

Simplifying, Regularizing and Strengthening Sum-Product Network Structure Learning: Supplementary Material

Antonio Vergari and Nicola Di Mauro and Floriana Esposito

University of Bari “Aldo Moro”, Bari, Italy
{antonio.vergari,nicola.dimauro,floriana.esposito}@uniba.it

Extended Results

	NLTCs	MSNBC	KDDCup2k	Plants	Audio	Jester	Netflix	Accidents	Retail	Pumsb-star	DNA	Kosarek	MSWeb	Book	EachMovie	WebKB	Reuters-52	BBC	Ad
LearnSPN																			
ρ	15	10	15	5	5	15	10	10	15	20	15	15	15	15	10	15	10	10	5
λ	0.4	0.4	0.2	0.4	0.6	0.6	0.6	0.4	0.2	0.6	0.6	0.6	0.4	0.4	0.8	0.6	0.4	0.6	0.8
m	50	100	10	10	100	100	50	10	50	10	100	10	500	100	100	10	50	10	10
α	0.5	0.1	0.5	0.2	2.0	1.0	2.0	0.2	2.0	0.1	0.1	0.5	0.5	1.0	1.0	0.5	0.2	0.1	0.1
SPN-B																			
ρ	5	5	5	5	5	5	5	10	10	10	15	5	5	5	5	5	5	5	5
m	10	50	10	10	50	10	50	10	50	10	50	10	10	50	10	10	10	10	10
α	2.0	1.0	2.0	1.0	2.0	2.0	2.0	0.2	2.0	0.1	0.1	2.0	1.0	2.0	2.0	2.0	1.0	1.0	0.1
SPN-BT																			
ρ	5	5	5	5	5	5	5	10	10	10	10	10	5	10	5	10	5	5	5
m	10	50	10	100	500	50	500	100	50	100	100	500	50	10	10	100	10	10	10
α	2.0	1.0	2.0	1.0	2.0	2.0	2.0	0.2	2.0	0.1	0.1	2.0	0.5	2.0	2.0	2.0	1.0	1.0	0.1
MT																			
k	20	30	20	30	22	16	30	30	4	26	6	6	10	4	6	2	4	2	2

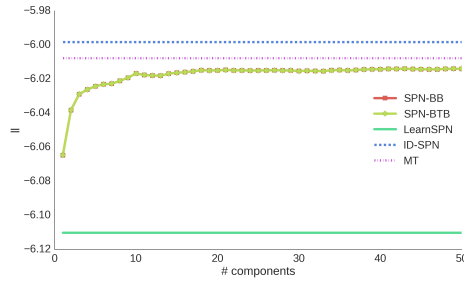
Table 1: Best parameter configurations on each dataset, found on the validation set portion for LearnSPN, SPN-B, SPN-BT and MT.

	LearnSPN	SPN-B	SPN-BT	ID-SPN	SPN-BB	SPN-BTB	MT
NLTCS	183	114	108	310	1004	1041	290
MSNBC	7354	5922	5774	46266	7963	8275	8645
KDDCup2k	5800	3946	4525	32067	10633	12148	39857
Plants	1156	1420	276	18833	7515	3023	7414
Audio	1146	1561	131	21009	10466	4113	6566
Jester	1289	789	687	10412	17753	8566	3064
Netflix	1331	2369	152	30294	14871	4768	11402
Accidents	839	1006	722	15472	9198	8280	14073
Retail	253	173	161	4041	2317	2505	320
Pumsb-star	230	1010	612	20952	7819	7790	18533
DNA	140	87	122	3040	2022	3049	228
Kosarek	225	1515	416	17799	16583	3642	18782
MSWeb	369	1711	1444	19682	16881	18273	36076
Book	457	832	389	61248	27473	16630	5918
EachMovie	307	939	946	118782	22506	29355	12100
WebKB	307	1960	490	45451	29265	34952	931
Reuters-52	2538	5256	4333	70864	36109	135970	15082
BBC	1136	2645	2479	61471	56271	121007	1324
Ad	668	1585	1356	87522	72632	58777	6850

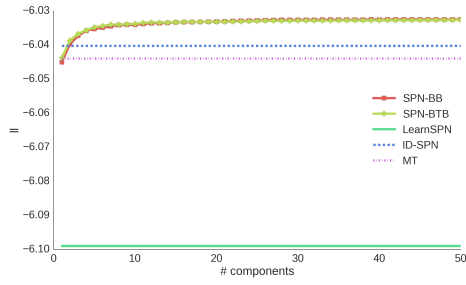
Table 2: Times (in seconds) taken to learn the best models on each dataset for LearnSPN, SPN-B, SPN-BT, SPN-BB, SPN-BTB and MT and with default parameters values for ID-SPN. Experiments run on a 4-core Intel Xeon E312xx (Sandy Bridge) @2.0 GHz with 8Gb of RAM and Ubuntu 14.04.1, kernel 3.13.0-39.

