Iterative Join Graph Propagation

Vibhav Gogate

Stat methods class

Adapted from Robert Mateescu's slides

The University of Texas at Dallas

What is IJGP?

- IJGP is an approximate algorithm for belief updating in Bayesian networks
- IJGP is a version of join-tree clustering which is both *anytime* and *iterative*
- IJGP applies message passing along a join-graph, rather than a join-tree
- Empirical evaluation shows that IJGP is almost always superior to other approximate schemes (IBP, MC)

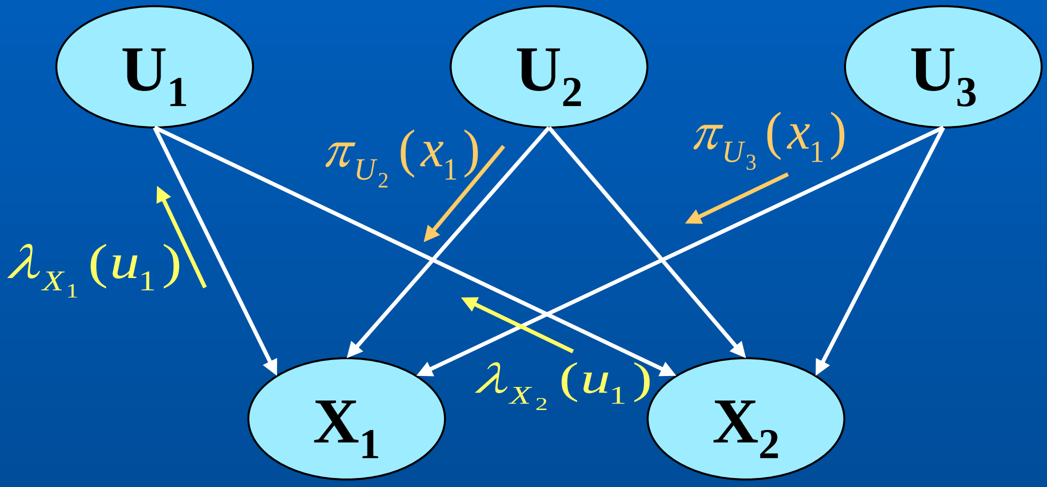
Outline

- IBP - Iterative Belief Propagation
- MC - Mini Clustering
- IJGP - Iterative Join Graph Propagation
- Empirical evaluation
- Conclusion

Iterative Belief Propagation - IBP

- Belief propagation is exact for poly-trees
- IBP - applying BP iteratively to cyclic networks

One step:
update BEL(U_1)



- No guarantees for convergence
- Works well for many coding networks

Mini-Clustering - MC

Cluster Tree Elimination | **Mini-Clustering, i=3**

1 — **A B C**: ***p(a), p(b|a), p(c|a,b)***

$$h_{(1,2)}(b,c) = \sum_{a} p(a) \cdot p(b \mid a) \cdot p(c \mid a,b)$$

BC

$$h^{1}_{(1,2)}(b,c) = \sum_{a} p(a) \cdot p(b \mid a) \cdot p(c \mid a,b)$$

2 — **B C D**: ***p(d|b), h$_{(1,2)}$(b,c)*** | **C D F**: ***p(f|c,d)***

$$h_{(2,3)}(b,f) = \sum_{c,d} p(d \mid b) \cdot h^{1}_{(1,2)}(b,c) \cdot p(f \mid c,d)$$

BF

sep(2,3) = {B,F}

elim(2,3) = {C,D}

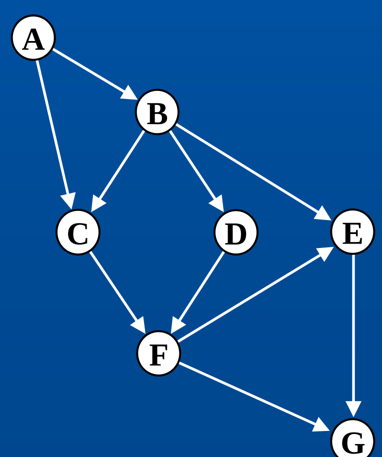
$$h^{1}_{(2,3)}(b) = \sum_{c,d} p(d \mid b) \cdot h^{1}_{(1,2)}(b,c)$$

$$h^{2}_{(2,3)}(f) = \sum_{c,d} p(f \mid c,d)$$

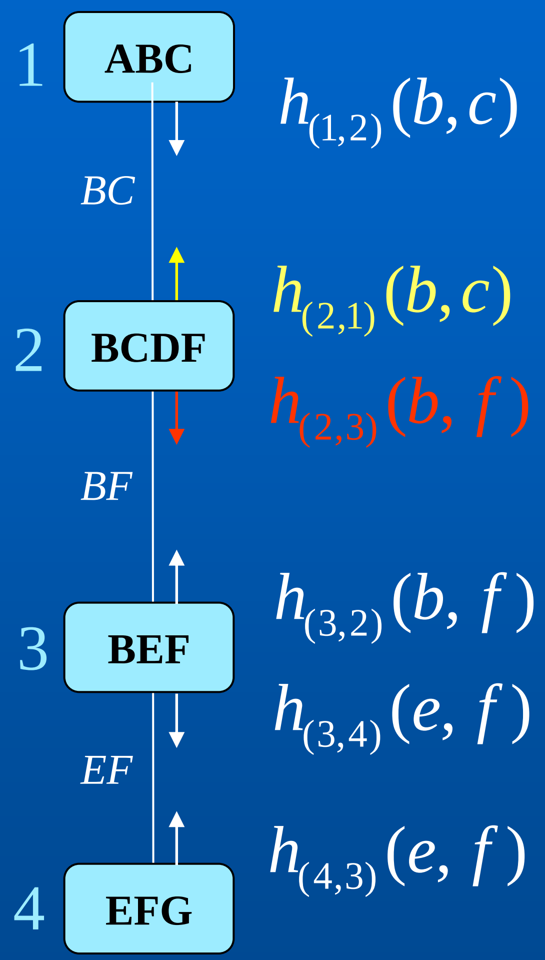
3 — **B E F**: ***p(e|b,f)***

EF

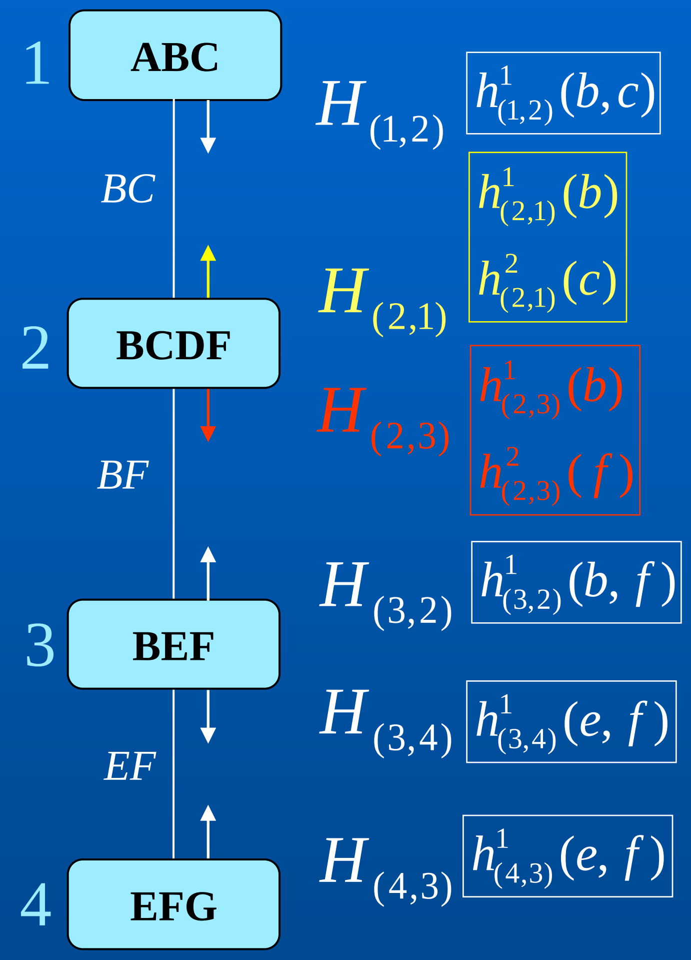
4 — **E F G**: ***p(g|e,f)***



CTE vs. MC



Cluster Tree Elimination



Mini-Clustering, i=3

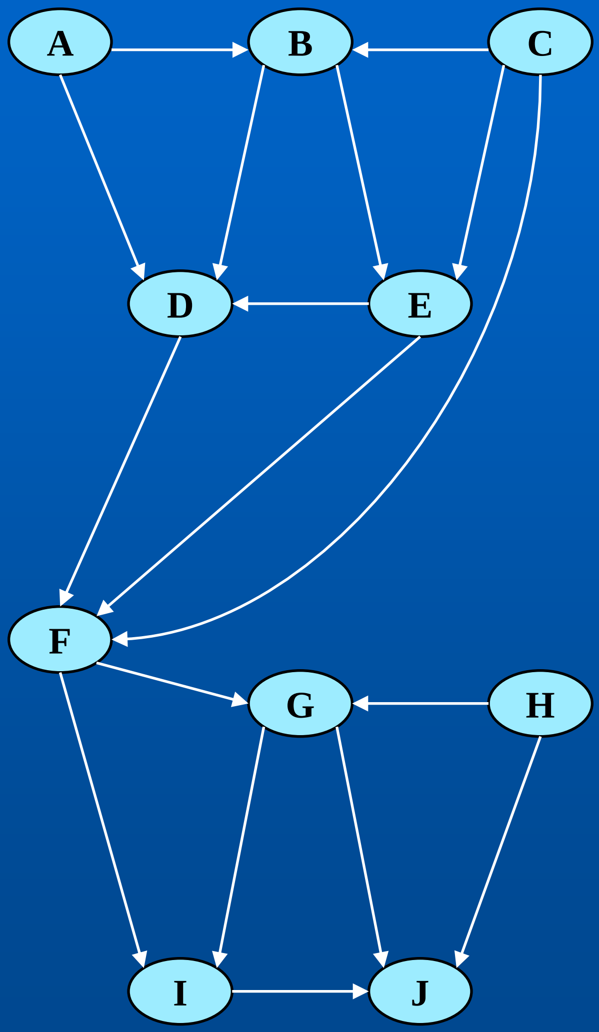
IJGP - Motivation

- IBP is applied to a loopy network iteratively
  - not an anytime algorithm
  - when it converges, it converges very fast
- MC applies bounded inference along a tree decomposition
  - MC is an anytime algorithm controlled by i-bound
- IJGP combines:
  - the iterative feature of IBP
  - the anytime feature of MC

