Avoiding Repetition in Repeated Inference on Probabilistic Relational Models: The Lifted Junction Tree Algorithm – Evaluation Setting

Tanya Braun BRAUN@IFIS.UNI-LUEBECK.DE
Ralf Möller MOELLER@IFIS.UNI-LUEBECK.DE
University of Lübeck, Ratzeburger Allee 160
23562 Lübeck, Germany

1. Input Description for the Empirical Evaluation

During the empirical evaluation, we vary the following, (i) largest domain size n, (ii) number of parclusters n_J, and (iii) lifted width $(w_g, w_\#)$ in which w_g is the ground width and $w_\#$ the counting width. In the following, we provide a description of the basic model as well as the variations w.r.t. n, n_J, w_g, and $w_\#$, the evidence setup, and the fusion test model.

2. Basic Setting

The basic input model is G_{ex} with boolean ranges, $n = 1000$, which we use for each logvar, $n_J = 3$, $w_g = 3$, and $w_\# = 1$. Explicitly, we have $\mathbf{L} = \{D, W, M, X\}$, $\Phi = \{\phi_0, \phi_1, \phi_2, \phi_3\}$, and $\mathbf{R} = \{Epid, Nat, Man, Sick, Travel, Treat\}$. The domain sizes are $n = |\mathcal{D}(D)| = |\mathcal{D}(W)| = |\mathcal{D}(M)| = |\mathcal{D}(X)| = 1000$. The PRVs are $Epid$, $Sick(X)$, $Nat(D)$, $Man(W)$, $Travel(X)$, and $Treat(X, M)$ with boolean range. The parfactors in G_{ex} are $g_0 = \phi_0(Epid)$, $g_1 = \phi_1(Epid, Nat(D), Man(W))$, $g_2 = \phi_2(Epid, Sick(X), Travel(X))$, and $g_3 = \phi_3(Epid, Sick(X), Treat(X, M))$ with random potentials. When FOKC is involved, evidence is empty as to not put FOKC at a disadvantage. Otherwise, there exists observations for 10% of the instances of each one-logvar PRV.

3. Varying the Domain Size

G_{ex} remains the same, while $n = |\mathcal{D}(D)| = |\mathcal{D}(W)| = |\mathcal{D}(M)| = |\mathcal{D}(X)|$ varies. The last line holds the basic setting. $\max|\mathcal{A}|$ refers to the largest number of arguments.

Table 1: Key numbers about the input models used when varying n without FOKC

n	n_J	w_g	$w_\#$	$\lvert\mathbf{E}\rvert$	$\lvert G\rvert$	$\lvert gr(G)\rvert$	$\lvert G_i\rvert$ in	$\max\lvert\mathcal{A}\rvert$
5	3	3	1	4	4	11	$\{1, 2\}$	3
10	3	3	1	8	4	211	$\{1, 2\}$	3
15	3	3	1	12	4	301	$\{1, 2\}$	3
20	3	3	1	16	4	821	$\{1, 2\}$	3
50	3	3	1	40	4	5051	$\{1, 2\}$	3
100	3	3	1	80	4	20101	$\{1, 2\}$	3

n	n_J	w_g	$w_\#$	$\lvert\mathbf{E}\rvert$	$\lvert G\rvert$	$\lvert gr(G)\rvert$	$\lvert G_i\rvert$ in	$\max\lvert\mathcal{A}\rvert$
150	3	3	1	120	4	45151	$\{1,2\}$	3
200	3	3	1	160	4	80201	$\{1,2\}$	3
250	3	3	1	200	4	125251	$\{1,2\}$	3
300	3	3	1	240	4	180301	$\{1,2\}$	3
350	3	3	1	280	4	245351	$\{1,2\}$	3
400	3	3	1	320	4	320401	$\{1,2\}$	3
450	3	3	1	360	4	405451	$\{1,2\}$	3
500	3	3	1	400	4	500501	$\{1,2\}$	3
550	3	3	1	440	4	605551	$\{1,2\}$	3
600	3	3	1	480	4	720601	$\{1,2\}$	3
650	3	3	1	520	4	845651	$\{1,2\}$	3
700	3	3	1	560	4	980701	$\{1,2\}$	3
750	3	3	1	600	4	1125751	$\{1,2\}$	3
800	3	3	1	640	4	1280801	$\{1,2\}$	3
850	3	3	1	680	4	1445851	$\{1,2\}$	3
900	3	3	1	720	4	1620901	$\{1,2\}$	3
950	3	3	1	760	4	1805951	$\{1,2\}$	3
1000	3	3	1	800	4	2001001	$\{1,2\}$	3

