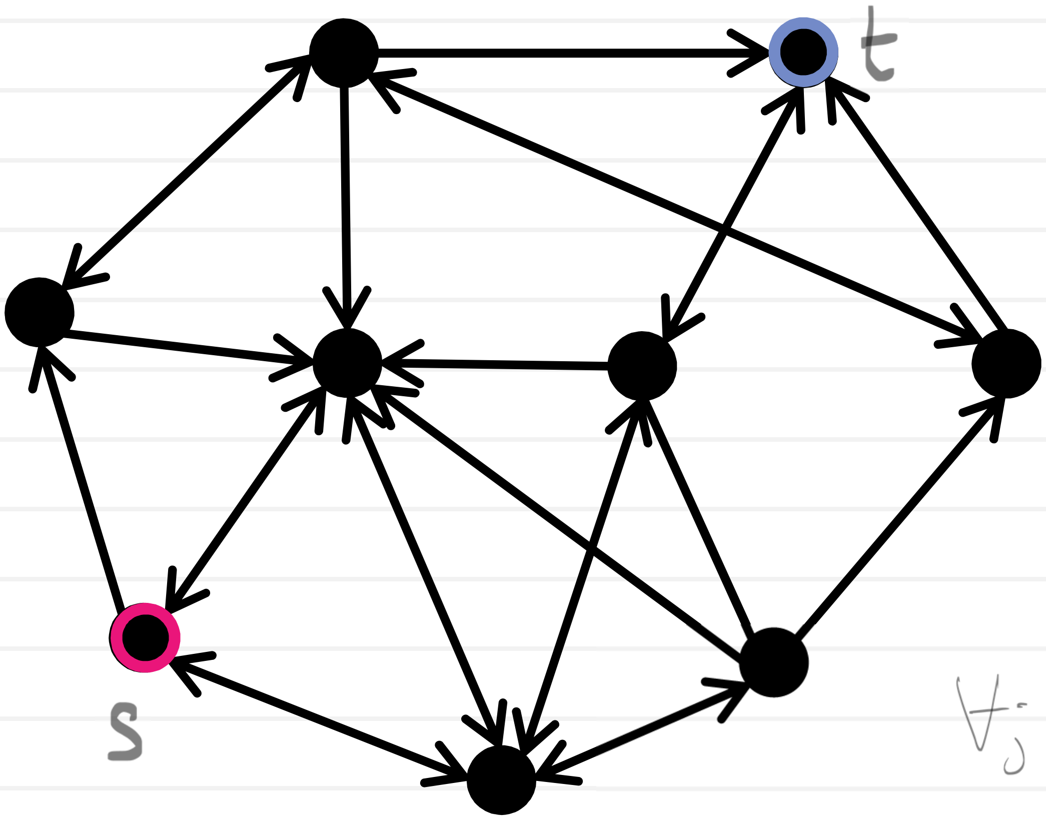
Lecture 18: Mincut/maxflow duality. Dual decomposition.

Maximum flow (maxflow)

Maximum flow problem: given arc capacities, a single source vertex s, and a single sink vertex t, find a maximal flow (that maximizes the inflow at s and outflow at t).



$$\max_{b,f} \; b$$

$$\text{s.t.:} \quad 0 \le f_{ij} \le u_{ij}$$

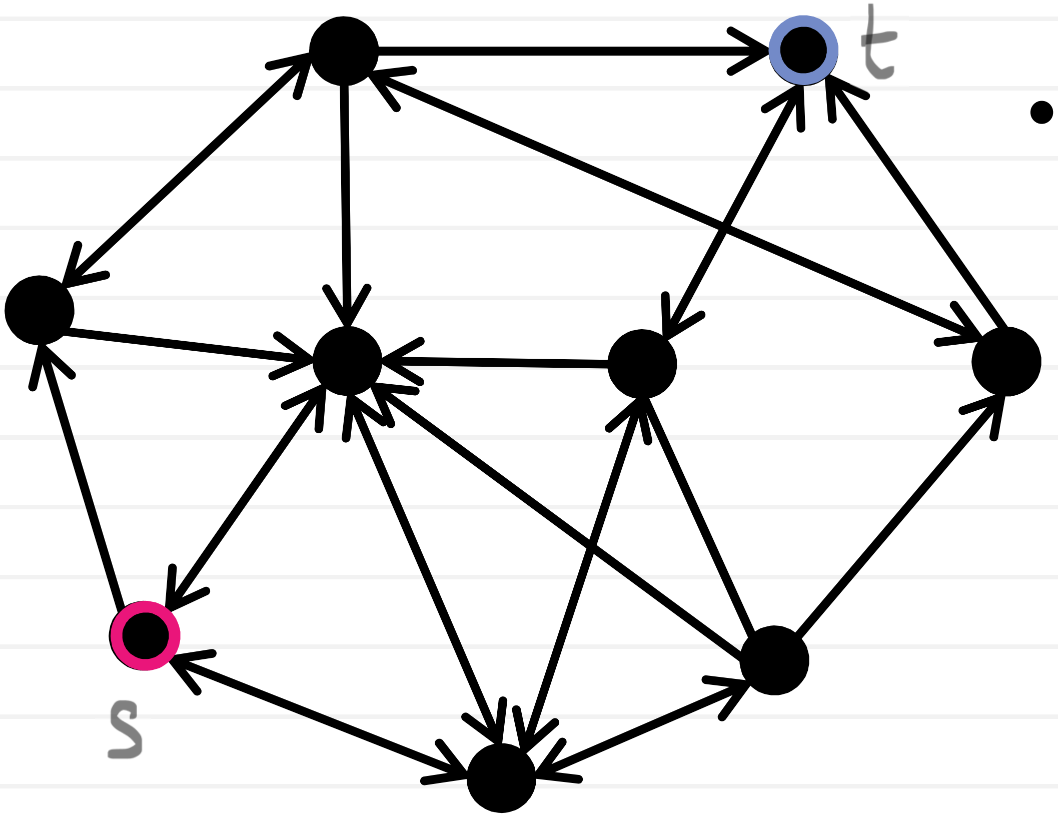
$$b + \sum_{i \in I(s)} f_{is} = \sum_{i \in O(s)} f_{si}$$

$$\sum_{i \in I(t)} f_{it} = \sum_{i \in O(t)} f_{ti} + b$$

$$\forall j \neq s,t \quad \sum_{i \in I(j)} f_{ij} = \sum_{i \in O(j)} f_{ji}$$

Maxflow interpretation

- Arc capacities are maximum amount of water that can be pumped.



- We try to pump as much water as possible from s to t.

