





























**Table 5: Clustering quality assessment for TRACCLUS and ppTRACCLUS on the Deer data set. For all scores larger values are better (best marked in bold).**

Data Set	Score	Experiment 1		Experiment 2	
		TRACCLUS	ppTRACCLUS	TRACCLUS	ppTRACCLUS
Deer	$(\epsilon', \text{minLns})$	(400, 3)	$(1 \times 10^6, 3)$	(282, 3)	$(550 \times 10^3, 3)$
	# of Clusters	1	1	2	2
	Noise	1	480	20	1333
	SC	N/A	N/A	0.089	<b>0.36</b>
	SC <sub>noise</sub>	N/A	N/A	0.089	<b>0.34</b>
	DBCW	N/A	N/A	0.47	<b>0.79</b>

Nevertheless, having only one cluster does not necessarily mean that the quality of the clustering algorithm is low. It simply says that the algorithm found only one group of similar elements.

Tab. 6 contains the results of Experiment 2 of TRACCLUS, ppTRACCLUS, and ppTRACCLUS' on the Travel data set.

**Table 6: Results of Experiment 2 of the clustering quality assessment for TRACCLUS, ppTRACCLUS, and ppTRACCLUS' on the Travel data set. The best results are marked in bold.**

	Experiment 2		
	TRACCLUS	ppTRACCLUS	ppTRACCLUS'
$(\epsilon', \text{minLns})$	$(47 \times 10^3, 3)$	$(13.5 \times 10^9, 3)$	$(13.5 \times 10^9, 3)$
# of Clusters	1	1	2
Noise	0	359	361
SC	N/A	N/A	<b>0.82</b>
SC <sub>noise</sub>	N/A	N/A	<b>0.81</b>
DBCW	N/A	N/A	<b>0.98</b>

## D.2 Runtime and Communication Costs

Tab. 7 contains the runtimes (in seconds) and Tab. 8 contains the communication costs (in MB) of the SED, the approx. distance of TRACCLUS, and clustering averaged over 10 experiments. The SED distance is applied on 2-dimensional data, the approx. TRACCLUS distance on 4-dimensional line segments. *maxIterations* is set to 4.

**Table 7: Runtimes for ppDBSCAN/ppTRACCLUS.**

Data Set Size n	Runtimes (s)		
	Squared Euclidean Distance	Appr. TRACCLUS-Distance	Density-based Clustering
100	2.36	4.36	10.9
200	8.04	14.82	58.43
300	17.35	28.94	171.12
400	29.03	45.39	391.69
500	44.17	64.98	750.284
600	61.04	86.91	1317.70
700	79.76	113.07	2043.01
800	102.99	144.09	2970.52
900	125.82	177.48	4187.54
1000	157.07	208.92	5682.425

**Table 8: Communication Costs for ppDBSCAN/ppTRACCLUS.**

Data Set Size n	Communication (MB)		
	Squared Euclidean Distance	Appr. TRACCLUS-Distance	Density-based Clustering
100	144	328	460
200	575	1317	3566
300	1294	2968	11912
400	2300	5280	28090
500	3594	8253	54694
600	5176	11889	94315
700	7044	16185	149548
800	9201	21143	222983
900	11645	26762	317214
1000	14376	33043	434835