Table 3: Key numbers about the input models used when varying n with FOKC

n	n_J	w_g	$w_\#$	$\lvert\mathbf{E}\rvert$	$\lvert G\rvert$	$\lvert gr(G)\rvert$	$\lvert G_i\rvert$ in	$\max\lvert\mathcal{A}\rvert$
2	3	3	1	0	4	11	$\{1,2\}$	3
4	3	3	1	0	4	37	$\{1,2\}$	3
6	3	3	1	0	4	79	$\{1,2\}$	3
8	3	3	1	0	4	137	$\{1,2\}$	3
10	3	3	1	0	4	211	$\{1,2\}$	3
12	3	3	1	0	4	301	$\{1,2\}$	3
14	3	3	1	0	4	407	$\{1,2\}$	3
16	3	3	1	0	4	529	$\{1,2\}$	3
18	3	3	1	0	4	667	$\{1,2\}$	3
20	3	3	1	0	4	821	$\{1,2\}$	3
50	3	3	1	0	4	5051	$\{1,2\}$	3
100	3	3	1	0	4	20101	$\{1,2\}$	3
150	3	3	1	0	4	45151	$\{1,2\}$	3
200	3	3	1	0	4	80201	$\{1,2\}$	3
250	3	3	1	0	4	125251	$\{1,2\}$	3
300	3	3	1	0	4	180301	$\{1,2\}$	3
350	3	3	1	0	4	245351	$\{1,2\}$	3
400	3	3	1	0	4	320401	$\{1,2\}$	3
450	3	3	1	0	4	405451	$\{1,2\}$	3
500	3	3	1	0	4	500501	$\{1,2\}$	3

n	n_J	w_g	$w_\#$	$\lvert\mathbf{E}\rvert$	$\lvert G\rvert$	$\lvert gr(G)\rvert$	$\lvert G_i\rvert$ in	$\max\lvert\mathcal{A}\rvert$
550	3	3	1	0	4	605551	$\{1,2\}$	3
600	3	3	1	0	4	720601	$\{1,2\}$	3
650	3	3	1	0	4	845651	$\{1,2\}$	3
700	3	3	1	0	4	980701	$\{1,2\}$	3
750	3	3	1	0	4	1125751	$\{1,2\}$	3
800	3	3	1	0	4	1280801	$\{1,2\}$	3
850	3	3	1	0	4	1445851	$\{1,2\}$	3
900	3	3	1	0	4	1620901	$\{1,2\}$	3
950	3	3	1	0	4	1805951	$\{1,2\}$	3
1000	3	3	1	0	4	2001001	$\{1,2\}$	3
2000	3	3	1	0	4	2001001	$\{1,2\}$	3
3000	3	3	1	0	4	2001001	$\{1,2\}$	3
4000	3	3	1	0	4	2001001	$\{1,2\}$	3
5000	3	3	1	0	4	2001001	$\{1,2\}$	3
6000	3	3	1	0	4	2001001	$\{1,2\}$	3
7000	3	3	1	0	4	2001001	$\{1,2\}$	3
8000	3	3	1	0	4	2001001	$\{1,2\}$	3
9000	3	3	1	0	4	2001001	$\{1,2\}$	3
10000	3	3	1	0	4	2001001	$\{1,2\}$	3

In the following, the tables show two values in the **E** column for number of observations, 0 for runs with FOKC and the other values for runs without FOKC.

4. Varying the Number of Parclusters

When varying n_J from 2 to 11, w_g and $w_\#$ remain fixed. Each new parcluster has a lifted width of $(3,0)$, leaving the overall lifted width at $(3,1)$. G_{ex} appears where $n_J = 3$ and $n = 1000$. A small jump in $\lvert gr(G)\rvert$ occurs when additional PRVs trigger a new parcluster. A larger jump occurs if an additional logvar triggers a new parcluster.