Dual to maxflow

$$\max_{b,f} \; b$$

$$\text{s.t.:} \; -f_{ij} \le 0 \quad d'_{ij}$$

$$f_{ij} \le u_{ij} \quad d_{ij}$$

$$x_s \quad \sum_{i \in O(s)} f_{si} - \sum_{i \in I(s)} f_{is} - b = 0$$

$$x_t \quad \sum_{i \in O(t)} f_{ti} + b - \sum_{i \in I(t)} f_{it} = 0$$

$$\forall_j \neq s, t$$

$$x_j \quad \sum_{i \in O(j)} f_{ji} - \sum_{i \in I(j)} f_{ij} = 0$$

$$L(b, f, x, d, d') =$$

$$= b(-1 + x_t - x_s) +$$

$$+ \sum_{ij} f_{ij} (d_{ij} - d'_{ij} + x_i - x_j) +$$

$$\sum_{ij} -u_{ij} d_{ij}$$

$$\min \sum_{ij} u_{ij} d_{ij}$$

$$\text{s.t.:} \; d, d' \geq 0$$

$$x_t - x_s = 1$$

$$d_{ij} - d'_{ij} + x_i - x_j = 0$$

Dual to maxflow

$$\min \sum_{ij} u_{ij} d_{ij}$$

s.t.: $d, d' \geq 0$

$$x_t - x_s = 1$$

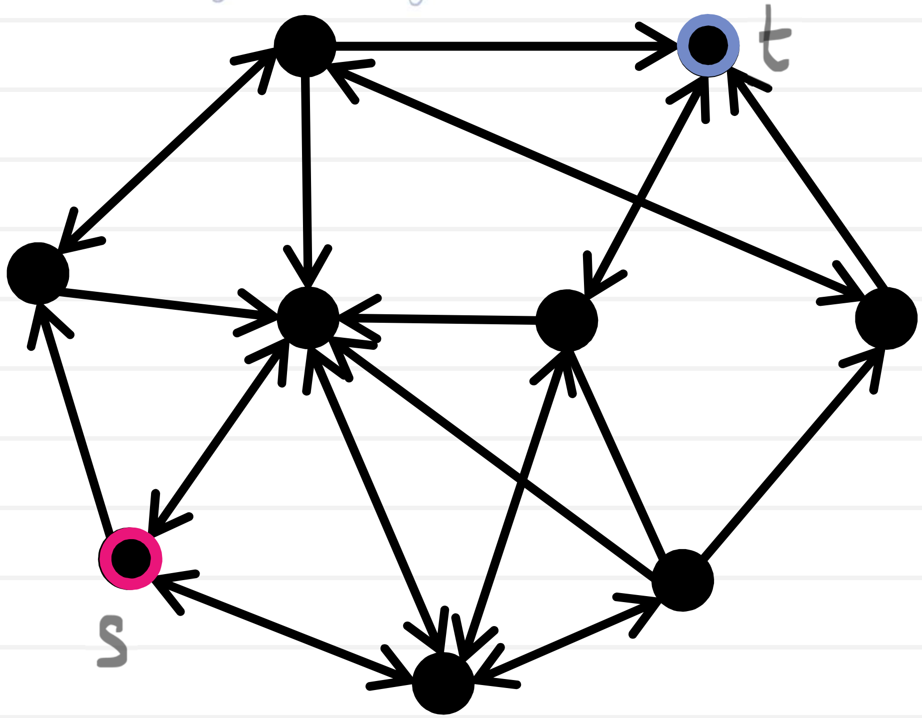
$$d_{ij} - d'_{ij} + x_i - x_j = 0$$

$$\min \sum_{ij} u_{ij} d_{ij}$$

s.t.: $d_{ij} \geq 0$

$$x_t - x_s = 1$$

$$d_{ij} \geq x_j - x_i$$



$$\min \sum_{ij} u_{ij} d_{ij}$$

s.t.: $d_{ij} \geq x_j - x_i$

$$d_{ij} \geq 0$$

$$x_s = 0 \quad x_t = 1$$

Dual to maxflow

$$\min \sum_{ij} u_{ij} d_{ij}$$

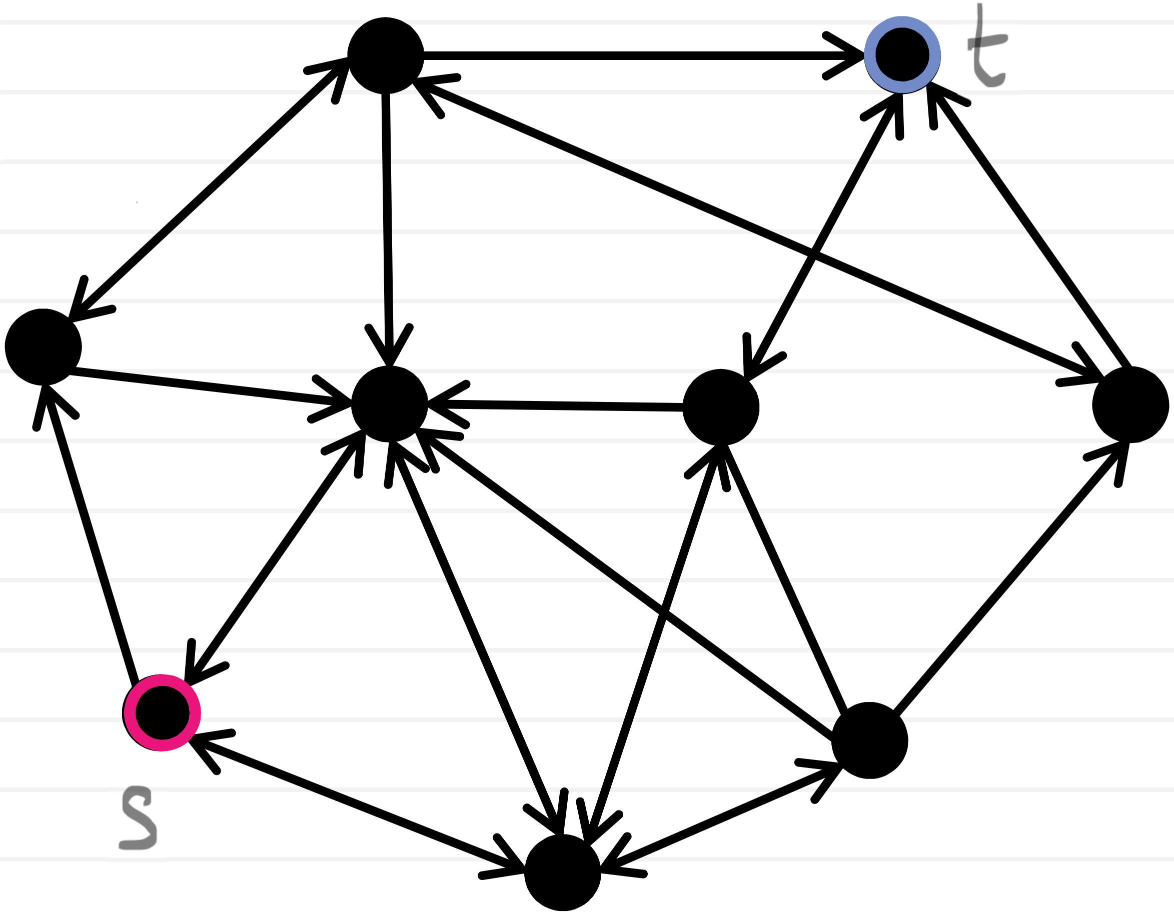
$$\text{s.t.: } d_{ij} \geq x_j - x_i$$

$$d_{ij} \geq 0$$

$$x_s = 0 \quad x_t = 1$$

$$\min \sum_{ij} u_{ij} \max(x_j - x_i, 0)$$

$$\text{s.t.: } x_s = 0 \quad x_t = 1$$



Rounding the optimum

$$\hat{x} = \arg\min \sum_{ij} u_{ij} \max(x_j - x_i, 0) \quad \text{s.t.}: x_s = 0 \quad x_t = 1$$

$$\tau \sim U(0;1)$$

$$\tilde{x} = \hat{x} > \tau$$

$$E_{\tau \sim U(0;1)} \max(\tilde{x}_j - \tilde{x}_i, 0) = E_{\tau \sim U(0;1)} [\hat{x}_i \le \tau \le \hat{x}_j] \le$$

$$\le \max(\hat{x}_j - \hat{x}_i, 0)$$

$$E_{\tau \sim U(0;1)} \sum_{ij} u_{ij} \max(\tilde{x}_j - \tilde{x}_i, 0) \le \sum_{ij} u_{ij} \max(\hat{x}_j - \hat{x}_i, 0)$$

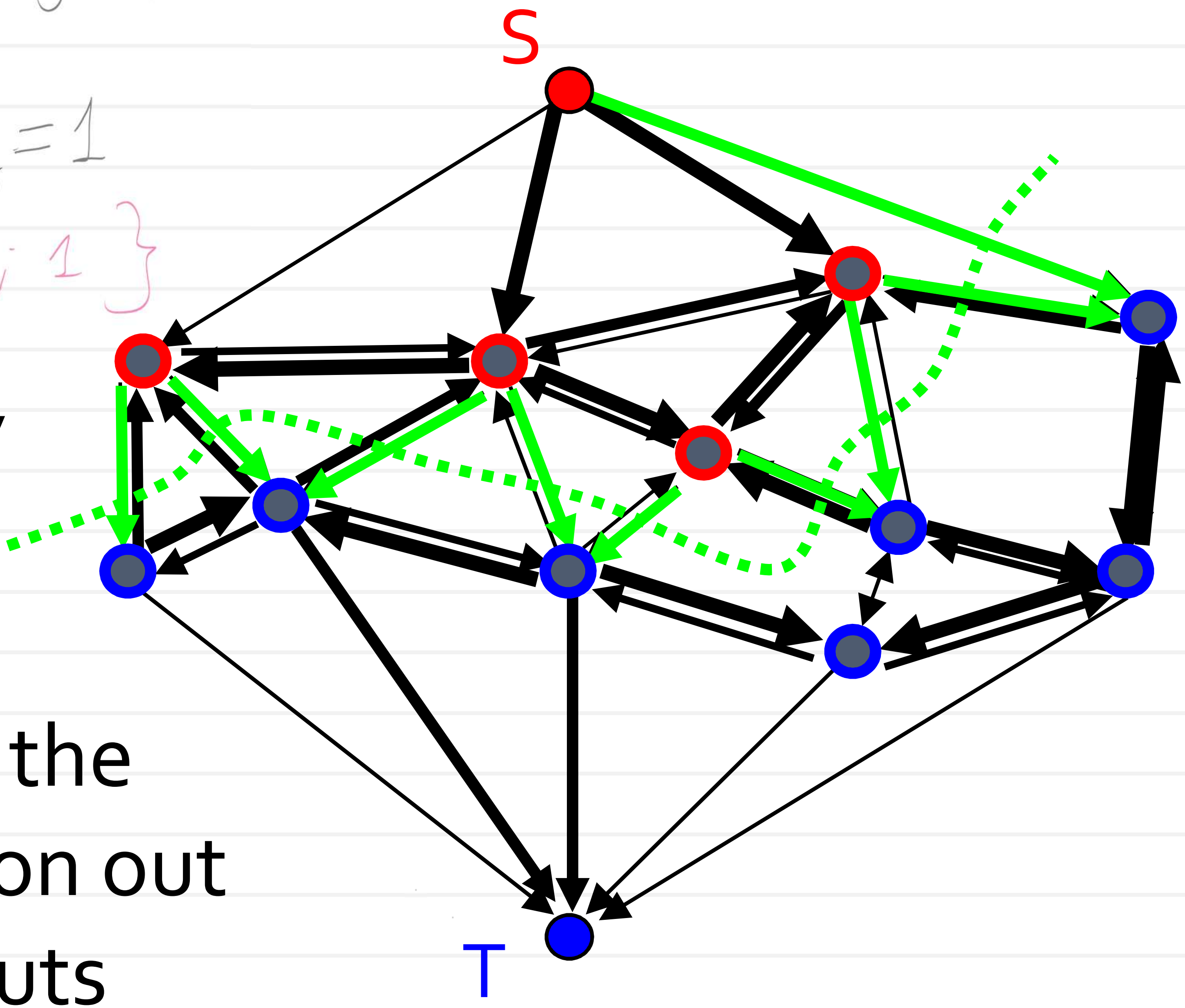
$$\forall \tau \quad \tilde{x} = \arg\min \sum_{ij} \max(x_j - x_i, 0)$$

Dual to maxflow

$$\min \sum_{ij} u_{ij} \max(x_j - x_i, 0)$$

$$\text{s.t.: } x_s = 0 \quad x_t = 1$$

$$x_i \in \{0; 1\}$$



- *st-cut* = any partition of vertices
- *st-mincut* = the best partition out of $2^{|V|-2}$ st-cuts

