Families of practices. A bottom-up approach to differentiate how French candidates made use of Twitter during the 2014 European Campaign

Dario Compagno

Families of practices. A bottom-up approach to differentiate how French candidates made use of Twitter during the 2014 European Campaign

Dario Compagno
(CREM, Université de Lorraine)

Tweets from the Campaign Trail: Researching Candidates' Use of Twitter during the European Parliamentary Elections, A. Frame, A. Mercier, G. Brachotte et C. Thimm (Eds), Peter Lang, Berlin, 2016.

Abstract

The aim of this study is to understand how French candidates made use of Twitter during the 2014 European elections. In particular, we are interested in identifying a few families of practices, or "Twitter styles", based on the media logics operators found in the tweets. To this end we adopted a bottom-up approach, based on unsupervised learning (clustering). As result we found three main "poles" defining the field of Twitter practices: some candidates tweet as a means for getting in direct contact with citizens, others diffuse hyperlinks to websites, while others still make an extensive use of the participative operators allowed by Twitter, such as hashtags, retweets and mentions. The results of this study permit to understand which practices are more common and to characterize them in accord to other variables such as the preferred twitting interface of a candidate, his or her political party and the content of the words used in his or her tweets.

1 Corpus and methodology

The corpus used in this study was extracted from the TEE2014 database[1]. The database collects all the tweets produced by Belgian, English, French, German, Italian and Spanish candidates to the European Elections 2014 during the campaign. Our corpus includes all messages twitted by French candidates. In total, around 72.000 tweets by 467 users have been collected. We decided not to consider for our analysis 106 candidates who twitted less than 15 tweets during the campaign, and 4 candidates who twitted more than 1000 tweets, leaving a more homogeneous corpus of 357 candidates.

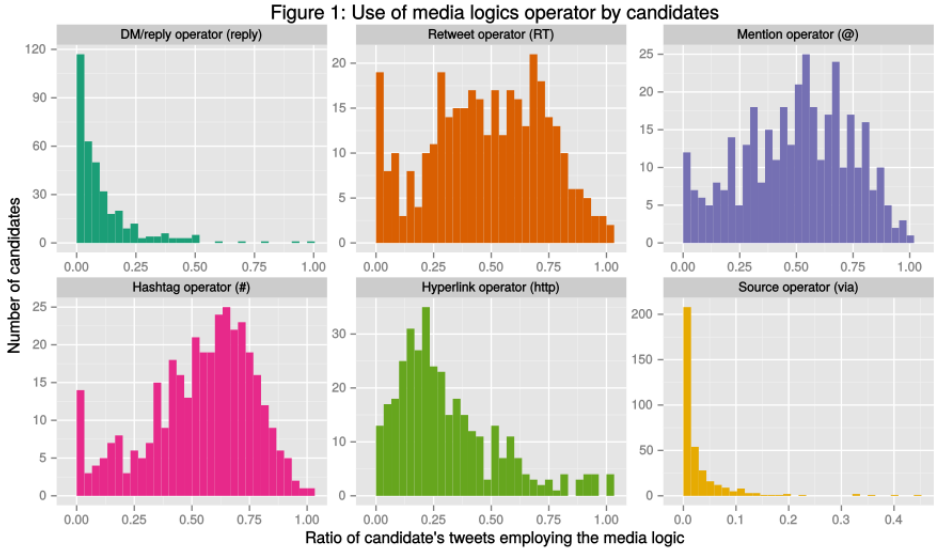
Data treatment, analysis and plots in this study have been realized with the open source statistical software R (R Core Team 2015), including the *dplyr* (Wickham & Francois 2015), *ggplot2* (Wickham 2009) and *cluster* (Maechler et al. 2015) packages[2].

Our study focuses on candidates instead than on individual tweets. Therefore we began by summarizing for each candidate the proportion of tweets including a media logic operator (Thimm 2012, Einspänner, Dang-Anh & Thimm 2013). To calculate these proportions we have preferentially considered the meta-data present in the TEE2014 database, which in turn are based on the Twitter REST API (Twitter Inc. 2015) and on dedicated extraction algorithms. However we have recalculated these

[1] TEE2014 Project: http://msh-dijon.u-bourgogne.fr/recherche-msh/les-programmes/547-twitter-aux-elections-europeennes-une-etude-contrastive-internationale-des-utilisations-de-twitter-par-les-candidats-aux-elections-au-parlement-europeen-en-mai-2014-tee2014.html

[2] The script and the aggregated data are available for reproducibility by contacting the author.

Figure 1: Use of media logics operator by candidates

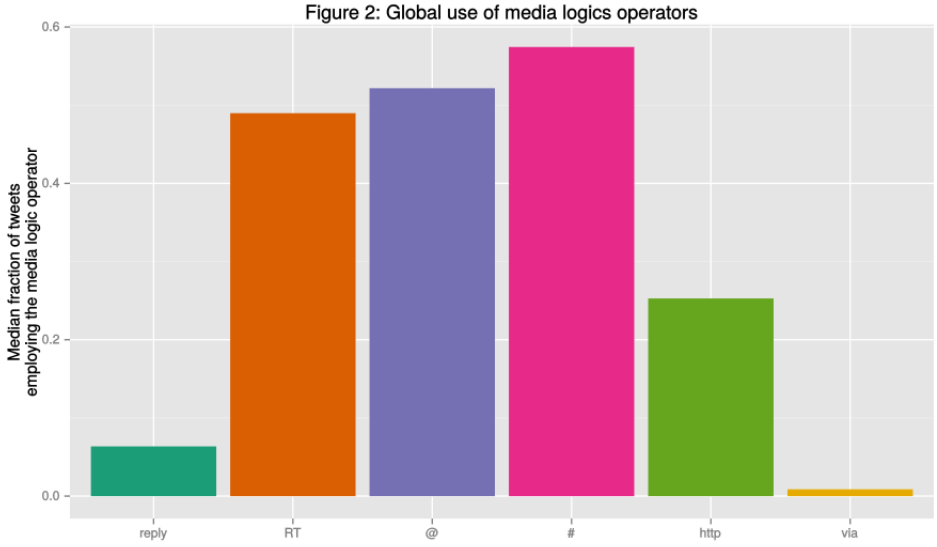


meta-data whenever needed by the ends of our research. The media logics operators used in this analysis are the following:

