

Quelle modélisation de l'espace politique français sur Twitter ?

Nicolas HERVÉ

INA

27 jan. 2021

Démarche globale d'analyse de l'écosystème médiatique :

- analyse de la propagation des informations sur Internet et dans les médias
- choix éditoriaux guidés par des considérations économiques, journalistiques ou politiques ?
- influence des discussions sur Twitter sur ces choix ? [Cagé et al., 2020]

Objectifs :

- construire un espace politique latent pour analyser des contenus : hashtags, URLs, médias, événements, images, ...
- valider, interpréter et visualiser cet espace
- quantifier la polarisation politique de ces contenus

Deux approches possibles

- [Barberá, 2015] : AFC sur matrice d'adjacence
- [Garimella and Weber, 2017] : simple combinaison linéaire pondérée

Travaux existants en France : [Briatte and Gallic, 2015]

Espace politique latent

Polarisation
politique



LA RÉPUBLIQUE
En Marche !

MOUVEMENT
DEMOCRATE



les
IR
Républicains



Propagation de la polarisation

Contenus
tweetés

#hashtag1

#hashtag2

#hashtag3

Espace politique latent



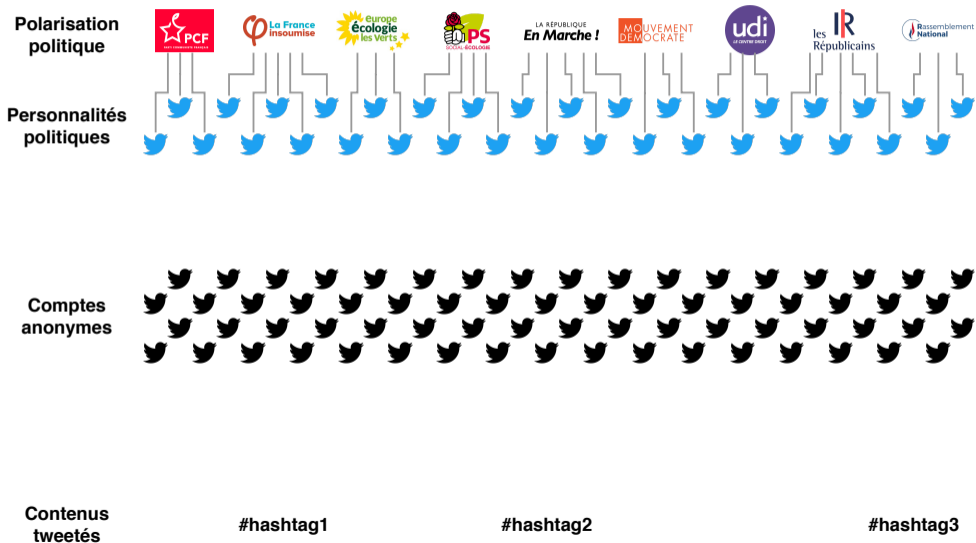
Contenus
tweetés

#hashtag1

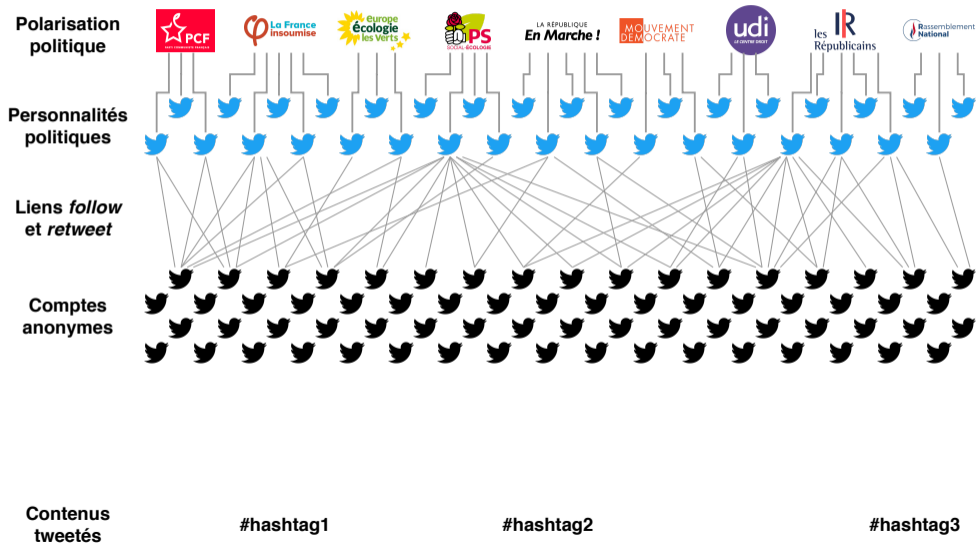
#hashtag2

#hashtag3

Espace politique latent



Espace politique latent



Espace politique latent

Polarisation politique



Personnalités politiques



Liens *follow* et *retweet*



Comptes anonymes



Contenus tweetés

#hashtag1

#hashtag2

#hashtag3

Espace politique latent

Polarisation politique



Personnalités politiques



Liens follow et retweet



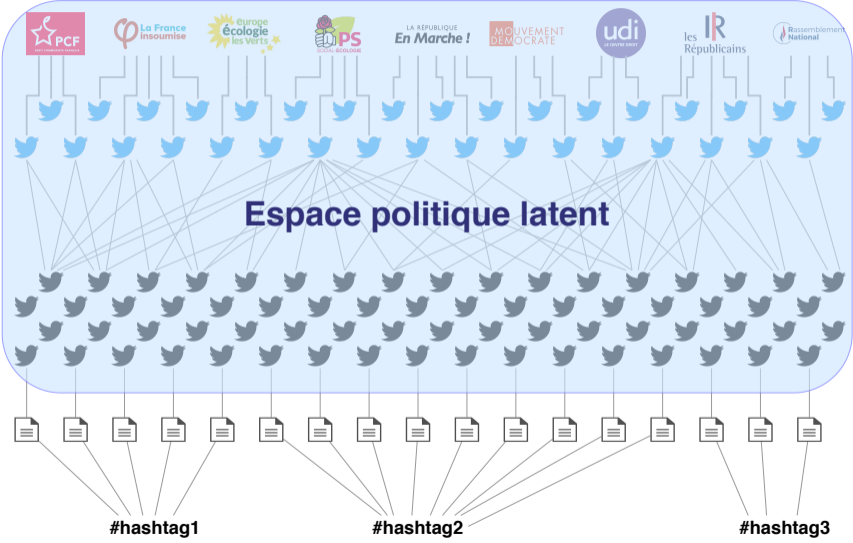
