Lab 3 addition - Plot on full data set, Fall 2017
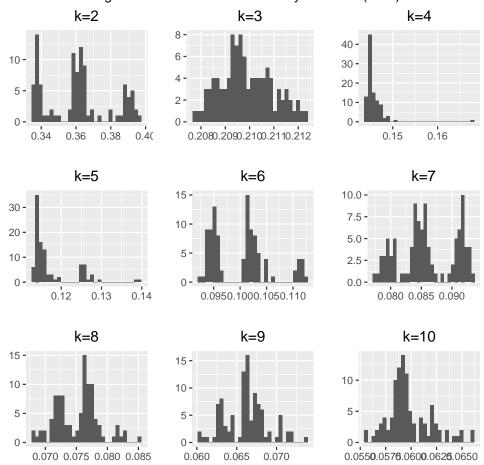
Melody Liu

October 23, 2018

I also ran the full data set with similarity function in CPP. It took me around 7 hours in my local machine. Results and plots will be shared in the report.

I plotted the two graphs with results generated from the full data set. Very similar pattern of cumulative distributions for increasing values of k was observed while distribution of similarity measure differs from the previous plot. k=2 still achieved the highest similarity score which is optimal for the choice of k.

Histogram of the correlation similarity measure (CPP)

Cumulative distributions for increasing values of k (CPP)....