Recovering st-mincut

$$\max_{b,f} \; b$$

$$\text{s.t.}: \; -f_{ij} \leq 0 \qquad d'_{ij}$$

$$f_{ij} \leq u_{ij} \qquad d_{ij}$$

$$\min \sum_{ij} u_{ij} d_{ij}$$

$$\text{s.t.}: \; d_{ij} \geq x_j - x_i$$

$$d_{ij} \geq 0$$

$$x_s \quad \sum_{i \in O(s)} f_{si} - \sum_{i \in I(s)} f_{is} - b = 0$$

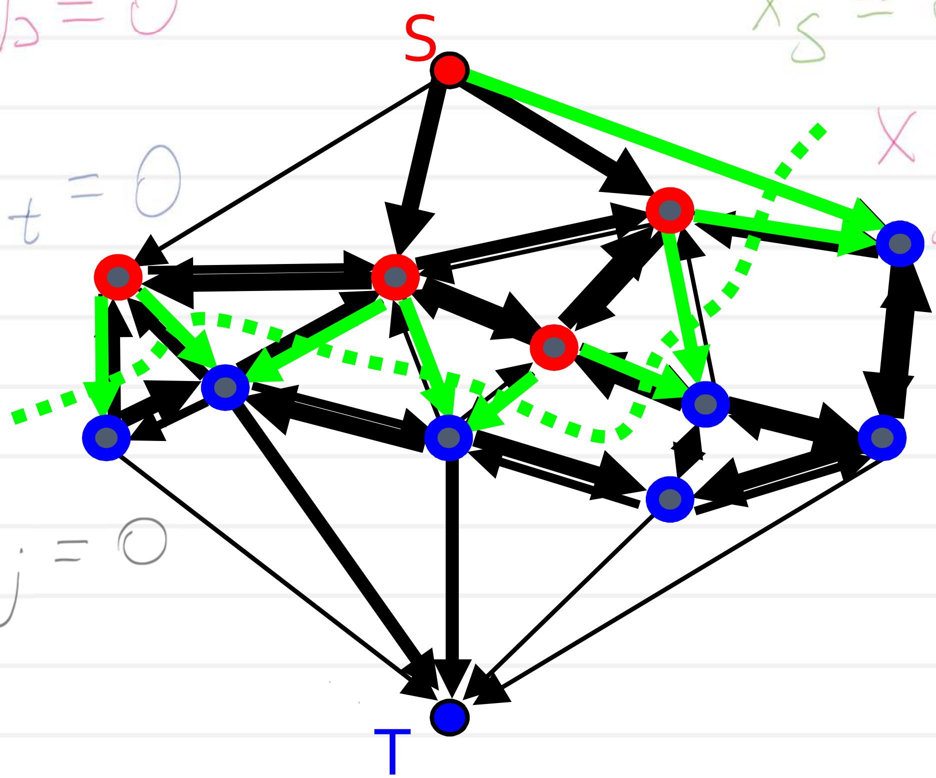
$$x_s = 0 \quad x_t = 1$$

$$x_t \quad \sum_{i \in O(t)} f_{ti} + b - \sum_{i \in I(t)} f_{it} = 0$$

$$x_j \in \{0;1\}$$

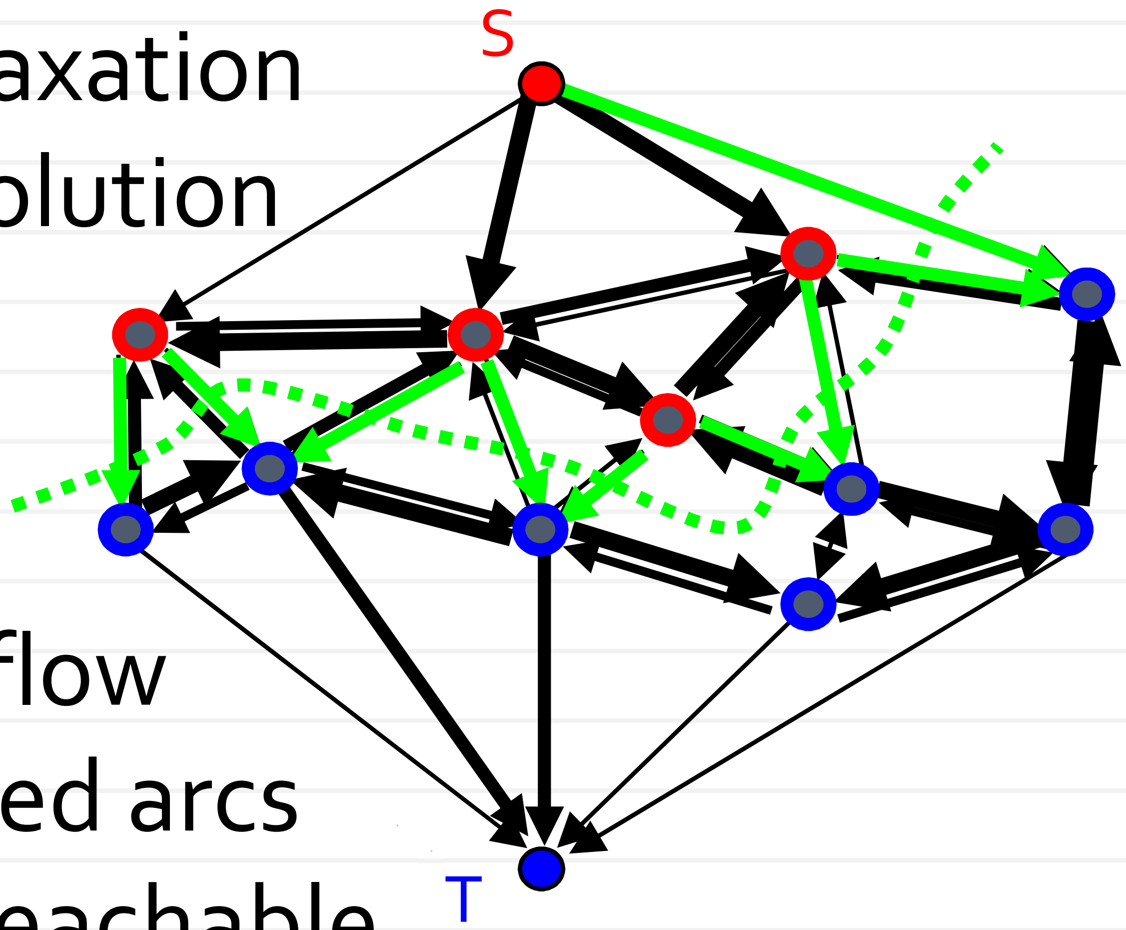
$$\forall j \neq s,t$$

$$x_j \quad \sum_{i \in O(j)} f_{ji} - \sum_{i \in I(j)} f_{ij} = 0$$



