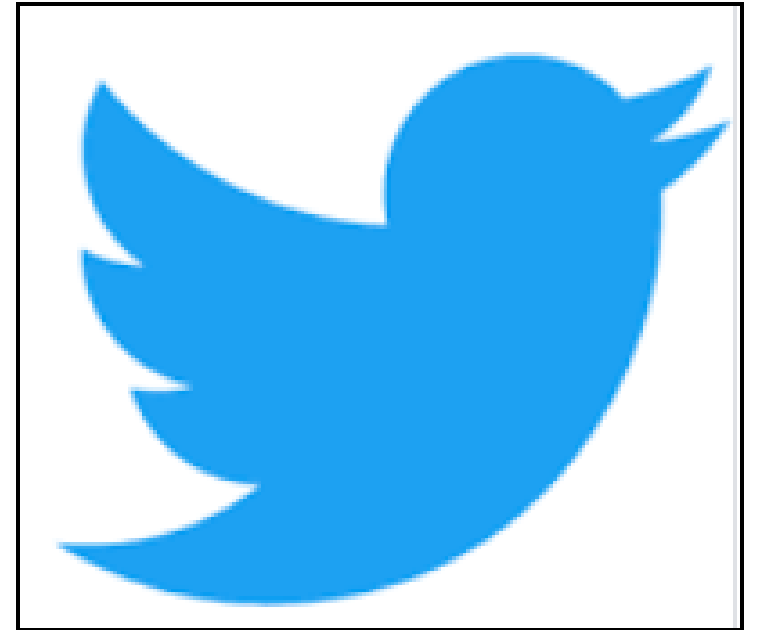


Projet Fouille de Données

Analyse de tweets de la campagne présidentielle 2017

Christine Fouque
Lucas Boscherie