Table 5: Key numbers about the input models used when varying n_J

n	n_J	w_g	$w_\#$	$\lvert\mathbf{E}\rvert$	$\lvert G\rvert$	$\lvert gr(G)\rvert$	$\lvert G_i\rvert$ in	$\max\lvert\mathcal{A}\rvert$
10	2	3	1	$8,0$	3	111	$\{1,2\}$	3
10	3	3	1	$8,0$	4	211	$\{1,2\}$	3
10	4	3	1	$10,0$	5	221	$\{1,2\}$	3
10	5	3	1	$12,0$	6	231	$\{1,2\}$	3
10	6	3	1	$12,0$	7	331	$\{1,2\}$	3
10	7	3	1	$14,0$	8	341	$\{1,2\}$	3
10	8	3	1	$14,0$	9	441	$\{1,2\}$	3
10	9	3	1	$16,0$	10	451	$\{1,2\}$	3
10	10	3	1	$18,0$	11	461	$\{1,2\}$	3
10	11	3	1	$18,0$	12	561	$\{1,2\}$	3

n	n_J	w_g	$w_\#$	$\lvert\mathbf{E}\rvert$	$\lvert G\rvert$	$\lvert gr(G)\rvert$	$\lvert G_i\rvert$ in	$\max\lvert\mathcal{A}\rvert$
100	2	3	1	80, 0	3	10101	{1, 2}	3
100	3	3	1	80, 0	4	20101	{1, 2}	3
100	4	3	1	100, 0	5	20201	{1, 2}	3
100	5	3	1	120, 0	6	20301	{1, 2}	3
100	6	3	1	120, 0	7	30301	{1, 2}	3
100	7	3	1	140, 0	8	30401	{1, 2}	3
100	8	3	1	140, 0	9	40401	{1, 2}	3
100	9	3	1	160, 0	10	40501	{1, 2}	3
100	10	3	1	180, 0	11	40601	{1, 2}	3
100	11	3	1	180, 0	12	50601	{1, 2}	3
1000	2	3	1	800, 0	3	1001001	{1, 2}	3
1000	3	3	1	800, 0	4	2001001	{1, 2}	3
1000	4	3	1	1000, 0	5	2002001	{1, 2}	3
1000	5	3	1	1200, 0	6	2003001	{1, 2}	3
1000	6	3	1	1200, 0	7	3003001	{1, 2}	3
1000	7	3	1	1400, 0	8	3004001	{1, 2}	3
1000	8	3	1	1400, 0	9	4004001	{1, 2}	3
1000	9	3	1	1600, 0	10	4005001	{1, 2}	3
1000	10	3	1	1800, 0	11	4006001	{1, 2}	3
1000	11	3	1	1800, 0	12	5006001	{1, 2}	3

5. Varying the Ground Width

Varying w_g from 2 to 11 means that n_J and $w_\#$ appear fixed and each parcluster has a ground width of w_g. G_{ex} appears where $w_g = 3$ and $n = 1000$.

Table 7: Key numbers about the input models used when varying w_g

n	n_J	w_g	$w_\#$	$\lvert\mathbf{E}\rvert$	$\lvert G\rvert$	$\lvert gr(G)\rvert$	$\lvert G_i\rvert$ in	$\max\lvert\mathcal{A}\rvert$
10	3	2	1	6, 0	4	31	{2}	2
10	3	3	1	8, 0	4	121	{1, 2}	3
10	3	4	1	12, 0	10	631	{3, 4}	3
10	3	5	1	12, 0	16	1231	{5, 7}	3
10	3	6	1	18, 0	25	2131	{8, 9}	3
10	3	7	1	24, 0	34	2761	{11, 12}	3
10	3	8	1	24, 0	43	3661	{14, 15}	3
10	3	9	1	24, 0	52	4561	{17, 18}	3
10	3	10	1	30, 0	61	5191	{20, 21}	3
10	3	11	1	36, 0	70	5821	{23, 24}	3
100	3	2	1	60, 0	4	301	{2}	2
100	3	3	1	80, 0	4	10201	{1, 2}	3
100	3	4	1	120, 0	10	60301	{3, 4}	3
100	3	5	1	120, 0	16	120301	{5, 7}	3