1. The *retweet operator* "RT": identifies retweets;
2. The *hashtag operator* "#": identifies tweets containing hashtags;
3. The *hyperlink operator* "http": identifies tweets containing links to webpages;
4. The direct message or reply operator "reply", referred to as *DM/reply operator* from now on: identifies tweets starting with an "@" character;
5. The *source operator* "via": identifies tweets ending with "via @" followed by a twitter account;
6. The *mention operator* "@": identifies tweets including mentions in their text that are not retweet, DM/reply or source operators.

Other meta-data considered in the study are the candidate's *party name* and the *twitting interface* most often used by him or her. The choice to focus just on media logics operators to perform a clusterisation of the candidates in our corpus has been determined by past works on the TEE2014 database (Mercier & Villa 2015, Thimm, Einspänner & Gautier 2015) and by a previous exploratory research confirming the importance of media logics operators as features to differentiate among the candidates' twitting practices (Compagno 2015). The aim of figure 1 is to show how many times each media logic operator has been used by French candidates. The ratio of tweets by a candidate that make use of a media logic operator is plotted on the x-axis of the facets: therefore a value of 0 means that a candidate never used a certain operator, a value of 1 that he or she used it in every tweet, and a value of 0.5 that we find that operator in one every two tweets produced by the candidate. On the y-axis we find the number of candidates distributed according to the ratio of their tweets employing each of the six operators.

We can observe that the ratios of retweet and mention operators are centred and approximatively normally distributed. The hashtag operator is slightly skewed to the left, suggesting that hashtags are used in tweets slightly more often than not. The hyperlink operator is instead more heavily skewed to



the right, suggesting that there is a large number of tweets in our corpus not pointing to webpages. This is even more true for the DM/reply and for the source operators: the vast majority of tweets are not replies nor direct messages, and just a very small number of tweets in our corpus points to a source at the end of their text. However, we decided to isolate source operators, instead of considering them just as a kind of mention, because they correlate with the presence of hyperlinks in tweets, and would have acted as confounders in clusterisation. Figure 2 permits to more easily compare these different distributions by plotting the median fraction of tweets per candidate that employ the operators. It is evident that retweet, mention and hashtag operators are found approximately in one tweet every two. Instead hyperlinks and especially DM/reply and source operators are much more rare and found respectively in one tweet every four, every fifteen and every one hundred.

2 Exploration and clustering

In order to explore the relationships among the different media logics operators, we performed a principal components analysis (PCA) using chi2 distances[3], the results of which are shown on the left half of figure 3. PCA creates a space of representation in which observations (in this case candidates) can be positioned. The position of each point is determined by its preference to one or another media logic operator. PCA therefore helps to understand if operators are often used by the same candidates or if on the contrary they characterize mutually exclusive uses of Twitter. The first two factors of PCA, used as coordinates for our plot, explain respectively the 40% and the 20% of the point variability.

PCA shows that hyperlink and mention operators tend to be mutually exclusive: if a candidate makes a large use of the first, he or she will not often use the second, and vice versa. Source operators are related with hyperlink ones: in fact, the are often added automatically to the end of the tweets produced by tweet buttons in news and other websites.

Hashtag and retweet operators tend to be used by the same family of candidates. We can see that this third family does not make use of direct messages and replies, nor publish many links to webpages.

[3]The function used to convert euclidean distances into chi2 ones has been written by Brieuc Conan Gruez, Université de Lorraine, that we would like to thank.

