

Dataset	20% CS						KMEANS	GT
	PCCC	PCCC-R	COPKM	CSC	DILS	LCC		
Appendicitis	0	0	–	0	0	20	19	0
Breast Cancer	0	0	0	2,930	0	–	587	0
Bupa	0	0	0	1,116	0	0	875	0
Circles	0	0	0	0	0	–	468	0
Ecoli	0	0	–	136	13	141	198	0
Glass	0	0	–	157	1	63	215	0
Haberman	0	0	0	745	0	0	473	0
Hayesroth	0	0	–	124	0	24	106	0
Heart	0	0	0	27	0	0	219	0
Ionosphere	0	0	0	1,075	0	–	463	0
Iris	0	0	0	0	0	0	1	0
Led7Digit	0	0	–	1,445	98	449	327	0
Monk2	0	0	0	82	0	–	895	0
Moons	0	0	0	16	0	0	225	0
Movement Libras	0	0	–	273	15	98	147	0
Newthyroid	0	0	0	29	0	50	193	0
Saheart	0	0	0	1,914	0	–	928	0
Sonar	0	0	0	416	0	0	231	0
Soybean	0	0	0	2	0	1	0	0
Spectfheart	0	0	0	463	0	0	390	0
Spiral	0	0	–	0	0	0	432	0
Tae	0	0	–	106	0	63	120	0
Vehicle	0	0	0	5,635	373	–	3,088	0
Wine	0	0	0	4	0	0	23	0
Zoo	0	0	0	29	0	1	4	0
Mean	0	0	0 *	669	20	48*	425	0

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

Table W16: 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 20% 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.