Cyntia Pérez

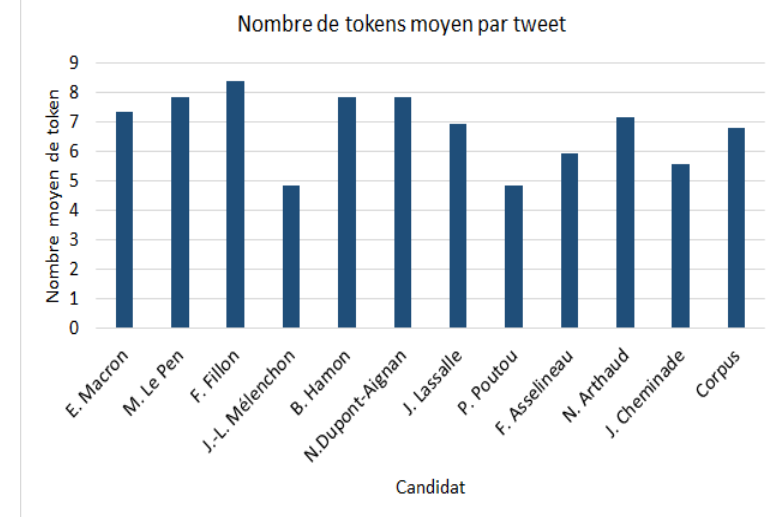
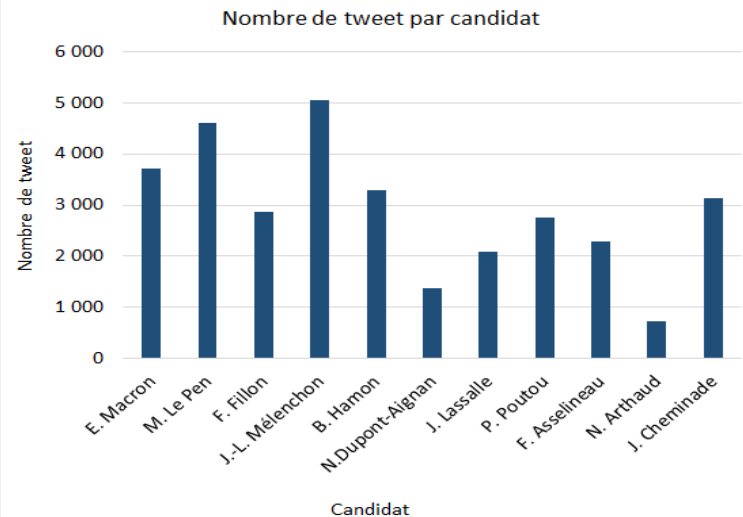
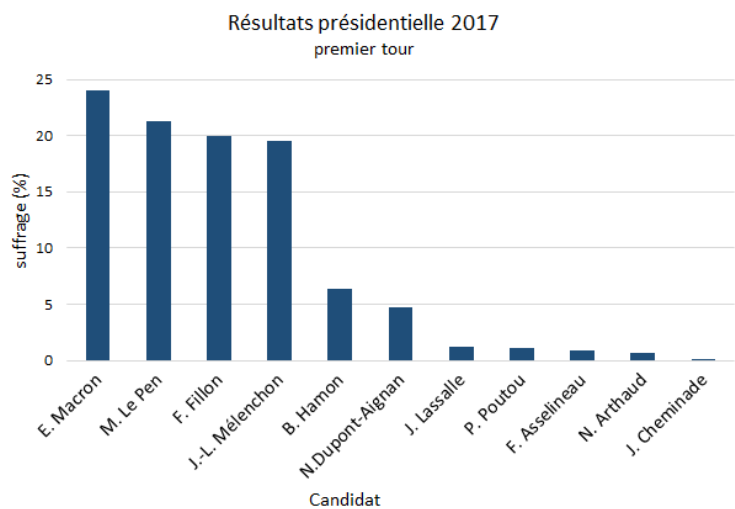
Master 1 Medas – CNAM Nantes