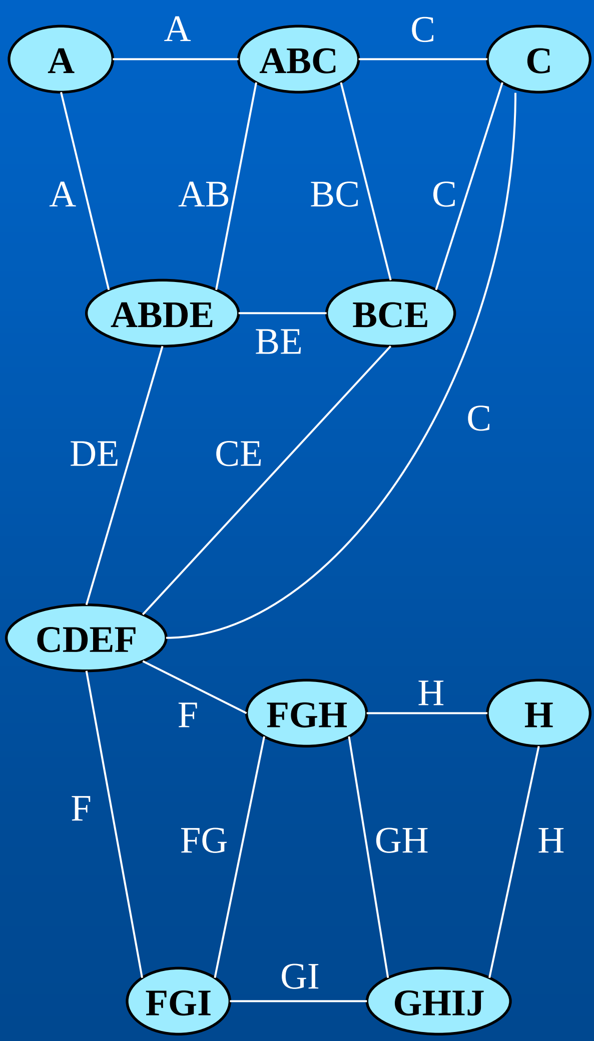
IJGP - The basic idea

- Apply Cluster Tree Elimination to any join-graph
- We commit to graphs that are minimal I-maps
- Avoid cycles as long as I-mapness is not violated
- Result: use *minimal arc-labeled* join-graphs

