

Quelle modélisation de l'espace politique français sur Twitter ?

Nicolas HERVÉ*

*Institut National de l'Audiovisuel - Ina
nherve@ina.fr,
<http://www.herve.name>

Dans le cadre général de l'analyse de la propagation des informations sur Internet et dans les médias, il est utile de considérer la dimension politique comme un des critères à observer et quantifier. Nous cherchons par exemple à savoir si les choix éditoriaux des principaux médias d'information sont en partie guidés par des considérations économiques, journalistiques ou politiques et à caractériser en quoi les discussions concernant l'actualité sur Twitter influent sur ces choix, notamment selon la polarisation politique des individus qui prennent part à ces échanges (Cagé et al., 2020) ?

Nous présentons ici la construction et la validation de l'espace politique latent permettant l'analyse de la polarisation de ces échanges ¹. Les publications sur ce sujet se situent majoritairement dans le cadre de l'espace politique bipolaire des USA. Nous basons nos travaux sur les approches développées par Barberá (2015) et Garimella et Weber (2017). Elles nécessitent toutes les deux l'existence d'un ensemble de comptes politiques de référence pour lesquels le positionnement politique est connu. Cette polarisation est ensuite propagée aux comptes Twitter anonymes en utilisant les liens structurels de *follow*. Intuitivement, plus un compte Twitter va suivre de comptes politiques d'un même parti, plus on va considérer qu'il en est proche idéologiquement. L'approche naïve de Garimella et Weber (2017) utilise une simple combinaison linéaire pondérée et fournit un positionnement sur un axe classique droite-gauche. L'utilisation de l'AFC, comme le propose Barberá (2015), est plus évoluée, elle ne nécessite aucun *a priori* sur les partis politiques et infère automatiquement leurs positionnements relatifs ainsi que l'importance de chacun des comptes politiques de référence. L'espace obtenu comporte n dimensions. Parmi les rares travaux spécifiques à la France, Briatte et Gallic (2015) utilisent cette dernière approche. En plus de la démarche de Barberá, nous proposons une approche similaire à celle de Garimella, spécifique à l'espace politique français. De plus, nous étendons ces deux modélisations en utilisant les informations de *retweet*. Il n'y a en effet aucune raison pour que l'action de suivre un compte politique particulier soit plus porteuse d'information sur le positionnement idéologique d'un utilisateur que lorsqu'il *retweete* ses messages. Nous avons donc 4 modélisations possibles que nous appliquons à l'espace francophone de Twitter. Nous introduisons de plus une mesure du degré de politisation d'un compte Twitter.

Pour nos expérimentations, nous avons constitué une liste de 4 824 comptes politiques de référence répartis selon 13 tendances. Nous pouvons ainsi établir la polarisation de 184 229 comptes en utilisant les liens de *follow* et de 168 393 avec les liens de *retweet*. En observant les deux premières dimensions des espaces de Barberá, on remarque que des clusters relativement cohérents sont formés. Il est en revanche difficile d'interpréter la signification des axes déterminés par l'AFC. Pour l'espace basé sur les *follow*, il semble que les comptes soient distribués sur un *manifold* de dimension intrinsèque plus faible qui préserve un classement des partis politiques selon l'axe traditionnel droite-gauche. On retrouve le même ordonnancement des partis sur l'espace basé sur les *retweets*, mais les comptes anonymes y sont plus uniformément répartis et l'espace est clairement tripolaire. En revanche, sur les deux espaces de Garimella, le positionnement sur l'axe droite-gauche est très cohérent.

1. Une version longue et détaillée de cette publication avec de nombreux graphiques ainsi que la suite de nos travaux sont disponibles en ligne : <http://www.herve.name/twitter-polarization>

Quelle modélisation de l'espace politique français sur Twitter ?