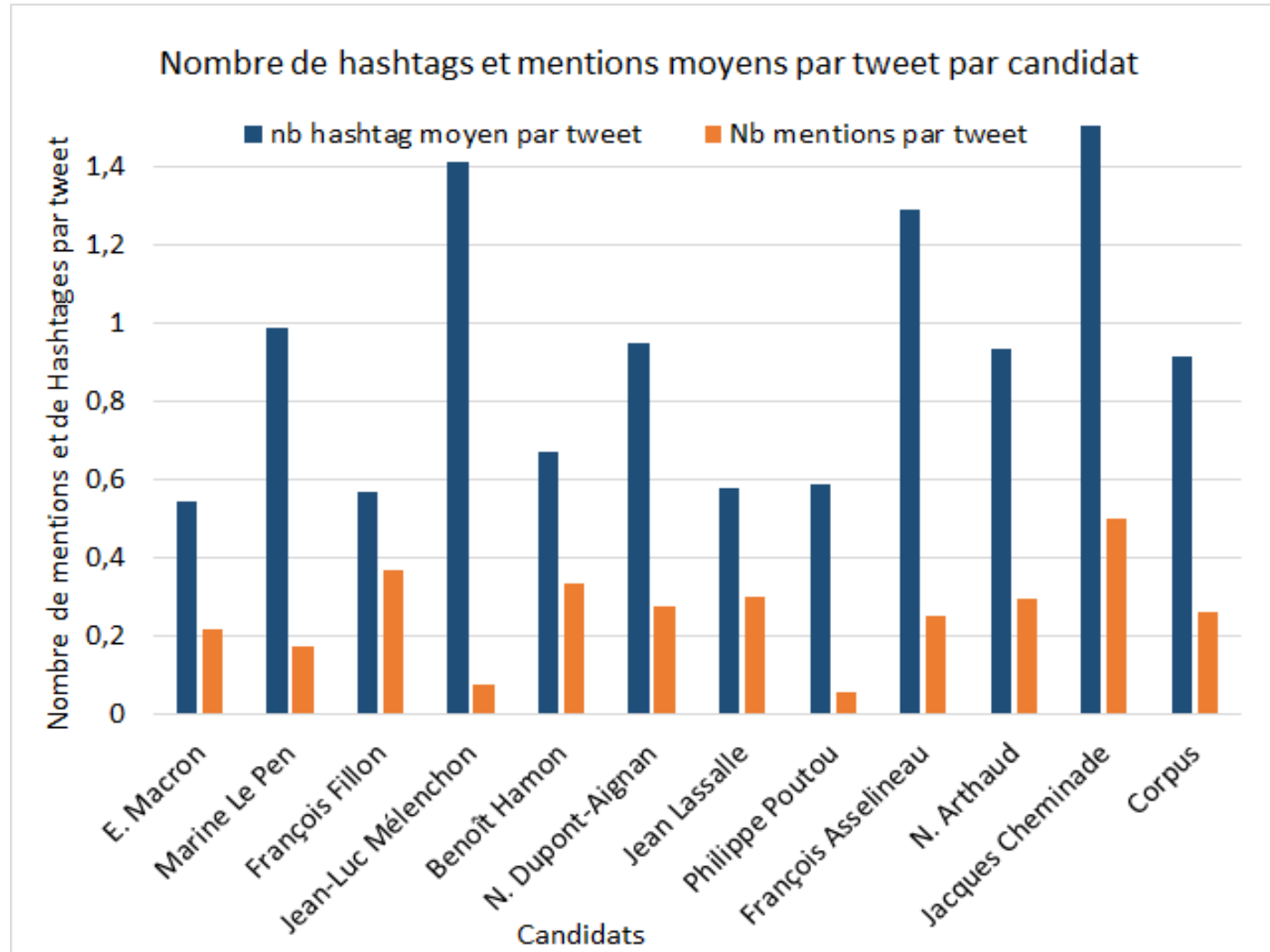
Introduction

- 🐦 Twitter peut être vu comme un indicateur intéressant pour connaître les réactions de ses utilisateurs sur plusieurs sujets : sociaux, politiques, économiques, environnementaux, etc.
- 🐦 Twitter est devenu un moyen de communication central pour l'univers politique permettant d'échanger directement avec les électeurs.
- 🐦 Nos analyses permettront de faire sortir les thèmes principaux, grâce aux techniques d'étiquetage et de lemmatisation. Ce sont les outils élémentaires de chaque traitement automatique de la langue.

