

# Analysis of Polarization in the News

EVAN DUFRAISSE<sup>1,2</sup>, JULIEN TOURILLE<sup>1</sup>, ARMELLE BRUN<sup>2</sup>, ADRIAN POPESCU<sup>1</sup>,

<sup>1</sup>Université Paris-Saclay, CEA, List, F-91120, Palaiseau, France

<sup>2</sup>Université de Lorraine - CNRS - Loria, Vandoeuvre les Nancy Cedex, France

{*evan.dufraisse,julien.tourille,adrian.popescu*}@cea.fr, *armelle.brun@loria.fr*

## 1 INTRODUCTION

Digital platforms have democratized news production and facilitated access to a wider range of sources. Many news providers are driven more by short-term financial goals and/or political interests, while readers are driven, at least in part, by the goal of making sense of the world through access to reliable information. This misalignment is exacerbated by the fact that social networks, which are also driven by profit, often act as intermediaries between sources and readers. The selection of news that social network users read is largely determined by recommendation algorithms that prioritize articles that are assumed to match users' preferences, regardless of quality and veracity. Existing work [2] shows that recommenders lead to polarization in the long run, a phenomenon that is detrimental to public debate and ultimately to democracy. We present work in progress from the BOOM project<sup>1</sup>, which focuses on understanding and characterizing polarization, and then proposing recommendation strategies that reduce it. The focus of the presented work is the automatic and fine-grained characterization of the stance of news articles using Twitter as a source of political knowledge.

## 2 AUTOMATIC ANALYSIS OF POLITICAL NEWS BASED ON SOCIAL MEDIA

The French media landscape is vast, with around 400 different media outlets reported by the Montaigne Institute in 2019<sup>2</sup>. In theory, this wide range of options gives citizens access to a variety of viewpoints and helps them make better sense of the world in an unbiased way. In practice, the access to news is to a large extent mediated by recommendation algorithms, which suggest items that are similar to previous consumptions. This exposure to a limited range of opinions can lead to polarization, especially for political topics that are prone to controversy.

The development of algorithms that model the content of news articles is a first step towards news recommendation systems that account for user preferences, but also promote diversity among recommendations. Automatically determining the political stance of news typically requires the use of machine learning algorithms that are trained on labeled data. This approach has several drawbacks. First, annotating data is expensive and must be on a large scale. Second, these annotations quickly become outdated due to the dynamic nature of news, where topics discussed and opinions expressed change rapidly over time. Finally, most labeled datasets use a simple annotation scheme (3 or 5 labels on a political scale from left to right), which tends to oversimplify the stance expressed in the news.

The use of social media, and Twitter in particular, as a dynamic source of up-to-date political information is appealing, but also challenging since data need to be organized. Twitter was successfully used for different types of political analyses [1], including community detection, controversy detection, information dissemination analysis, and polarization quantification, among others. It was equally used to automatically determine the political stance of news outlets. For instance, the authors of [4] infer the positioning of several US news outlets along eight different polarizing topics

<sup>1</sup><http://boom.loria.fr/>

<sup>2</sup><https://www.institutmontaigne.org/publications/media-polarization-la-francaise>

by using graphs of retweets from vocal users. While interesting, this work has several important limitations: (1) the eight topics were pre-determined by manually selecting characteristic hashtags, and are static, (2) they force the political stance on each topic to be represented on a five points scale between pro and con; and (3) the stance is only inferred at news outlet level, implying that individual articles necessarily have the same stance on a topic.

We propose to address the problem differently and attempt to qualify stance more dynamically, without resorting to the classical left/right predefined classification. Such an approach is motivated by recent events which show that classical political positioning becomes irrelevant for polarizing issues, such as vaccination, COVID-19, or pension reform. Our final objective is to position the political stance of news articles, and this entails two steps. First, topic detection will be performed by clustering article representations obtained by using the TF-IDF weighted sum of fasttext word embeddings. Second, the alignment between individual news articles and automatically detected Twitter political communities will be predicted using either clustering or graph neural networks. Such an approach is interesting since both detection and article mapping to communities can be updated to follow the pace of news. These tasks are ongoing and results will be detailed in future work.