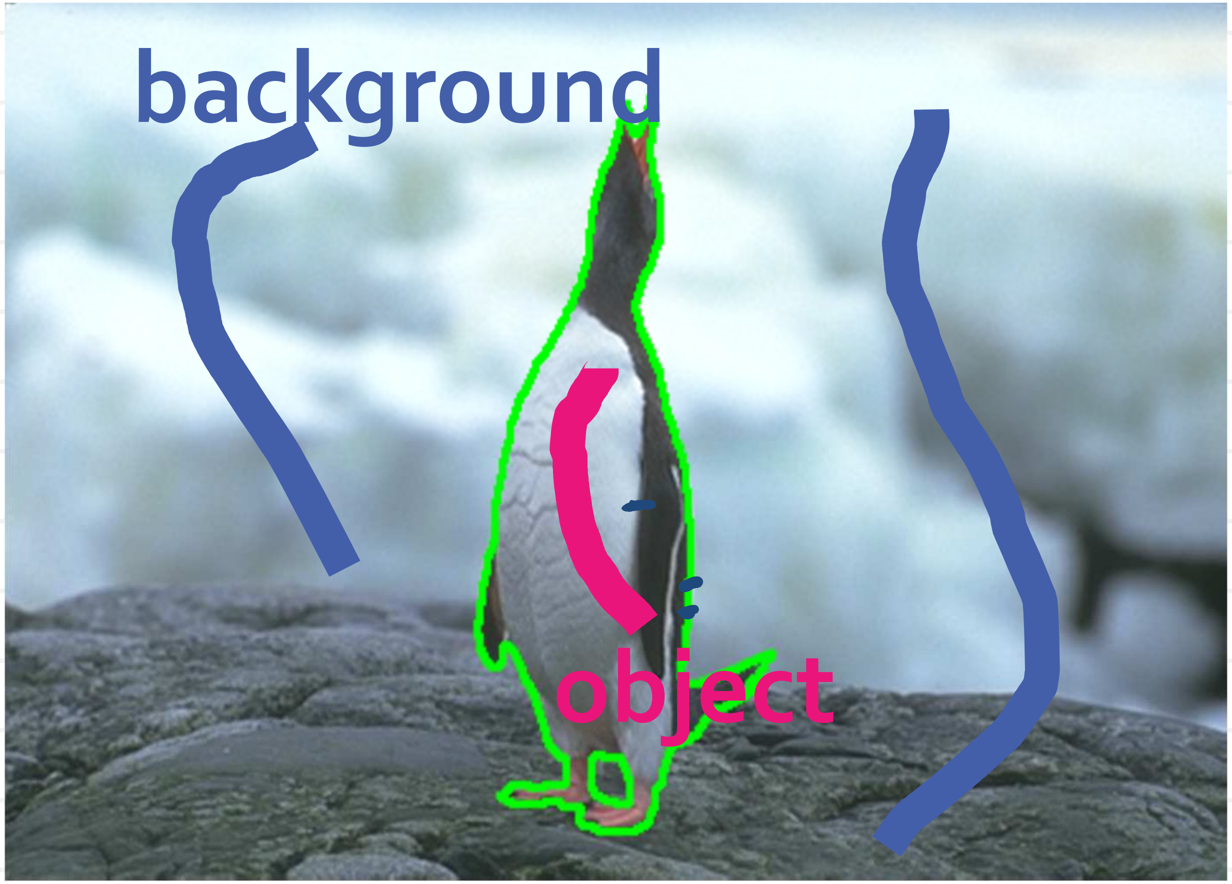
Recovering st-mincut

1. Solve the LP relaxation
2. Threshold the solution



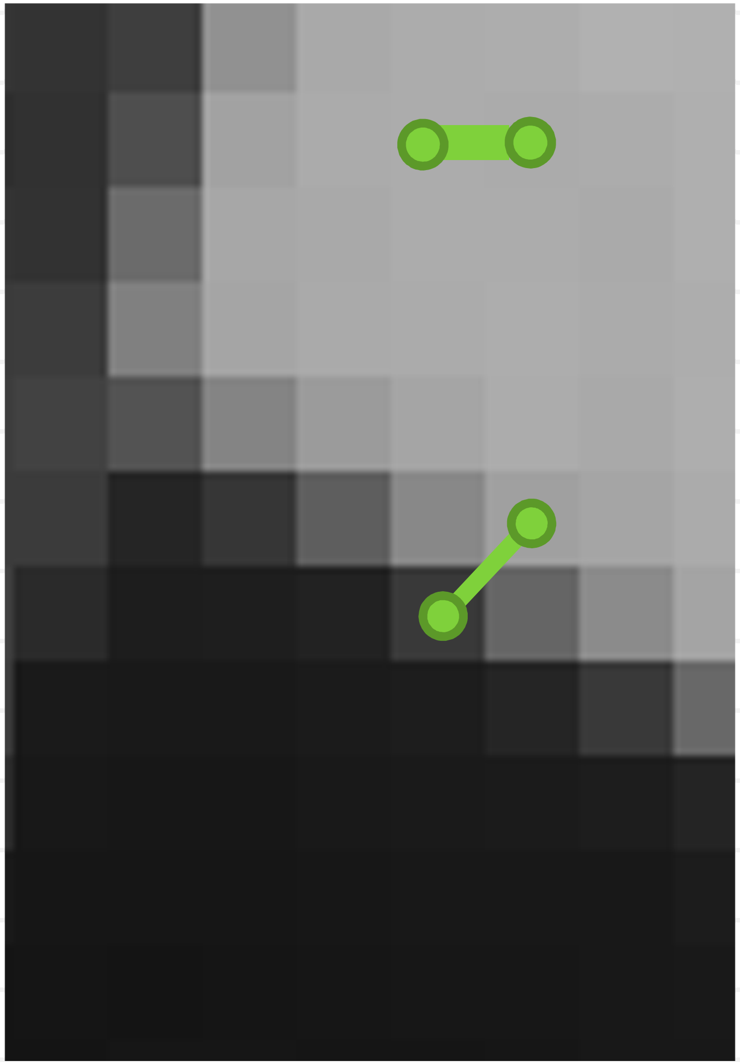
1. Find maximum flow
2. Remove saturated arcs
3. For all vertices reachable from *s* set $x_j = 0$
4. For all others set $x_j = 1$

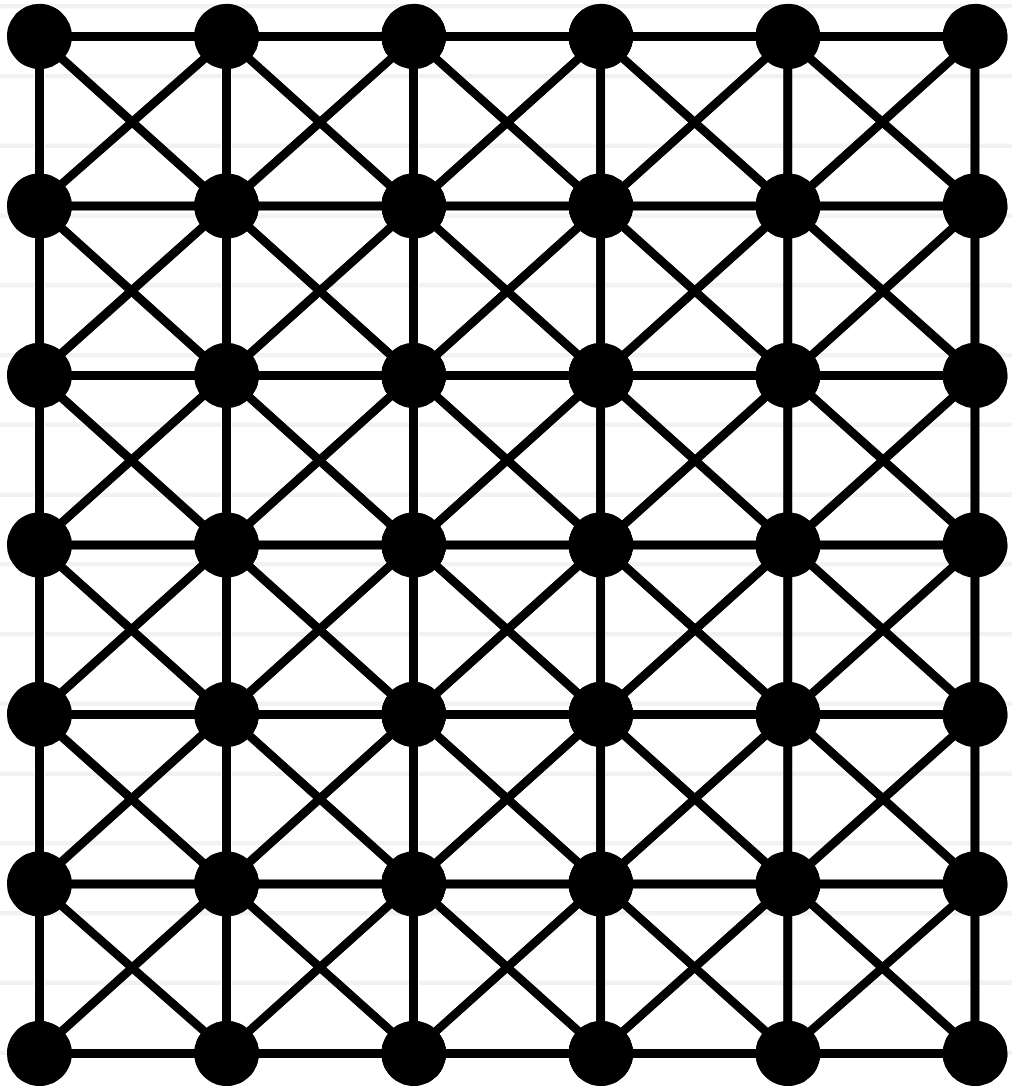
Segmentation via st-mincut



[Boykov & Jolly '01]

Segmentation via st-mincut

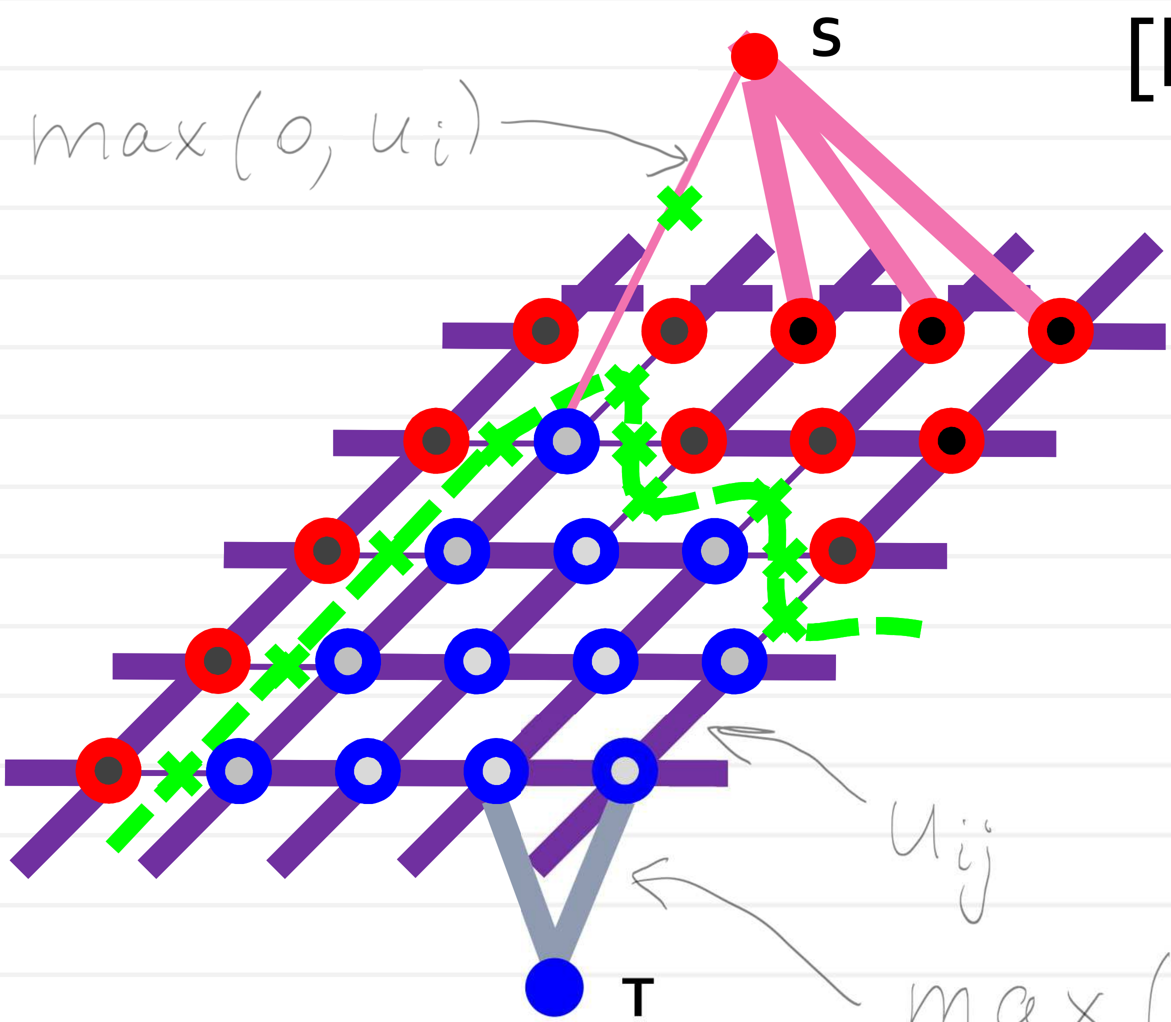




$$u_{ij} = \exp\left(-\frac{\|I_p - I_q\|^2}{2\sigma^2}\right)\frac{1}{\|p-q\|^2}$$

Segmentation via mincut

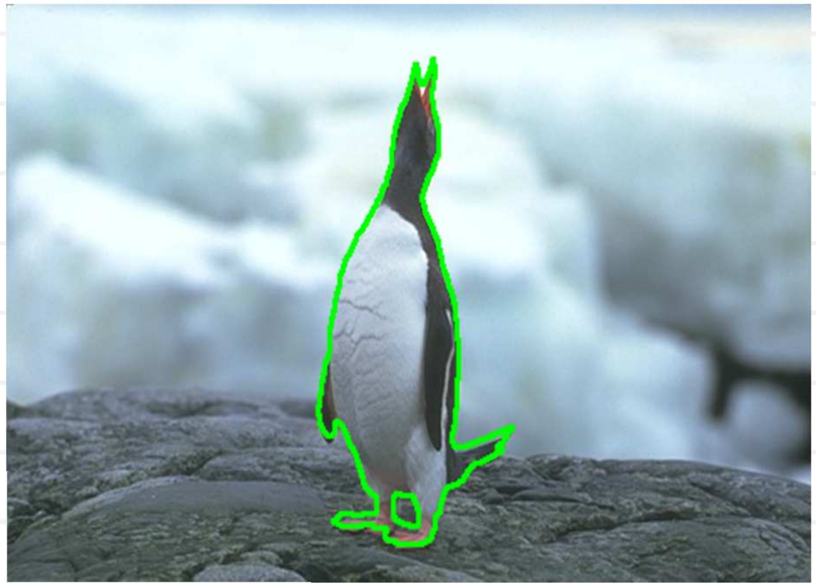
[Boykov & Jolly '01]