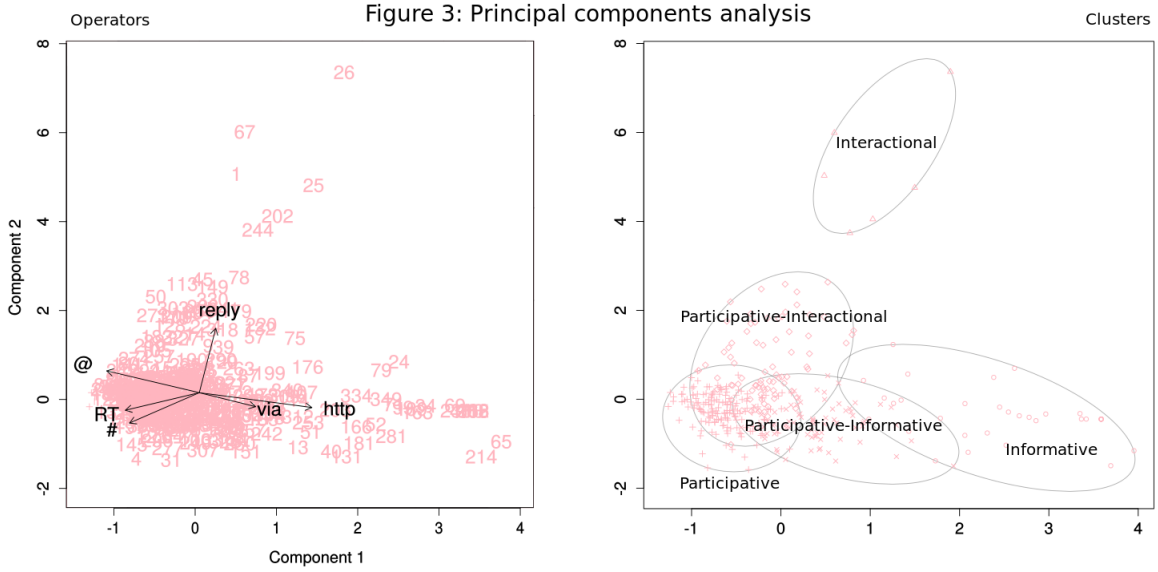


Figure 3: Principal components analysis

Conversely, those who make heavy use of DM/reply operators or of hyperlink operators do not insert many hashtag or retweet operators in their tweets.

We begin to form an idea of the main trends in the use of Twitter by French candidates in the 2014 campaign: the majority of them, represented at the lower-left corner of figure 3, used hashtags and retweets but did not tweet many hyperlinks nor engaged in direct conversations. We notice two other minoritarian tendencies: some candidates, at the lower-right side of the diagram, used Twitter mainly to diffuse hyperlinks, while some other candidates, at the top, got more involved in direct exchanges.

To get a better understanding of this insight we performed an agglomerative hierarchical clustering on our data. This technique produces a segmentation of the observations into a predetermined number of classes, according to the preferences that the candidates show about the use of media logics operators. After exploring several different partitions, we ended up with a five-classes clustering. The clusters are plotted on the right half of figure 3, on PCA coordinates. We notice three "extreme" groups of users and two intermediary ones.

The clusters are variable in size, as shown in figure 4. The largest, in the middle, that we called "Participative", includes 164 candidates out of 357. The two small clusters at the extremes of the diagram, named "Interactional" and "Informative", include respectively 6 and 33 candidates. The remaining two intermediary, medium-sized clusters named "Participative-Interactional" and "Participative-Informative" share what remains in two groups of 78 and 76 candidates each.

3 Interpretation of the clusters

Let us now enter into the details of the interpretation of these clusters. Figure 5 plots for each cluster the median fraction of tweets employing the media logic operators. This profiling helps us to understand what the candidates in each group have in common, and therefore to interpret these results. We chose not to plot the source operator at this stage because it is not useful to differentiate the five clusters.

We named the first cluster on the left "Interactional" because the candidates in it used Twitter mainly to engage into direct conversations with other users. More than the 70% of tweets produced by the candidates in this group are direct messages or replies. On the contrary hashtag, hyperlink and

Figure 4: Size of the clusters

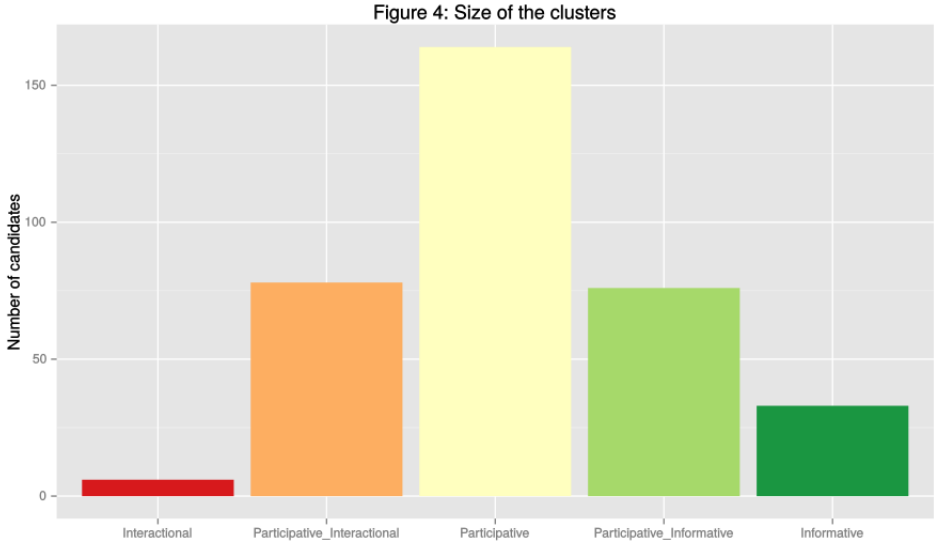
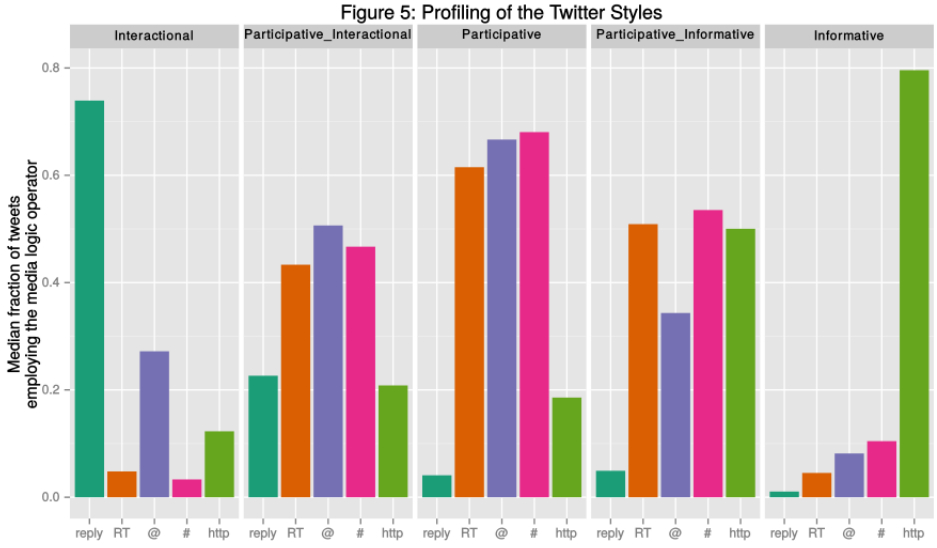