IJGP - Example

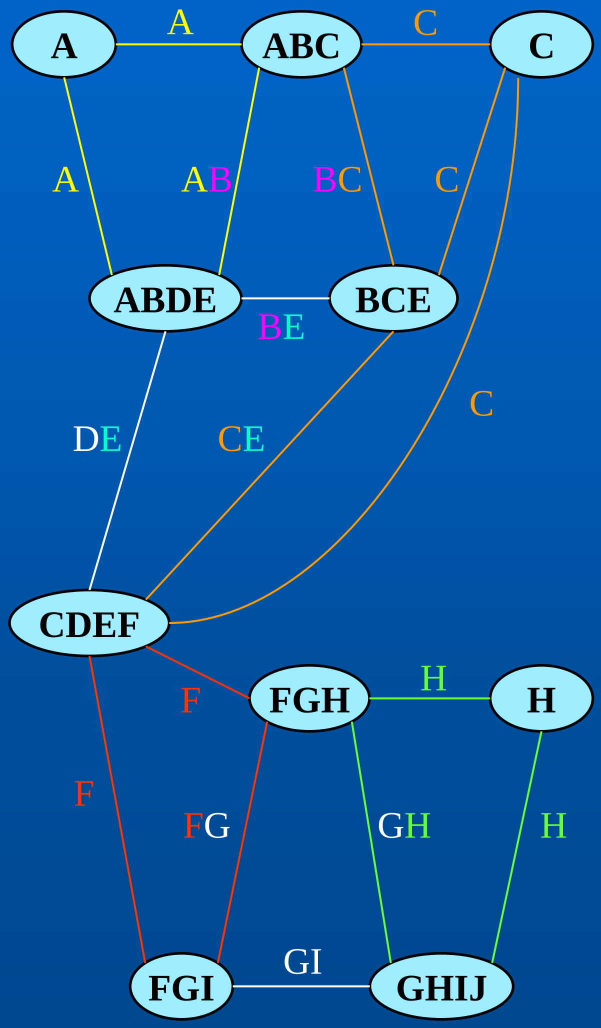


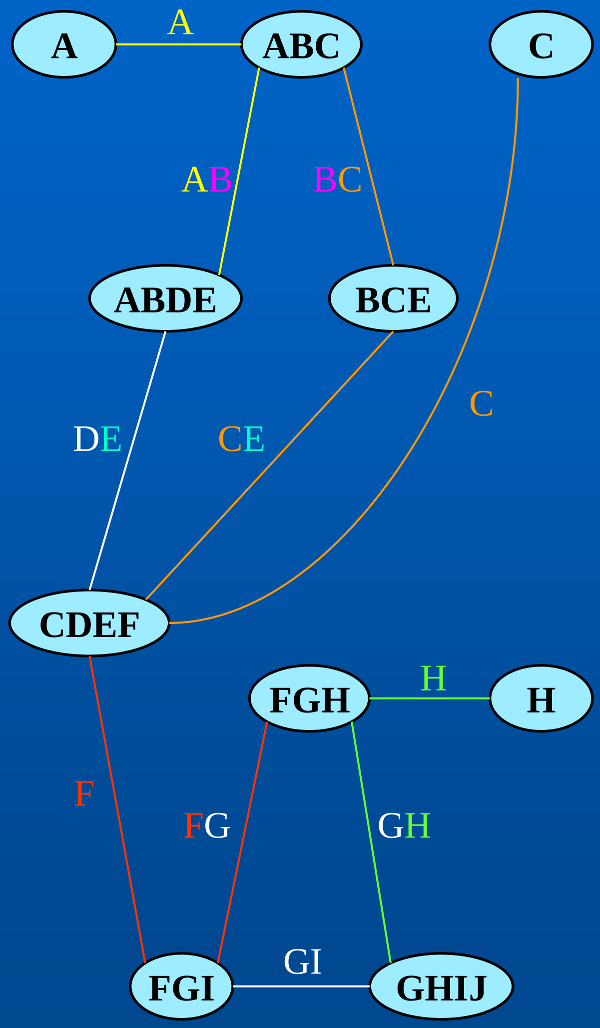
a) Belief network



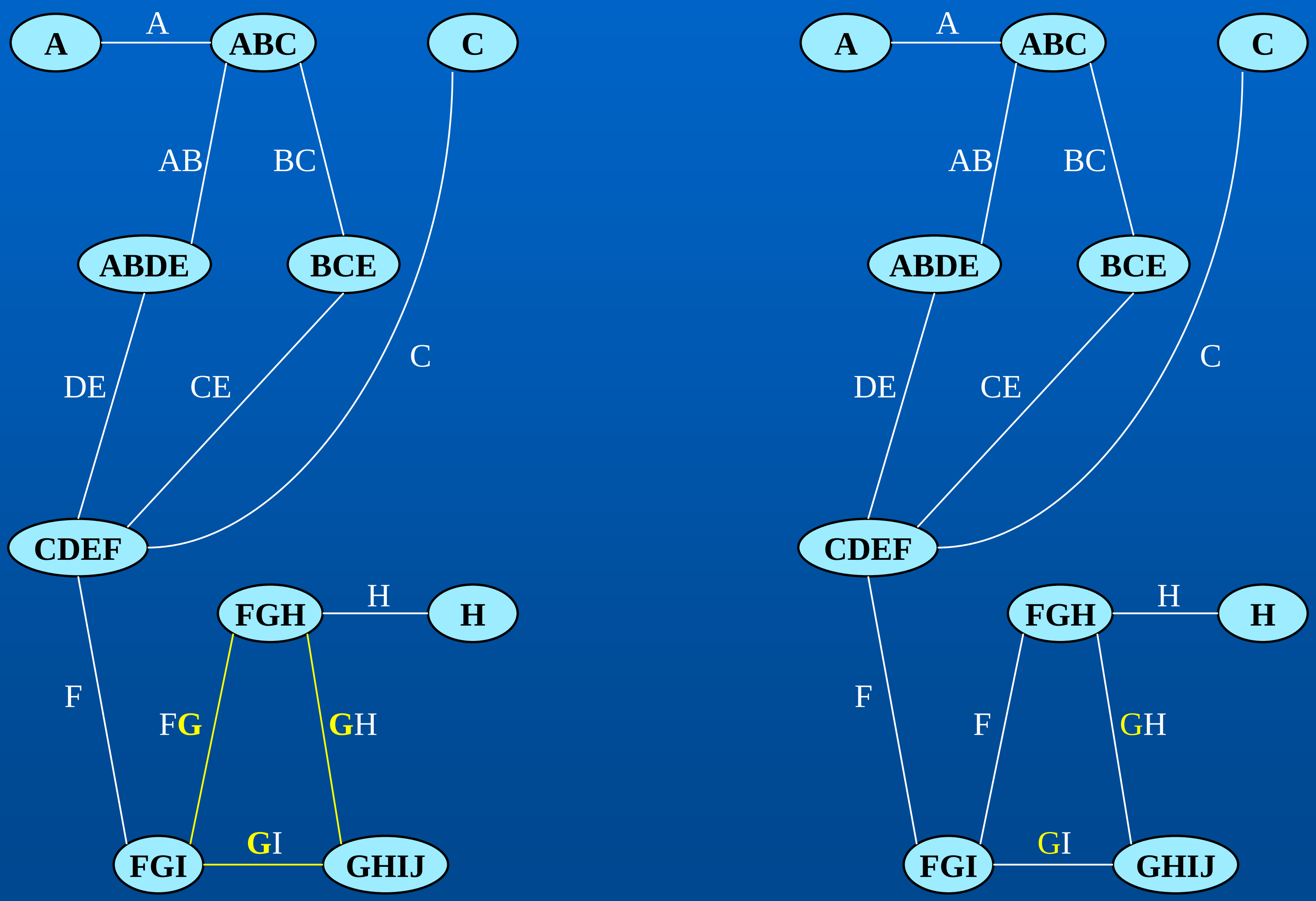
a) The graph IBP works on

Arc-minimal join-graph

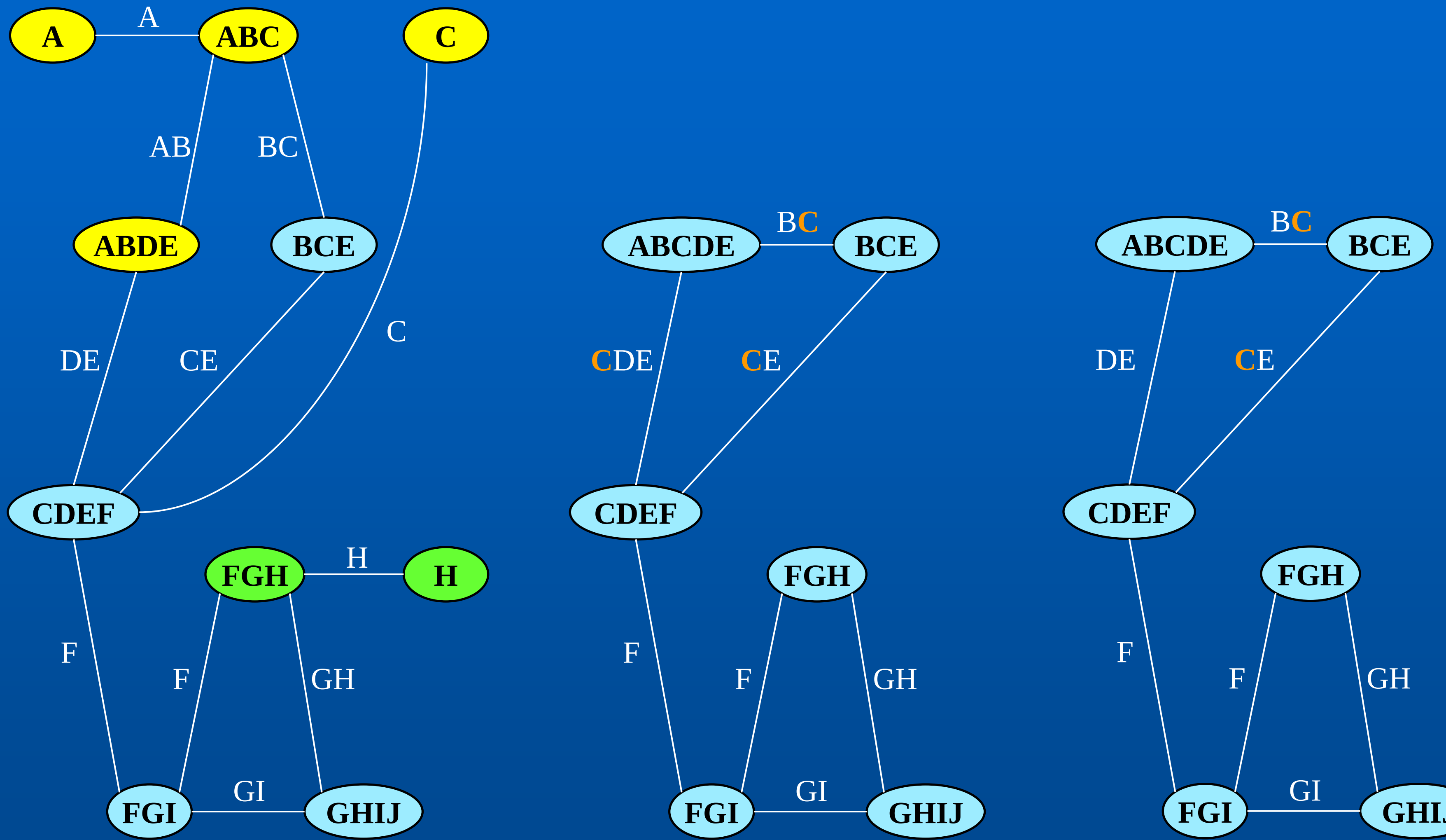




Minimal arc-labeled join-graph



Join-graph decompositions

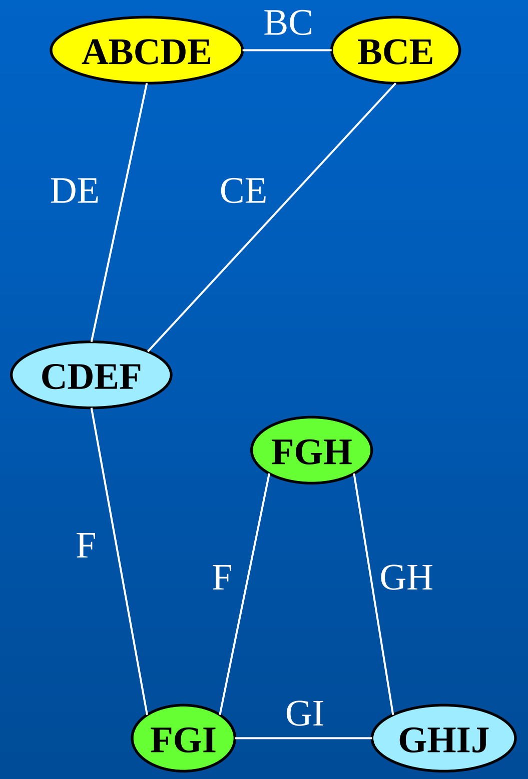


a) Minimal arc-labeled join graph

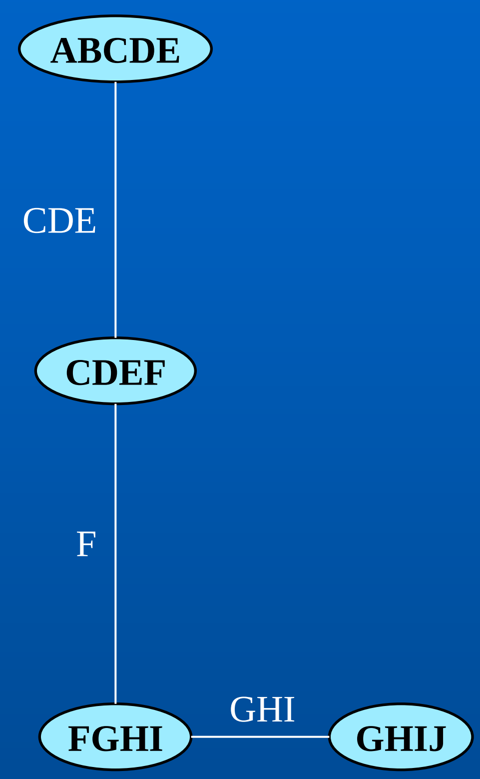
b) Join-graph obtained by collapsing nodes of graph a)

c) Minimal arc-labeled join graph

Tree decomposition

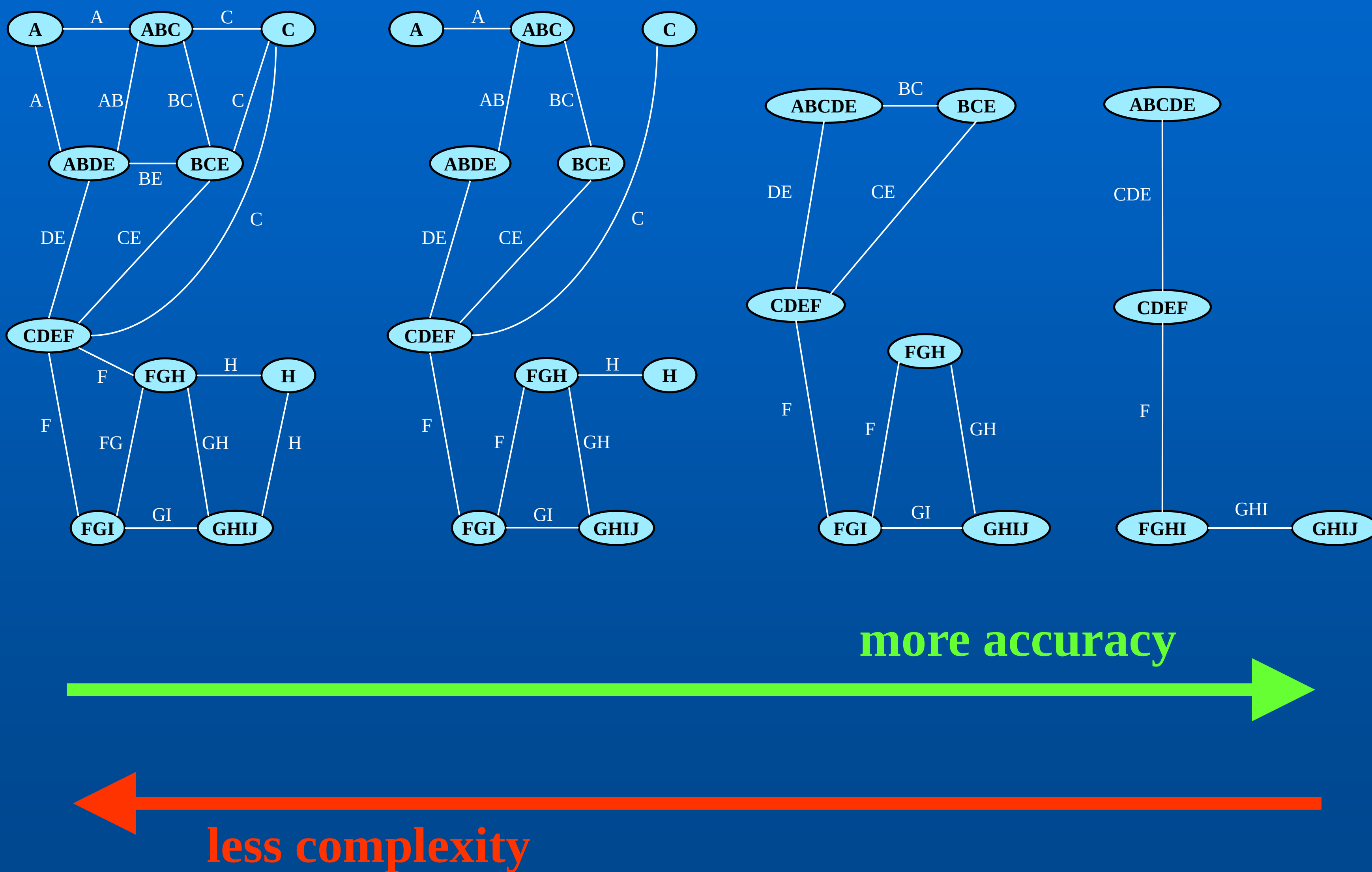


a) Minimal arc-labeled join graph

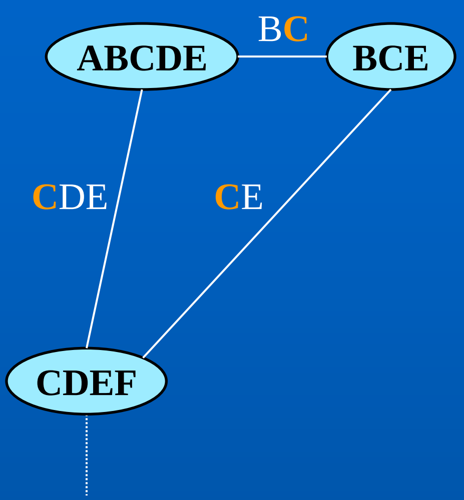


a) Tree decomposition

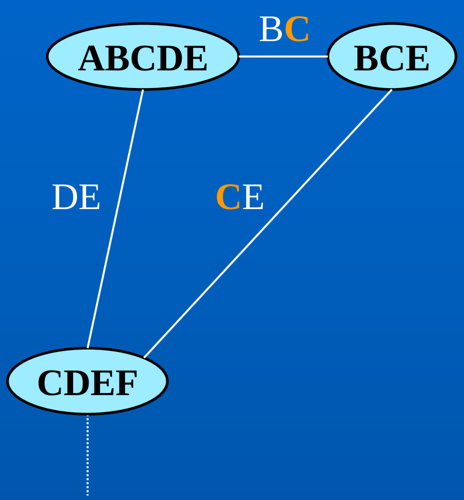
Join-graphs


A
ABC
C
ABDE
BCE
CDEF
FGH
H
FGI
GHIJ
A
AB
BC
C
BE
DE
CE
C
F
F
FG
GH
H
GI
H
A
ABC
C
ABDE
BCE
CDEF
FGH
H
FGI
GHIJ
A
AB
BC
DE
CE
C
F
F
GH
H
GI
ABCDE
BCE
CDEF
FGH
FGI
GHIJ
BC
DE
CE
F
F
GH
GI
ABCDE
CDEF
FGHI
GHIJ
CDE
F
GHI
more accuracy
less complexity

Minimal arc-labeled decomposition



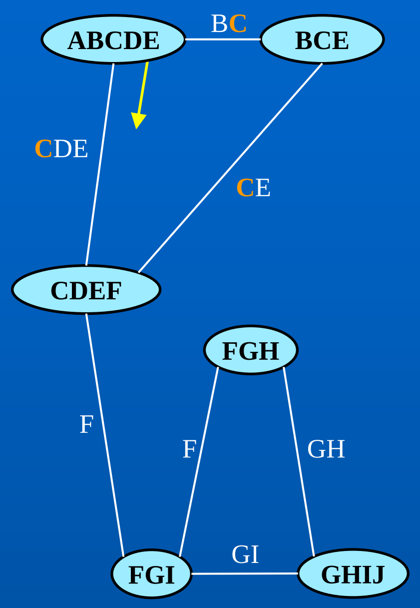
a) Fragment of an arc-labeled join-graph

