MAP Estimation in Binary MRFs using Bipartite Multi-Cuts

Sashank J. Reddi,
Sunita Sarawagi, Sundar Vishwanathan
IIT Bombay

MAP Estimation

$$G = (\mathcal{V}, \mathcal{E}),\ x \in \{0,1\}^n$$

$$E(\mathbf{x}|\theta) = \sum_{i \in \mathcal{V}} \theta_i(x_i) + \sum_{(i,j) \in \mathcal{E}} \theta_{ij}(x_i, x_j) + \theta_{const}$$

$$\mathbf{x}^* = \operatorname{argmin}_{\mathbf{x} \in \{0,1\}^n} E(\mathbf{x}|\theta)$$

Popular Approach: LP Relaxation

$$\min_{\mu} \sum_{i,x_i} \theta_i(x_i)\mu_i(x_i) + \sum_{(i,j),x_i,x_j} \theta_{ij}(x_i,x_j)\mu_{ij}(x_i,x_j)$$

$$\sum_{x_j} \mu_{ij}(x_i,x_j) = \mu_i(x_i) \quad \forall (i,j) \in \mathcal{E}, \forall x_i \in \{0,1\}$$

$$\sum_{x_i} \mu_i(x_i) = 1 \quad \forall i \in \mathcal{V}$$

$$\mu_{ij}(x_i,x_j) \geq 0 \quad \forall (i,j) \in \mathcal{E}, \forall x_i, x_j \in \{0,1\}$$

- Implementations: TRW (tree message-passing), QPBO (graph-cut), MPLP (belief propagation)

Graph-cut Approximation

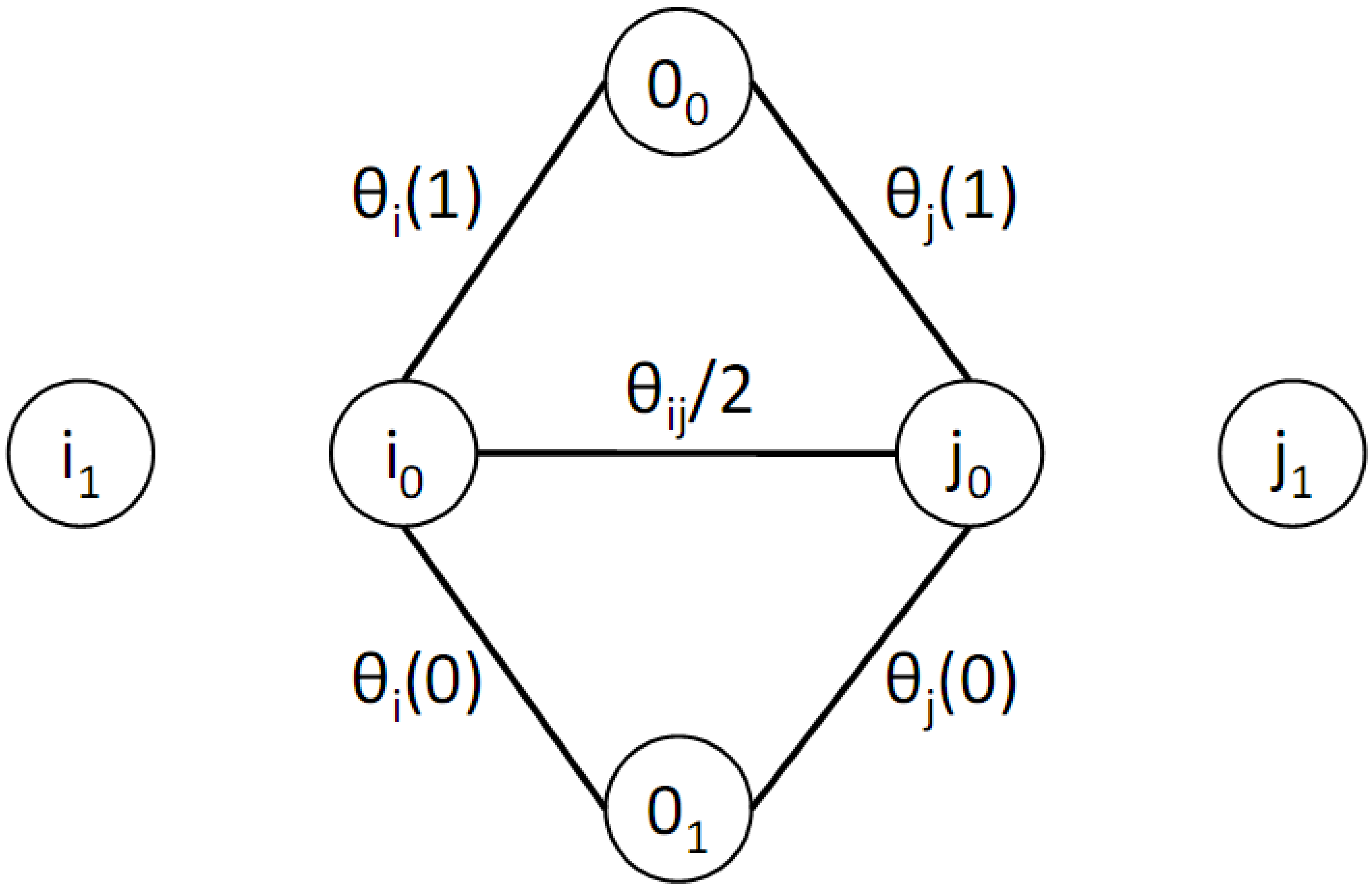
- Redefine E(x|θ) such that:
  - Symmetric weights $\theta_{ij}(x_i, x_j) = \theta_{ij}(\overline{x}_i, \overline{x}_j)$
  - Zero-normalized $\min_{x_i} \theta_i(x_i) = 0$ and $\min_{x_i, x_j} \theta_{ij}(x_i, x_j) = 0$