Figure 5: Profiling of the Twitter Styles



retweet operators were very rarely used. Mention operators are found in one tweet every four. This twitting style is highly unconventional, exploiting to the largest extent one minor function of Twitter (that of sending messages addressed to individual users). And in fact only a very small number of candidates, six, used Twitter in this fashion. With a metaphor, we could say that tweets are seen by the users of this group as a substitute for SMS, almost as if Twitter was a phone.

Moving to the other extreme of the diagram, to the farthest right, we find the "Informative" group. These candidates used Twitter almost exclusively as a means to diffuse information produced elsewhere, in the form of hyperlinks. They made a very modest if any use of all other media logics operators. With another metaphor, we could say that Twitter is seen by this group's members as a kind of RSS publisher and reader: they post links to websites but do not make use of other logics specific to Twitter. This cluster is the second smallest, including 33 candidates, and represents a second "deviant" use of Twitter – opposite to the first interactional one. Both these twitting styles constitute radical exceptions to the most common use of Twitter, and they play an important role in shaping the spectrum of alternative uses allowed by this social medium.

At the centre of the diagram we find the largest cluster. The candidates in it made heavy use of hashtag, retweet and mention operators. We named this cluster "Participative" because these operators are meant to promote participation in larger exchanges and discourses: hashtags are anchors to certain events or topics, retweets are instruments of Twitter's polyphony, repeating what was said by others, and mentions invite specific users (and their followers) as interlocutors. Participative candidates rarely engage in direct conversations and do not publish many links to external resources. They play with the most specific logics of Twitter: retweets, hashtags and mentions were born on this social medium and appear to define its most common use. We could say that these candidates practice "pure twitting", with no immediate interactional nor informative ends.

The "Participative_Interactional" cluster is an intermediary between pure interaction and pure participation. A fifth of the tweets produced by the users in this group were direct messages or replies, that is thrice as much as the global median. In addition to this, hashtag, retweet and mention operators are found each in 40 to 50 percent of their tweets. These candidates participate to the "meaning game" proper to Twitter but also interact individually with other users.

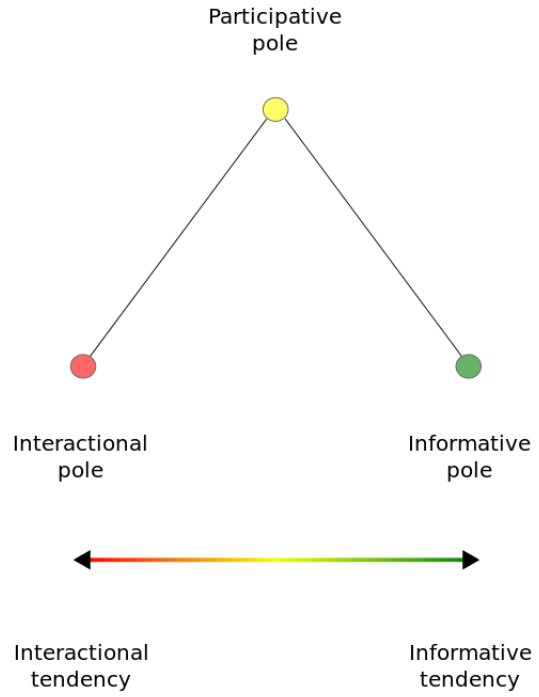
The last remaining group, "Participative_Informative", is also an intermediary one. One every two tweets produced by the members of this cluster contains an hyperlink. Hashtag and retweet operators are found with the same frequency. Mentions are used slightly less, around a third of the times. Candidates in this cluster use hyperlinks double as much as the global median, and also participate with hashtags, retweets and mentions.

4 Structuring the Twitting practices of French candidates

These considerations could be thought as a way to give a structural description of the norms regulating the use of Twitter by French candidates during the 2014 campaign. The statistical tools make some of the semiotic norms visible, allowing the researcher to access new aspects of this practice. François Rastier (2011) refers to the discourse norms made visible by corpus analysis as "new observables". In our case, the clusters built on the use of the media logics operators help to define the norms that play a role for this field of twitting practices.

The five families of twitting practices are organized around *three poles*: the *participative*, the *informative* and the *interactional*. The three "pure" clusters are each close to one pole and far from the other two, while the two intermediary clusters are each close to two poles and far from the remaining third. It should be noticed that no candidates adopted a style close to the informative and to the interactional poles while being far from the participative one. It means that, in our corpus, informative and interactional uses of Twitter are mutually exclusive (there is no "Interactional-Informative" style).