$\max(0, u_i)$

u_{ij}

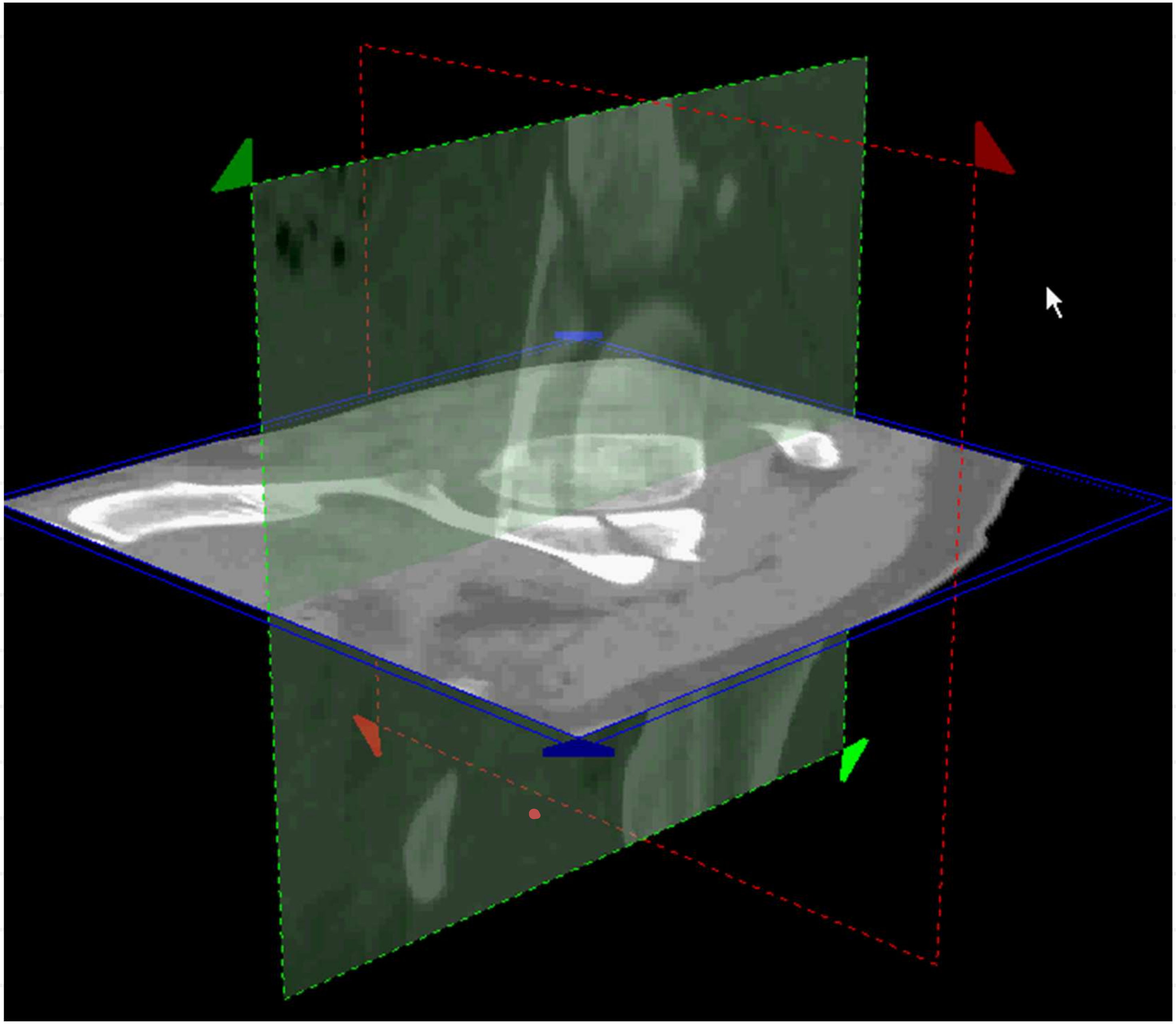
$\max(0, -u_i)$



$$\min_{x\in\{0;1\}} \sum_i u_i x_i + \lambda \sum_{ij} u_{ij} |x_i - x_j|$$

Segmentation example

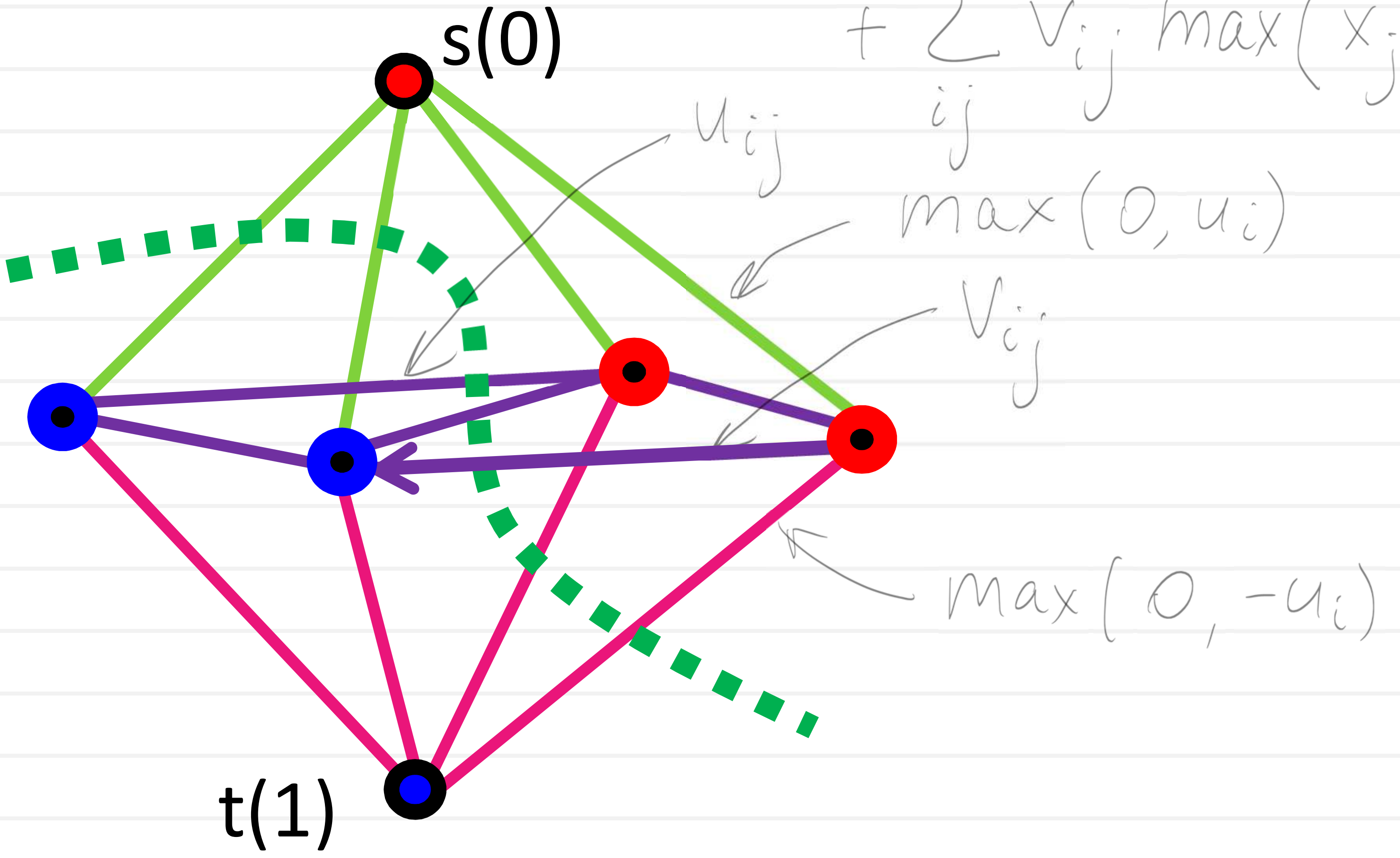
Video from [Boykov, Kolmogorov, 2003]



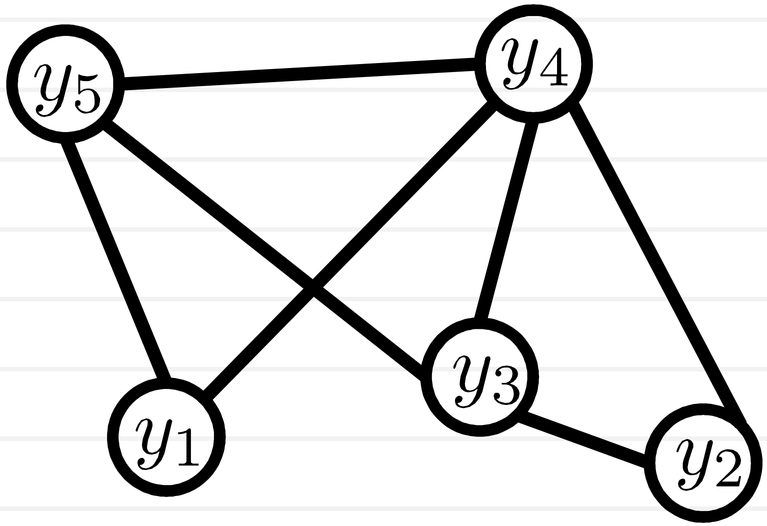
Binary optimization with pairwise terms

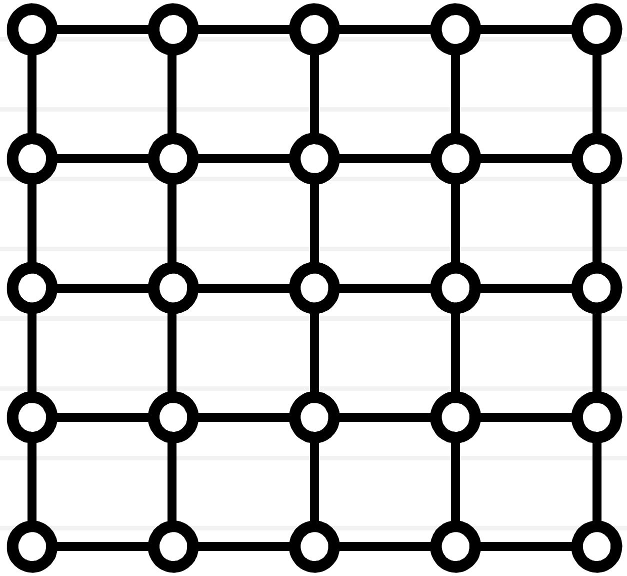
$$\min_{x \in \{0;1\}} \sum_i u_i x_i + \sum_{ij} u_{ij} |x_i - x_j| +$$