- Any energy function can be redefined into canonical normal form

Graph Construction $H=(V_H,E_H)$

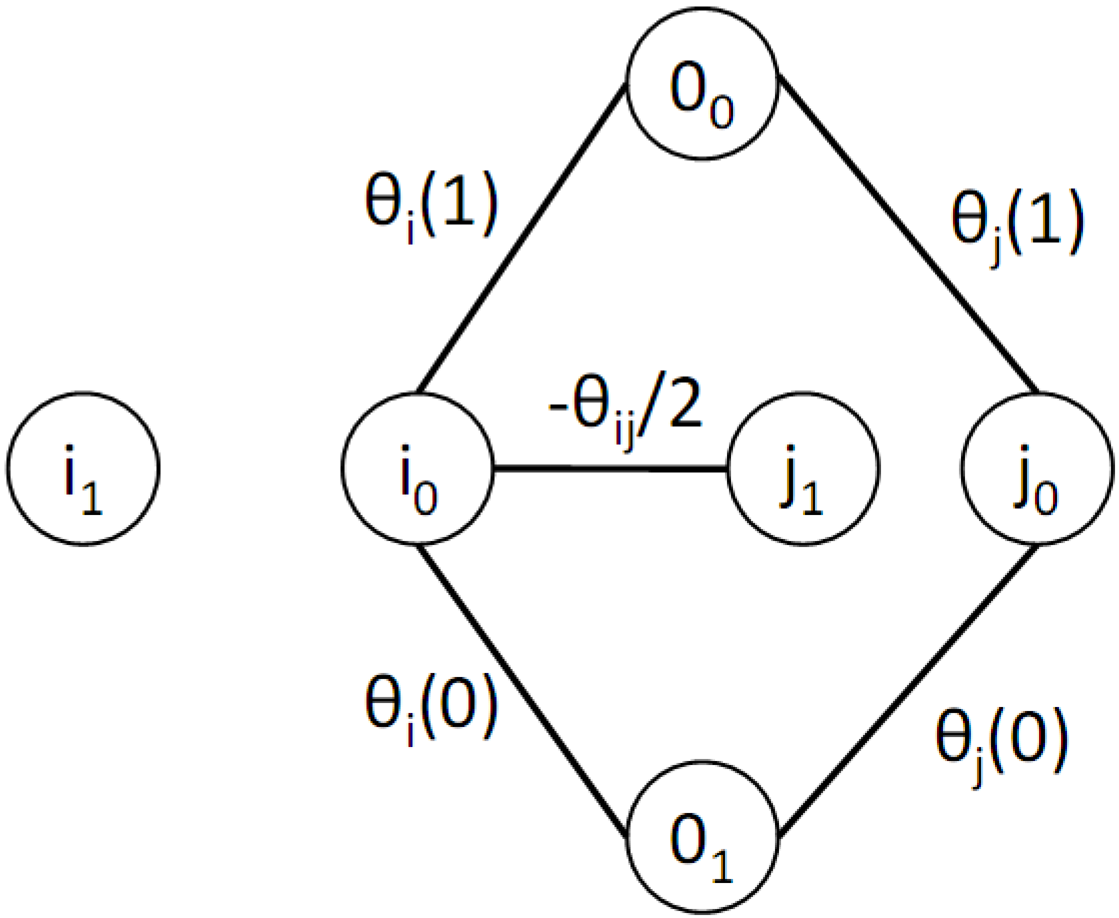
- Include a x_0 pseudo-variable so that vertex weights can be redefined as edge weights
- Create a new set of vertices i_0 and i_1 for each x_i in $0 \le i \le n$
- For each vertex $i \in V$: add an edge $(0_0,i_0)$ if its weight is non-zero, else connect $(0_1,i_0)$
- For each edge $(i, j) \in E$: add an edge (i_0,j_0) if the edge is submodular, else add edge (i_0,j_1)