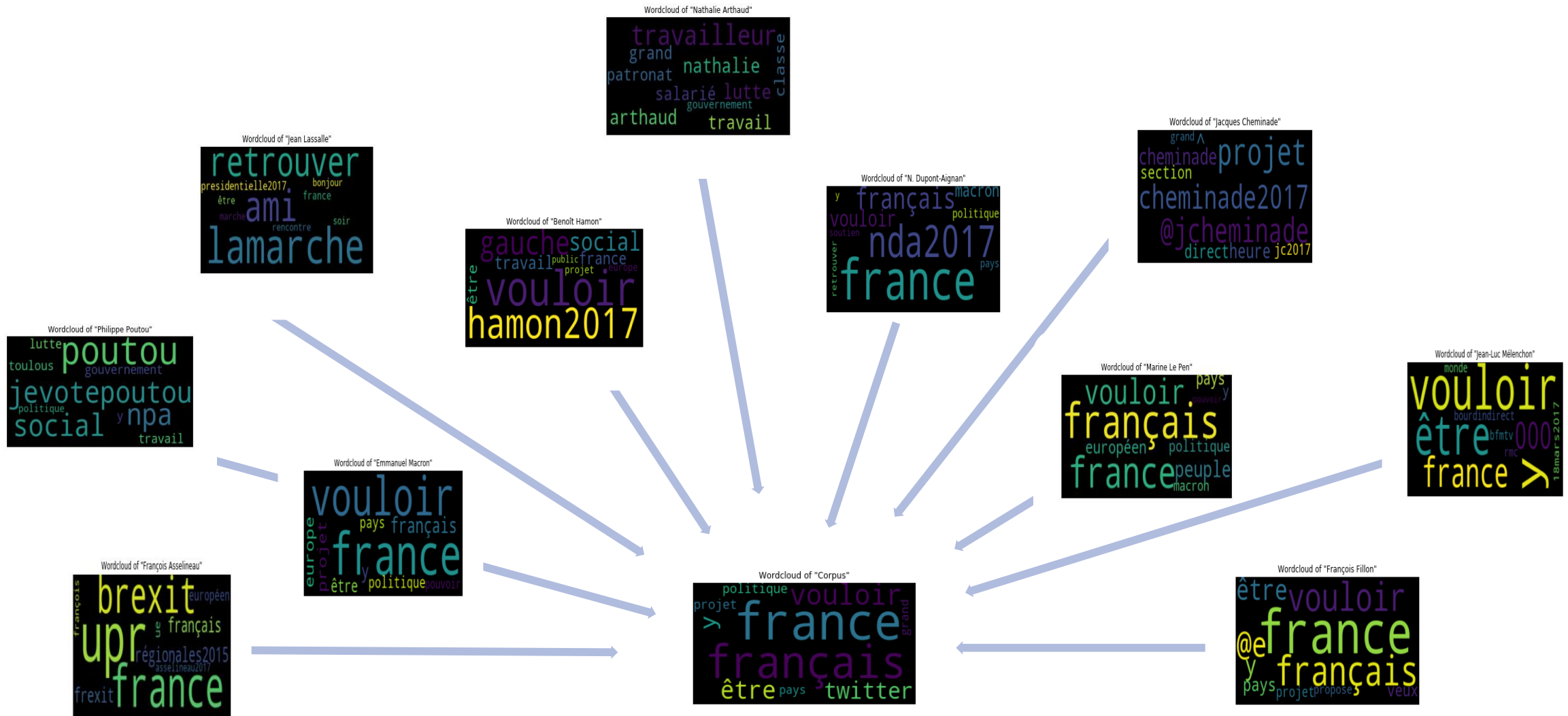
Premiers chiffres clés