	LearnSPN SPN-B	LearnSPN SPN-BT	LearnSPN ID-SPN	LearnSPN SPN-BB	LearnSPN SPN-BTB	LearnSPN MT	SPN-B SPN-BT
NLTCS	2.60e-01	2.59e-01	8.03e-02	1.63e-01	1.63e-01	1.10e-01	1.50e-18
MSNBC	1.86e-05	1.82e-05	1.24e-05	8.87e-06	1.26e-05	1.49e-04	1.83e-150
KDDCup2k	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00
Plants	9.41e-01	5.47e-01	2.13e-01	1.48e-02	5.16e-03	8.11e-01	1.69e-17
Audio	4.59e-01	6.10e-01	2.60e-01	1.86e-01	1.33e-01	7.37e-01	1.82e-02
Jester	2.11e-01	1.96e-01	1.43e-01	2.51e-01	1.19e-01	5.97e-01	6.44e-01
Netflix	6.70e-04	1.14e-01	4.18e-07	4.06e-04	7.51e-07	4.30e-03	2.02e-15
Accidents	1.32e-08	5.73e-10	1.34e-74	5.43e-24	7.15e-28	2.76e-05	3.66e-08
Retail	3.38e-01	3.36e-01	1.03e-01	1.52e-01	1.51e-01	7.65e-02	4.03e-145
Pumsb-star	1.55e-09	1.16e-12	9.39e-24	4.31e-15	1.54e-18	1.56e-05	3.91e-31
DNA	1.80e-24	2.29e-22	2.20e-26	2.30e-49	2.42e-08	4.69e-20	9.81e-01
Kosarek	3.13e-01	3.05e-01	2.14e-01	2.74e-01	1.89e-01	1.31e-01	2.85e-18
MSWeb	3.83e-02	3.37e-02	1.40e-02	1.84e-03	6.02e-04	6.00e-02	1.30e-06
Book	7.34e-01	6.83e-01	5.72e-01	6.39e-01	4.45e-01	7.30e-01	3.72e-03
EachMovie	7.89e-01	7.86e-01	8.73e-01	5.51e-01	5.58e-01	5.94e-01	1.16e-01
WebKB	6.51e-01	5.23e-01	2.77e-01	2.50e-01	8.98e-02	9.99e-01	1.44e-09
Reuters-52	2.50e-01	2.37e-01	3.20e-01	4.35e-02	4.35e-02	7.23e-01	3.77e-01
BBC	6.61e-01	7.65e-01	4.79e-01	6.57e-03	7.90e-03	1.71e-02	2.61e-01
Ad	1.71e-06	3.63e-07	5.30e-01	8.73e-14	2.12e-14	7.86e-06	6.21e-03
	SPN-B ID-SPN	SPN-B SPN-BB	SPN-B SPN-BTB	SPN-B MT	SPN-BT ID-SPN	SPN-BT SPN-BB	SPN-BT SPN-BTB
NLTCS	2.01e-08	3.00e-11	2.94e-11	4.30e-11	1.98e-08	1.44e-23	1.37e-23
MSNBC	2.96e-71	4.11e-28	8.77e-22	6.59e-34	1.40e-71	2.08e-27	3.40e-22
KDDCup2k	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00
Plants	8.90e-12	7.30e-102	2.74e-126	4.52e-08	3.27e-03	9.16e-60	6.83e-95
Audio	3.19e-61	6.37e-181	5.98e-199	6.08e-25	1.52e-53	1.74e-118	2.47e-200
Jester	1.22e-69	1.61e-102	7.40e-04	4.72e-35	5.42e-73	1.42e-110	1.12e-03
Netflix	4.53e-143	1.17e-203	6.19e-252	1.28e-82	6.20e-114	6.75e-92	1.63e-221
Accidents	9.77e-281	2.71e-70	1.08e-99	3.74e-11	1.66e-279	1.75e-61	7.97e-93
Retail	7.23e-01	9.91e-29	5.60e-29	5.03e-07	6.81e-01	1.01e-28	7.80e-29
