

# French Presidential Election Candidates Tweets

Emma Kerinec    Nicolas Derumigny

ENS Lyon

11 May 2017



# Introduction

- Dataset: tweets of the 11 French presidential election candidates.
- Language: Python
- Package: Sklearn, Numpy, Matplotlib



# Working on words

## Preprocessing

- Keep only relevant words → words with more than 5 letters.
- Distinguish hashtags with words.
- Find common points between candidates without semantic analysis



# Simple Data

Compute the most used words for a given candidate.



# Distance between tweets

## Distance of sets of words

Measure the proportion of words that are different between two set of words  $S_1$  and  $S_2$ :

$$d(S_1, S_2) = \frac{1}{2} \cdot \left( \sum_{w \in S_1, w \notin S_2} f(w) + \sum_{w \in S_2, w \notin S_1} f(w) \right)$$



# Distance between two candidates

**Figure:** Distance between candidates, projected on the distance with ??? on  $x$  and the distance with ??? on  $y$

- result in a 11-dimensional space

# Kmeans

**Figure:** Clustering candidates by similarities between used words and hashtags.

# Hierarchical clustering

**Figure:** Clustering candidates by similarities between used words and hashtags.



# Evolution over 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

```
Rule: ('#fillon',) → ('français', 'dupontaignan') , 0.149
Rule: ('#fillon', 'dlf_officiel') → ('#macron',) , 0.222
Rule: ('#fillon', 'dlf_officiel') → ('dupontaignan',) , 0.988
Rule: ('#le79inter',) → ('dupontaignan',) , 1.000
Rule: ('#legrandjury',) → ('dupontaignan',) , 1.000
...
Rule: ('judiciaire',) → ('casier',) , 0.937
Rule: ('judiciaire',) → ('vierge',) , 0.875
Rule: ('judiciaire',) → ('vierge', 'casier') , 0.875
Rule: ('élection',) → ('dupontaignan', 'dlf_officiel') , 0.556
```

- Lots of auto-citations
- Very few real political expressions, except "Front National"



# Conclusion

Thank you for listening.

Does anybody have questions?