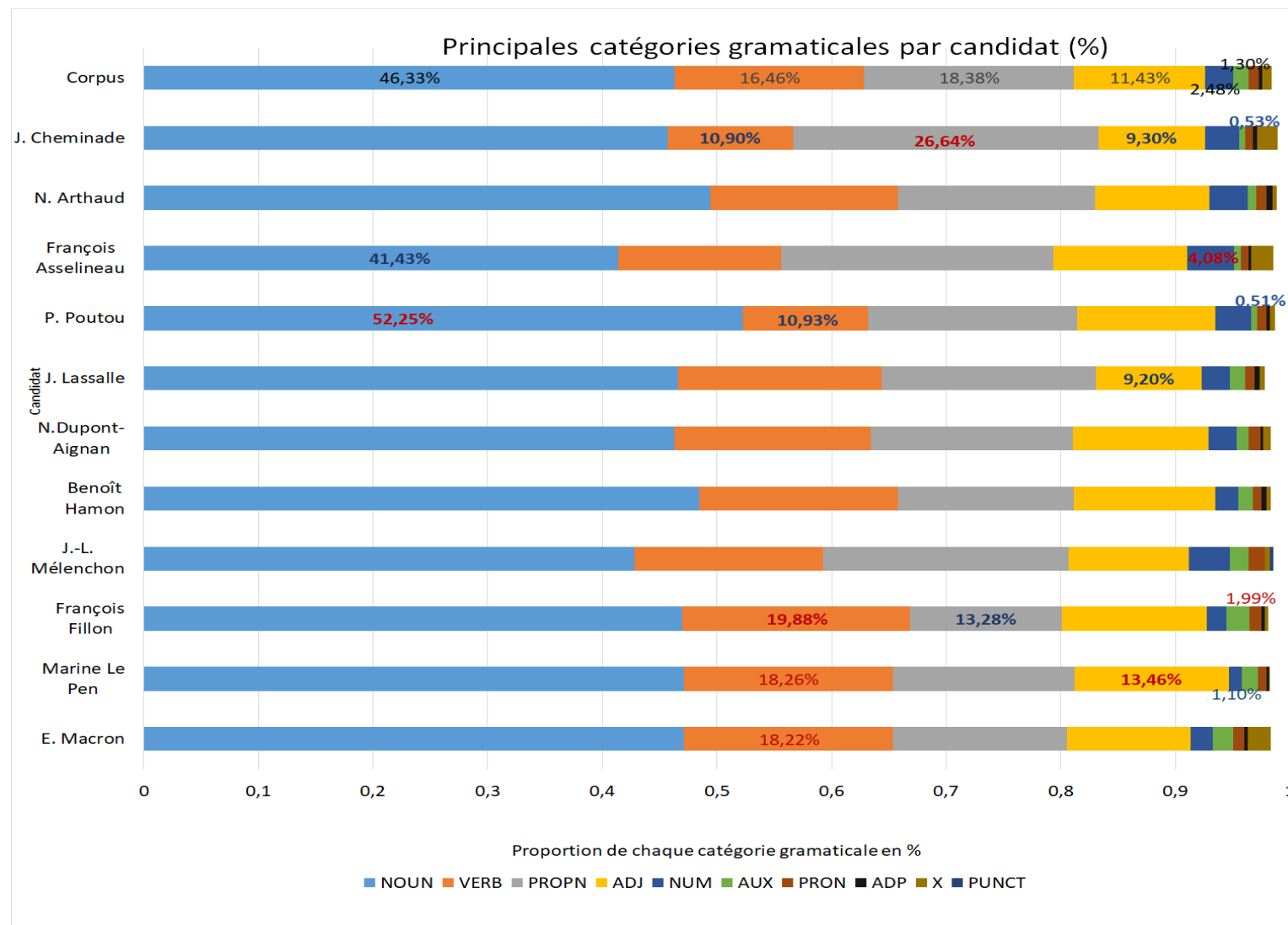
Moyenne de Hashtags & Mentions par candidat