Pumsb-star	1.85e-56	1.30e-10	4.19e-35	1.59e-30	8.50e-33	1.91e-01	6.39e-09
DNA	2.15e-01	2.72e-55	7.41e-08	2.46e-104	2.05e-02	1.18e-58	1.54e-06
Kosarek	1.38e-17	5.20e-05	5.29e-66	6.13e-02	4.28e-18	2.07e-11	1.11e-40
MSWeb	5.82e-07	5.48e-31	8.98e-108	4.14e-19	3.51e-10	2.48e-08	3.48e-55
Book	1.39e-03	2.74e-17	3.74e-39	2.33e-02	6.25e-01	9.31e-25	8.94e-18
EachMovie	4.83e-02	1.23e-22	1.01e-17	9.30e-31	4.39e-02	1.97e-22	2.98e-18
WebKB	4.85e-29	3.31e-11	1.11e-75	2.54e-05	4.40e-12	2.50e-01	2.27e-36
Reuters-52	5.99e-01	4.63e-60	3.66e-59	5.03e-57	4.68e-01	2.24e-60	2.68e-60
BBC	3.49e-01	8.16e-25	3.95e-22	6.46e-25	2.71e-01	4.33e-27	1.07e-24
Ad	8.61e-37	4.48e-37	1.81e-40	1.35e-01	1.31e-40	5.23e-30	1.01e-35
	SPN-BT MT	ID-SPN SPN-BB	ID-SPN SPN-BTB	ID-SPN MT	SPN-BB SPN-BTB	SPN-BB MT	SPN-BTB MT
NLTCS	4.39e-11	6.25e-03	6.25e-03	2.58e-03	4.27e-97	7.40e-03	7.82e-03
MSNBC	1.79e-34	5.88e-95	5.34e-06	1.45e-36	0.00e+00	3.02e-91	2.07e-18
KDDCup2k	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00
Plants	4.35e-18	6.21e-27	6.83e-49	6.39e-33	1.20e-24	3.50e-93	1.05e-123
Audio	1.05e-18	2.02e-06	6.36e-14	1.44e-14	1.67e-06	2.15e-32	6.78e-58
Jester	2.56e-37	2.44e-01	2.61e-110	1.30e-09	1.55e-220	9.97e-08	4.80e-57
Netflix	1.93e-53	1.09e-08	4.19e-01	3.62e-20	1.14e-24	1.63e-02	4.88e-22
Accidents	1.31e-15	3.78e-248	1.75e-226	7.16e-307	6.64e-50	2.12e-112	1.96e-141
Retail	7.06e-07	2.10e-04	1.82e-04	6.56e-01	1.15e-13	1.33e-03	1.13e-03
Pumsb-star	1.32e-54	9.62e-43	1.71e-18	6.72e-157	8.03e-110	7.75e-80	9.17e-120
DNA	2.11e-110	5.24e-25	1.03e-12	1.12e-152	7.23e-42	2.43e-153	1.69e-124
Kosarek	8.13e-15	1.50e-10	2.57e-48	9.33e-02	3.96e-70	1.53e-02	1.23e-36
MSWeb	2.69e-22	4.21e-48	1.37e-92	4.37e-32	3.08e-164	4.56e-76	8.12e-97
Book	1.64e-07	2.11e-33	8.15e-14	9.54e-11	6.17e-101	3.11e-11	6.78e-33
EachMovie	2.23e-30	3.78e-17	2.51e-15	4.01e-24	1.30e-01	2.79e-51	7.66e-50
WebKB	1.43e-21	1.09e-08	2.12e-05	8.16e-36	1.75e-38	7.68e-13	3.36e-53
Reuters-52	2.73e-60	5.84e-34	1.33e-33	3.85e-72	5.43e-11	1.29e-112	9.45e-113
BBC	3.88e-26	3.93e-17	3.81e-15	1.91e-34	2.23e-01	1.16e-38	2.03e-38
Ad	1.99e-02	2.15e-63	5.39e-67	1.04e-35	4.32e-35	5.28e-17	1.06e-18

