

Analysis of Clusters - OPTICS 2

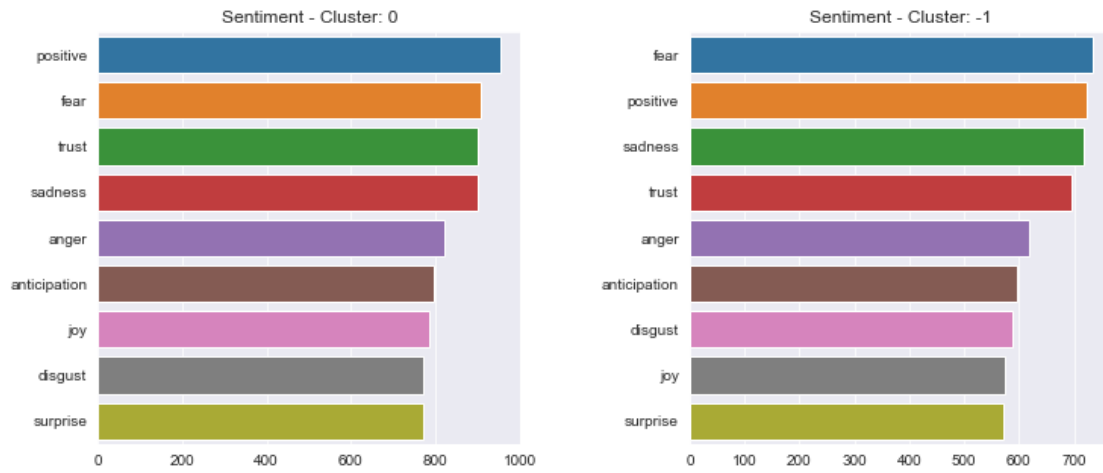
October 20, 2020

1 Analysis of Clusters - OPTICS 2

This run of OPTICS used default sklearn parameters with a max epsilon of 0.05 and min samples per cluster of 200. The algorithm used the cosine dissimilarity between document vectors as distance metric. Two clusters were produced, both of similar size. The growth of these wasn't tested with addition of more data. This seemed to be the most promising clustering, although low number of clusters and low max epsilon parameter may be cause for concern.

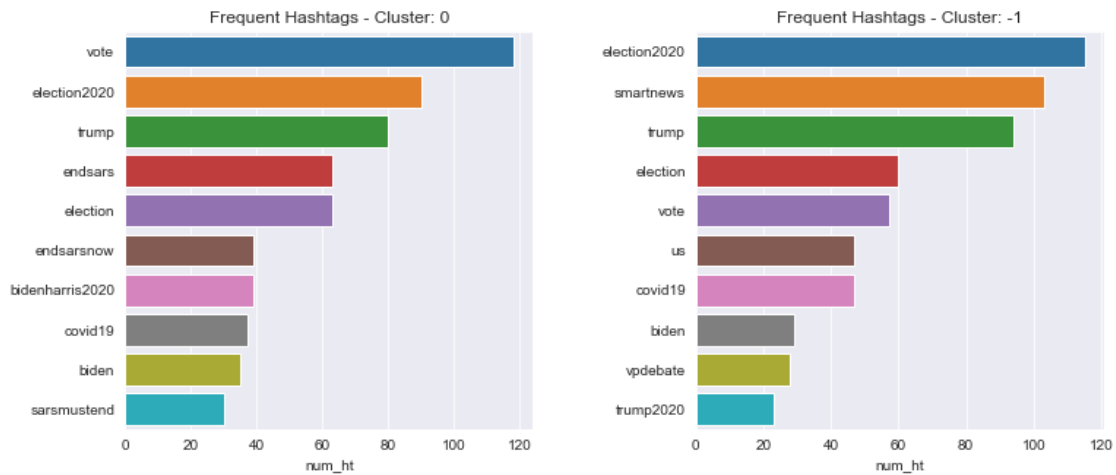


2 Cluster Sentiment



Not a significant difference in ordering identified.