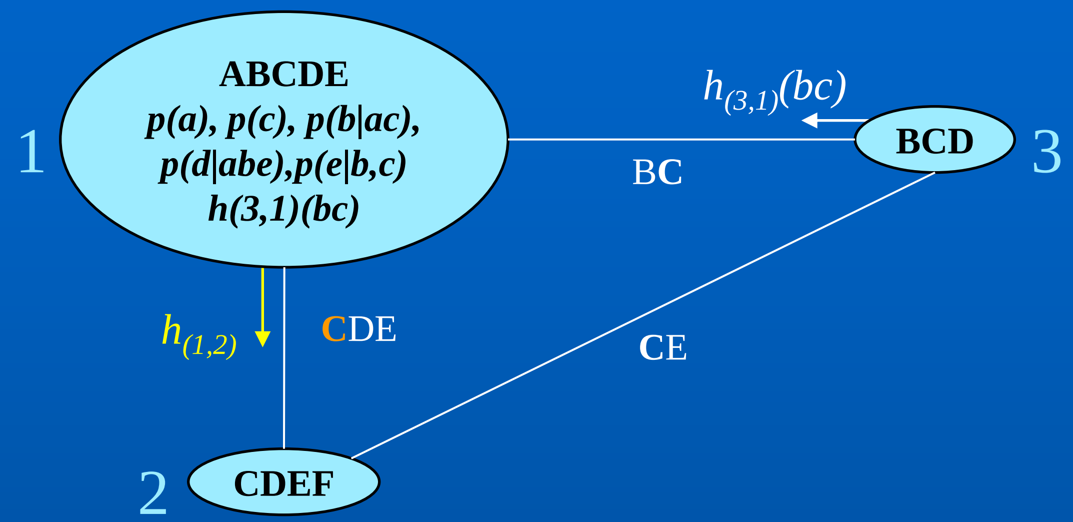


a) Shrinking labels to make it a *minimal* arc-labeled join-graph

- Use a DFS algorithm to eliminate cycles relative to each variable

Message propagation





Minimal arc-labeled:

sep(1,2)={D,E}

elim(1,2)={A,B,C}

$$h_{(1,2)}(de) = \sum_{a,b,c} p(a)\,p(c)\,p(b\,|\,ac)\,p(d\,|\,abe)\,p(e\,|\,bc)h_{(3,1)}(bc)$$

Non-minimal arc-labeled:

sep(1,2)={C,D,E}

elim(1,2)={A,B}

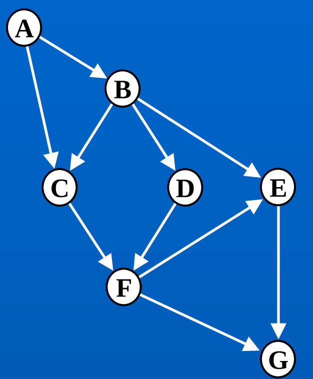
$$h_{(1,2)}(cde) = \sum_{a,b} p(a)\,p(c)\,p(b\,|\,ac)\,p(d\,|\,abe)\,p(e\,|\,bc)h_{(3,1)}(bc)$$

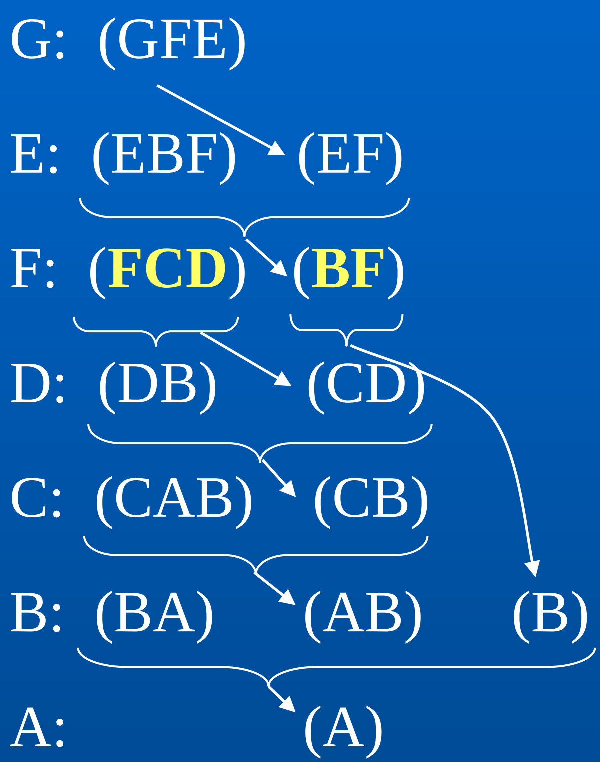
Bounded decompositions

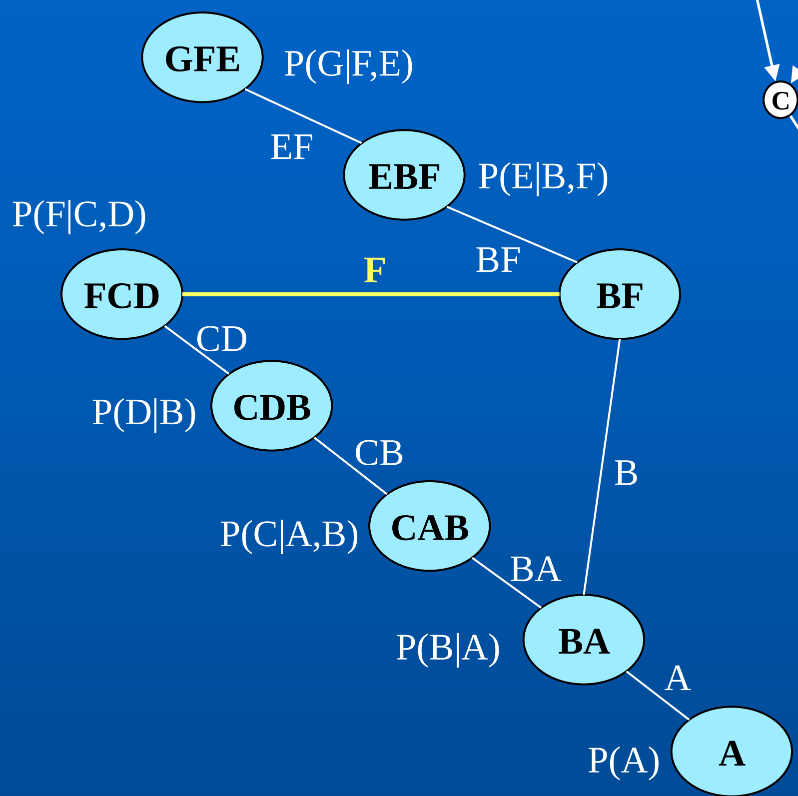
- We want arc-labeled decompositions such that:
  - the cluster size (internal width) is bounded by i (the accuracy parameter)
  - the width of the decomposition as a graph (external width) is as small as possible

- Possible approaches to build decompositions:
  - partition-based algorithms - inspired by the mini-bucket decomposition
  - grouping-based algorithms

Partition-based algorithms







a) schematic mini-bucket(*i*), *i*=3

b) arc-labeled join-graph decomposition

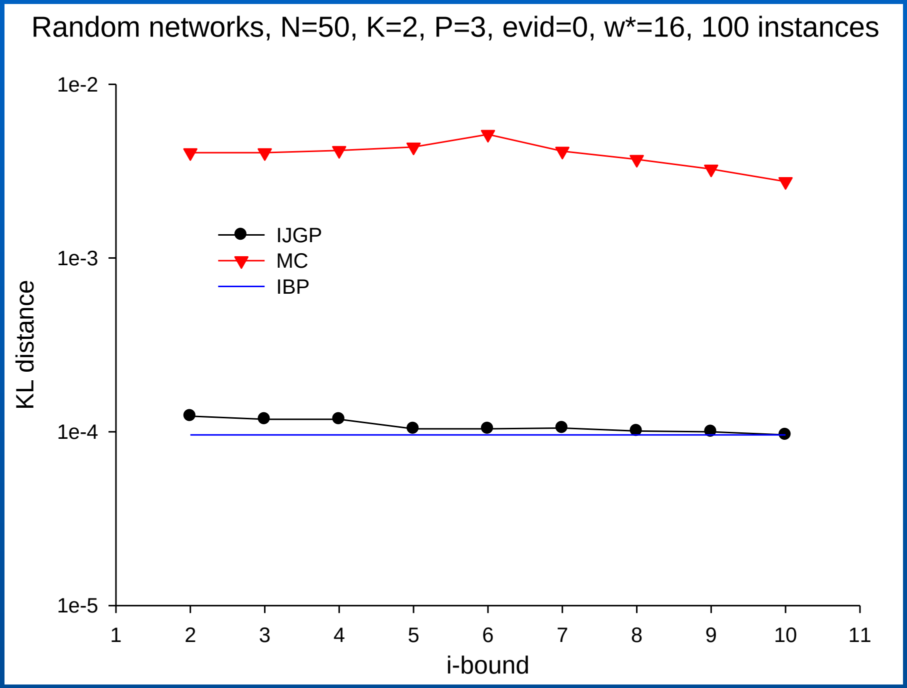
IJGP properties

- IJGP(i) applies BP to min arc-labeled join-graph, whose cluster size is bounded by i
- On join-trees IJGP finds exact beliefs
- IJGP is a Generalized Belief Propagation algorithm (Yedidia, Freeman, Weiss 2001)
- Complexity of one iteration:
  - time: $O(deg \bullet (n+N) \bullet d^{i+1})$
  - space: $O(N \bullet d^{\theta})$

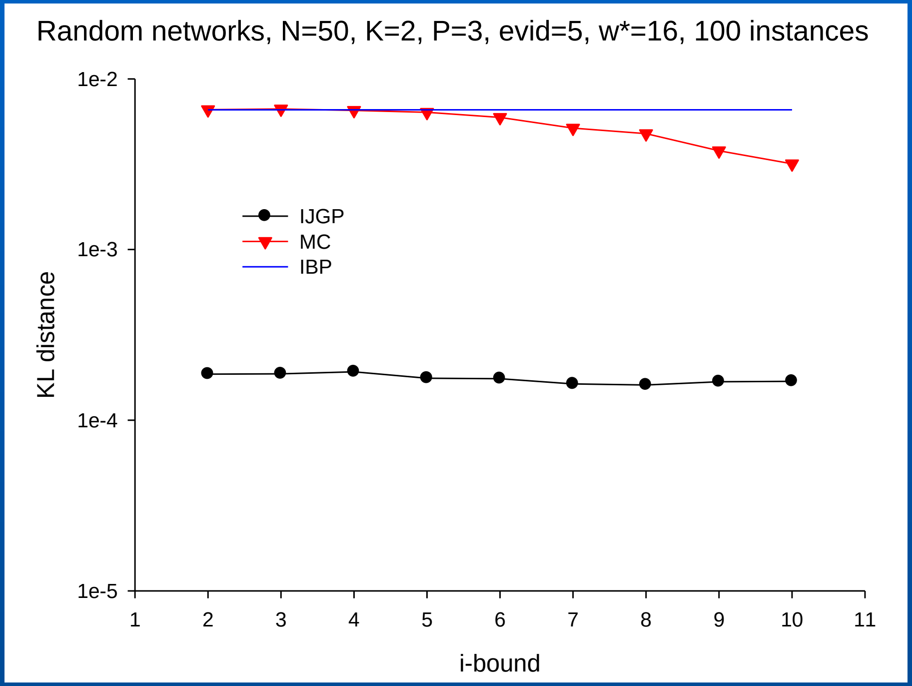
Empirical evaluation

- Algorithms:
  - Exact
  - IBP
  - MC
  - IJGP
- Measures:
  - Absolute error
  - Relative error
  - Kullback-Leibler (KL) distance
  - Bit Error Rate
  - Time
- Networks (all variables are binary):
  - Random networks
  - Grid networks (MxM)
  - CPCS 54, 360, 422
  - Coding networks