$$+ \sum_{ij} v_{ij} \max(x_j - x_i, 0)$$



Binary optimization with pairwise terms





General case (binary variables, pairwise energy):

$$\min \sum_i E_i(x_i) + \sum_{ij} E_{ij}(x_i, x_j)$$

$$x \in \{0;1\}$$

$E_i(0)$
$E_i(1)$

Implementing a pairwise term

$$\sum_i E_i(x_i) + \sum_{ij} E_{ij}(x_i, x_j)$$

$$x \in \{0; 1\}$$

	$x_j = 0$	$x_j = 1$
$x_i = 0$	$A = E_{ij}(0,0)$	$B = E_{ij}(0,1)$
$x_i = 1$	$C = E_{ij}(1,0)$	$D = E_{ij}(1,1)$

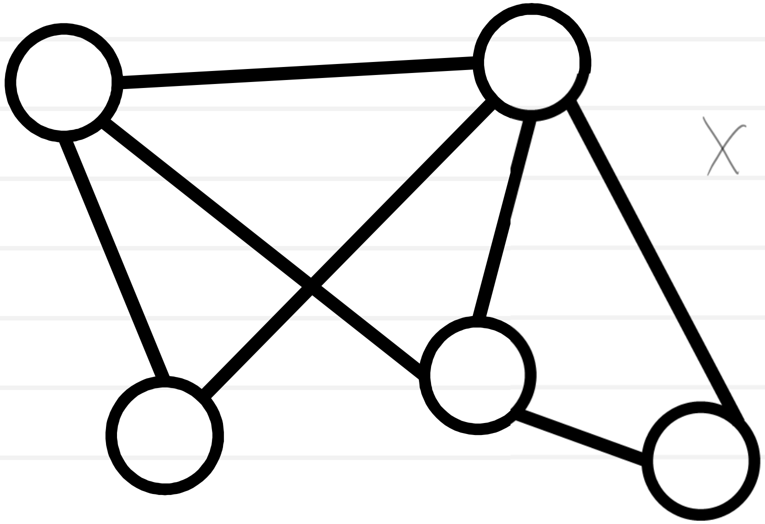
$$= \begin{array}{|c|c|}\hline A & A \\ \hline A & A \\ \hline\end{array} + \begin{array}{|c|c|}\hline 0 & 0 \\ \hline C\text{-}A & C\text{-}A \\ \hline\end{array} + \begin{array}{|c|c|}\hline 0 & D\text{-}C \\ \hline 0 & D\text{-}C \\ \hline\end{array} + \begin{array}{|c|c|}\hline 0 & B\text{+}C\text{-}A\text{-}D \\ \hline 0 & 0 \\ \hline\end{array}$$

Sufficient condition (*submodularity*):

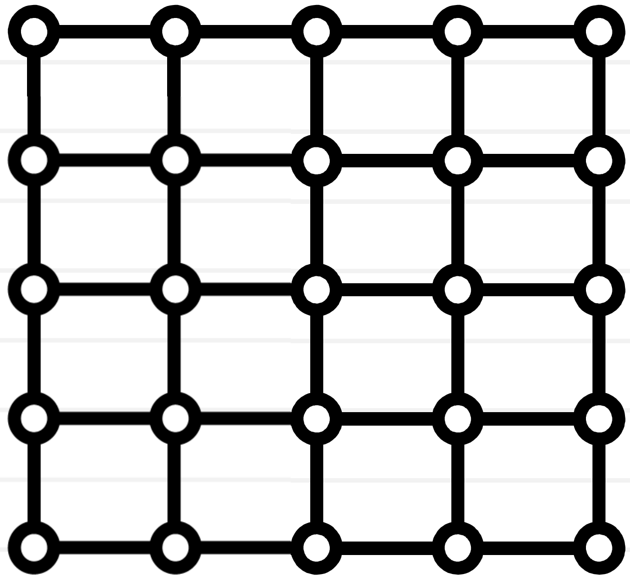
$$A + D \leq B + C$$

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

Binary optimization with pairwise terms



$x \in \{0;1\}$



For $\sum_i E_i(x_i) + \sum_{ij} E_{ij}(x_i, x_j)$

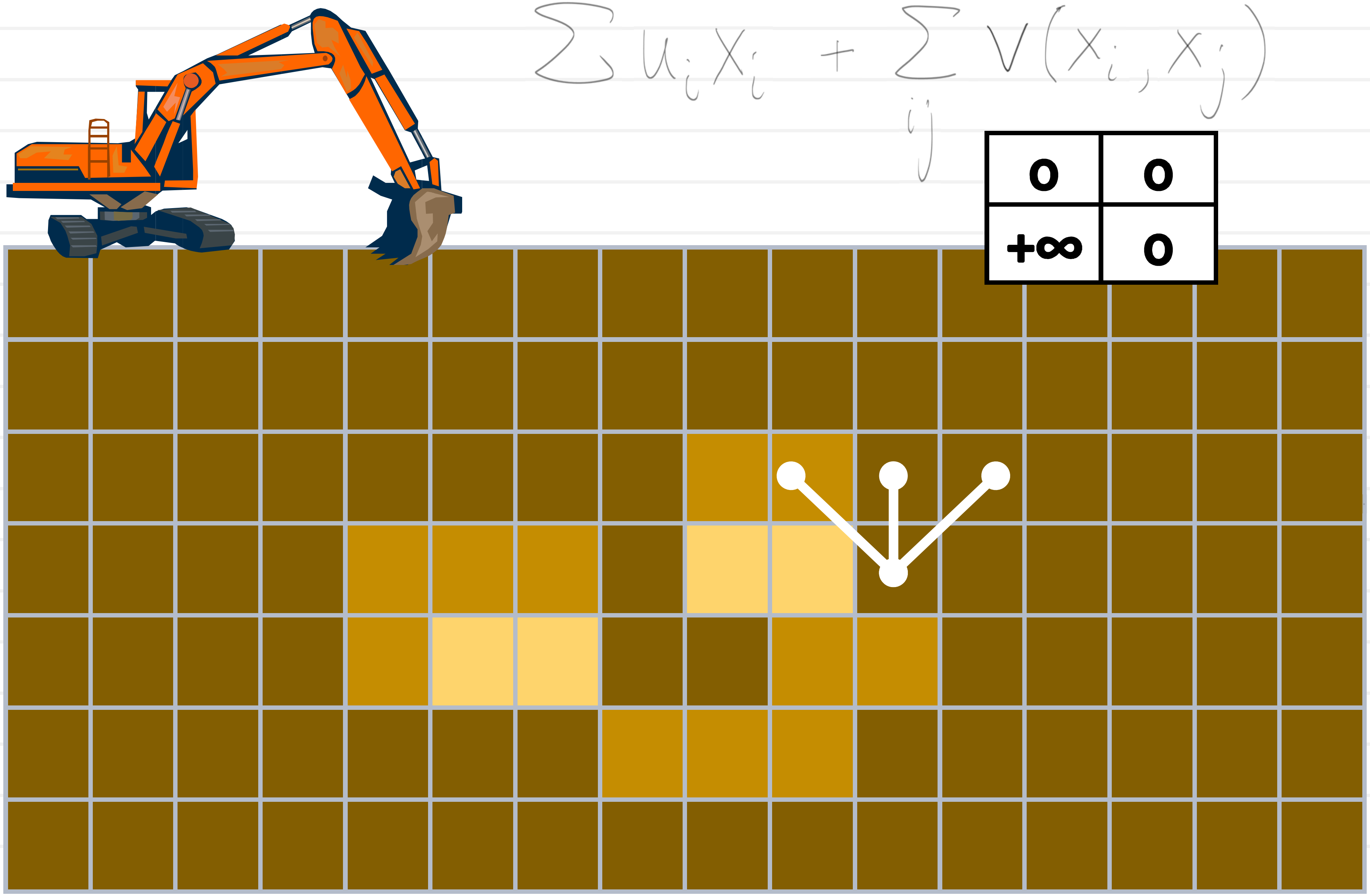
if $E_{ij}(0,0) + E_{ij}(1,1) \le E_{ij}(0,1) + E_{ij}(1,0)$

then we can find a global optimum exactly and efficiently

Open pit mining

$$\sum_i u_i x_i + \sum_{ij} V(x_i, x_j)$$

0	0
+∞	0

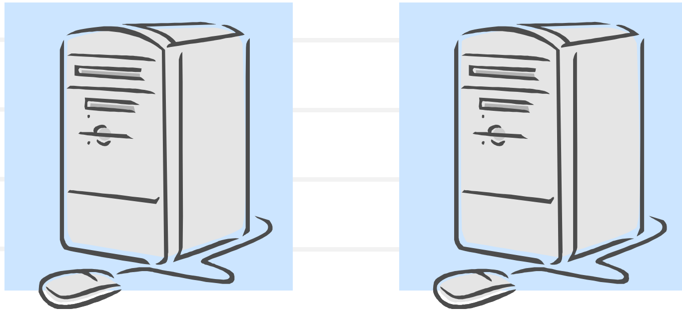


But how to find mincuts in huge graphs??

