

Dataset	15% CS						KMEANS	GT
	PCCC	PCCC-R	COPKM	CSC	DILS	LCC		
Appendicitis	0	0	–	1	1	18	22	0
Breast Cancer	0	0	0	1,676	0	0	337	0
Bupa	0	0	0	608	0	–	464	0
Circles	0	0	0	0	0	0	253	0
Ecoli	0	0	–	92	7	38	114	0
Glass	0	0	–	272	1	36	125	0
Haberman	0	0	0	394	0	12	252	0
Hayesroth	0	0	–	67	0	31	54	0
Heart	0	0	0	141	0	39	145	0
Ionosphere	0	0	0	554	0	318	272	0
Iris	0	0	0	0	0	0	1	0
Led7Digit	0	0	–	895	27	–	178	0
Monk2	0	0	0	13	0	0	539	0
Moons	0	0	0	0	0	40	128	0
Movement Libras	0	0	0	128	3	34	77	0
Newthyroid	0	0	0	76	2	49	87	0
Saheart	0	0	0	123	0	–	495	0
Sonar	0	0	–	247	0	27	135	0
Soybean	0	0	0	13	0	0	1	0
Spectfheart	0	0	0	263	0	0	225	0
Spiral	0	0	–	2	0	21	239	0
Tae	0	0	–	49	2	40	62	0
Vehicle	0	0	0	852	211	–	1,725	0
Wine	0	0	0	0	0	0	14	0
Zoo	0	0	0	9	0	0	1	0
Mean	0	0	0 *	259	10	33*	238	0

*Nan values (–) are ignored when computing the sum.

Table W15: Average number of cannot-link constraint violations of the PCCC and the PCCC-R algorithms and the four state-of-the-art algorithms (COPKM, CSC, DILS, LCC) for the constraint sets of size 15% CS. The lowest values are stated in bold. The column KMEANS reports the average number of cannot-link constraint violations obtained with the k-means algorithm. The hyphen indicates that the respective algorithm returned no solution within the time limit of 1,800 seconds.