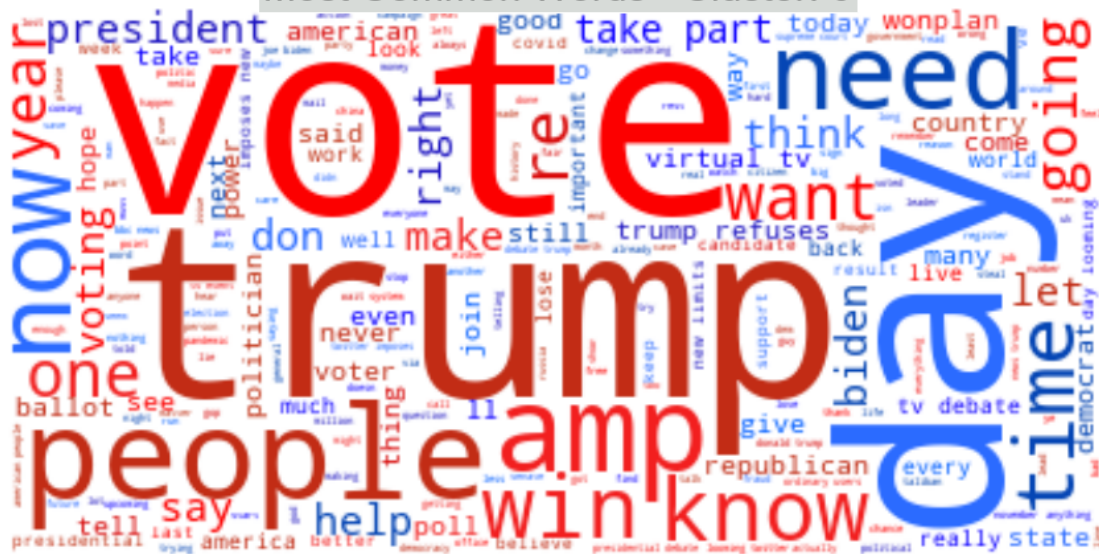
3 Cluster Hashtags



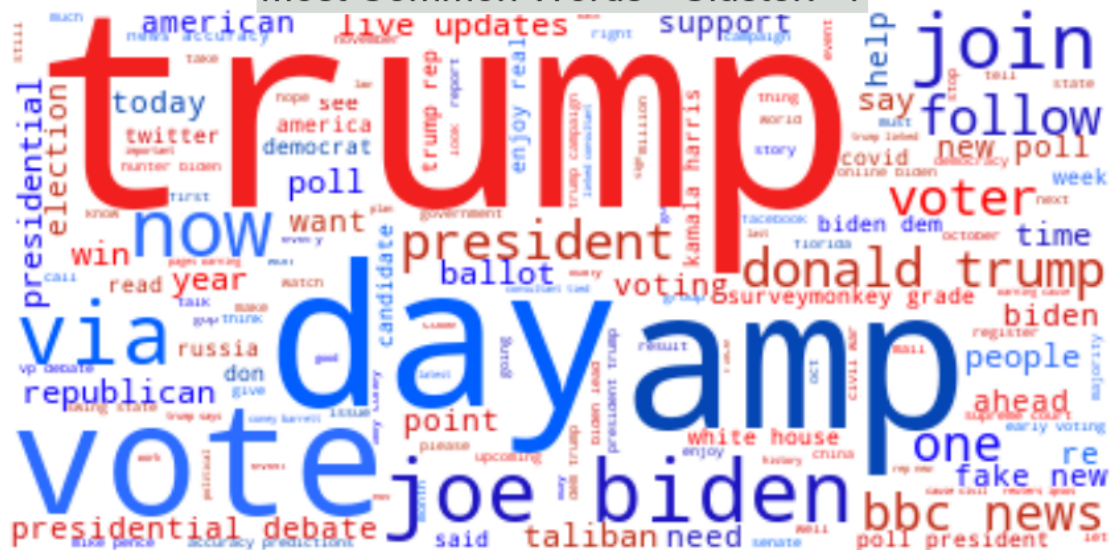
It difficult to conclude anything from these, beyond that sars based tweets seem mostly in Cluster 0. It is unclear why so many sars tweets showed up in the scrape (sars hashtags are in support of police brutality protests in Nigeria - it could be linking into the police brutality discussions surrounding the US election).