Random networks - KL at convergence

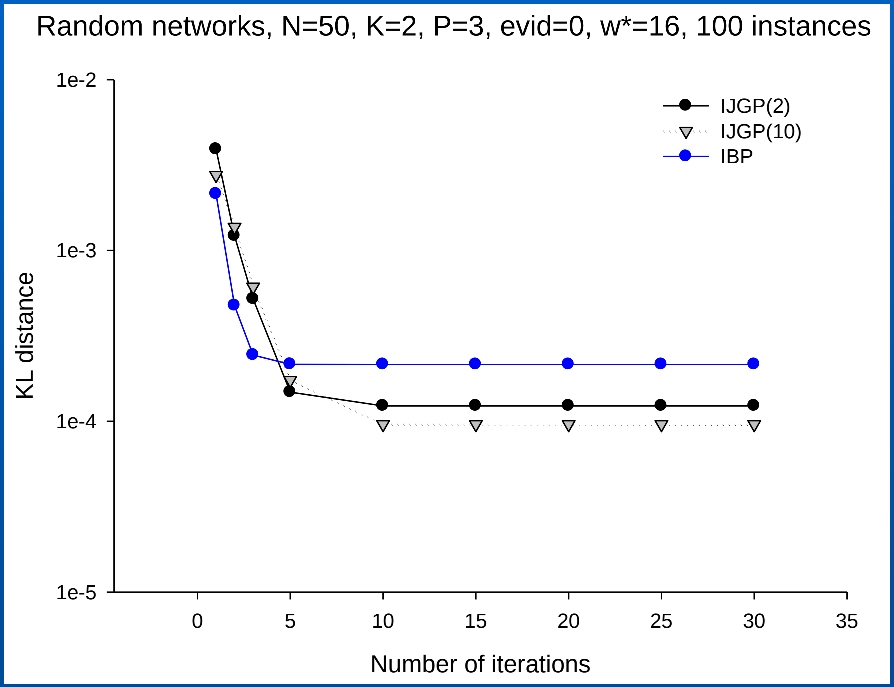


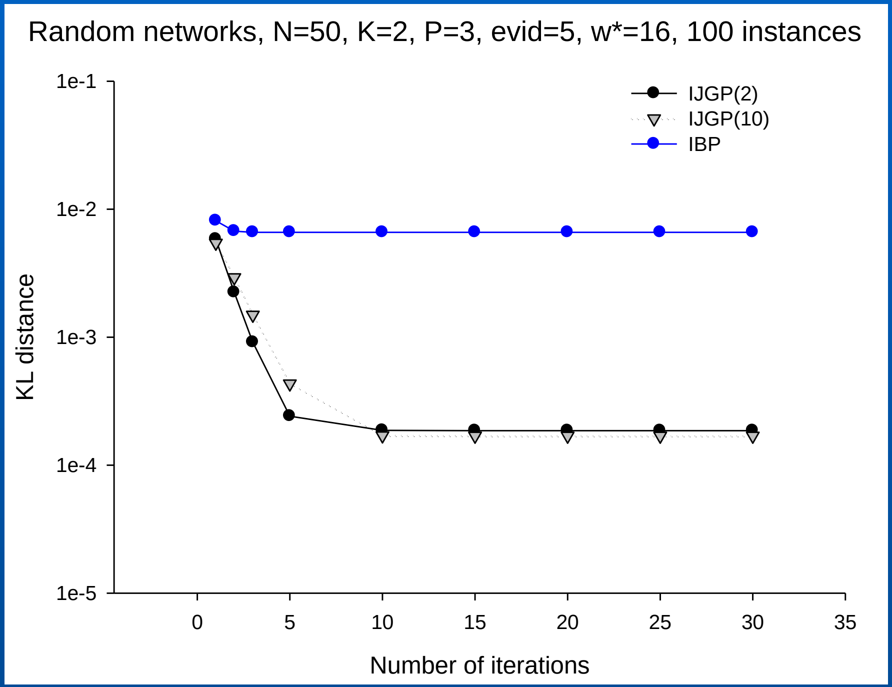
evidence=0



evidence=5

Random networks - KL vs. iterations

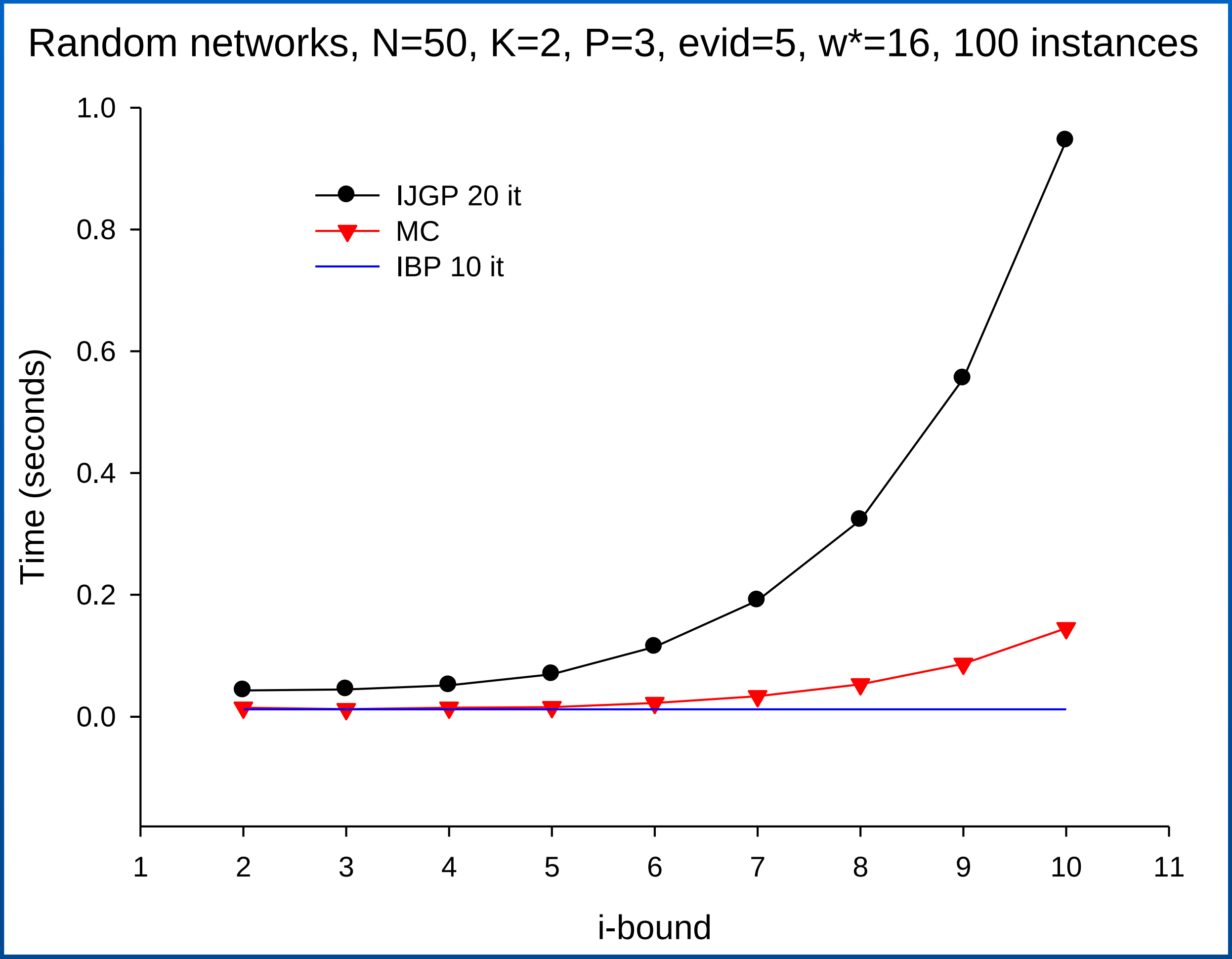




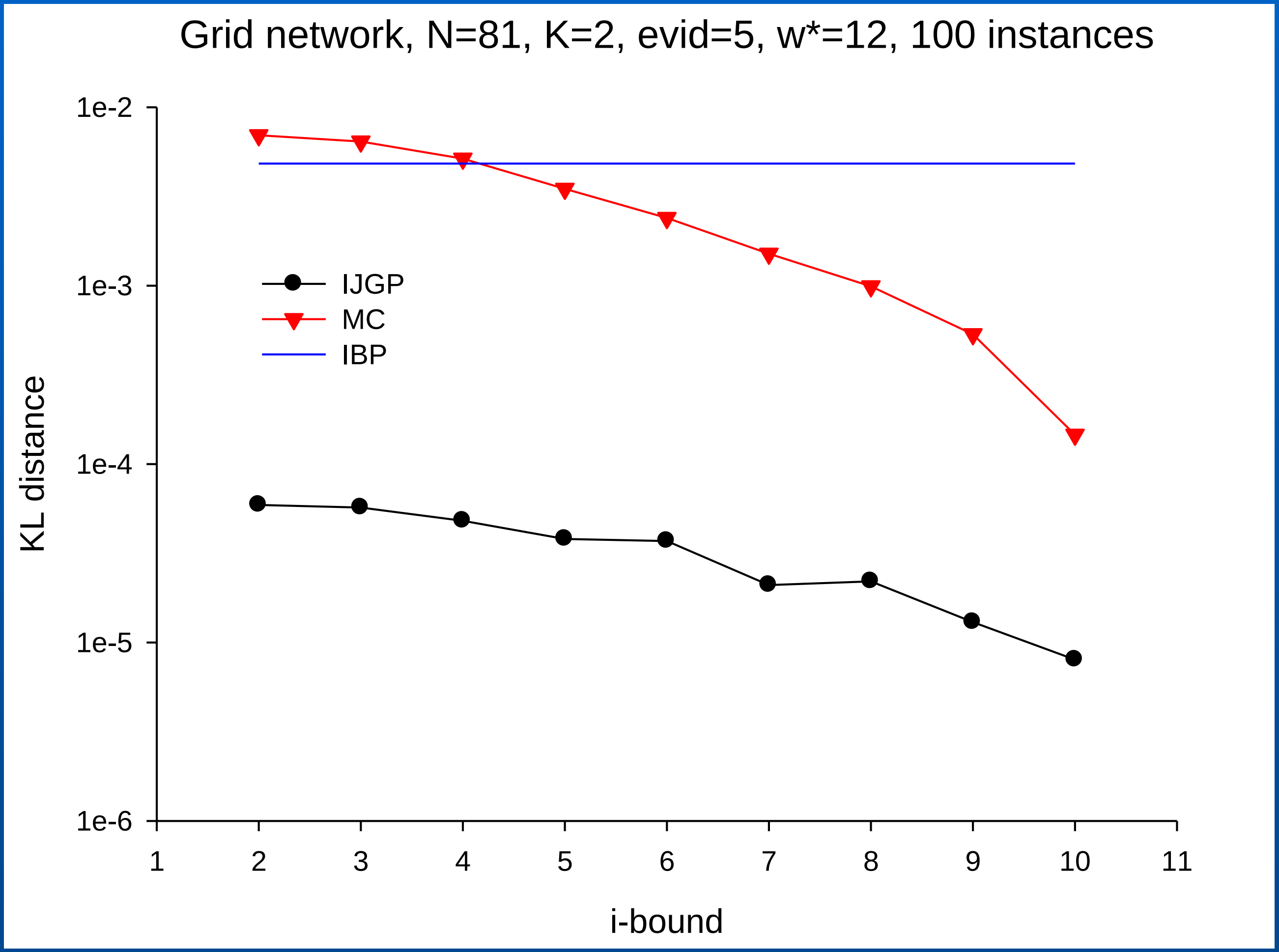
evidence=0

evidence=5

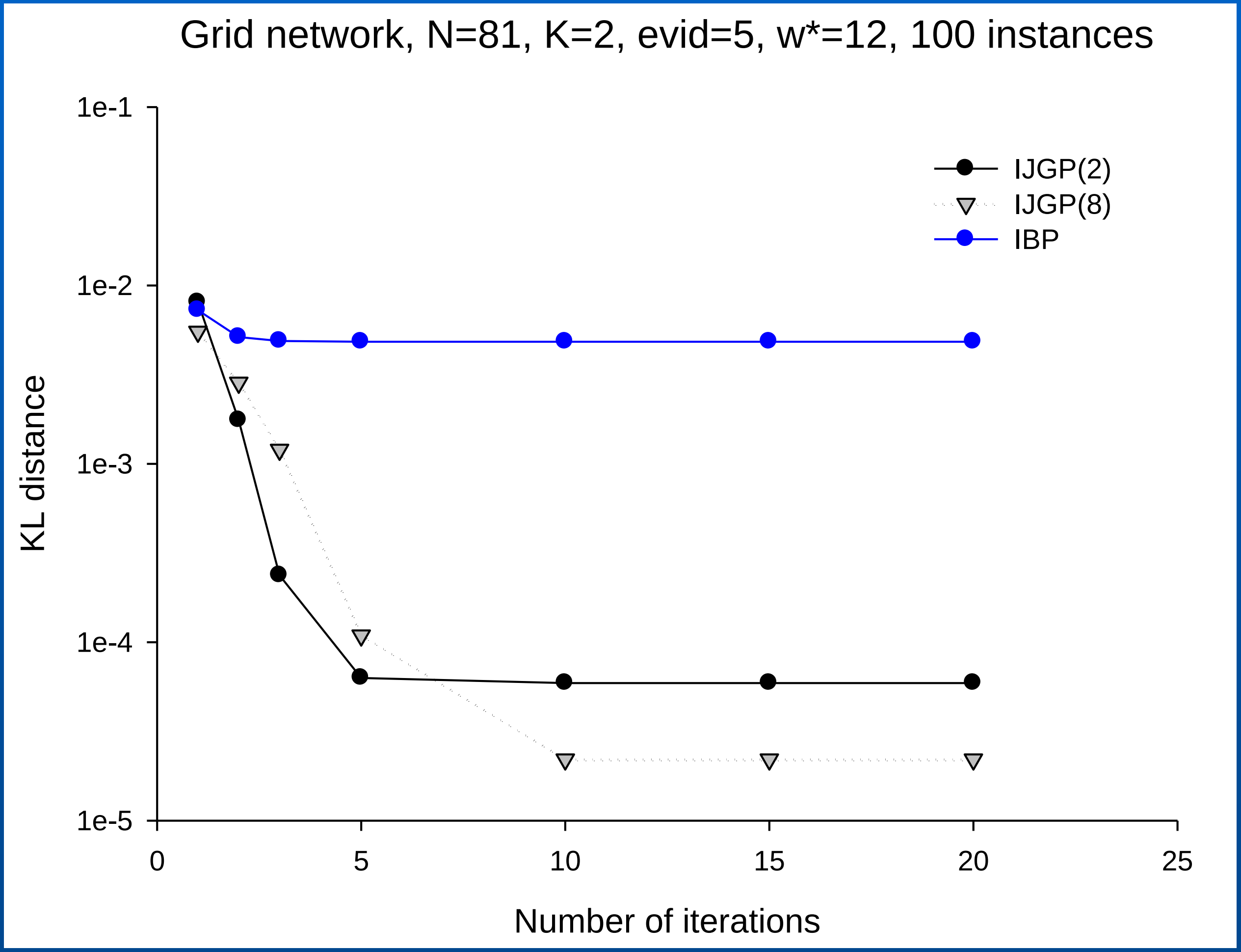
Random networks - Time



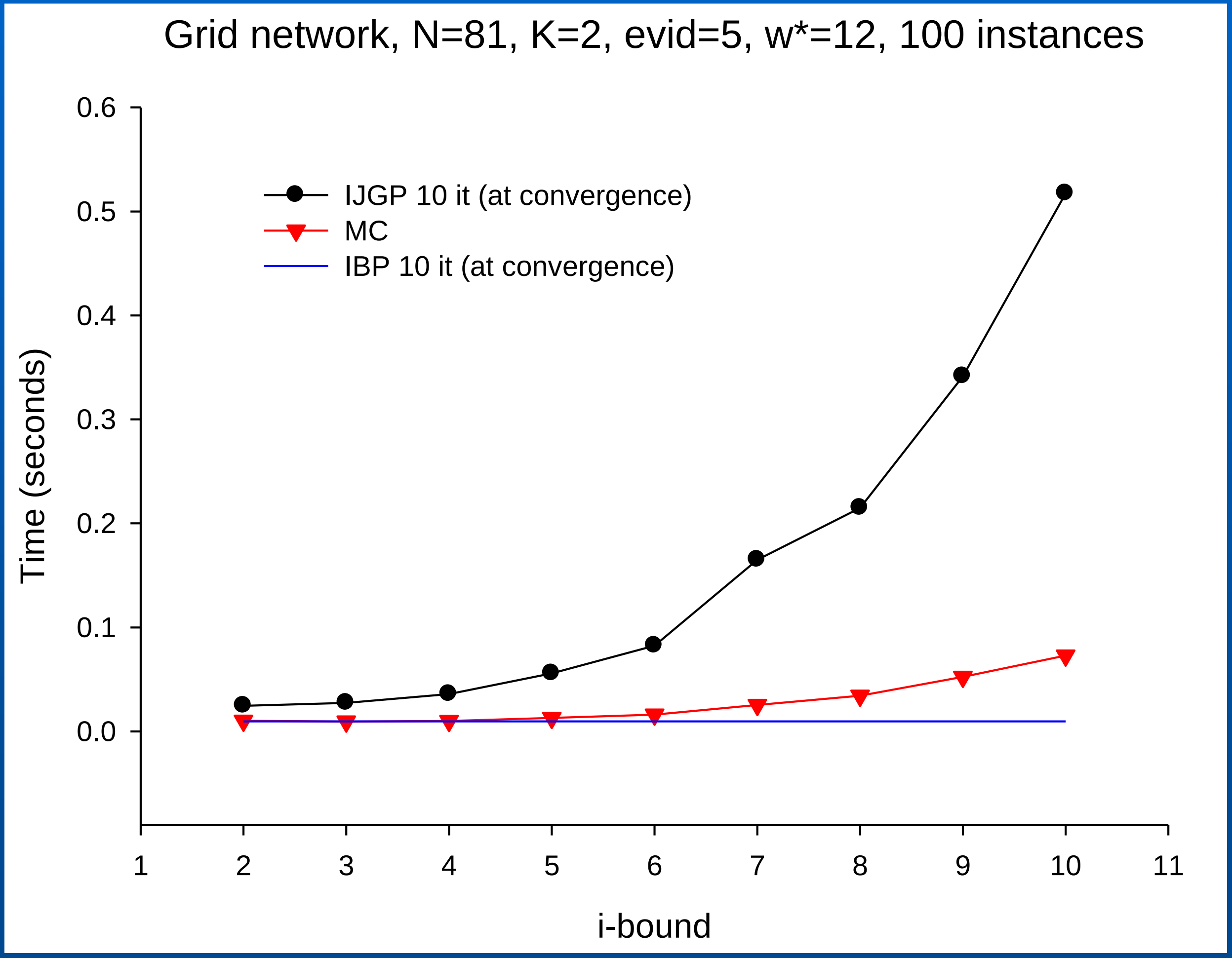
Grid 81 – KL distance at convergence



Grid 81 – KL distance vs. iterations



Grid 81 – Time vs. i-bound



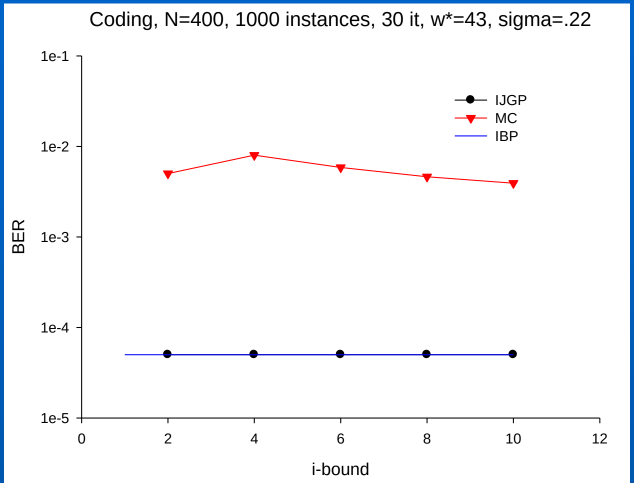
Coding networks

Bit Error Rate

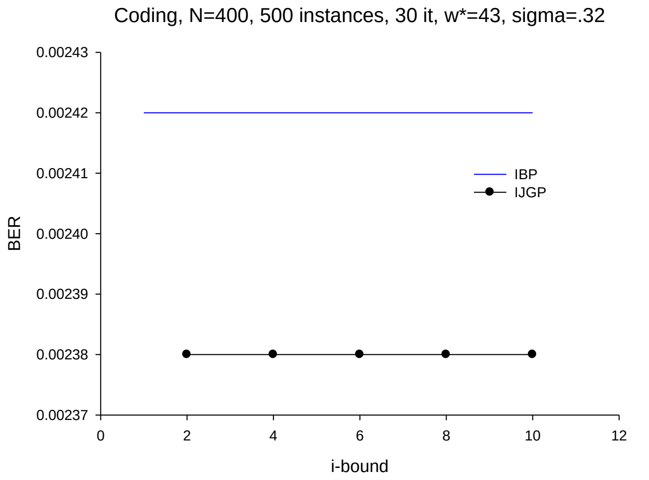
σ		i-bound 2	4	6	8	10	IBP
0.22	IJGP	0.00005	0.00005	0.00005	0.00005	0.00005	0.00005
	MC	0.00501	0.00800	0.00586	0.00462	0.00392	
0.28	IJGP	0.00062	0.00062	0.00062	0.00062	0.00062	0.00064
	MC	0.02170	0.02968	0.02492	0.02048	0.01840	
0.32	IJGP	0.00238	0.00238	0.00238	0.00238	0.00238	0.00242
	MC	0.04018	0.05004	0.04480	0.03878	0.03558	
0.40	IJGP	0.01202	0.01188	0.01194	0.01210	0.01192	0.01220
	MC	0.08726	0.09762	0.09272	0.08766	0.08334	
0.51	IJGP	0.07664	0.07498	0.07524	0.07578	0.07554	0.07816
	MC	0.15396	0.16048	0.15710	0.15452	0.15180	
0.65	IJGP	0.19070	0.19056	0.19016	0.19030	0.19056	0.19142
	MC	0.21890	0.22056	0.21928	0.21904	0.21830	
Time							
	IJGP	0.36262	0.41695	0.86213	2.62307	9.23610	0.019752
	MC	0.25281	0.21816	0.31094	0.74851	2.33257	