Graph Construction



$$\theta_{ij} = \theta_{ij}(0,1) + \theta_{ij}(1,0) - \theta_{ij}(0,0) - \theta_{ij}(1,1) > 0$$

Graph Construction



$$\theta_{ij} = \theta_{ij}(0,1) + \theta_{ij}(1,0) - \theta_{ij}(0,0) - \theta_{ij}(1,1) < 0$$

Graph-cut Approximation

$$\min_{d_e, D(.)} \sum_{e \in \mathcal{E}_H} w_e d_e$$

$$\begin{aligned}
d_e + D(i_s) - D(j_t) &\geq 0 & \forall e = (i_s, j_t) \in \mathcal{E}_H \\
d_e + D(j_t) - D(i_s) &\geq 0 & \forall e = (i_s, j_t) \in \mathcal{E}_H \\
D(0_0) &= 0 & \\
D(i_s) &\in [0,1] & \forall i_s \in \mathcal{V}_H \\
d_e &\in [0,1] & \forall e \in \mathcal{E}_H \\
D(i_0) + D(i_1) &= 1 & \forall i \in \{0, \ldots, n\}
\end{aligned}$$

Notation notes:

$D(i_0) \equiv x_i$

$D(i_1) \equiv \bar{x}_i$

w_e = edge weight

d_e = path distance with an associated edge

- Solved exactly by finding s-t min-cut with (s,t) as $(0_s, 0_t)$

Bipartite Multi-cut

- Given an undirected graph J = (N,A) with non-negative edge weights and k(s,t) pairs:
  - find the minimum cut which divides the graph into two regions that separate the (s,t) pairs
- LP relaxation given in Sreyash Kenkre and Sundar Vishwanathan, 2006
- Bipartite Multi-cut is a subset of the Multi-cut problem which allows for any number of regions

Bipartite Multi-cut LP

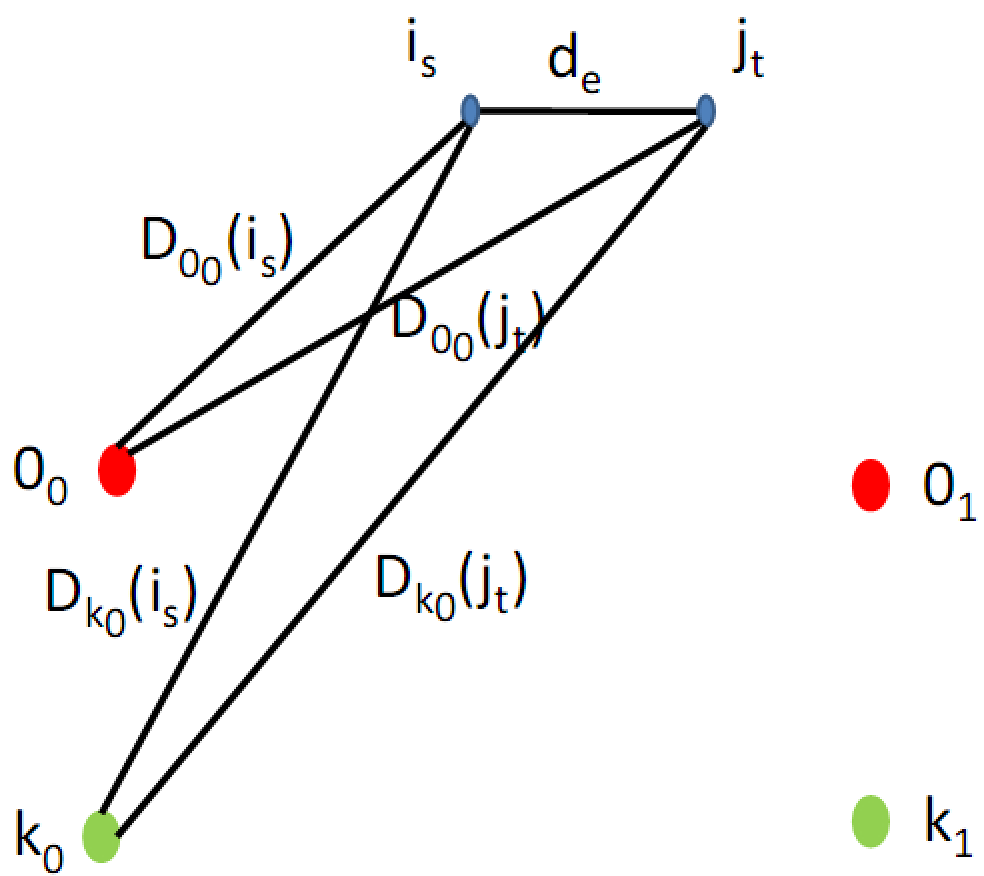
$$\begin{aligned}
\min \sum_{e \in \mathcal{E}_H} w_e d_e & & \\
D_{i_0}(i_1) &= 1 \\
d_e + D_u(i_s) - D_u(j_t) &\geq 0 \\
d_e + D_u(j_t) - D_u(i_s) &\geq 0 \\
D_{i_0}(j_0) &= D_{i_1}(j_1) \\
D_{i_0}(j_1) &= D_{i_1}(j_0) \\
D_u(i_s) &\geq 0 \\
d_e &\geq 0
\end{aligned}$$