Comptes anonymes



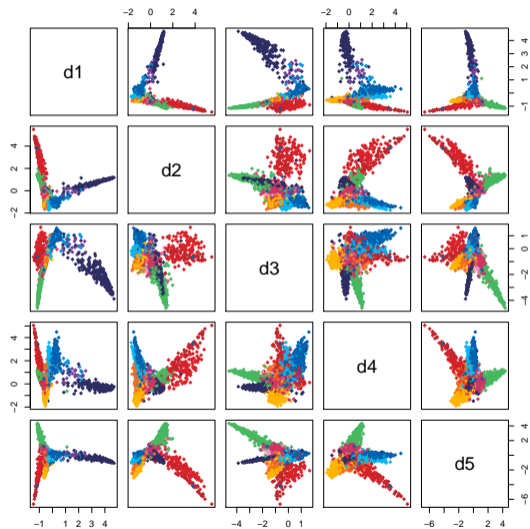
Tweets



Contenus tweetés

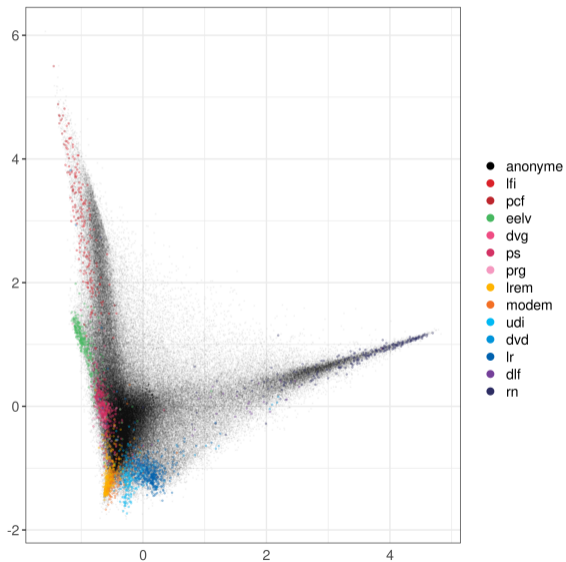


Espace Barberá/follow

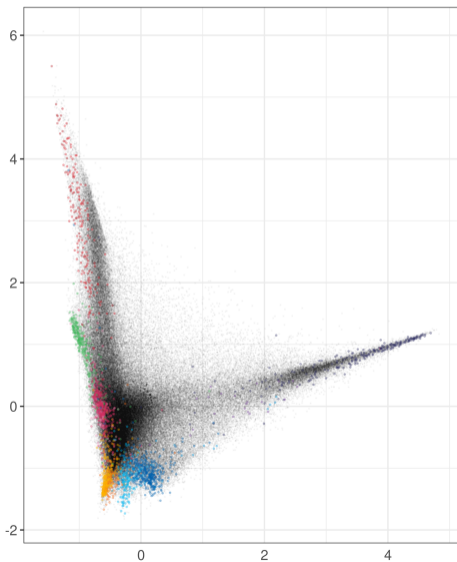


- 4 377 comptes politiques suivis par au moins 25 comptes anonymes
- 184 229 comptes actifs suivant au moins 5 comptes politiques
- ces comptes couvrent 12.5% des tweets captés, 8% de journalistes
- 2 235 comptes politiques sont également actifs

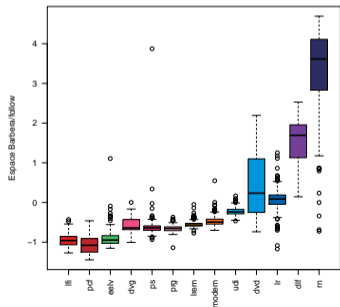
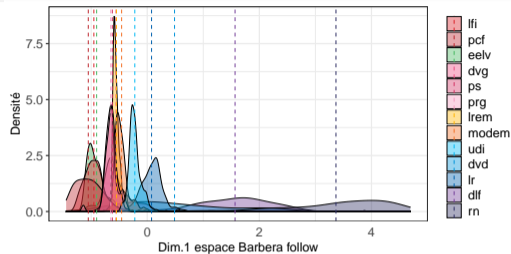
Espace Barberá/follow - interprétation



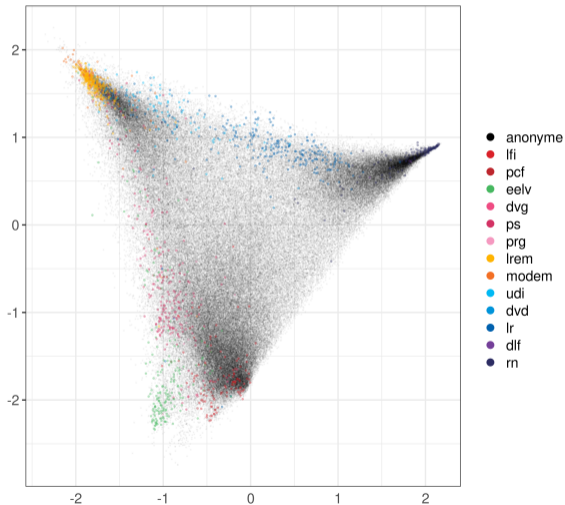
Espace Barberá/follow - interprétation



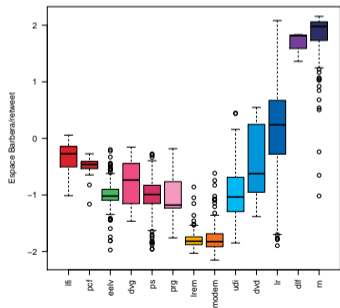
- anonyme
- lfi
- pcf
- eelv
- dvg
- ps
- prg
- lrem
- modem
- udi
- dvd
- lr
- dlf
- rn