### 3 PRELIMINARY RESULTS AND FUTURE WORK

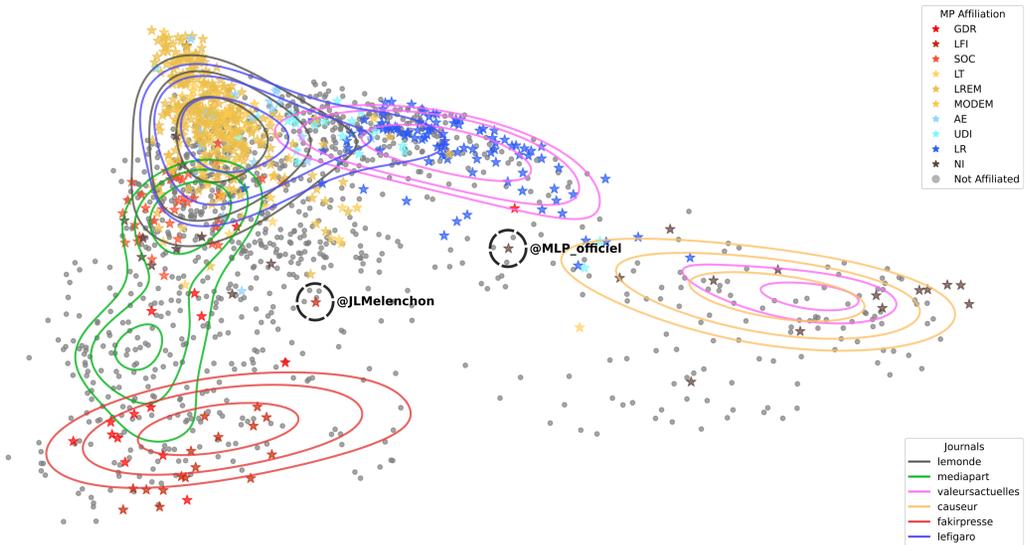


Fig. 1. Newspapers Distribution over Political Space

We describe below the mapping between community representations and news outlets, which are an intermediate step of our approach. Inspired by [3], we collect a list of Twitter accounts to obtain a comprehensive representation of the French political spectrum. This list includes: French MPs of the XVth legislature, economists, journalists, presidential candidates, NGOs, and accounts which are active on politically-relevant topics such as humanitarian aid, COVID-19, racism, immigration, feminism, etc. We collected data for accounts which have between 10,000 and 250,000 followers. Note that we did not apply this filter to MPs to reproduce [3]. The lower bound is applied to remove accounts which have low influence. The upper bound is needed due to limitation of the Twitter API, but also because of the risk of political misrepresentation for very popular accounts, as they have

followers that do not represent support and that come from across the political spectrum. This is illustrated in Figure 1 by the slight offset in the positioning of Jean-Luc Mélenchon and Marine Le Pen, both with over 2 million followers. The final dataset includes lists of followers from 1,672 seed accounts. We reduced the binary matrix of accounts and followers to a 2D representation using correspondence analysis. The obtained communities, illustrated in Figure 1, are in line with the results reported in [3] and confirm the usefulness of the approach.

We then collected, for all users, their tweets and retweets with URLs from 2016 until the results of the French presidential election of 2022. This raw dataset covers the entire XVth legislature, and include approximately two million tweets. We then selected only URLs with domain names of news outlets from the Montaigne Institute’s mapping of the French media landscape. This subset includes 197 domain names, and over 500,000 URLs. We then used the distribution of domain names tweeted by different accounts to fit a mixture of Gaussians over the 2D projection of the political spectrum. The result for a subset of journals is illustrated in Figure 1, with high densities being associated to the center of the contours.

These contour plots allow us to make several interesting observations. The alignment between newspapers and political communities follows intuition. Mainstream journals, such as Le Monde and Le Figaro, are associated to center-left and center-right groups, respectively. Causeur is clearly associated to far-right accounts, and Fakir to radical left. Interestingly, Mediapart and Valeurs Actuelles span over different political groups, toward the left and the right sides of the political spectrum, respectively. This might be indicative of agreement between political groups for a part of the topics, and of disagreement for others. Importantly, the results from Figure 1 confirm the fragmentation of the news space. Readers who favor a limited number of sources are likely to be subject to polarization, it is thus important to propose recommender systems which propose diversified points of view, while being attentive not to antagonize readers.

A more detailed analysis, at topic and article levels, is needed in order to fully operationalize our approach. We have started collecting the articles linked in tweets. A preliminary analysis suggests that more than 85% of the URLs are still alive and that these articles are at least partially retrievable. Using the topic modeling information extracted from these articles, we will study the positioning of news outlets in the political space per topic. This will lead to a finer-grained characterization of the political communities, and a better mapping of individual articles in the political space. Then, given a user’s current topic and her/his past consumption of news, we will be able to propose a diversification-driven recommendation strategy.

## REFERENCES

- [1] Despoina Antonakaki, Paraskevi Fragopoulou, and Sotiris Ioannidis. 2021. A survey of Twitter research : Data model , graph structure , sentiment analysis. *Expert Syst. Appl.* 164, July 2020 (2021), 114006. <https://doi.org/10.1016/j.eswa.2020.114006>
- [2] Katharina Ludwig, Alexander Grote, Andreea Iana, Mehwish Alam, Heiko Paulheim, Harald Sack, Christof Weinhardt, and Philipp Müller. 2023. Divided by the Algorithm? The (Limited) Effects of Content-and Sentiment-Based News Recommendation on Affective, Ideological, and Perceived Polarization. *Social Science Computer Review* (2023), 08944393221149290.
- [3] Pedro Ramaciotti Morales, Jean-Philippe Cointet, and Gabriel Muñoz Zolotoochin. 2022. Unfolding the Dimensionality Structure of Social Networks in Ideological Embeddings. In *Proceedings of IEEE/ACM ASONAM 2021* (Virtual Event, Netherlands). ACM, New York, NY, USA, 333–338. <https://doi.org/10.1145/3487351.3489441>
- [4] Peter Stefanov, Kareem Darwish, Atanas Atanasov, and Preslav Nakov. 2020. Predicting the Topical Stance and Political Leaning of Media using Tweets. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Online, 527–537. <https://doi.org/10.18653/v1/2020.acl-main.50>