La validation formelle de la polarisation politique des comptes Twitter n'est pas directement accessible faute de données nominatives disponibles (et c'est heureux). Aussi, comme pour les travaux dans d'autres pays, nous ne pouvons procéder qu'à une validation partielle et indirecte. Nous utilisons deux approches inédites : corrélation de la distribution des polarisations avec des données issues du sondage du Reuters Institute pour les utilisateurs français de Twitter ($\rho = 0.940$ pour l'espace Garimella/*retweet*) et analyse qualitative de la polarisation de hashtags que l'on sait être très partisans. Nous disposons en effet d'un corpus de 3.8 milliards de tweets en français sur une période de deux ans (Mazoyer et al., 2018). Ce corpus permet d'analyser la propagation de tous les types de contenus qui sont échangés sur Twitter. On associe à chaque *tweet* la polarisation politique du compte qui l'a posté. Nous nous focalisons sur les *hashtags*. Ils permettent aux utilisateurs d'encoder dans un token unique une signification, potentiellement partisane, qui est partagée par une communauté. L'interprétation de leur propagation est ainsi simplifiée et correspond à l'usage que nous souhaitons en faire ici : une première validation de l'espace politisé et un cadre d'analyse dans cet espace et dans le temps. Pour chaque espace, la représentation d'un *hashtag* est alors simplement la moyenne des *tweets* dans lesquels il apparaît. Nous avons extrait les principaux *hashtags* utilisés sur la période et regardons le positionnement de certains d'entre eux qui sont *a priori* partisans et que l'on s'attend à retrouver dans des régions bien précises des espaces latents. Un premier ensemble est constitué de tous les *hashtags* contenant 'soutien' ou 'demission', ils sont régulièrement utilisés pour marquer son soutien ou son opposition à une personne. Un second ensemble est lié au processus électoral, par nature partisan, avec les termes 'vote', 'europe', 'election', et 'municipal' afin de couvrir les deux élections qui se sont tenues sur la période de notre corpus ainsi que le *hashtag* #directan qui concerne l'actualité des débats à l'Assemblée Nationale. Parmi les *hashtags* très proches des zones identifiées des partis, on retrouve naturellement ceux qui ont été utilisés pour les campagnes électorales (#envideurope, #renaissanceeuropeenne, #votepourleclimat, #jevotainsoumis, ...) ou liés à certaines personnalités (#soutienfillon, #soutienmarine, ...). Plus au centre, on a des *hashtags* ayant suscité de nombreux *tweets* mais ne concernant que très peu de comptes politisés.

Si on ne s'intéresse qu'à une modélisation sur une dimension unique droite-gauche, alors l'utilisation de Barberá permet d'avoir un bon résultat avec les liens de *follow*. Toutefois, l'approche Garimella fournit des résultats équivalents pour une mise en œuvre plus simple et plus souple. Si on souhaite plutôt baser nos analyses sur les liens de *retweet*, alors il conviendra de conserver au moins deux dimensions de l'espace latent. Pour une même base de comptes politiques de référence, l'utilisation des informations de *follow* ou *retweet* conduit à caractériser des ensembles de comptes Twitter différents et donc à pouvoir observer des contenus différents. Il est important de bien cerner ce périmètre avant d'interpréter les polarisations induites pour les contenus. Nous avons présenté quelques prémices de résultats et de visualisations en considérant les *hashtags* partisans et avons observé un positionnement cohérent dans les espaces latents.

Références

- Barberá, P. (2015). Birds of the Same Feather Tweet Together : Bayesian Ideal Point Estimation Using Twitter Data. *Political Analysis* 23(1).
- Briatte, F. et E. Gallic (2015). Recovering the French Party Space from Twitter Data. In *Science Po Quanti*, Paris, France.
- Cagé, J., N. Hervé, et B. Mazoyer (2020). Social Media and Newsroom Production Decisions. SSRN Scholarly Paper ID 3663899, Social Science Research Network, Rochester, NY.
- Garimella, V. R. K. et I. Weber (2017). A long-term analysis of polarization on Twitter. In *11th International Conference on Web and Social Media, ICWSM 2017*.
- Mazoyer, B., J. Cagé, C. Hudelot, et M.-L. Viaud (2018). Real-time collection of reliable and representative tweets datasets related to news events. In *BroDyn 2018, co-located with ECIR 2018*.