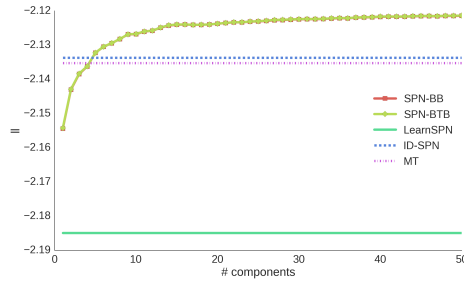
Table 3: Pairwise Wilcoxon signed rank test p-values for the comparisons of test log likelihoods for each algorithmic pair (columns) on all datasets (rows).



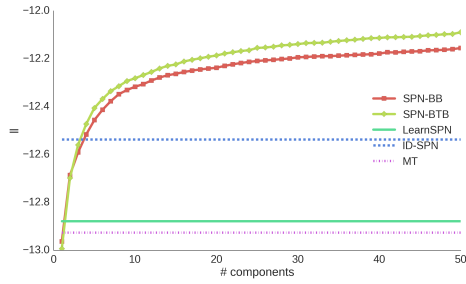
(a) NLTCS



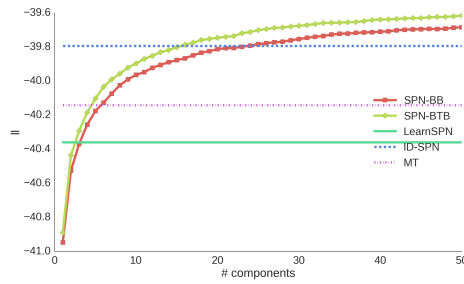
(b) MSNBC



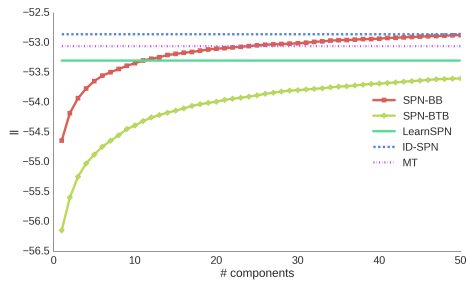
(c) KDDCup2k



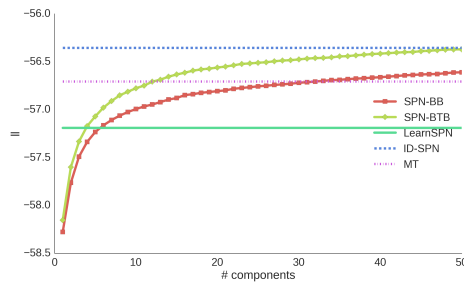
(d) Plants



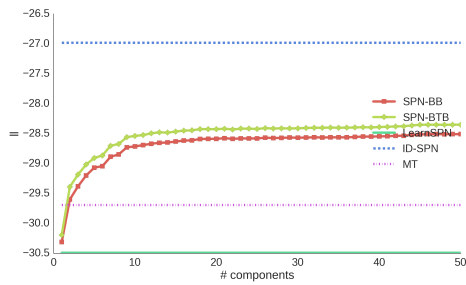
(e) Audio



(f) Jester



(g) Netflix



(h) Accidents

Fig. 1: Average test likelihood values (y-axis) for SPN-BB and SPN-BTB while increasing k (x-axis) on NLTCS, MSNBC, KDDCup2k, Plants, Audio, Jester, Netflix, Accidents. The best values for LearnSPN, ID-SPN and MT are reported as well.