$$\min_{d_e, D(.)} \sum_{e \in \mathcal{E}_H} w_e d_e$$

$$\begin{aligned}
d_e + D(i_s) - D(j_t) &\geq 0 \\
d_e + D(j_t) - D(i_s) &\geq 0 \\
D(0_0) &= 0 \\
D(i_s) &\in [0, 1] \\
d_e &\in [0, 1] \\
D(i_0) + D(i_1) &= 1
\end{aligned}$$

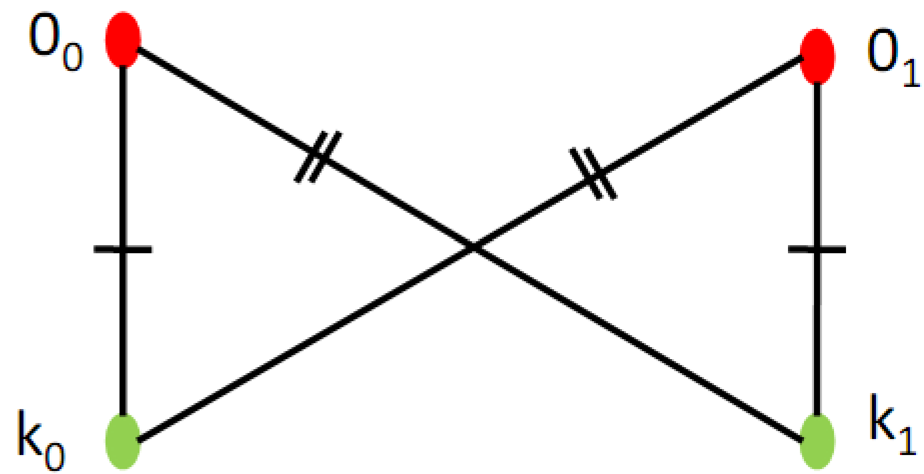
Bipartite Multi-cut LP

$$
\begin{array}{rcl}
\min \sum_{e \in \mathcal{E}_H} w_e d_e & & \\
D_{i_0}(i_1) & = & 1 \\
d_e + D_u(i_s) - D_u(j_t) & \geq & 0 \\
d_e + D_u(j_t) - D_u(i_s) & \geq & 0 \\
D_{i_0}(j_0) & = & D_{i_1}(j_1) \\
D_{i_0}(j_1) & = & D_{i_1}(j_0) \\
D_u(i_s) & \geq & 0 \\
d_e & \geq & 0
\end{array}
$$



Bipartite Multi-cut LP

$$
\begin{aligned}
\min \sum_{e \in \mathcal{E}_H} w_e d_e & & \\
D_{i_0}(i_1) &= 1 \\
d_e + D_u(i_s) - D_u(j_t) &\geq 0 \\
d_e + D_u(j_t) - D_u(i_s) &\geq 0 \\
D_{i_0}(j_0) &= D_{i_1}(j_1) \\
D_{i_0}(j_1) &= D_{i_1}(j_0) \\
D_u(i_s) &\geq 0 \\
d_e &\geq 0
\end{aligned}
$$