n	n_J	w_g	$w_\#$	$\lvert\mathbf{E}\rvert$	$\lvert G\rvert$	$\lvert gr(G)\rvert$	$\lvert G_i\rvert$ in	$\max\lvert\mathcal{A}\rvert$
100	3	6	1	180, 0	25	210301	{8, 9}	3
100	3	7	1	240, 0	34	270601	{11, 12}	3
100	3	8	1	240, 0	43	360601	{14, 15}	3
100	3	9	1	240, 0	52	450601	{17, 18}	3
100	3	10	1	300, 0	61	510901	{20, 21}	3
100	3	11	1	360, 0	70	571201	{23, 24}	3
1000	3	2	1	600, 0	4	3001	{2}	2
1000	3	3	1	800, 0	4	1002001	{1, 2}	3
1000	3	4	1	1200, 0	10	6003001	{3, 4}	3
1000	3	5	1	1200, 0	16	12003001	{5, 7}	3
1000	3	6	1	1800, 0	25	21003001	{8, 9}	3
1000	3	7	1	2400, 0	34	27006001	{11, 12}	3
1000	3	8	1	2400, 0	43	36006001	{14, 15}	3
1000	3	9	1	2400, 0	52	45006001	{17, 18}	3
1000	3	10	1	3000, 0	61	51009001	{20, 21}	3
1000	3	11	1	3600, 0	70	57012001	{23, 24}	3

6. Varying the Counting Width

When varying $w_\#$ from 0 to 9, n_J and w_g are fixed and each of the three parclusters has a counting width of $w_\#$. G_{ex} appears where $w_\# = 1$ and $n = 1000$.

Table 9: Key numbers about the input models used when varying $w_\#$

n	n_J	w_g	$w_\#$	$\lvert\mathbf{E}\rvert$	$\lvert G\rvert$	$\lvert gr(G)\rvert$	$\lvert G_i\rvert$ in	$\max\lvert\mathcal{A}\rvert$
10	3	3	0	6, 0	4	31	{2}	2
10	3	3	1	12, 0	4	301	{1, 2}	3
10	3	3	2	18, 0	4	3001	{1, 2}	4
10	3	3	3	24, 0	7	6001	{2, 3}	4
10	3	3	4	30, 0	10	9001	{3, 4}	4
10	3	3	5	36, 0	13	12001	{4, 5}	4
10	3	3	6	42, 0	16	15001	{5, 6}	4
10	3	3	7	48, 0	19	18001	{6, 7}	4
10	3	3	8	54, 0	22	21001	{7, 8}	4
10	3	3	9	60, 0	25	24001	{8, 9}	4
100	3	3	0	60, 0	4	301	{2}	2
100	3	3	1	120, 0	4	30001	{1, 2}	3
100	3	3	2	180, 0	4	3000001	{1, 2}	4
100	3	3	3	240, 0	7	6000001	{2, 3}	4
100	3	3	4	300, 0	10	9000001	{3, 4}	4
100	3	3	5	360, 0	13	12000001	{4, 5}	4
100	3	3	6	420, 0	16	15000001	{5, 6}	4
100	3	3	7	480, 0	19	18000001	{6, 7}	4

n	n_J	w_g	$w_\#$	$\lvert\mathbf{E}\rvert$	$\lvert G\rvert$	$\lvert gr(G)\rvert$	$\lvert G_i\rvert$ in	$\max\lvert\mathcal{A}\rvert$
100	3	3	8	540, 0	22	21000001	$\{7, 8\}$	4
100	3	3	9	600, 0	25	24000001	$\{8, 9\}$	4
1000	3	3	0	600, 0	4	3001	$\{2\}$	2
1000	3	3	1	1200, 0	4	3, 000, 001	$\{1, 2\}$	3
1000	3	3	2	1800, 0	10	$3 \cdot 1000^3 + 1$	$\{1, 2\}$	4
1000	3	3	3	2400, 0	16	$6 \cdot 1000^3 + 1$	$\{2, 3\}$	4
1000	3	3	4	3000, 0	25	$9 \cdot 1000^3 + 1$	$\{3, 4\}$	4
1000	3	3	5	3600, 0	34	$12 \cdot 1000^3 + 1$	$\{4, 5\}$	4
1000	3	3	6	4200, 0	43	$15 \cdot 1000^3 + 1$	$\{5, 6\}$	4
1000	3	3	7	4800, 0	52	$18 \cdot 1000^3 + 1$	$\{6, 7\}$	4
1000	3	3	8	5400, 0	61	$21 \cdot 1000^3 + 1$	$\{7, 8\}$	4
1000	3	3	9	6000, 0	70	$24 \cdot 1000^3 + 1$	$\{8, 9\}$	4