Espace Barberá/retweet



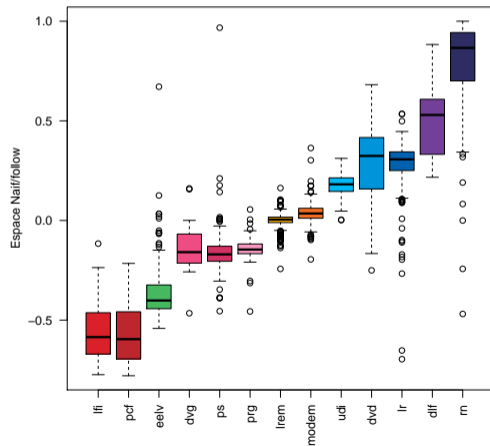
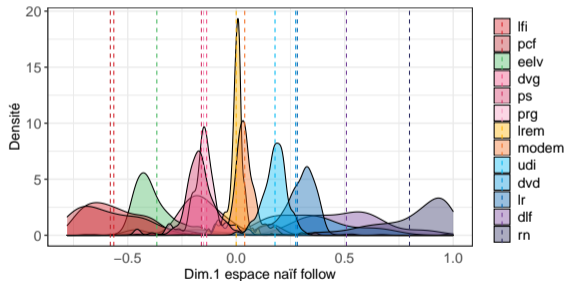
- 1 438 comptes politiques retweetés par au moins 25 comptes anonymes
- 168 393 comptes actifs ayant retweeté au moins 5 comptes politiques
- ces comptes couvrent 22.8% des tweets captés, 3% de journalistes



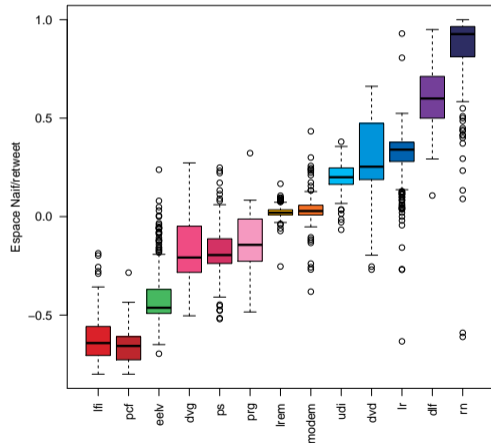
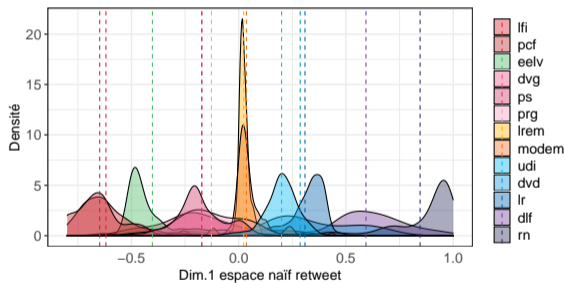
Espace Naïf/follow

lfi	pcf	eelv	dvg	ps	prg	lrem	modem	udi	dvd	lr	dlf	rn
0.1	0.1	0.25	0.4	0.4	0.4	0.5	0.5	0.6	0.6	0.7	0.9	1.0