Schema 1: Poles and tendencies

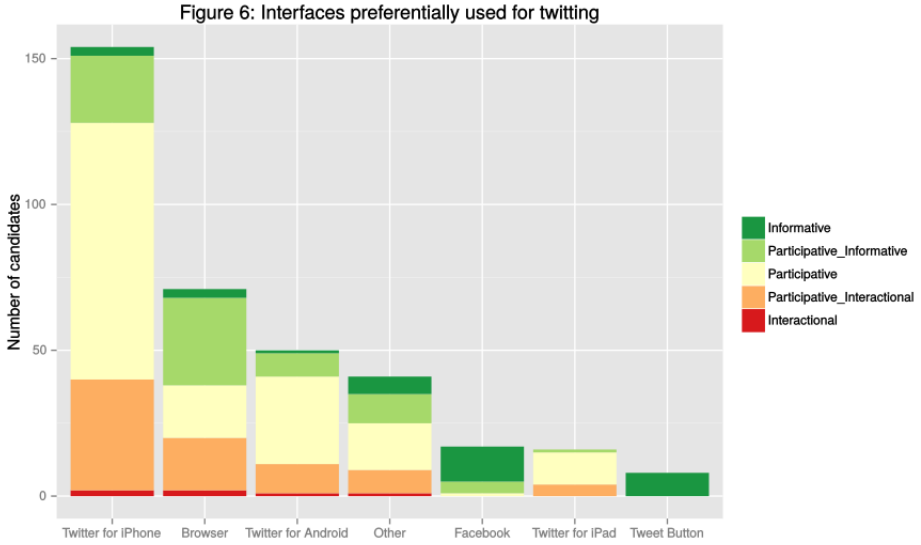


This configuration allows us to talk about *two opposite tendencies*: an interactional and an informative one. When referring to a tendency, we will talk about both pure and intermediary clusters (for example, both the informative and the participative-informative clusters constitute the informative tendency). Schema 1 resumes the configuration of poles and tendencies.

It should be noticed that even if Twitter could have been used by candidates at the same time as a means to send personalized messages (to interact) and messages with a broader reach (to inform), no candidate used it in this way. Also, the diffusion of hyperlinks is very rarely accompanied by personalized mentions, excluding from the observed field of practices the possibility of a "hybrid" between interaction and information. Twitter was not an instrument to send personalized links to websites.

Candidates preferred either a more personalized approach or a more impersonal one, therefore showing that they conceptualized the medium in two very different ways, exploiting different aspects of its potential. It would be interesting, in future research, to further explore the structure of different fields of practices, and compare them. This could result in a better understanding of how technical potentials are reduced to a smaller set whenever their use is filtered by social and semiotic norms. The patterns of this reduction would point to some of the logics of culture and communication, regulating human exchanges.

Twitter *per se*, as a technology, allows for a number of different uses: with Umberto Eco (1979) we could say that Twitter *model user* is broadly defined, permitting many potential uses to co-exist. However what matters the most in this case is the transition between the model user and the *empirical uses*: the individual uses of the social medium. What characterizes this transition are the norms,



made visible in our diagrams, that cause just a limited number of families of uses to actually exist. If it is true that every candidate behaves differently, and if it is also true that Twitter gives the same potential to all candidates, we still manage to identify a few main patterns. Therefore it is exactly *in the constitution of these patterns* that we may see the effect of the social and semiotic norms, responsible for reducing the larger virtual potential of Twitter to a small number of actual families of uses.

French political communication showed a difficulty in merging two tendencies: the informative and the interactional. Maybe French candidates consciously perceive these two tendencies as mutually exclusive, and believe that an hybridization would be seen as contradictory by citizens. Another hypothesis could be that merging information and interaction, for example by sending many personalized references to websites to individual users, would simply be too a costly activity to perform.

5 Associations between clusters and twitting interfaces

The five clusters have been calculated exclusively on the recurrence of media logics operators in the candidates' tweets. Therefore we may now look for further associations between these families of practices and the other variables at our disposal. This may help us to better understand the twitting styles and also to characterize the other variables' levels on the base of what we have discovered.

We are going to start by observing how the clusters are associated with the preferred twitting device of each candidate. This is important because if it is true that Twitter is the same platform for all users, different interfaces may produce different experiences of the medium, exposing users to subsets of the Twitter flux or making the use of some media logics operators easier. Figure 6 plots the number of candidates using each twitting interface most often. The colours of the bars indicate to which twitting style the candidates belong.

We notice immediately that one interface (Twitter for iPhone) is used much more frequently than all other ones (by 154 candidates in 357). The second most popular one (the Twitter website accessed with a web browser) is used by 71 candidates. Twitter for Android is used preferentially by 50 candidates, making it the third most popular. We find three other interfaces preferred each by 8 to 17 candidates (Facebook, Twitter for iPad, tweet buttons embedded in websites) and a higher number of interfaces

used each by less than 8 candidates (summing up to 41 candidates and represented as the column "Other" in the diagram).

The three major interfaces are used by candidates in all clusters. However, we see that the proportions may vary. The candidates using Twitter with an app for smartphones or tablets (either on iOs or on Android) tend to be more participative, and use Twitter to inform as much as to interact with other users. Candidates using iPads in particular use Twitter only rarely to distribute hyperlinks. Instead the candidates twitting from a web browser tend to be more informative than participative: as many of them are found in the informative cluster as the sum of those found in the participative and interactional ones.

This tendency becomes predominant for users twitting from a Facebook application or from tweet buttons found directly on websites: almost the totality of these candidates use Twitter prevalently to distribute hyperlinks. In particular, a small subset of 7 candidates in the informative cluster makes a noticeable use of the source operator (in one tweet every three). This subset uses mostly tweet buttons to tweet, rather than Facebook or other interfaces. Finally, candidates using prevalently other interfaces seem to make the most balanced use of Twitter, as the distribution of clusters in this group closely resembles the global one.

