

Dataset	Objects	Features	Clusters	5% CS						KMEANS
				PCCC-N2-S	PCCC-N2-S-RD	COPKM	LCC	CSC	DILS	
Banana	5,300	2	2	0	0	—	—	0	—	—
Letter	20,000	16	26	10,796	2,635	—	—	—	—	—
Shuttle	57,999	9	7	3,722	0	—	—	—	—	—
CIFAR10	60,000	3,072	10	629,094	0	—	—	—	—	—
CIFAR100	60,000	3,072	100	83,659	48,522	—	—	—	—	—
MNIST	70,000	784	10	355,448	0	—	—	—	—	—
Mean				180,453	8,526	—*	—*	0*	—*	—

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

Table W98: Average number of cannot-link constraint violations of the PCCC and the PCCC-N2-S algorithms for the constraint sets of size 5% 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.