4 Cluster Wordclouds

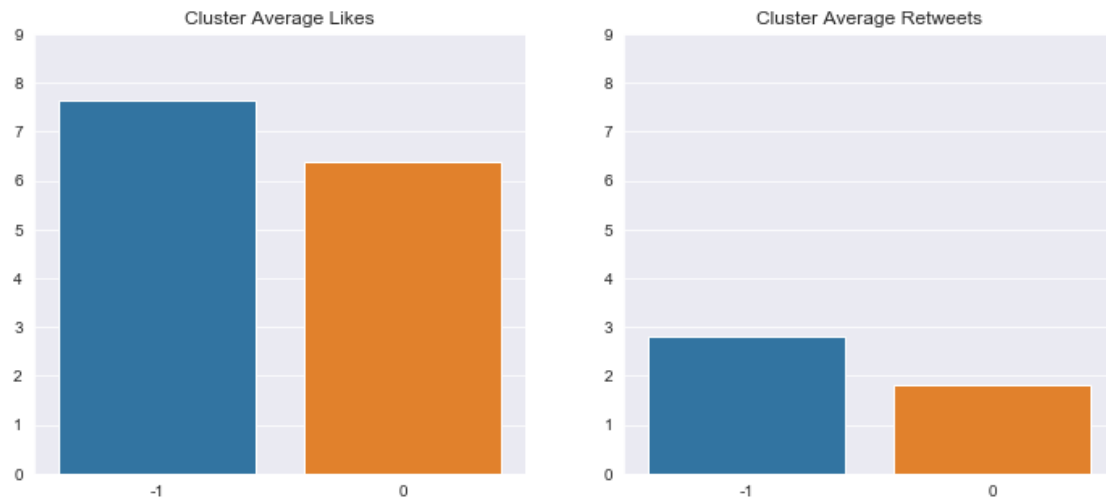
Most Common Words - Cluster: 0



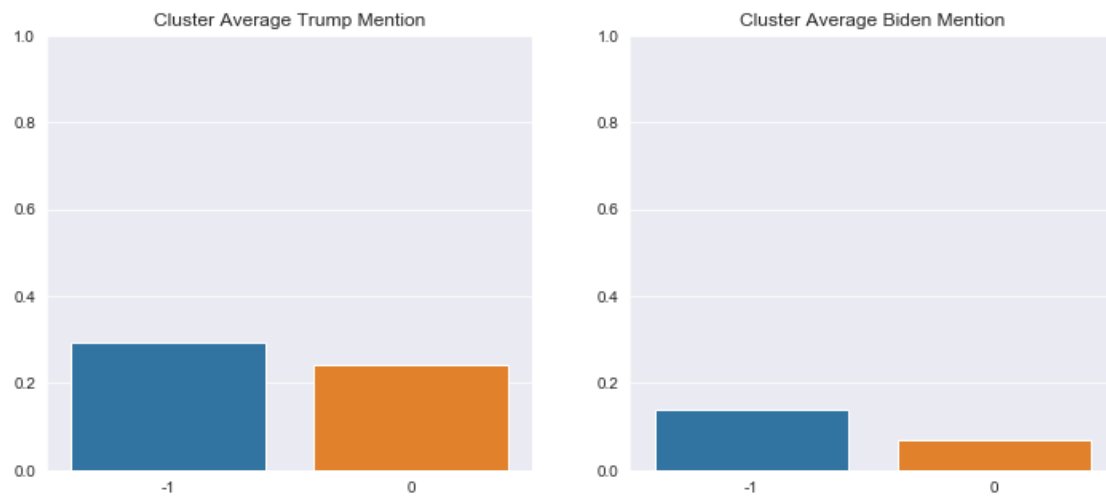
Most Common Words - Cluster: -1



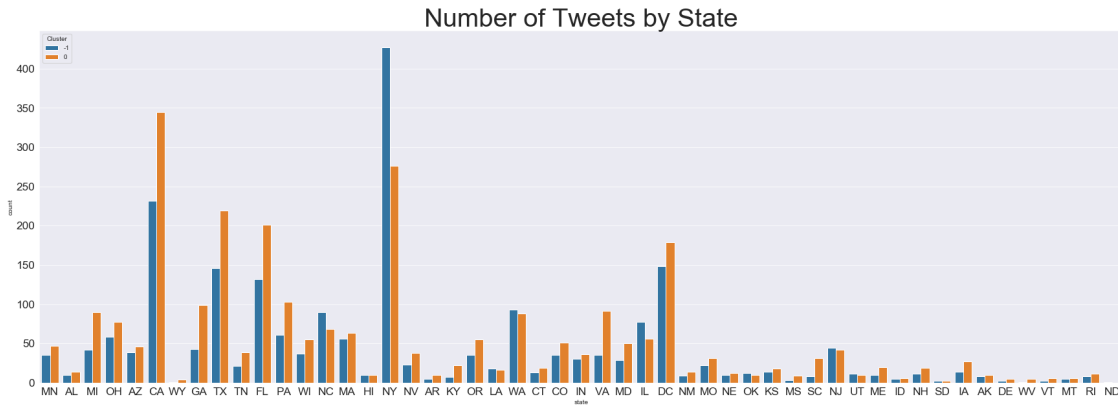
5 Cluster Likes and Retweets



6 Cluster Candidate Mentions



7 Cluster Locations



8 Overall Conclusions

While at the face of it, this looked like a more promising clustering, the clusters seemed to be broadly similar. This is likely a result of such a small max epsilon parameter being chosen, but also due to the general lack of clusterability within the tweets themselves.