7. Varying the Evidence Coverage

For evidence testing, we add evidence for the 1-logvar PRVs in G_{ex}, i.e., four evidence parfactors, in 10% steps. The line with $n = 1000$ and $|\mathbf{E}| = 0$ is G_{ex}.

Table 11: Key numbers about the input models used when varying evidence coverage

n	n_J	w_g	$w_\#$	$\lvert\mathbf{E}\rvert$	$\lvert G\rvert$	$\lvert gr(G)\rvert$	$\lvert G_i\rvert$ in	$\max\lvert\mathcal{A}\rvert$
10	3	3	1	0	4	211	$\{1, 2\}$	3
10	3	3	1	4	4	211	$\{1, 2\}$	3
10	3	3	1	8	4	211	$\{1, 2\}$	3
10	3	3	1	12	4	211	$\{1, 2\}$	3
10	3	3	1	16	4	211	$\{1, 2\}$	3
10	3	3	1	20	4	211	$\{1, 2\}$	3
10	3	3	1	24	4	211	$\{1, 2\}$	3
10	3	3	1	28	4	211	$\{1, 2\}$	3
10	3	3	1	32	4	211	$\{1, 2\}$	3
10	3	3	1	36	4	211	$\{1, 2\}$	3
10	3	3	1	36	4	211	$\{1, 2\}$	3
100	3	3	1	0	4	20101	$\{1, 2\}$	3
100	3	3	1	40	4	20101	$\{1, 2\}$	3
100	3	3	1	80	4	20101	$\{1, 2\}$	3
100	3	3	1	120	4	20101	$\{1, 2\}$	3
100	3	3	1	160	4	20101	$\{1, 2\}$	3
100	3	3	1	200	4	20101	$\{1, 2\}$	3
100	3	3	1	240	4	20101	$\{1, 2\}$	3
100	3	3	1	280	4	20101	$\{1, 2\}$	3
100	3	3	1	320	4	20101	$\{1, 2\}$	3
100	3	3	1	360	4	20101	$\{1, 2\}$	3
100	3	3	1	396	4	20101	$\{1, 2\}$	3

n	n_J	w_g	$w_\#$	$\lvert\mathbf{E}\rvert$	$\lvert G\rvert$	$\lvert gr(G)\rvert$	$\lvert G_i\rvert$ in	$\max\lvert\mathcal{A}\rvert$
1000	3	3	1	0	4	2001001	$\{1,2\}$	3
1000	3	3	1	400	4	2001001	$\{1,2\}$	3
1000	3	3	1	800	4	2001001	$\{1,2\}$	3
1000	3	3	1	1200	4	2001001	$\{1,2\}$	3
1000	3	3	1	1600	4	2001001	$\{1,2\}$	3
1000	3	3	1	2000	4	2001001	$\{1,2\}$	3
1000	3	3	1	2400	4	2001001	$\{1,2\}$	3
1000	3	3	1	2800	4	2001001	$\{1,2\}$	3
1000	3	3	1	3200	4	2001001	$\{1,2\}$	3
1000	3	3	1	3600	4	2001001	$\{1,2\}$	3
1000	3	3	1	3996	4	2001001	$\{1,2\}$	3

8. Unnecessary Groundings and Fusion

For fusion, we use a slight variation of G_{ex} with $n_J = 4$, which has unnecessary groundings with LJT. After fusion, $n_J = 3$ and $w_g = 5$. We vary the domain size n again.