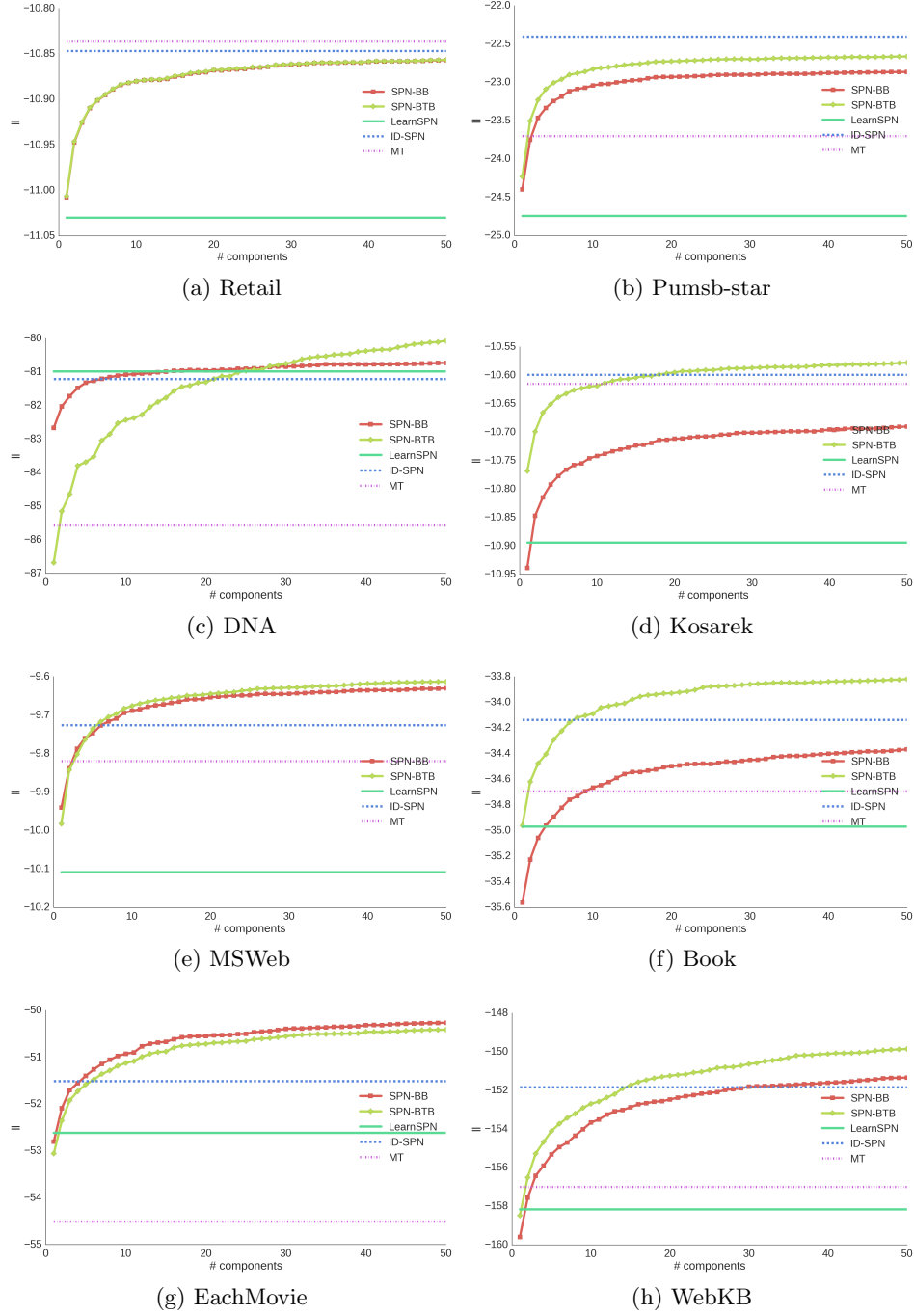


Fig. 2: Average test likelihood values (y-axis) for SPN-BB and SPN-BTB while increasing k (x-axis) on Retail, Pumsb-star, DNA, Kosarek, MSWeb, Book, Each-Movie, WebKB. The best values for LearnSPN, ID-SPN and MT are reported as well.

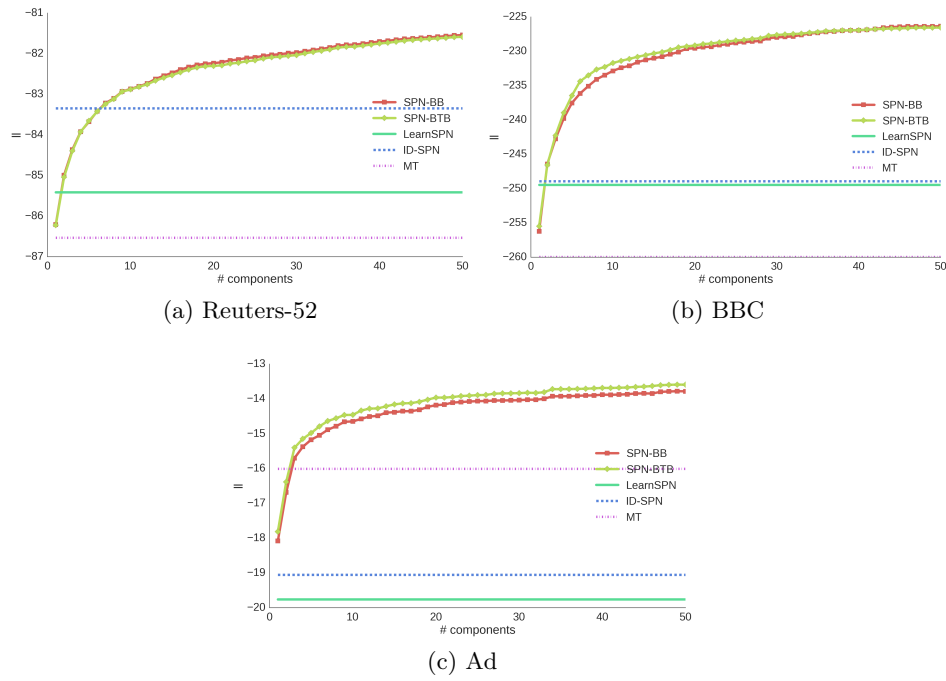


Fig. 3: Average test likelihood values (y-axis) for SPN-BB and SPN-BTB while increasing k (x-axis) on Reuters-52, BBC and Ad. The best values for LearnSPN, ID-SPN and MT are reported as well.