

Dataset	Objects	Features	Clusters	1% CS						KMEANS
				PCCC-N2-S	PCCC-N2-S-RD	COPKM	LCC	CSC	DILS	
Banana	5,300	2	2	<b>0</b>	<b>0</b>	—	—	<b>0</b>	—	—
Letter	20,000	16	26	<b>0</b>	<b>0</b>	—	—	—	—	—
Shuttle	57,999	9	7	61	<b>0</b>	—	—	—	—	—
CIFAR10	60,000	3,072	10	5,294	<b>4,191</b>	—	—	—	—	—
CIFAR100	60,000	3,072	100	<b>0</b>	<b>0</b>	—	—	—	—	—
MNIST	70,000	784	10	3,390	<b>964</b>	—	—	—	—	—
Mean				1,458	859	—*	—*	<b>0*</b>	—*	—

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

Table W97: Average number of cannot-link constraint violations of the PCCC and the PCCC-N2-S algorithms for the constraint sets of size 1% 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.