

BIG DATA COMPUTING 2020/21 – HOMEWORK 2

Do the tests with the parameters indicated in the first 3 columns of the following table. Fill the table reporting, for each test, 4 values: the approximate average silhouette coefficient of the input clustering (**approxSilhFull**); the time to compute approxSilhFull; the exact silhouette coefficient of the clustering sample (**exactSilhSample**); the time to compute exactSilhSample. The running times must not include the time to load the input file and to compute the sample. The values in the table for Uber_3_small.csv must be the average of 3 runs of the same experiment.

Table to be filled

Dataset	k	t	approxSilhFull	Time (ms) to compute approxSilhFull	exactSilhSample	Time (ms) to compute exactSilhSample
Uber_10_large.csv	10	2000	0.628	702745	0.526	11204
Uber_3_large.csv	3	2000	0.479	127136	0.464	2364
Uber_3_small.csv	3	40	0.911	111	0.715	10
Uber_3_small.csv	3	80	0.968	180	0.811	24
Uber_3_small.csv	3	1000	0.914	536	0.914	126