Do the interfaces facilitate a certain approach to the medium, or there is another explanation for the association between the candidates' twitting styles and their preferences for a certain interface? Candidates using tweet buttons and Facebook application to tweet may actually do not fully master the media logics operators specific to Twitter (hashtags, mentions) and therefore adopt an informative twitting style just because they tweet as a "residual", secondary activity, in which they are not so confident or competent. On the other hand, iPhones and other mobile devices do allow for a continuous scan of the Twitter flux, and so they may invite the user to retweet more often, or help to internalize the use of hashtags. Our study points to some potential paths that should be verified, either with experimental procedures, or with the analysis of dedicated datasets, accompanied by a semiotic comparison of the twitting interfaces (Zinna 2004).

6 Associations between clusters and political parties

The families of Twitter practices that we have identified are not homogeneously distributed among the political parties of the candidates. Figure 7 shows the number of candidates for each of the major parties (with at least 15 twitting candidates in our corpus), and the colour of the bars identifies the family to which the candidates belong.

Two of the largest parties, PS-PRG ("Parti socialiste - Parti radical de gauche") and UDI-MODEM ("Union des démocrates et indépendants – Mouvement démocrate"), show a similar distribution. Around the 75% of their candidates belongs to the participative cluster, and the remaining ones are equally divided between the informative and the interactional tendencies. The candidates of the UMP ("Union pour un mouvement populaire") party are more attracted by the interactional pole: around the 62% of them adopted a participative style, while the 31% was more interactional and the 7% more informative. The fourth largest party, EELV ("Europe écologie les verts"), was instead more attracted by the informative pole (51% participative, 31% informative tendency, 17% interactional tendency).

Among the four smaller parties, the Parti Pirate shows the most pronounce bias towards the interactional pole, with around the 75% of its candidates tending towards it; just one in fifteen candidates has been classified as purely participative and the other three show an informative tendency. The FDG ("Front de gauche") shows the opposite configuration, with 50% of participative candidates and 41% of informative ones. The candidates of the FN ("Front national") were more informative (48%) and almost equally participative and interactional (28% and 24% respectively). Last, Nous Citoyens was mainly participative (43%), and more interactional (35%) than informative (22%). Figure 8 resumes this information with a PCA plot.

Figure 7: Distribution of clusters among parties

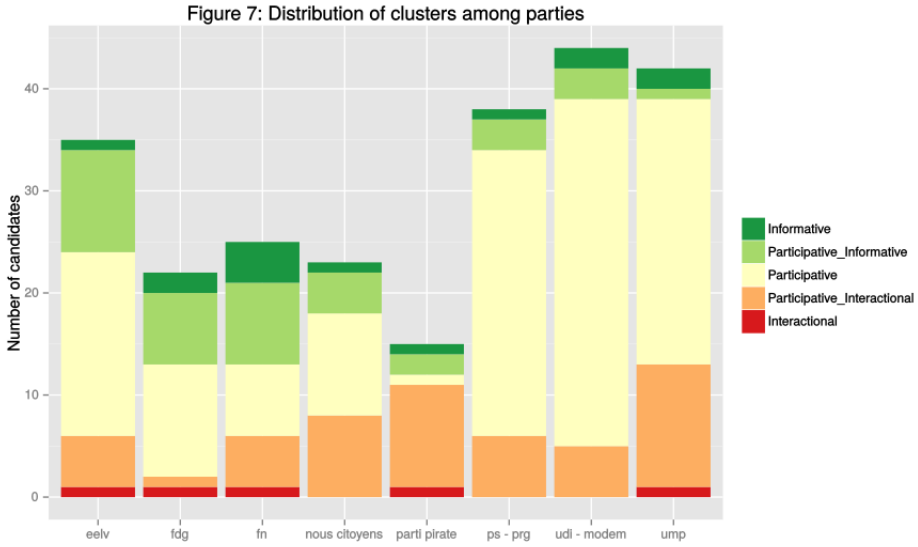


Figure 8: Principal components analysis (parties and tendencies)