Bipartite Multi-cut LP

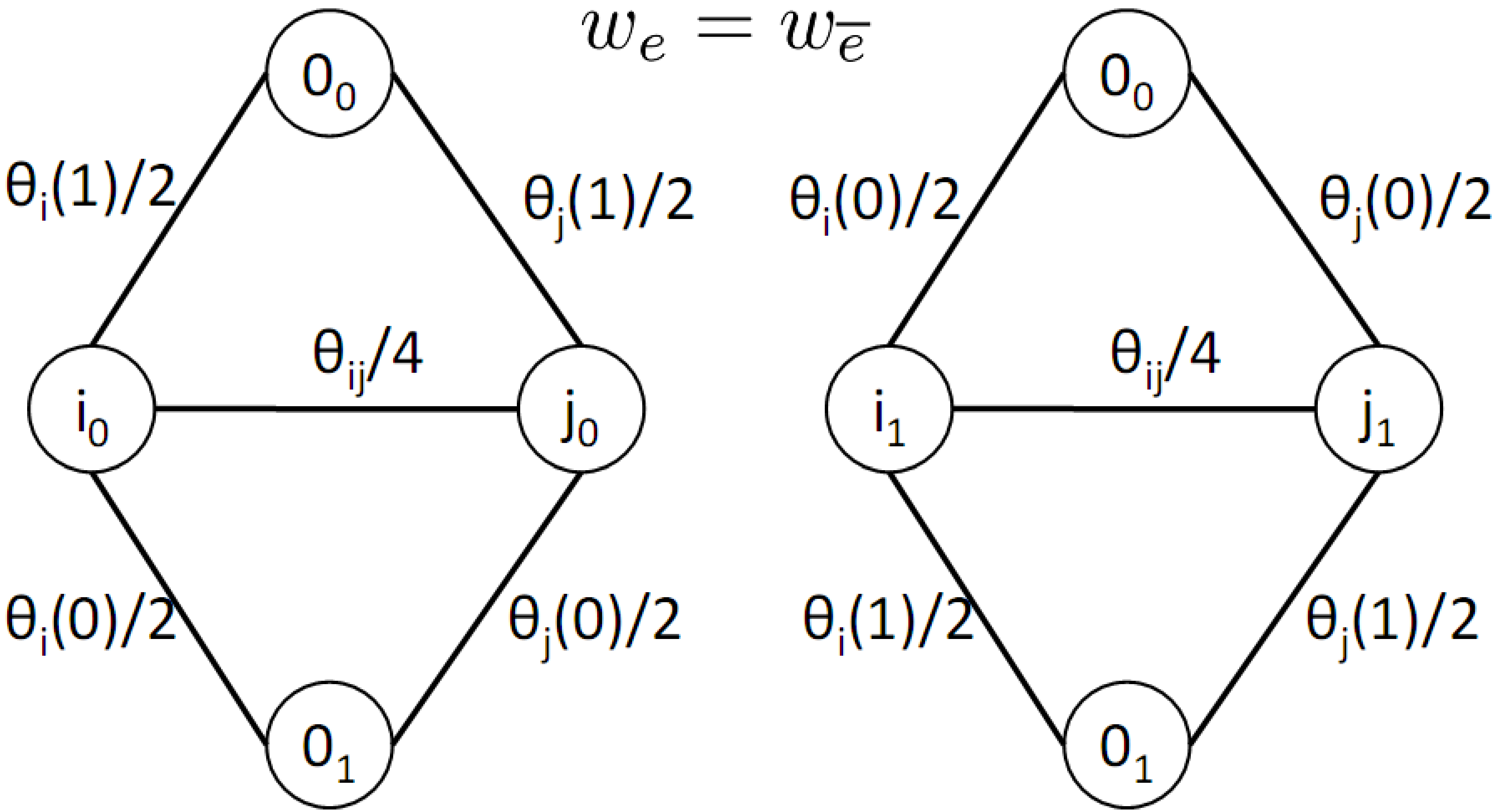
- Gives O(log k) performance where k = number of variables adjacent to non-submodular edges
- Constraints expand exponentially
  - Becomes infeasible for large problems
  - Does not scale beyond 30x30 grid and 50 node clique graphs after implementing a greedy heuristic to minimize the number of non-submodular edges
- Needs a combinatorial algorithm that exploits symmetry to be efficient