N=400, P=4, 500 instances, 30 iterations, w*=43

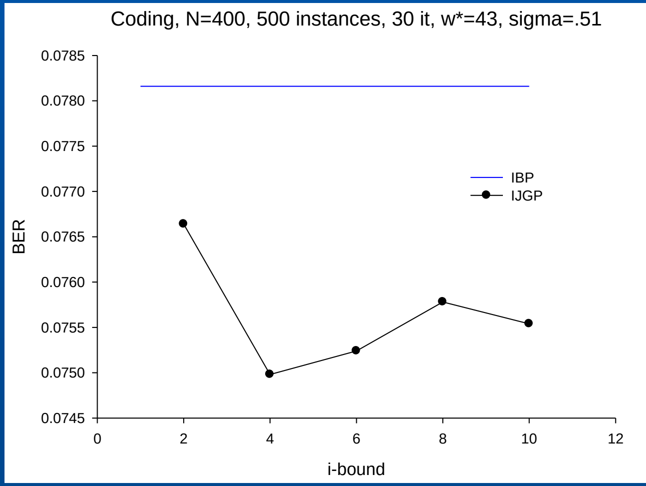
Coding networks - BER



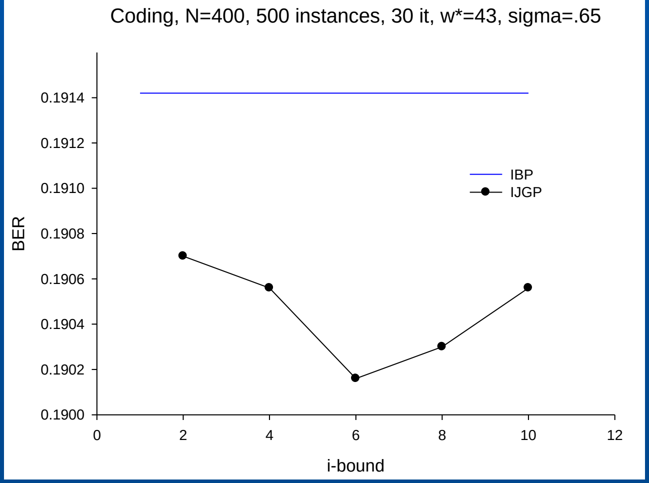
sigma=.22



sigma=.32

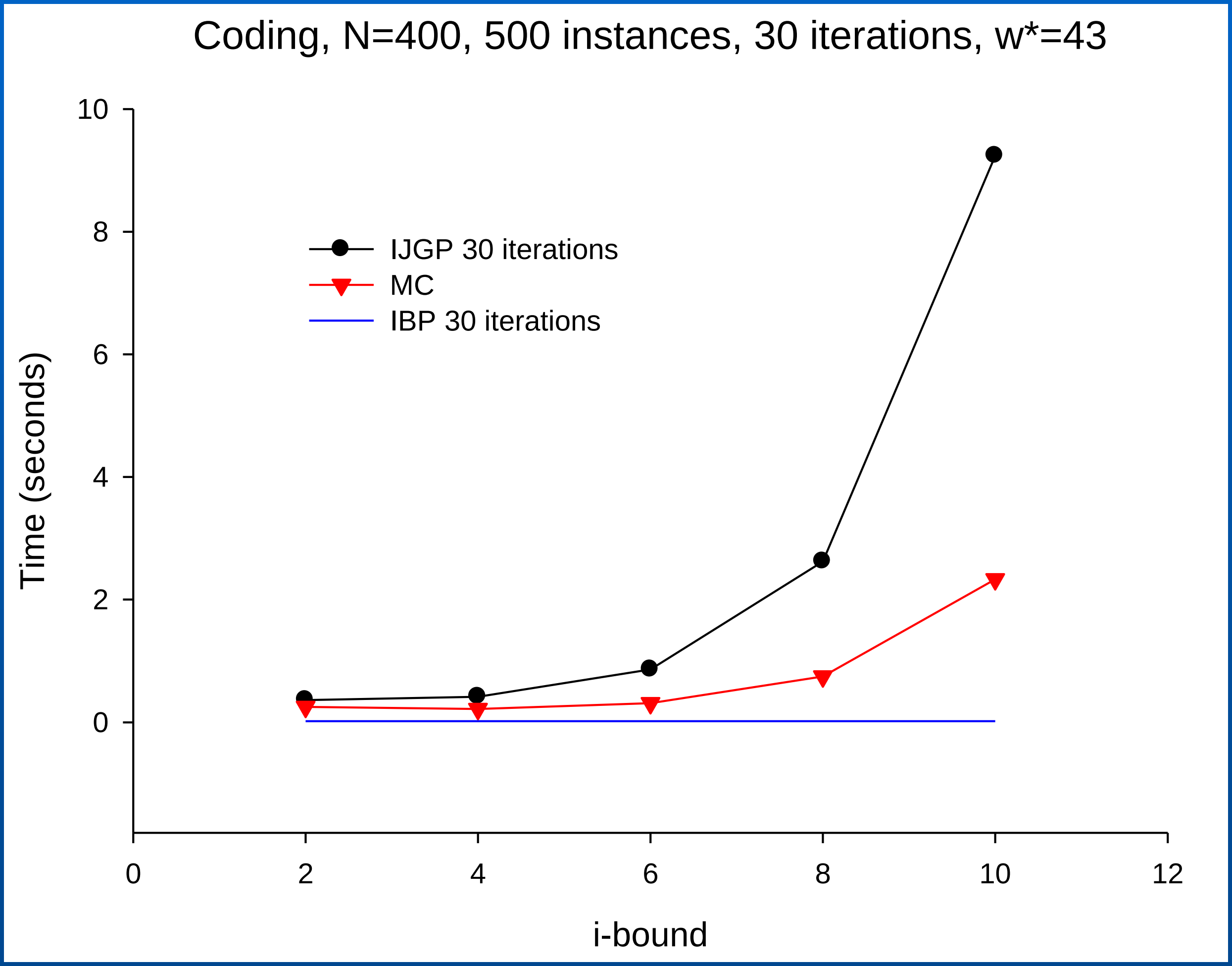


sigma=.51

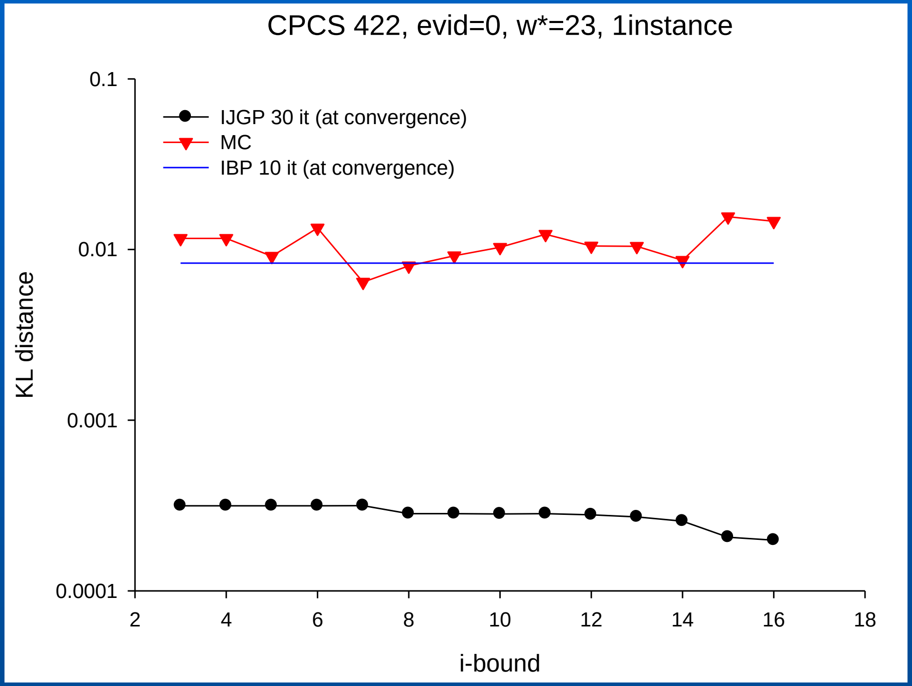


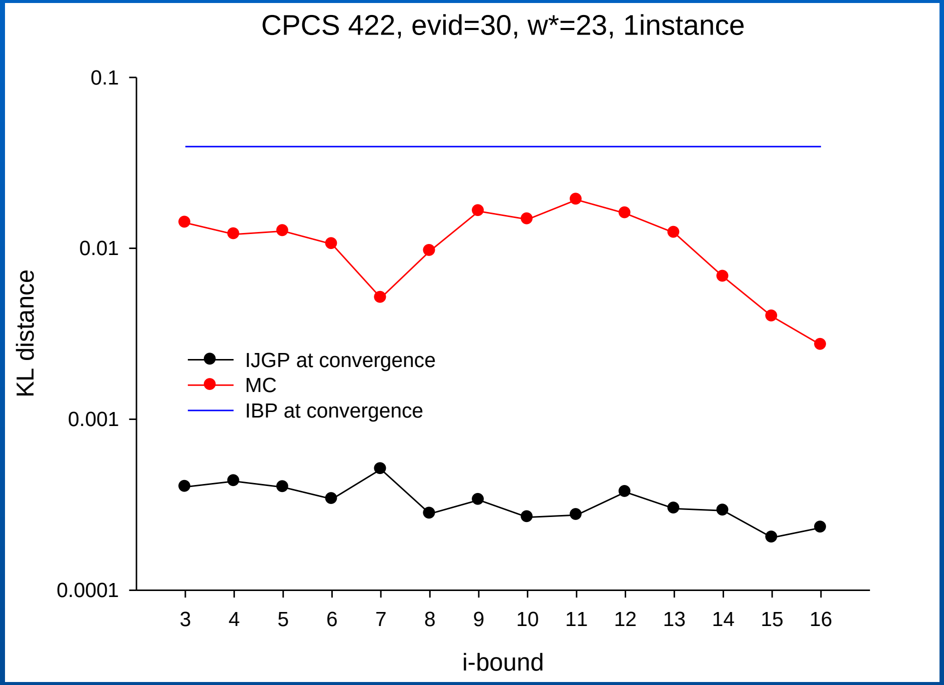
sigma=.65

Coding networks - Time



CPCS 422 – KL distance

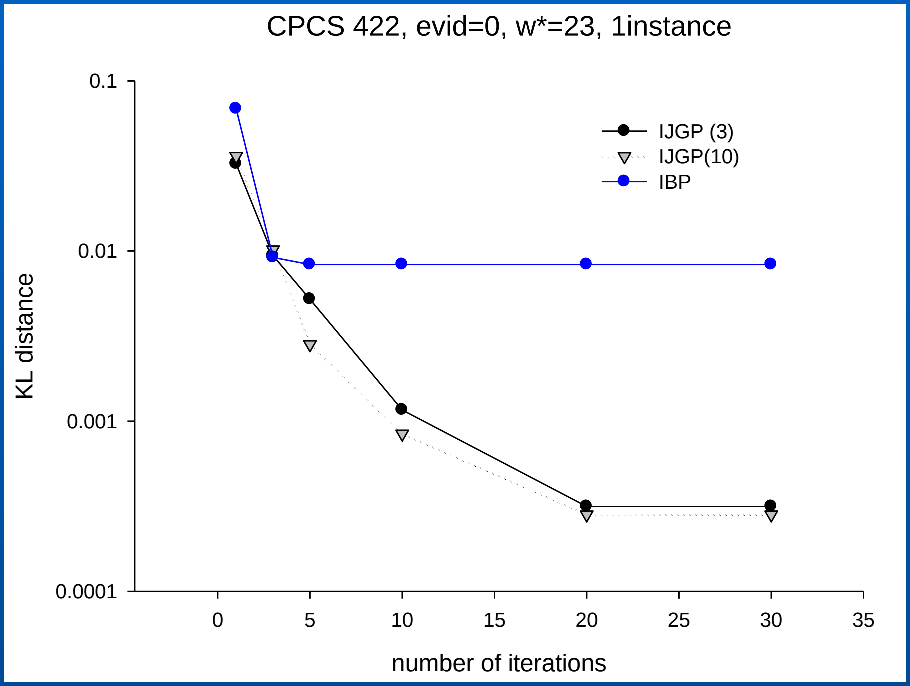




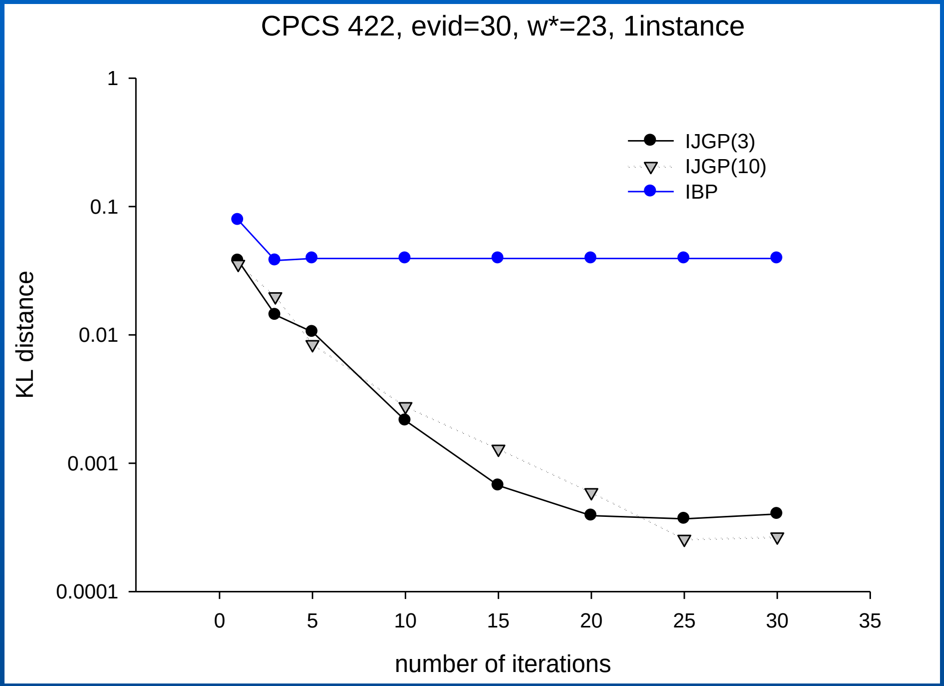
evidence=0

evidence=30

CPCS 422 – KL vs. iterations

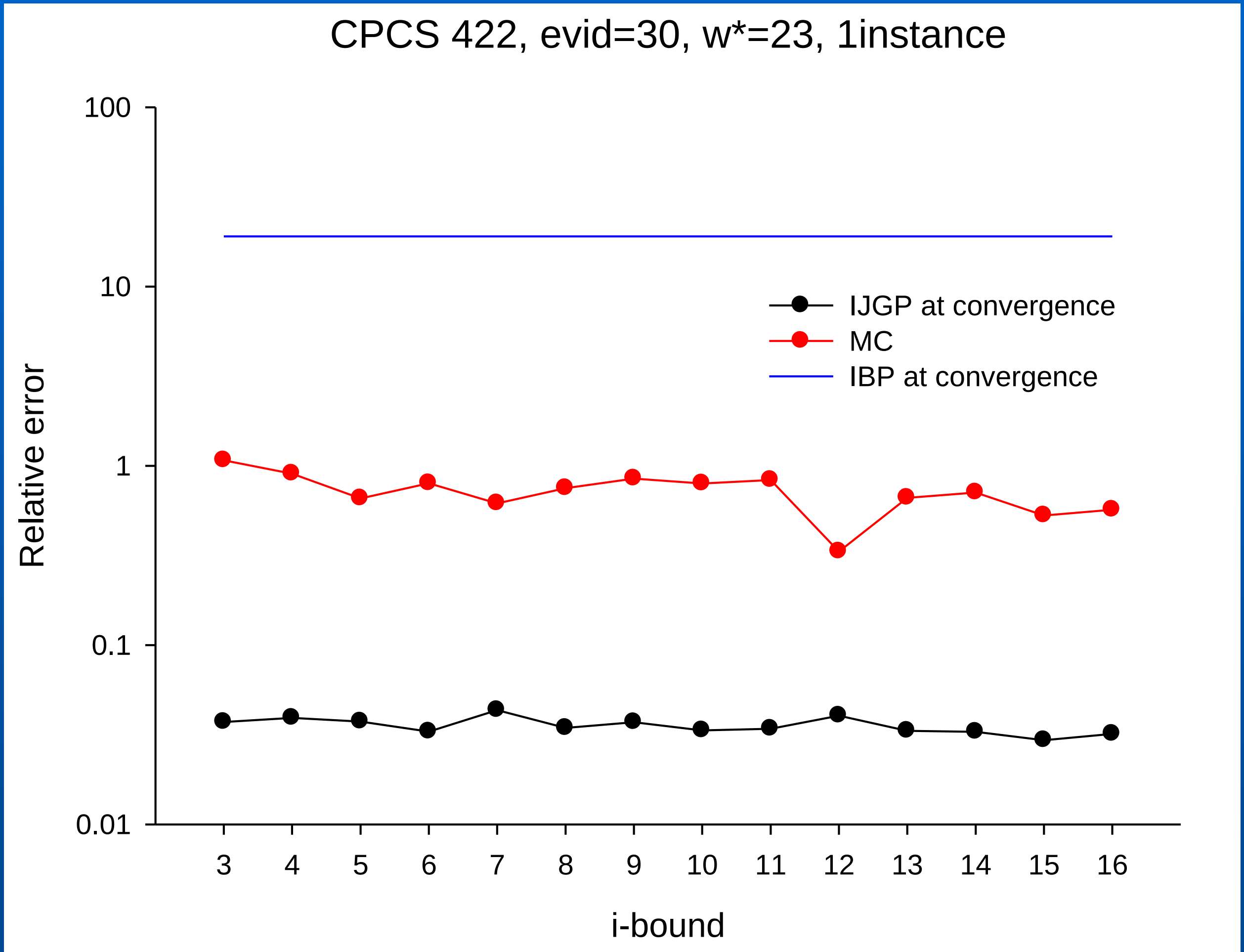


evidence=0