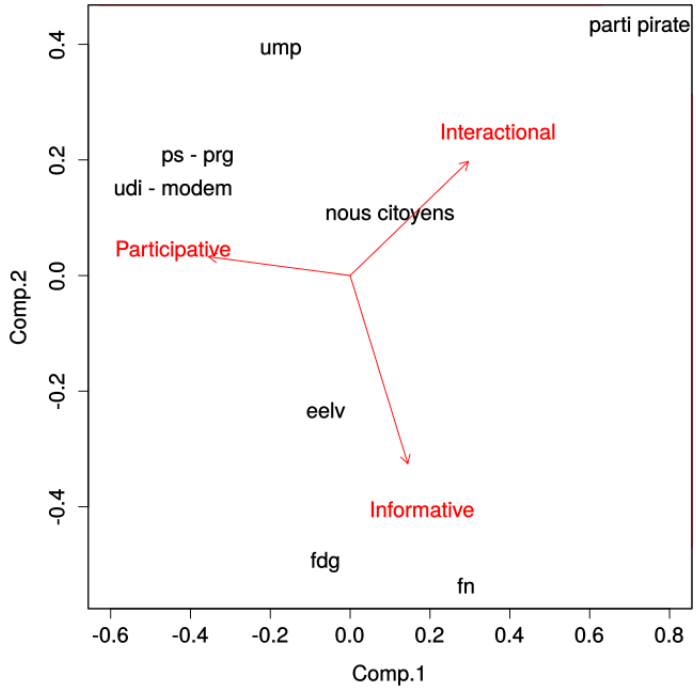
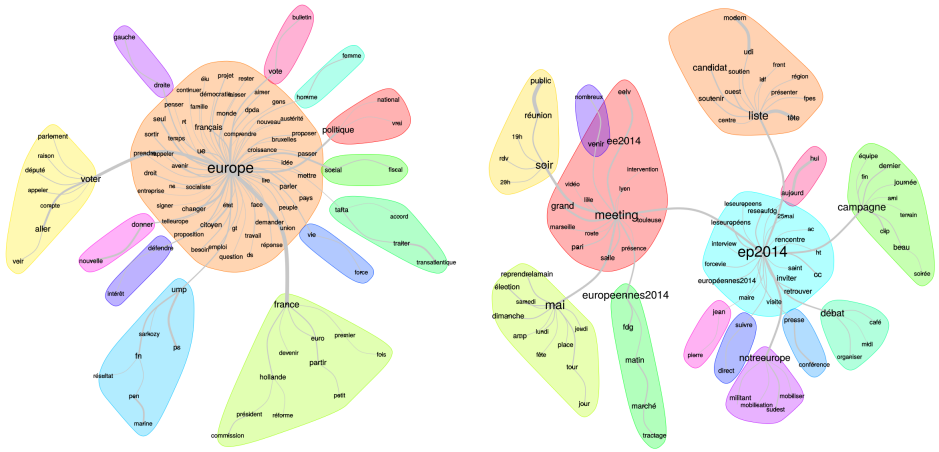


Figure 9: Lexical classes of the Tweets' texts



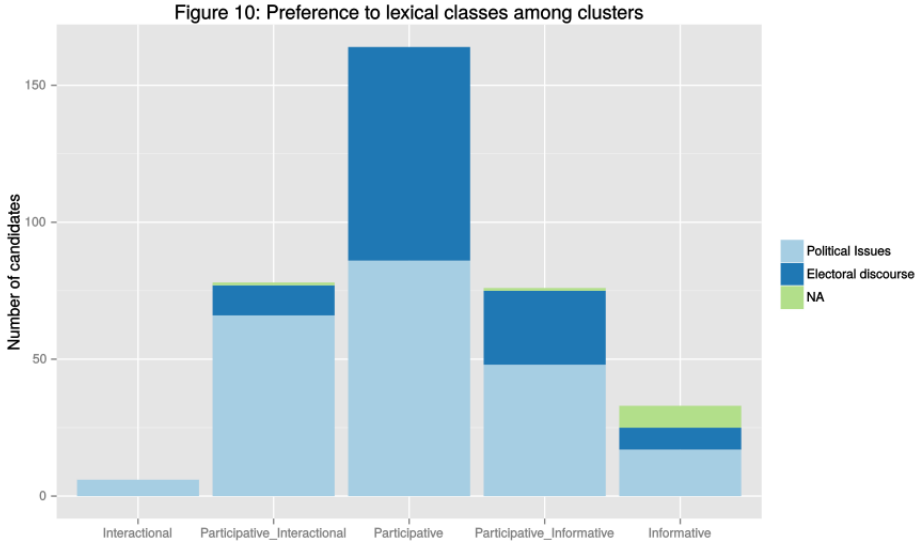
We can observe that the Parti Pirate is the most interactional, while the FN and the FDG are the most informative, and the PS-PRG and UDI-MODEM the most participative. The UMP is equally interactional and participative. It is not possible to associate a certain twitting tendency to the leftist or rightist orientation of a candidate. We remark that the most extremist parties (FN, FDG) adopt more informative styles, while the moderate ones (PS-PRG, UDI-MODEM, UMP) tend to be more participative. However it should be noticed that the moderate parties are also the largest ones.

7 Classification of tweets based on textual analysis

As last step, we wanted to know if our clusters could be further characterized on the basis of the discourses realized by the candidates in their tweets. Thanks to the software Iramuteq (Ratinaud 2009), we performed to this end a classification of the tweets' texts, based on the co-occurrence of words (Reinert 1983). As result of this classification, the tweets in our corpus have been preferentially related to one of two ample classes of co-occurring words. Similitude diagrams for the two classes, realized with Iramuteq, are shown in figure 9.

The left half of the figure describes the first of the two classes, that we have called "Political Issues". The words in this class are mainly related to political topics, such as the relationship between France and the European Union, the conduct of the French prime minister François Hollande, the relationships between the three main French parties PS, UMP and FN. Economic issues were also mentioned during the campaign and are included in this class, such as the TAFTA, unemployment and growth. The right half of figure 9 instead describes the lexical class we called "Electoral discourse". It includes hashtags and words used during the campaign to communicate about events, meeting, debates and the elections dates. The two lexical classes are intentionally broad, but give some hints about the contents most addressed by the different parties on Twitter. Based on the classification of the tweets realized with Iramuteq, we calculated the preference of each candidate's tweets for political issues or for the campaign itself. Figure 10 plots the preferences of the candidates in each cluster.

We notice that in the participative cluster political issues and the electoral discourse appear to have the same importance. All the other clusters instead show a preference for the political issues, and



this preference is more pronounced on the interactional side than on the informative side. Some of the tweets have not been categorized (they are coloured in green in figure 10), because in most cases they only included a hyperlink without any commentary. The majority of the candidates using most often uncategorised tweets belongs to the purely informative cluster. Figure 11 plots the preferences of each of the major political parties for political issues or for the electoral discourse.