Dual decomposition

How to decompose/parallelize?

$$\min f_1(x) + f_2(x)$$
$$\text{s.t.: } x \in D$$



$$\min f_1(x_1) + f_2(x_2)$$
$$\text{s.t.: } x_1 \in D \quad x_2 \in D$$
$$x_1 - x_2 = 0 \qquad \lambda$$

$$L(x_1, x_2, \lambda) = f_1(x_1) + f_2(x_2) + \lambda^T (x_1 - x_2)$$
$$x_1 \in D \quad x_2 \in D$$

$$d(\lambda) = \min_{x_1 \in D} (f_1(x_1) + \lambda^T x_1) + \min_{x_2 \in D} (f_2(x_2) - \lambda^T x_2)$$

Dual decomposition

How to decompose / parallelize?

$$d(\lambda) = \min_{x_1 \in D} (f_1(x_1) + \lambda^T x_1) + \min_{x_2 \in D} (f_2(x_2) - \lambda^T x_2)$$

$x_1 - x_2 = 0$ λ

$$\nabla d(\lambda) = \hat{x}_1(\lambda) - \hat{x}_2(\lambda)$$

Intuitive interpretation of subgradient ascent: trying to bring together two minima.

If $\hat{x}_1(\lambda) = \hat{x}_2(\lambda)$

then $d(\lambda) = f1(x)+f2(x)$. Strong duality!

Part-of-speech tagging

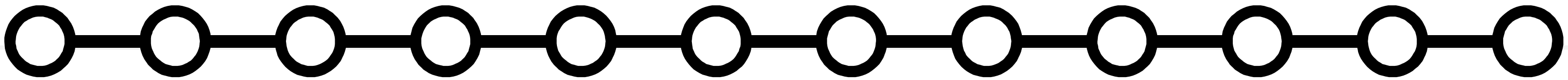
DT/ The JJ/ quick JJ/ brown NN/ fox ~~NNS~~/ jumps IN/ over DT/ the JJ/ lazy NN/ dog

VBZ

- -RRB- - Right bracket
- CD - Cardinal number
- EX - Existential there
- IN - Preposition
- JJR - Comparative adjective
- LS - List Item Marker
- NN - Singular noun
- NNP - Proper singular noun
- PDT - Predeterminer
- PRP - Personal pronoun
- RB - Adverb
- RBS - Superlative Adverb
- SYM - Symbol
- UH - Interjection
- VBD - Verb, past tense
- VBN - Verb, past participle
- VBZ - Verb, 3rd ps. sing. present
- WP - wh-pronoun
- WRB - wh-adverb
- CC - Coordinating conjunction
- DT - Determiner
- FW - Foreign word
- JJ - Adjective
- JJS - Superlative adjective
- MD - Modal
- NNS - Plural noun
- NNPS - Proper plural noun
- POS - Possesive ending
- PP$ - Possesive pronoun
- RBR - Comparative adverb
- RP - Particle
- TO - to
- VB - Verb, base form
- VBG - Verb, gerund/present participle
- VBP - Verb, non 3rd ps. sing. present
- WDT - wh-determiner
- WP$ - Possesive wh-pronoun

http://cogcomp.cs.illinois.edu/demo/pos/

Part-of-Speech tagging



$$E(y) = \sum_{i=1}^{w} E_i(y_i) + \sum_{i=1}^{w-1} E_{i,i+1}(y_i, y_{i+1})$$

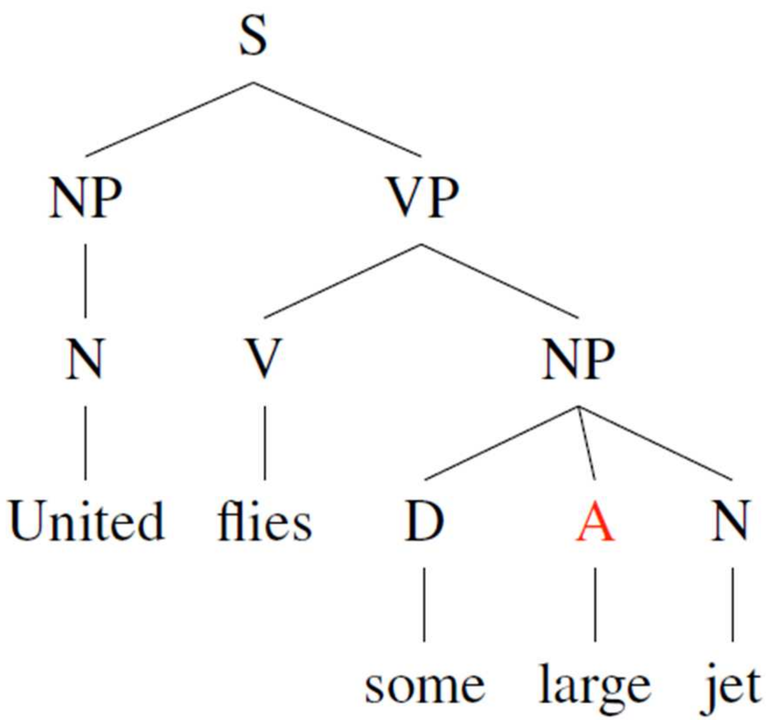
Dictionary information, capitalization, suffixes and prefixes, all of this for the preceding and subsequent words, etc.

Transition probabilities

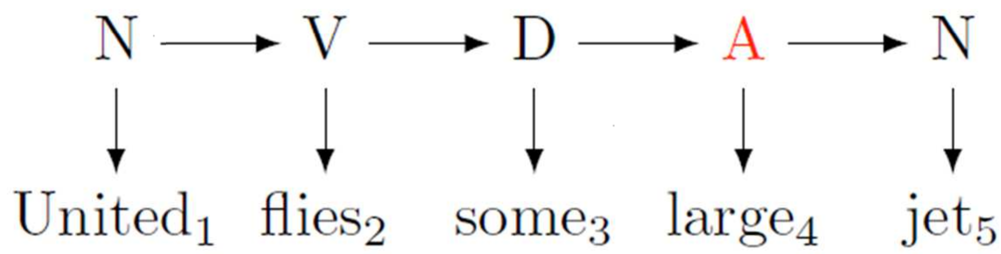
$$O(k,t) = \boxed{\begin{gathered}\min_{y_1 \ldots y_k} \sum_{i=1}^{k} E_i(y_i) + \sum_{i=1}^{k-1} E_{i,i+1}(y_i, y_{i+1}) \\ \text{s.t. } y_k = t\end{gathered}}$$

Dual decomposition in NLP

$$E(y) = E_{Tree}(y) + E_{CHAIN}(y)$$



[Rush & Collins//ACL 2011]

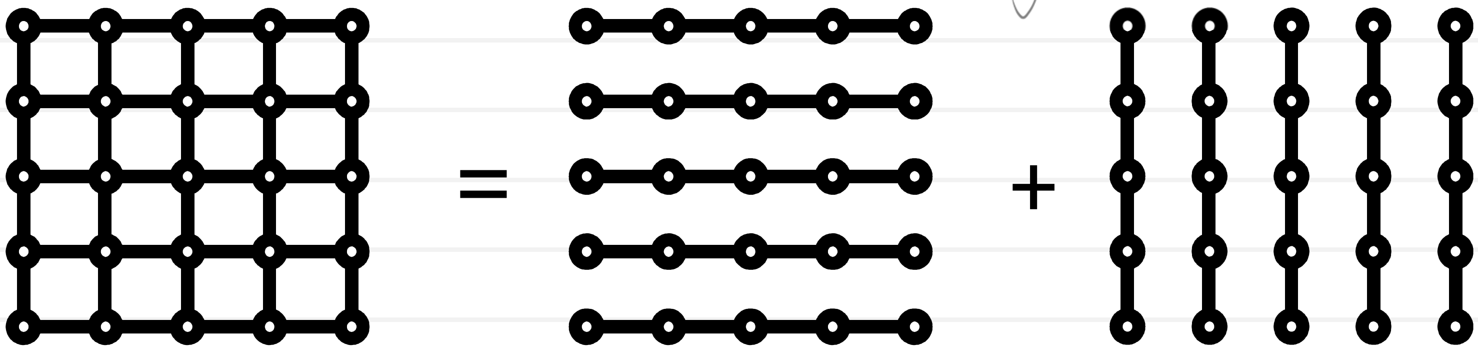


Context-free grammar

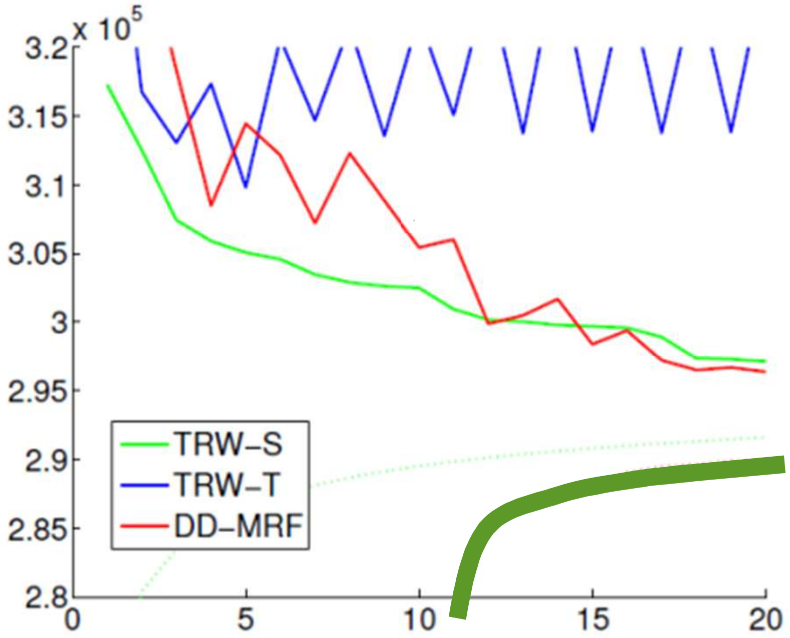
Conditional random fiels/
Hidden Markov Model

