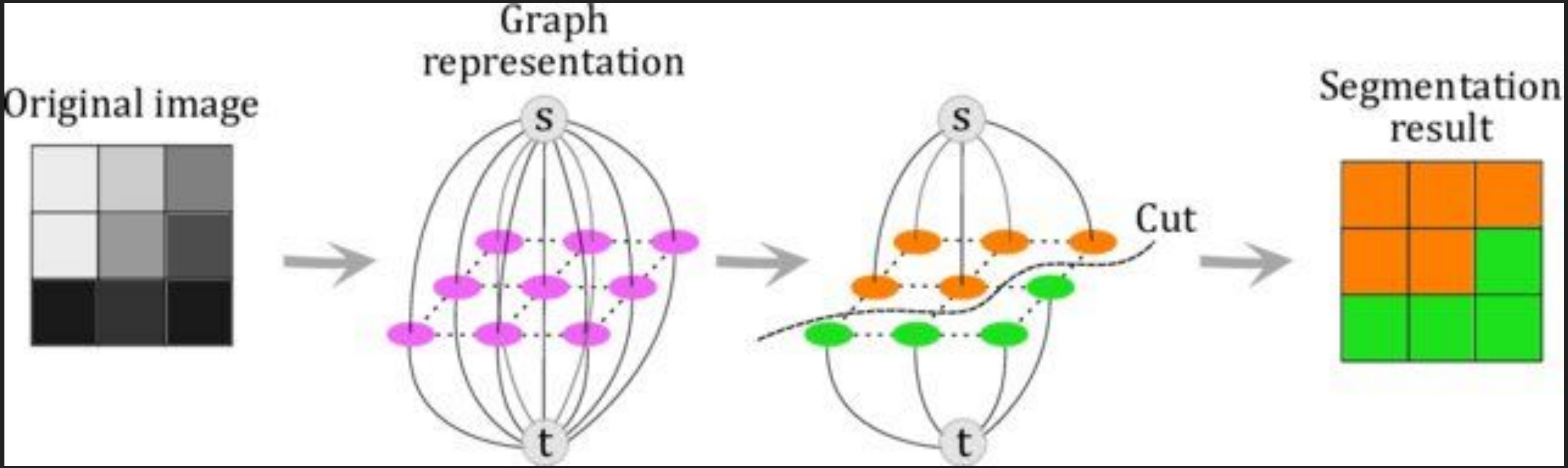
Submodular

Graph Cuts



https://www.mdpi.com/2073-8994/10/5/169

Graph Cuts for image segmentation



Gauriau, Romane. *Shape-based approaches for fast multi-organ localization and segmentation in 3D medical images*. Diss. Telecom ParisTech, 2015.

Non-submodular case

- Set negative weight to zero
- Roof Duality/QPBO

Multilabel case

- Multilabel graph cuts (only convex energy terms)
- Move-based algorithm
  - alpha-beta swap
  - alpha expansion
  - fusion moves