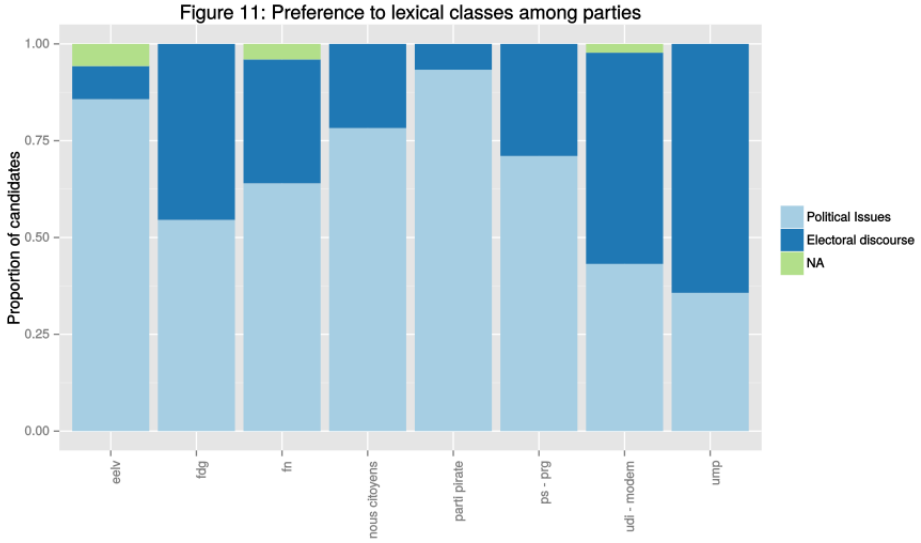
The Parti Pirate is the one most focusing on political issues, followed closely by the EELV. On the opposite, the 65% of the UMP candidates focused mostly on the electoral discourse, followed by the UDI-MODEM (55% of the candidates). This information should be crossed with the associations between clusters and political parties seen above. The Parti Pirate and the UMP were the most interactional parties; however, it seems that the kind of interaction they perform differs between the two. Our research appears to show that the members of the Parti Pirate use Twitter prevalently to engage in one-to-one conversations about political issues (a *personalized political debate*), while those of the UMP use Twitter to realize a *personalized electoral discourse*, inviting individual users to meetings and to other occasions. We could talk in these two cases of sub-syles of twitting, defined within the interactional tendency. Textual analysis could be exploited even further to identify more twitting sub-styles.

Conclusion

We have found tendencies in our corpus that show how semiotic and social norms reduce and channel the technical potentials of a social medium. Twitter gives space to many different individual uses, but just a few families of uses, or "twitting styles", are actually realized. These styles are not homogeneously distributed among the political parties of the candidates. We have also found further associations between the twitting styles and some other variables, such as the kind of discourse realized in the tweets' texts or the favoured twitting interface.

The next step of our research will be to perform a similar analysis to other corpora extracted from the TEE2014 database. We want to know if Belgian, English, German, Italian and Spanish candidates use Twitter in a comparable way as the French. Some of the twitting styles we found in France may not exist elsewhere, or their importance may vary from one country to the other. After having extended

Figure 11: Preference to lexical classes among parties



our research to a larger corpus we may be able to produce a better structural description of the norms guiding the use of Twitter for electoral communication in Europe today.

References

[1] Compagno, D. (2015). "Families of practices. Employing data mining techniques to differentiate how French candidates make use of Twitter". *Twitter at the European Elections 2014*, Metz, February 2015.

[2] Eco, U. (1979). *The role of the reader*. Indiana UP, Bloomington.

[3] Einspänner, J., Dang-Anh, M. and Thimm, C. (2013). "Computer-assisted content analysis of Twitter data". In: Bruns, A., Weller, K., Burgess, J., Mahrt, M. and Puschmann, C. (Eds.), *Twitter and Society*. Peter Lang, New York, pp. 97-108.

[4] Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M. and Hornik, K. (2015). *cluster: Cluster Analysis Basics and Extensions*. R package version 2.0.1.

[5] Mercier, A. & Villa, M. (2015). "The figure of the enemy during the 2014 European tweet-campaign in France and Italy". *Twitter at the European Elections 2014*, Metz, February 2015.

[6] Rastier, F. (2011). *La mesure et le grain. Sémantique de corpus*. Champion, Paris.

[7] Ratinaud, P. (2009). *IRaMuTeQ: Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires*. http://www.iramuteq.org/

[8] Reinert, M. (1983). "Une méthode de classification descendante hiérarchique : application à l'analyse lexicale par contexte", *Les cahiers de l'analyse des données*, vol. VIII, n°2, pp. 187-198.

[9] R Core Team (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

[10] Thimm, C. (2012). "Political deliberation and the internet: Forms and Functions of civic Participation on Twitter". *Communiquer dans un monde de normes. L'information et la communication dans les enjeux contemporains de la mondialisation*, Lille, March 2012.

[11] Thimm, C., Einspänner, J. & Gautier, L. (2015). "German political communication on Twitter during the 2014 European Elections". *Twitter at the European Elections 2014*, Metz, February 2015.

[12] Twitter Inc. (2015). *Twitter documentation*. https://dev.twitter.com/overview/documentation

[13] Wickham, H. (2009). *ggplot2: elegant graphics for data analysis*. Springer, New York.

[14] Wickham, H. & Francois, R. (2015). *dplyr: A Grammar of Data Manipulation*. R package version 0.4.1.

[15] Zinna, A. (2004) "L'objet et ses interfaces", *EC*.