

French Presidential Election Candidates Tweets

Emma Kerinec Nicolas Derumigny

ENS Lyon

11 May 2017



- tweets of the 11 French presidential election candidates.
- work in Python, use of



Preprocessing

- use only relevant words.
- distinguish hashtags with words.



Compute the most used words for a given candidate.

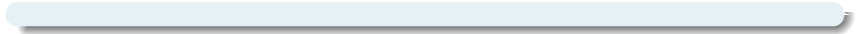


Distance between two set of tweets:

$$d(set1, set2) = \frac{1}{2} \cdot \left(\sum_{word \in set1, word \notin set2} (freq(word)) + \left(\sum_{word \in set2, word \notin set1} (freq(word)) \right) \right)$$



Distance between two candidates



Kmeans

Clustering candidates by similarities between used words, or tweets.



Hierarchical clustering

Can do the same by using Hierarchical clustering.



Time

Cluster the candidates for some time periods:

- before and after the beginning presidential campaign
- during the first and the second part of the presidential campaign



A priori algorithm

Compute for each candidates the frequent words patterns.



Difficulties



End

Thank you for your listening.

Does anybody have any question?