Modification to Graph H

- Construct an alternate graph H such that the weights $w_e = w_{\bar{e}}$ for $e = (i_s, j_t)$ and its complementary edges $\bar{e} = (i_{\bar{s}}, j_{\bar{t}})$
- This defines a symmetric graph such that for every path P the complementary path $\bar{P}$ is the path obtained by reversing the order of the edges in P.
- Increases number of vertices by 2n+2

Modification to Graph H

$$w_e = w_{\overline{e}}$$



$$\theta_{ij} = \theta_{ij}(0,1) + \theta_{ij}(1,0) - \theta_{ij}(0,0) - \theta_{ij}(1,1)$$

Garg's Algorithm

- Garg's algorithm provides an estimated solution to min-cut given an error bound ε by simultaneously solving the primal and dual LP relaxations

Multi-cut LP: Primal

$$\min_d \sum_{e \in \mathcal{E}_H} w_e d_e$$

$$\sum_{e \in P} d_e \geq 1 \quad \forall P \in \mathcal{P}$$

$$d_e \geq 0 \quad \forall e \in \mathcal{E}_H$$

Multi-cut LP: Dual

$$\max_f \sum_{P \in \mathcal{P}} f_P$$

$$\sum_{P \in \mathcal{P}_e} f_P \leq w_e \quad \forall e \in E_H$$

$$f_P \geq 0 \quad \forall P \in \mathcal{P}$$

BMC-Sym LP

- The BMC LP as defined on a symmetric graph (where $w_e = w_{\bar{e}}$)

$$\min \sum_{e \in \mathcal{E}_H} w_e d_e$$

$$\sum_{e \in P} d_e \geq 1 \quad \forall P \in \mathcal{P}$$

$$d_e \geq 0, \;\; d_e = d_{\overline{e}} \quad \forall e \in \mathcal{E}_H$$

Combinatorial Algorithm

- **Theorem:** When H is symmetric, the BMC-Sym LP, BMC LP, and Multi-cut LP are equivalent.
  - Therefore, modify Garg's algorithm so that it can exploit the symmetry in the graph using the $d_e = d_{\bar{e}}$ condition from BMC-Sym and use this as a solution to BMC

Proof summary

Any feasible solution to one LP is a feasible solution to another without changing its objective value

Algorithm summary

Primal Step

- Find shortest path P between $(i_0, i_1)\ \forall (i_0, i_1) \in ST$
- Update $d_e = d_e(1 + \frac{\epsilon f_P}{w_e})\ \forall e \in P$
- Update complementary path

Dual Step

- Let $f_P = \min w_e$
- Update the flow in the path by f_P
- Update flow in complementary path

Algorithm summary

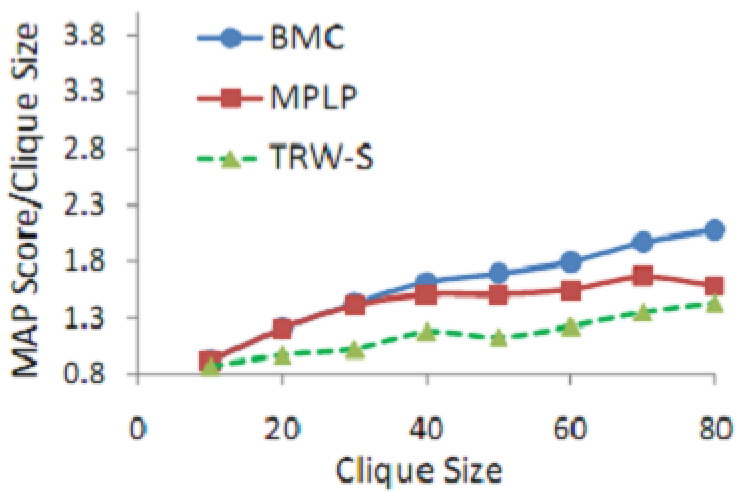
Input: Graphical model G with reparameterized energy function E, approximation guarantee ϵ
Create symmetric graph H from G and E
Initialize $d_e = \delta$ (δ derived from ϵ as shown in Section 3.1), and $f = 0$, $f_e = 0$,
x=arbitrary initial labeling of graphical model G.
Define: Primal objective $P(\{d_e\}) = \sum_e w_e d_e / \min_{P \in \mathcal{P}} \sum_{e \in P} d_e$
Define: Dual objective $D(f, \{f_e\}) = f / (\max_e f_e / w_e)$
while $\min\left(E(\mathbf{x}) - \theta_{const}, P(\{d_e\})\right) > (1+\epsilon) D(f, \{f_e\})$ **do**
  P = Shortest path between (i_0, i_1) $\forall (i_0, i_1) \in$ ST
  if $(\sum_{e \in P} d_e < 1)$ **then**
    With $f_P = \min_{e \in P} w_e$ update $f = f + f_P,\ f_e = f_e + f_P,\ d_e = d_e(1 + \frac{\epsilon f_P}{w_e})\ \forall e \in P$.
    Repeat above for the complement path $\overline{P}$
    $\mathbf{x}' =$ current solution after rounding, $\mathbf{x} =$better of $\mathbf{x}$ and $\mathbf{x}'$
  end if
end while
Return bound = $D(f, \{f_e\}) + \theta_{const}$, MAP = **x**.

