

Dataset	5% CS								KMEANS	GT
	PCCC	PCCC-N2-S	PCCC-N5-S	PCCC-N2-S-RD	COPKM	CSC	DILS	LCC		
n500-k2	0.54	0.54	0.54	0.54	–	-0.15	0.51	0.54	0.54	0.54
n500-k5	0.65	0.65	0.65	0.65	0.60	-0.17	0.32	0.49	0.70	0.61
n500-k10	0.53	0.53	0.53	0.53	0.45	-0.41	-0.07	0.52	0.56	0.50
n500-k20	0.39	0.39	0.39	0.39	0.36	-0.53	-0.21	0.35	0.39	0.21
n500-k50	0.37	0.37	0.37	0.37	0.36	-0.72	-0.45	0.35	0.37	0.07
n500-k100	0.35	0.35	0.35	0.35	0.32	-0.73	-0.64	0.31	0.37	-0.11
n1000-k2	0.54	0.54	0.54	0.54	0.54	0.54	0.53	0.54	0.54	0.54
n1000-k5	0.63	0.63	0.63	0.63	–	-0.45	0.13	0.59	0.66	0.60
n1000-k10	0.53	0.53	0.53	0.53	0.50	-0.34	-0.11	0.44	0.55	0.51
n1000-k20	0.35	0.35	0.35	0.36	0.35	-0.37	-0.23	0.33	0.39	0.24
n1000-k50	0.33	0.33	0.33	0.34	0.32	-0.70	-0.33	0.34	0.36	0.10
n1000-k100	0.36	0.36	0.36	0.37	0.34	-0.77	-0.47	–	0.37	-0.06
n2000-k2	0.55	0.55	0.55	0.55	0.55	0.55	0.55	0.29	0.55	0.55
n2000-k5	0.62	0.62	0.62	0.62	0.62	-0.23	-0.09	0.51	0.70	0.60
n2000-k10	0.53	0.53	0.53	0.53	0.49	-0.31	-0.16	0.48	0.55	0.51
n2000-k20	0.32	0.32	0.32	0.33	0.29	-0.30	-0.19	0.32	0.39	0.24
n2000-k50	0.31	0.31	0.31	0.31	0.30	-0.57	-0.26	0.32	0.35	0.10
n2000-k100	0.30	0.30	0.30	0.31	0.30	-0.78	-0.33	0.24	0.34	-0.05
n5000-k2	0.54	0.54	0.54	0.54	–	0.54	0.25	0.51	0.54	0.54
n5000-k5	0.61	0.61	0.61	0.61	0.59	0.14	-0.06	0.54	0.67	0.60
n5000-k10	0.51	0.51	0.51	0.51	–	-0.10	-0.06	0.07	0.57	0.51
n5000-k20	0.25	0.26	0.25	0.29	–	-0.13	-0.10	0.18	0.38	0.23
n5000-k50	-0.11	0.25	0.25	0.26	0.22	-0.22	-0.14	0.25	0.35	0.10
n5000-k100	-0.16	0.26	0.26	0.26	0.24	-0.54	-0.20	0.25	0.34	-0.05
Mean	0.41	0.44	0.44	0.45	0.11*	-0.28	-0.08	0.32*	0.48	0.32

*Nan values (–) are replaced with -1 before computing the mean.

Table W86: Average Silhouette coefficients of the versions of the PCCC algorithm and the four state-of-the-art algorithms (COPKM, CSC, DILS, LCC) obtained with constraint sets of size 5% CS. Higher values indicate better separated clusters. The highest values are stated in bold. The column KMEANS reports the average Silhouette coefficients that were obtained with the unconstrained k-means algorithm. The column GT reports the Silhouette coefficients of the ground truth assignment. The hyphen indicates that the respective algorithm returned no solution within the time limit of 3,600 seconds.