Table 13: Key numbers about the input models used for evaluating fusion

n	n_J	w_g	$w_\#$	$\lvert\mathbf{E}\rvert$	$\lvert G\rvert$	$\lvert gr(G)\rvert$	$\lvert G_i\rvert$	$\max\lvert\mathcal{A}\rvert$
2	$4\to3$	$3\to5$	1	0	5	15	$\{1,2\}\to\{2,3\}$	3
4	$4\to3$	$3\to5$	1	0	5	53	$\{1,2\}\to\{2,3\}$	3
6	$4\to3$	$3\to5$	1	0	5	115	$\{1,2\}\to\{2,3\}$	3
8	$4\to3$	$3\to5$	1	0	5	201	$\{1,2\}\to\{2,3\}$	3
10	$4\to3$	$3\to5$	1	0	5	311	$\{1,2\}\to\{2,3\}$	3
12	$4\to3$	$3\to5$	1	0	5	445	$\{1,2\}\to\{2,3\}$	3
14	$4\to3$	$3\to5$	1	0	5	603	$\{1,2\}\to\{2,3\}$	3
16	$4\to3$	$3\to5$	1	0	5	785	$\{1,2\}\to\{2,3\}$	3
18	$4\to3$	$3\to5$	1	0	5	991	$\{1,2\}\to\{2,3\}$	3
20	$4\to3$	$3\to5$	1	0	5	1221	$\{1,2\}\to\{2,3\}$	3
50	$4\to3$	$3\to5$	1	0	5	7551	$\{1,2\}\to\{2,3\}$	3
100	$4\to3$	$3\to5$	1	0	5	30101	$\{1,2\}\to\{2,3\}$	3
150	$4\to3$	$3\to5$	1	0	5	67651	$\{1,2\}\to\{2,3\}$	3
200	$4\to3$	$3\to5$	1	0	5	120201	$\{1,2\}\to\{2,3\}$	3
250	$4\to3$	$3\to5$	1	0	5	187751	$\{1,2\}\to\{2,3\}$	3
300	$4\to3$	$3\to5$	1	0	5	270301	$\{1,2\}\to\{2,3\}$	3
350	$4\to3$	$3\to5$	1	0	5	367851	$\{1,2\}\to\{2,3\}$	3
400	$4\to3$	$3\to5$	1	0	5	480401	$\{1,2\}\to\{2,3\}$	3
450	$4\to3$	$3\to5$	1	0	5	607951	$\{1,2\}\to\{2,3\}$	3
500	$4\to3$	$3\to5$	1	0	5	750501	$\{1,2\}\to\{2,3\}$	3
550	$4\to3$	$3\to5$	1	0	5	908051	$\{1,2\}\to\{2,3\}$	3
600	$4\to3$	$3\to5$	1	0	5	1080601	$\{1,2\}\to\{2,3\}$	3
650	$4\to3$	$3\to5$	1	0	5	1268151	$\{1,2\}\to\{2,3\}$	3

n	n_J	w_g	$w_\#$	$\|\mathbf{E}\|$	$\|G\|$	$\|gr(G)\|$	$\|G_i\|$	$\max\|\mathcal{A}\|$
700	$4 \to 3$	$3 \to 5$	1	0	5	1470701	$\{1,2\} \to \{2,3\}$	3
750	$4 \to 3$	$3 \to 5$	1	0	5	1688251	$\{1,2\} \to \{2,3\}$	3
800	$4 \to 3$	$3 \to 5$	1	0	5	1920801	$\{1,2\} \to \{2,3\}$	3
850	$4 \to 3$	$3 \to 5$	1	0	5	2168351	$\{1,2\} \to \{2,3\}$	3
900	$4 \to 3$	$3 \to 5$	1	0	5	2430901	$\{1,2\} \to \{2,3\}$	3
950	$4 \to 3$	$3 \to 5$	1	0	5	2708451	$\{1,2\} \to \{2,3\}$	3
1000	$4 \to 3$	$3 \to 5$	1	0	5	3001001	$\{1,2\} \to \{2,3\}$	3

For more information regarding the input models, please send an email to: `braun@ifis.uni-luebeck.de`.