- Converges in $O(\varepsilon^{-2} k m^2)$
- Improvement gives $O(\varepsilon^{-2} m^2)$ (Fleischer 1999)

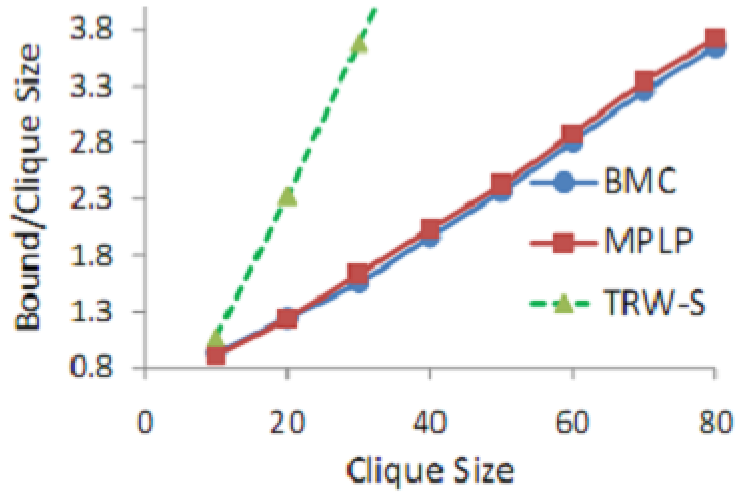
Results

- Compared TRW-S (baseline), MPLP (1000 iterations), BMC ($\varepsilon = 0.02$)
- Data sets:
  - Clique graph based binary MRFs
  - Maxcut instances from BiqMac library

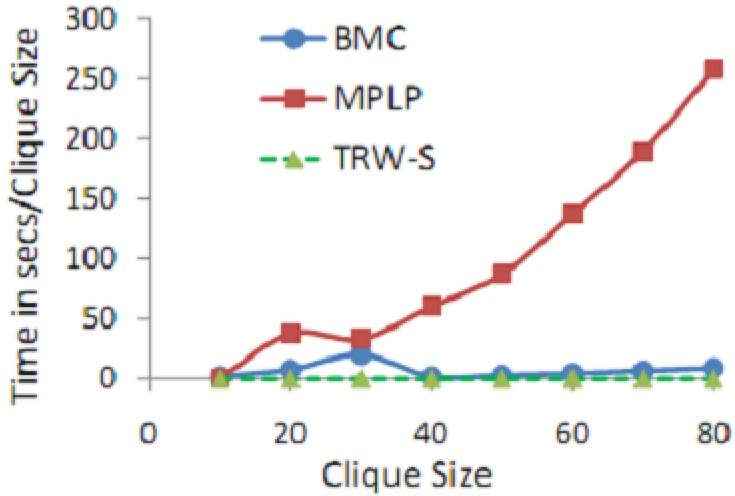
Results



MAP Score/Clique Size
3.8
3.3
2.8
2.3
1.8
1.3
0.8
0
20
40
60
80
Clique Size
BMC
MPLP
TRW-S