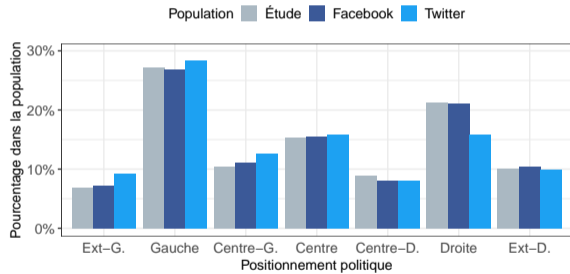
Table: Poids attribués aux partis politiques



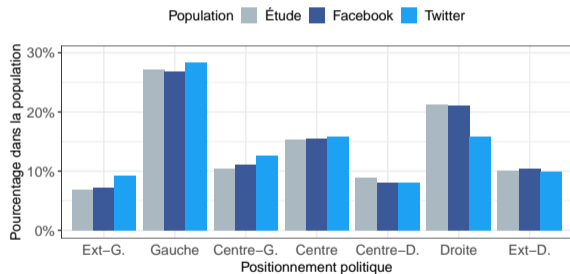
Espace Naïf/retweet



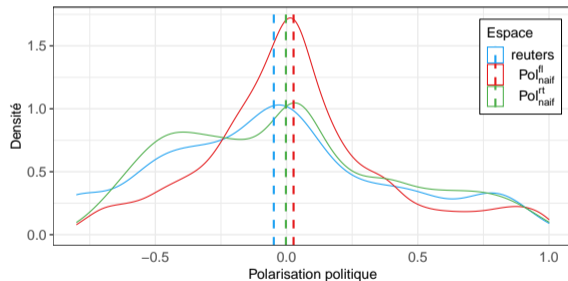
Validation partielle



Validation partielle

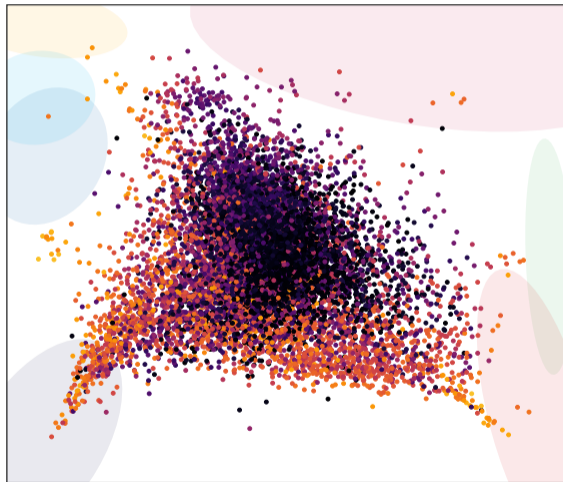


Corrélation avec les résultats d'un sondage :
sur Pol_{naif}^{rt} ($\rho = 0.940$) et Pol_{naif}^{flw}
($\rho = 0.909$)



Visualisation des hashtags - Barberá/follow (carte de Kohonen)

Pct. comptes
politisés



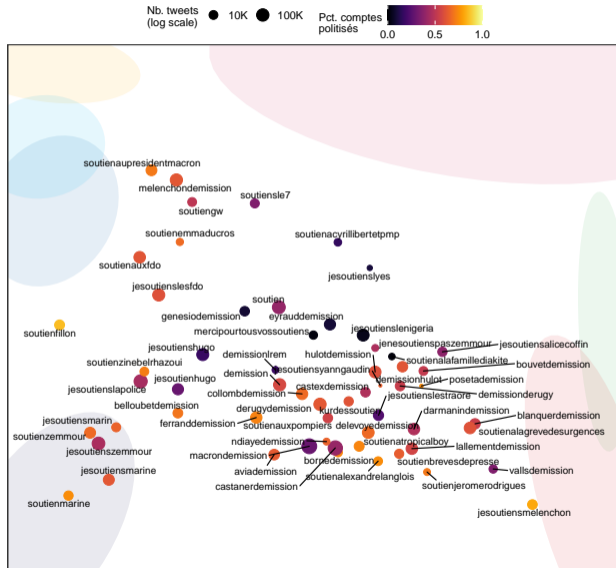
Captation de Twitter francophone selon
[Mazoyer et al., 2018]

