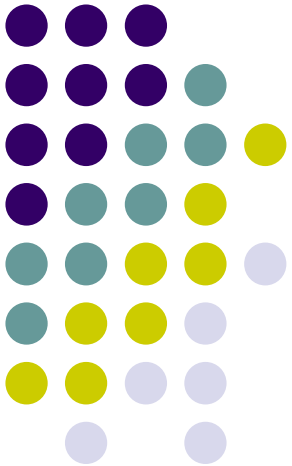
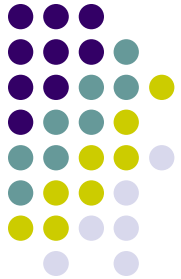
Big Data Analytics

Third Assignment



Deadline: 1 February 2021



1) Create in Quanteda a corpus starting from the folder that includes the Irish party manifestoes of 1992 and 1997

- Analyze such texts via Wordscores, using as reference texts the 1992 party programs. Party reference scores: DL 4.5; Lab 6.88; FF 13.13; FG 15; PDs 17.63 (such scores refer to a left-right economic scale). Run the analysis to obtain raw scores. Plot and comment your results
- Analyze such texts via Wordfish placing DL 92 to the left of FG 92
- Compare the results obtained via Wordfish with the results you get via Wordscores and explain the differences you get

Deadline: 1 February 2021

- Now use the results you got via the Wordfish analysis as the reference scores for DL, FF and PDs (all from 1992 elections) and estimate once again the Wordscores raw scores, while do not add any reference score this time for FG and Lab
- Which are the main differences with respect to the previous analysis? Is the correlation between Wordfish and Wordscores scores increasing or decreasing? Why?

Deadline: 1 February 2021

2) Make any query you like in Twitter either via streaming or via a geographical analysis or both. Then create a DfM and analyze your results as you like (according to what you learnt up to now)