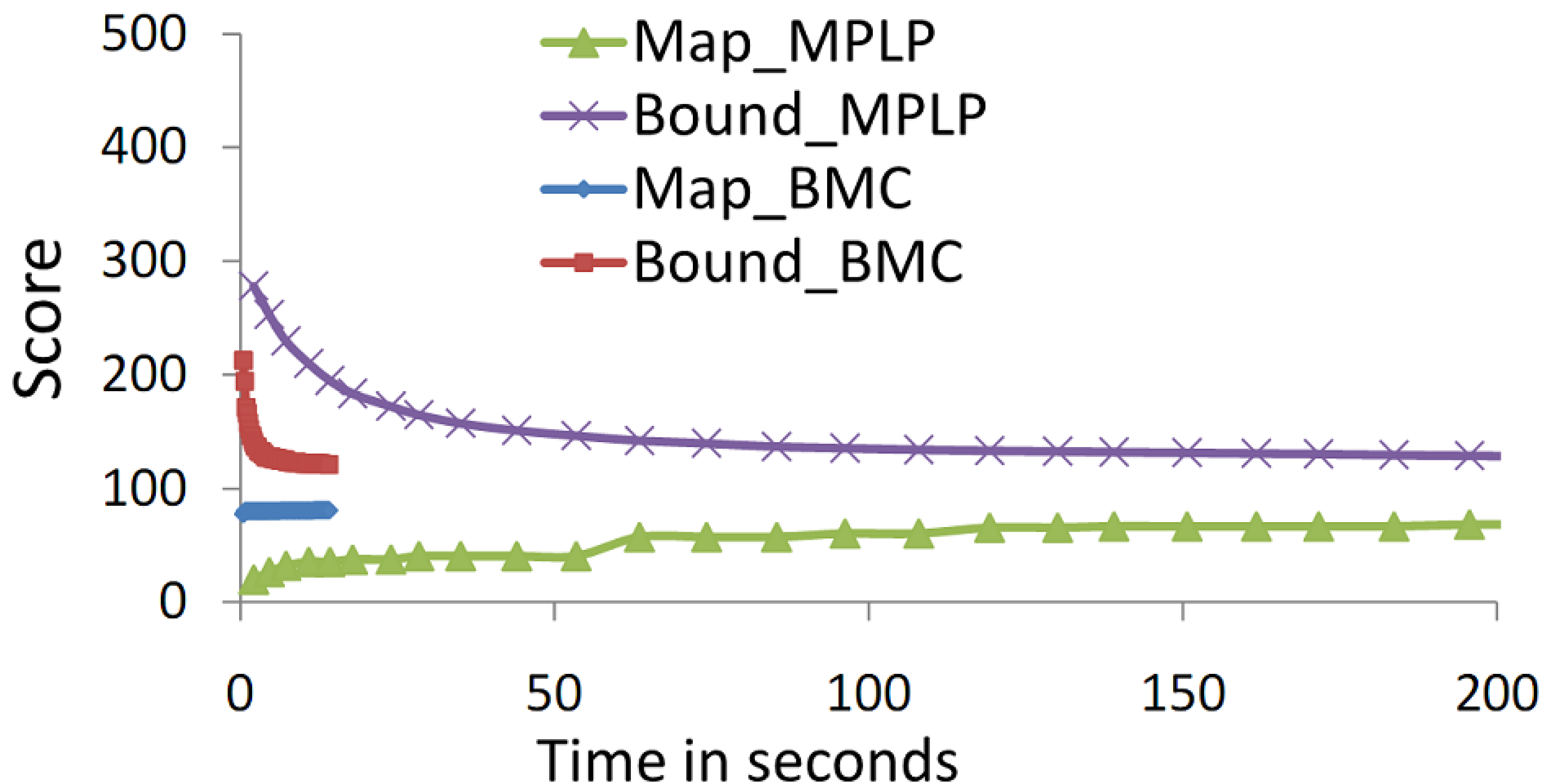


Bound/Clique Size
3.8
3.3
2.8
2.3
1.8
1.3
0.8
0
20
40
60
80
Clique Size
BMC
MPLP
TRW-S



Time in secs/Clique Size
300
250
200
150
100
50
0
0
20
40
60
80
Clique Size
BMC
MPLP
TRW-S

Results



Map_MPLP
Bound_MPLP
Map_BMC
Bound_BMC
Score
500
400
300
200
100
0
0
50
100
150
200
Time in seconds

Results

		Bound			Time in seconds		
Graph	density	BMC	MPLP	TRW-S	BMC	MPLP	TRW-S
pm1s	0.1	131	200	257	45	43	0.005
pw01	0.1	2079	2397	2745	48	46	0.006
w01	0.1	720	1115	1320	46	41	0.004
g05	0.5	1650	1720	2475	761	317	0.021
pw05	0.5	9131	9195	13696	699	1139	0.021
w05	0.5	2245	2488	6588	737	1261	0.021
pw09	0.9	16493	16404	24563	106	2524	0.041
w09	0.9	4073	4095	11763	123	2671	0.053
pm1d	0.99	842	924	2463	12	1307	0.047