Nuages de mots



Catégories Grammaticales

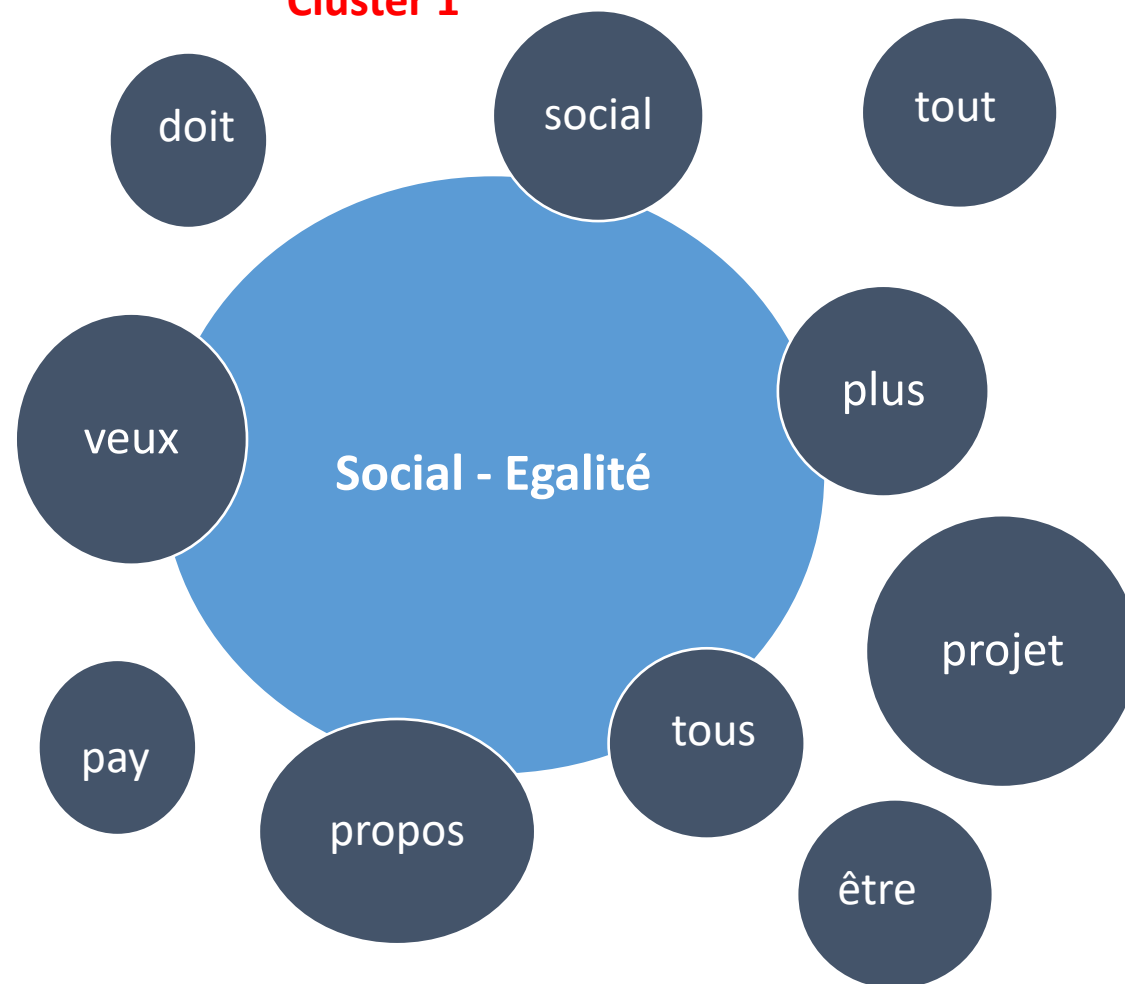


« Top Stemmes » sur le Cluster Global

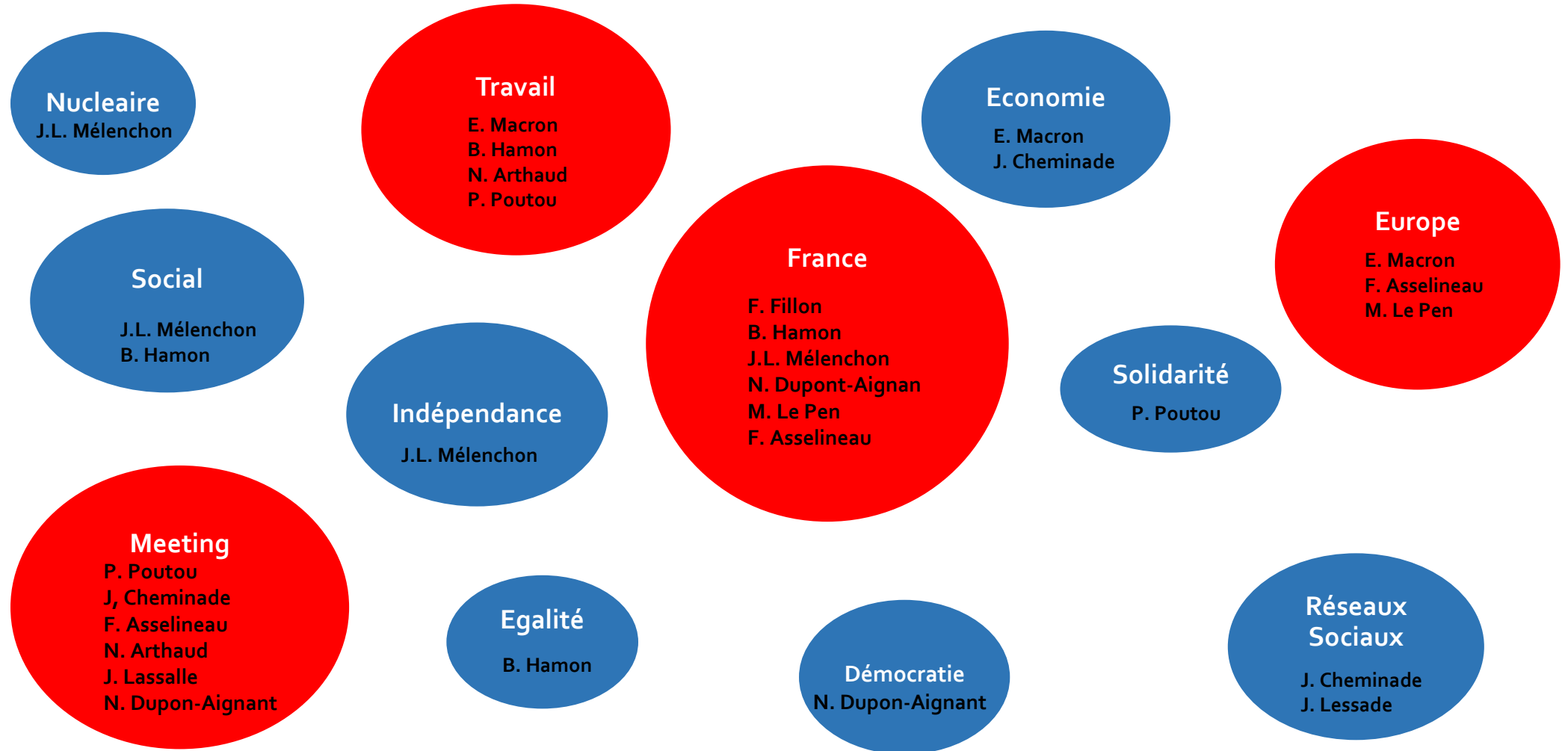
Cluster 0



Cluster 1



Clusters de Thématiques Abordées



Thématiques des Tweets de chaque Candidat

Nathalie Arthaud	Jacques Cheminade	Jean Lassalle	Jean-Luc Mélenchon	François Asselineau	Benoit Hamon	Marine Le Pen	Emmanuel Macron	Philippe Poutou	Nicolas Dupont-Aignan	François Fillon
travail	projet	facebook	franc	européen	franc	européen	travail	droit	franc	franc
lutt	économ	amis	droit	programm	gauch	peupl	europ	social	soutien	social
salari	modernis	photo	social	réunion	droit	national	econom	meeting	démocrat	lutt
ouvri	financi	Itinérair	nucléair	franc	travail	franc	franc	travail	débat	projet
manifest	youtub	march	européen	publicu	européen	pouvoir	projet	solidar	réunion	redress
franc	facebook	soir	constitu	soir	inégal	mélenchon	publique	public	publi	national
travailleur	meeting	franc	indépend		femm	grand	renouvel	lutt	polit	pay

Apprentissage Supervisé

Sur-apprentissage



Modèle	K-Neighbors Classifier	K-Neighbors Classifier	Multinomial NB	Multinomial NB	SVM	Multi-Layer Perceptron	Multi-Layer Perceptron	Multi-Layer Perceptron
Accuracy	28 %	64 %	58 %	62 %	73 %	16 %	51 %	16 %
Paramètres	k_neighbors = 10	k_neighbors = 5	default	default	default	hidden_layers = 5	hidden_layers = 10	hidden_layers = 20
Input	Retweet, Nb de tokens	Tweet (countVectorizer + tfidf transform)	Tweet (countVectorizer + tfidf transform)	Tweet (countVectorizer)	Tweet (countVectorizer + tfidf transform)	Tweet (countVectorizer + tfidf transform)	Tweet (countVectorizer + tfidf transform)	Tweet (countVectorizer + tfidf transform)