Dual decomposition in Markov Random Fields

$$E(x) = \sum E_i(x_i) + \sum_{ij \in N} E_{ij}(x_i, x_j)$$



(a) Estimated disparity

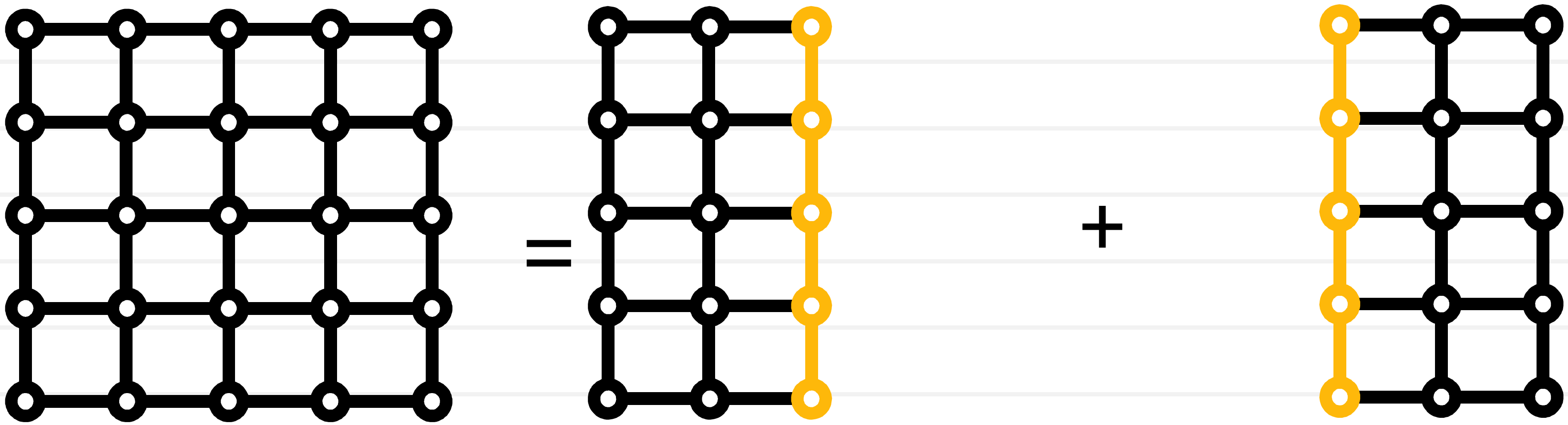


(b) Energy and lower bound plots

[Komodakis,PAMI 07]

Distributed segmentation

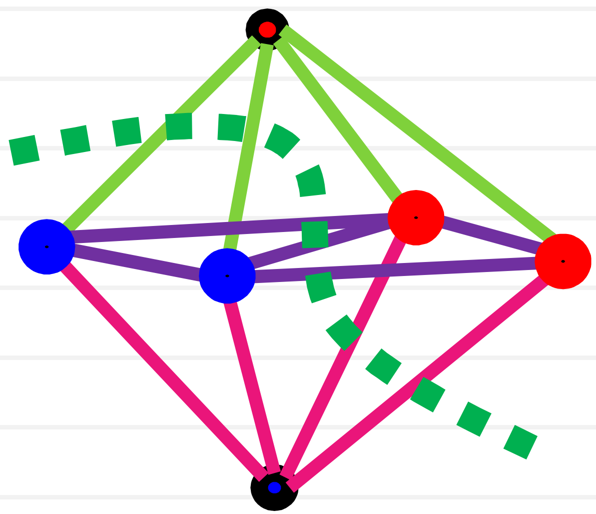
Binary submodular energy on a very large graph:



$E(\mathbf{y})$ $E^1(\mathbf{y}^1)$ $E^2(\mathbf{y}^2)$

$$\hat{\mathbf{y}}^1 = \arg\min \left(E^1(\mathbf{y}^1) + \sum_{p \in \mathcal{V}} \lambda_p y_p^1 \right)$$

$$\hat{\mathbf{y}}^2 = \arg\min \left(E^2(\mathbf{y}^2) - \sum_{p \in \mathcal{V}} \lambda_p y_p^2 \right)$$

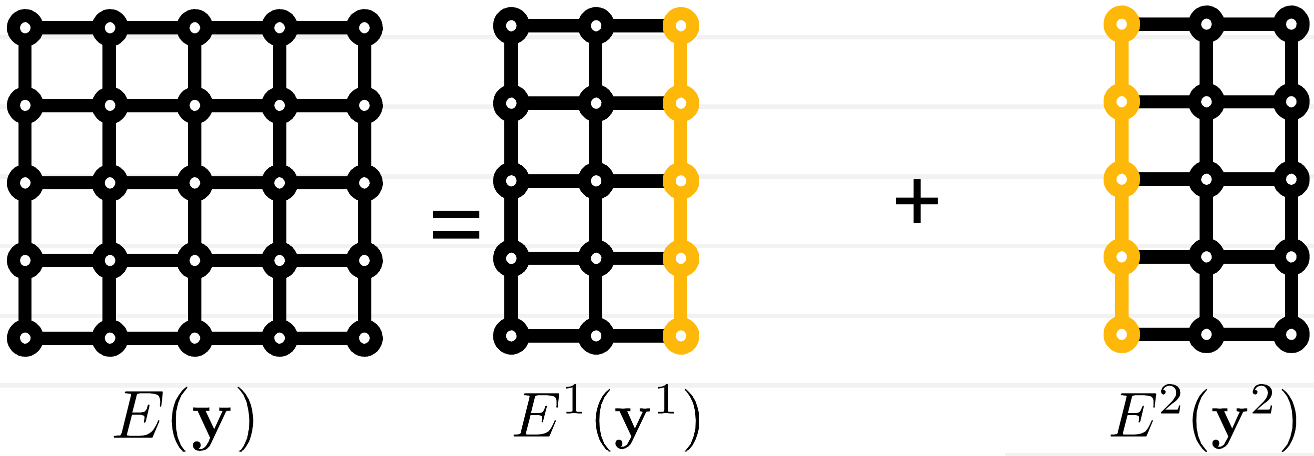


Key moment: at each moment, only few weights in the graph are affected.

[Strandmark&Kahl,CVPR2010]

Parallelizing large mincut

Binary submodular energy on a very large graph:



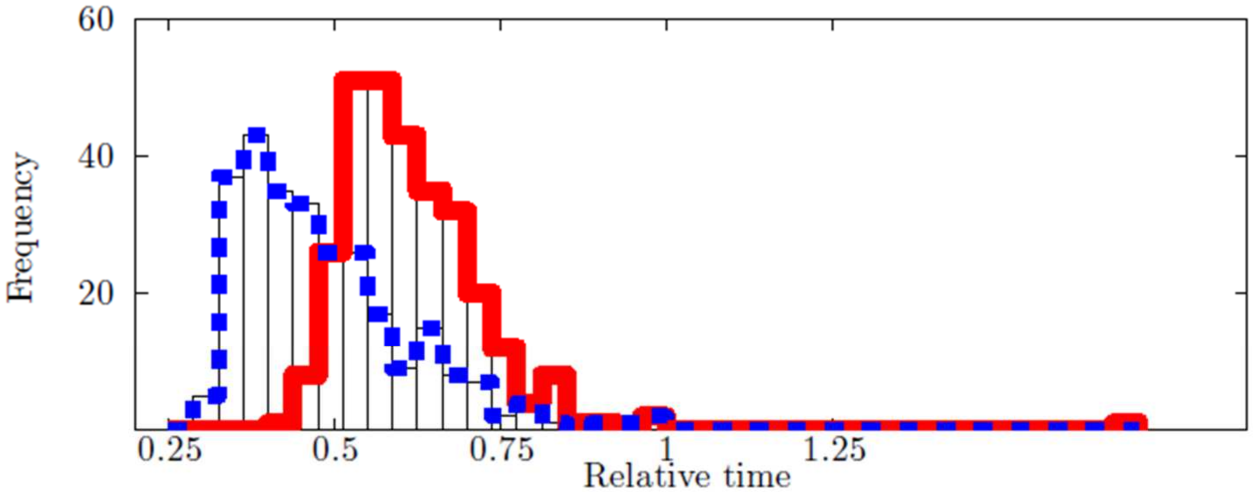


Figure 4: Relative times with 2 (red) and 4 (dotted blue) computational threads for the 301 images in the Berkeley segmentation database, using 4-connectivity. The medians are 0.596 and 0.455.

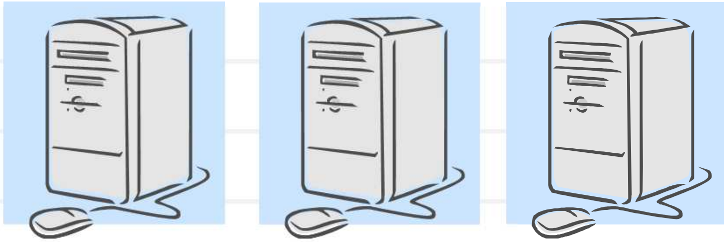
Key advantage:
memory

[Strandmark&Kahl,CVPR2010]

Dual decomposition (multiple parts)

How to decompose/parallelize?

$$\min \; f_1(x) + f_2(x) + f_3(x)$$
$$\text{s.t: } \; x \in D$$



$$\min \; f_1(x_1) + f_2(x_2) + f_3(x_3) + \dots$$
$$\text{s.t: } \; x_1 \in D \quad x_2 \in D \quad x_3 \in D$$
$$x_1 = X \quad x_2 = X \quad x_3 = X$$

$$L(x_i, \lambda_i, X) = \sum f_i(x_i) + \sum \lambda_i^T x_i - \sum \lambda_i^T X$$

$$\max_{\lambda_i} \; d(\lambda_i) = \sum \min_{x_i \in D} \left(f_i(x_i) + \lambda_i^T x_i \right)$$
$$\text{s.t: } \; \sum \lambda_i = 0$$

Dual decomposition (multiple parts)

$$\max_{\lambda_i} d(\lambda_i) = \sum \min_{x_i \in D} \left(f_i(x_i) + \lambda_i' x_i\right)$$

$$\text{s.t.} \quad \sum \lambda_i = 0$$

Subgradient: $(\hat{x}_1, \hat{x}_2, \hat{x}_3, \ldots)$

Projected subgradient:

$$(\hat{x}_1 - \hat{x}_0, \hat{x}_2 - \hat{x}_0, \hat{x}_3 - \hat{x}_0, \ldots)$$

$$\hat{x}_0 = \frac{1}{N} \sum \hat{x}_i$$

Intuitive interpretation of projected subgradient ascent: trying to bring together all minima.

One-slide summary of the course