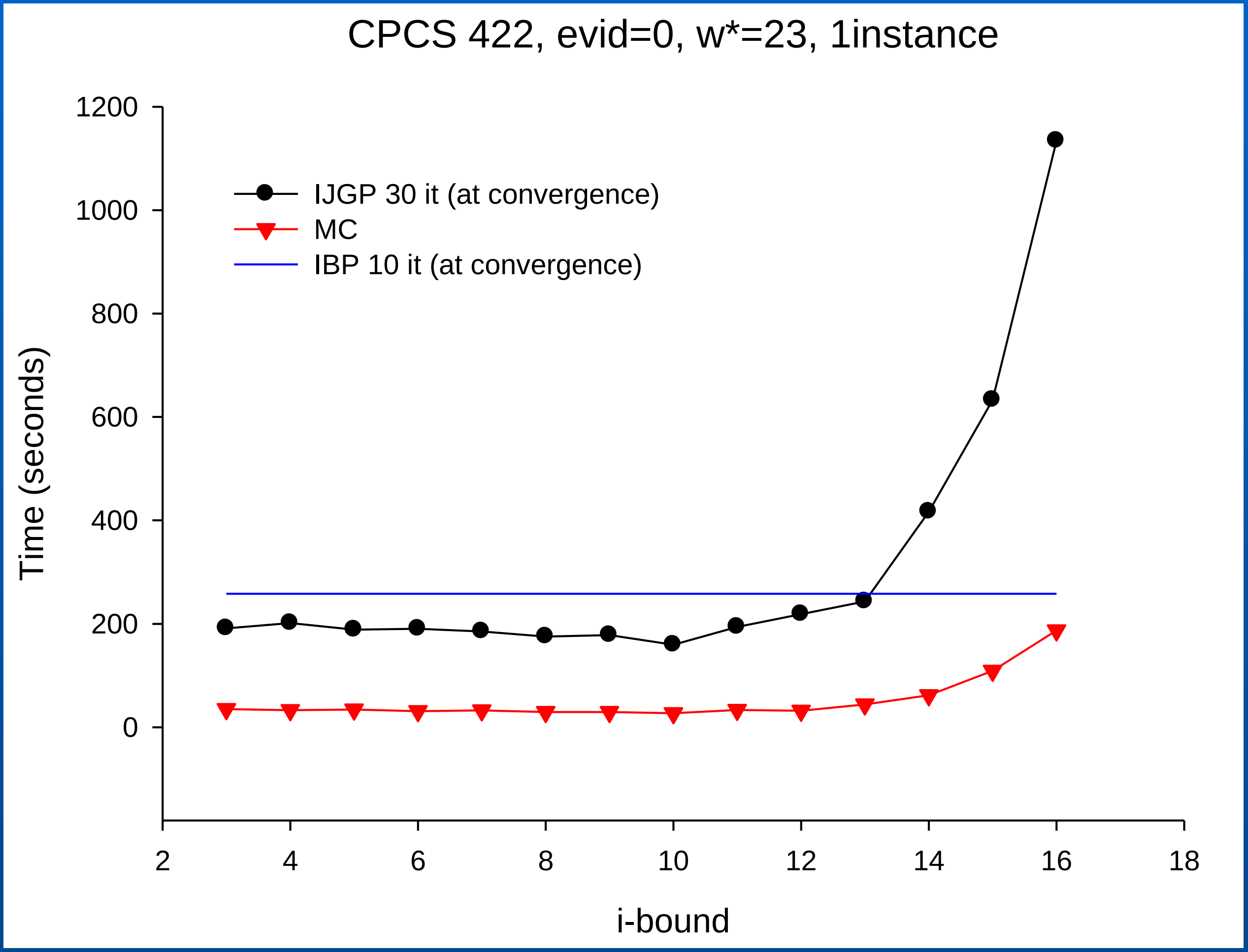
evidence=30

CPCS 422 – Relative error



evidence=30

CPCS 422 – Time vs. i-bound



evidence=0

Conclusion

- IJGP borrows the iterative feature from IBP and the anytime virtues of bounded inference from MC
- Empirical evaluation showed the potential of IJGP, which improves with iteration and most of the time with i-bound, and scales up to large networks
- IJGP is almost always superior, often by a high margin, to IBP and MC
- Based on all our experiments, we think that IJGP provides a practical breakthrough to the task of belief updating