Matrice de Confusion – Modèle SVM -

vrai \ prédiction	Benoît Hamon	Emmanuel Macron	François Asselineau	François Fillon	Jacques Cheminade	Jean Lassalle	Jean-Luc Mélenchon	Marine Le Pen	N. Dupont-Aignan	Nathalie Arthaud	Philippe Poutou	Total
Benoît Hamon	803	113	5	43	6	3	41	47	0	0	11	1072
Emmanuel Macron	103	816	4	55	5	19	39	84	2	0	38	1165
François Asselineau	34	39	587	15	13	16	33	45	2	0	35	819
François Fillon	113	133	1	589	3	1	41	128	2	1	3	1015
Jacques Cheminade	41	29	10	11	793	7	35	31	0	1	52	1010
Jean Lassalle	38	62	4	13	8	449	24	36	3	0	41	678
Jean-Luc Mélenchon	59	37	2	20	3	2	1497	51	1	0	12	1684
Marine Le Pen	62	67	8	57	4	0	24	1270	1	1	2	1496
N. Dupont-Aignan	50	60	5	27	1	2	28	132	115	0	4	424
Nathalie Arthaud	28	28	0	6	4	0	30	44	1	74	34	249
Philippe Poutou	59	34	8	9	11	4	39	28	4	4	719	919
Total	587	1305	629	802	845	500	1790	1849	131	81	940	9459

Matrice de Confusion (% ligne)

vrai \ prédiction	Benoît Hamon	Emmanuel Macron	François Asselineau	François Fillon	Jacques Cheminade	Jean Lassalle	Jean-Luc Mélenchon	Marine Le Pen	N. Dupont-Aignan	Nathalie Arthaud	Philippe Poutou	Nb de tweets
Benoît Hamon	75%	11%	0%	4%	1%	0%	4%	4%	0%	0%	1%	1072
Emmanuel Macron	9%	70%	0%	5%	0%	2%	3%	7%	0%	0%	3%	1165
François Asselineau	4%	5%	72%	2%	2%	2%	4%	5%	0%	0%	4%	819
François Fillon	11%	13%	0%	58%	0%	0%	4%	13%	0%	0%	0%	1015
Jacques Cheminade	4%	3%	1%	1%	79%	1%	3%	3%	0%	0%	5%	1010
Jean Lassalle	6%	9%	1%	2%	1%	66%	4%	5%	0%	0%	6%	678
Jean-Luc Mélenchon	4%	2%	0%	1%	0%	0%	89%	3%	0%	0%	1%	1684
Marine Le Pen	4%	4%	1%	4%	0%	0%	2%	85%	0%	0%	0%	1496
N. Dupont-Aignan	12%	14%	1%	6%	0%	0%	7%	31%	27%	0%	1%	424
Nathalie Arthaud	11%	11%	0%	2%	2%	0%	12%	18%	0%	30%	14%	249
Philippe Poutou	6%	4%	1%	1%	1%	0%	4%	3%	0%	0%	78%	919

Conclusion

- 🐦 Nous pouvons utiliser les tweets pour extraire les opinions et thématiques abordées dans une campagne présidentielle.
- 🐦 Les personnages politiques représentent un courant politique et ils vont le défendre à travers leurs discours (neutre, positive, négatif) en envoyant de tweets à leurs « followers ».
- 🐦 Il faut maîtriser la fouille de données pour pouvoir connaître, décrire et analyser linguistiquement et statistiquement les messages politiques envoyés sur cette plateforme de microblogging.