- 07/2018 à 07/2020, ~ 5 millions tweets / jour, ~ 3.8 milliards au total
- 1.2M comptes actifs (min. 200 tweets captés, min. 50% fr)

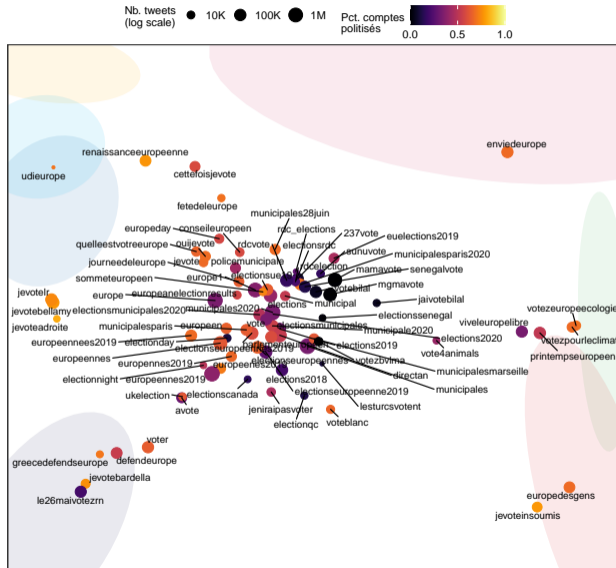
Hashtags sélectionnés dans notre corpus :

- top 100 des hashtags sur une journée ou top 1 000 sur la période
- total de 13 377 hashtags, détectés dans 13.15% des tweets
- visualisation des 12 267 tweetés par au moins 10 comptes politisés

Visualisation des hashtags partisans



Visualisation des hashtags partisans



Conclusions

L'espace francophone sur Twitter est clairement tripolaire

Conclusions

L'espace francophone sur Twitter est clairement tripolaire

Modélisation sur une dimension unique droite-gauche :

- Barberá (AFC) : bon résultat avec les liens de *follow* mais pas avec les *retweet*
- Garimella (naïf) : résultats équivalents, mise en œuvre plus simple et plus souple

Conclusions

L'espace francophone sur Twitter est clairement tripolaire

Modélisation sur une dimension unique droite-gauche :

- Barberá (AFC) : bon résultat avec les liens de *follow* mais pas avec les *retweet*
- Garimella (naïf) : résultats équivalents, mise en œuvre plus simple et plus souple

Pour une même base de comptes politiques de référence :

- informations de *follow* ou *retweet* : caractériser des ensembles de comptes Twitter différents
- donc observer des contenus différents
- bien cerner ce périmètre avant d'interpréter les polarisations induites pour les contenus

- [Barberá 2015] BARBERÁ, Pablo: Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data. In: *Political Analysis* 23 (2015), Nr. 1
- [Briatte and Gallic 2015] BRIATTE, François ; GALLIC, Ewen: Recovering the French Party Space from Twitter Data. In: *Science Po Quanti*. Paris, France, Mai 2015
- [Cagé et al. 2020] CAGÉ, Julia ; HERVÉ, Nicolas ; MAZOYER, Beatrice: Social Media and Newsroom Production Decisions / Social Science Research Network. Rochester, NY, Juli 2020 (ID 3663899). – SSRN Scholarly Paper
- [Garimella and Weber 2017] GARIMELLA, Venkata Rama K. ; WEBER, Ingmar: A long-term analysis of polarization on Twitter. In: *11th International Conference on Web and Social Media, ICWSM 2017*, 2017
- [Mazoyer et al. 2018] MAZOYER, Béatrice ; CAGÉ, Julia ; HUDELLOT, Céline ; VIAUD, Marie-Luce: Real-time collection of reliable and representative tweets datasets related to news events. In: *BroDyn 2018 (ECIR 2018)*, 2018

<http://www.herve.name/twitter-polarization>