CESifo WORKING PAPERS

8326
2020

Original Version: May 2020
This Version: July 2023

Social Media and the Dynamics of Protests

Pierre C. Boyer, Thomas Delemotte, Germain Gauthier, Vincent Rollet, Benoît Schmutz

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 8326*

Social Media and the Dynamics of Protests

Abstract

This paper provides quantitative evidence on the heterogeneous effects of social media on protest dynamics. On the one hand, social media enables the development of online communities of protesters that keep movements alive. On the other hand, social media is fertile ground for political polarization and radicalization. Using data from the 2018-2019 Yellow Vest uprising in France, we show that local street protests triggered the creation of large communities of protesters on Facebook. However, these communities progressively became more antagonistic, negative, and ideologically segregated. While moderate discussants left the discussions, those who remained radicalized. Facebook's recommender algorithm likely contributed to this pattern by consistently showcasing radical content.

JEL-Codes: F150, J400, J600, J800, C830.

Keywords: protests, social media, Yellow Vest movement, NLP techniques, collective action.

*Pierre C. Boyer**
CREST, École Polytechnique / Palaiseau / France
pierre.boyer@polytechnique.edu

Thomas Delemotte
CREST, ENSAE
Palaiseau / France
thomas.delemotte@ensae.fr

Germain Gauthier
CREST, École Polytechnique
Palaiseau / France
germain.gauthier@polytechnique.edu

Vincent Rollet
MIT Massachusetts Institute of Technology
Cambridge / MA / USA
vrollet@mit.edu

Benoît Schmutz
CREST, École Polytechnique
Palaiseau / France
benoit.schmutz@polytechnique.edu

*corresponding author

The authors thank Micael Castanheira, Luisa Dörr, Elie Gerschel, Julien Grenet, Sophie Hatte, Fanny Henriet, Clément Imbert, Francis Kramarz, James Lo, Nolwenn Loisel, Andrew Lonsdale, Phoebe Mac Donald, Clément Malgouyres, Isabelle Méjean, Vincent Pons, Audrey Rain, Anasuya Raj, Clemence Tricaud, Lionel Wilner, and Ekaterina Zhuravskaya, as well as many seminar and conference participants for their comments. The authors gratefully acknowledge the Investissements d'Avenir (ANR-11-IDEX-0003/Labex Ecodec/ANR-11-LABX-0047), ANR-19-CE41-0011-01 and ANR-20-CE41-0013-01 for financial support, the CASD (Centre d'Accès Sécurisé aux Données) and INSEE for the access to French administrative data, and Change.org for sharing their anonymized data.

1 Introduction

Since the early 2010s, social media have been instrumental in the emergence of protest movements worldwide (see, for reviews, Della Porta and Diani 2020; Zhuravskaya, Petrova and Enikolopov 2020). As a result, modern protest movements are now a combination of overlapping online and offline mobilizations. The many implications of this hybridization make it difficult to assess the impact of social media on the effectiveness of protest movements. For example, as argued by Tufekci (2017), recent movements are efficient at organizing protests from scratch, but such ability does not necessarily equate to actual capacity-building. The goal of this paper is to document one of the possible trade-offs regarding the role of social media in collective action problems: on the one hand, social media may help protest movements persist over time by nurturing active online communities of protesters; on the other hand, the way social media is used by protesters and structures discussions may be conducive to the radicalization and marginalization of those communities.

Our analysis is based on the Yellow vest movement, widely recognized as the most significant social unrest episode in recent French history.[1] Several features of this movement make it particularly well-suited to the study of the impact of social media on protests. First, it was sparked by an online petition and used Facebook to coordinate street protests across the country. Second, it outlived the initial day of protests (November 17, 2018 – hereafter, 11/17) and stayed active on various online platforms for several months, with hundreds of thousands of petition signatures and lively discussions on dedicated Facebook pages. Third, it turned into a full-blown protest against the government and lost support in the general population. These conflicting patterns suggest that the ease of coordination via social media

[1]For references, see Algan, Beasley, Cohen and Foucault (2019); Bendali and Rubert (2020); Boyer, Delemotte, Gauthier, Rollet and Schmutz (2020); Cointet, Morales, Cardon, Froio, Mogoutov, Ooghe and Plique (2021); Sebbah, Souillard, Thiong-Kay and Smyrnaios (2018).

may come at the cost of the ability to articulate a consistent message while preserving a broad supporter base. We use geolocated data on street protests, Facebook groups, and petition signatures on Change.org, combined with the textual analysis of discussions on Facebook pages, to study the interplay between online and offline mobilizations and its implications for the movement's dynamics.

We start by documenting the existence of a positive feedback loop between mobilization on Facebook and street protests. To that end, we use spatial regressions at the municipality level and focus on the first day of the protests: the 11/17 roadblocks. Those roadblocks had been organized on dedicated Facebook groups all over the country, as confirmed by the strong positive correlation between the presence of an early Facebook group and the occurrence of a roadblock. This correlation aligns with previous research on other settings (see, among others, Steinert-Threlkeld, Mocanu, Vespignani and Fowler 2015; Steinert-Threlkeld 2017; Clarke and Kocak 2020) and is consistent with the history of the movement's early days. However, the reverse direction of the link is novel and more unexpected: as our first main contribution, we show that those early roadblocks then spurred the creation of a new wave of Facebook groups. To establish causality, we use two different methods. First, we build the daily time series of group creations and we report a sharp discontinuity following the 11/17 roadblocks; second, we propose an instrumental variable strategy based on the spatial dispersion of roundabouts in French cities. Roundabouts are attractive protest locations because they enable demonstrators to block several roads simultaneously and are easy to set camp on. At the same time, they are widely recognized as architectural fads. Our estimates suggest that the presence of a local roadblock led to the creation of 1.2 additional local Facebook groups. We interpret this finding as evidence that protesters sought to continue interacting with each other online after the initial offline contact, resulting in the consolidation of the Yellow Vests' online infrastructure.

Despite the development of active Yellow vest online communities, the number of protesters in the street quickly subsided. To understand the unraveling of the movement, we turn to the analysis of a corpus of 2.8 million sentences posted on public Facebook pages. Using various Natural Language Processing (NLP) techniques, we show that messages related to organizational concerns and practical demands decreased over time, while those with more antagonistic content, such as insults or mentions of violence, increased. Similarly, the share of messages with a negative sentiment or a higher probability of being written by affiliates of extreme parties also increased. We conduct two quantitative exercises suggesting that the discussions radicalized partly because they were taking place on social media — these correspond to our second main contribution.

First, we show that discussants were disproportionately exposed to radical content. Since our dataset includes information on the ordering of comments displayed to users, we evaluate whether our radicalization measures correlate with the recommendations of Facebook's algorithm. We find that Facebook significantly modified the original comment ordering and that comments associated with our radicalization measures were more likely to appear prominently below Facebook posts. Second, we show that online discussions drove moderate discussants out and, conversely, increasingly attracted radical discussants. Using a de-identified panel of individual discussants and their posts, we can decompose the radicalization process we measure between an extensive margin (changes in the composition of the population of discussants) and an intensive margin (an individual-level increased tendency to post radical messages). Empirically, we find that both margins played an almost equally important role, although the effect of the extensive margin was slightly delayed compared to that of the intensive margin. This result is consistent with the notion that participation in these online communities is volatile and subject to the rise of echo chambers that limit the diversity of participants in favor of the most radically opinionated.

Overall, we view the findings presented in this paper as a cautionary tale on the impact of social media on the effectiveness of protest movements. Many studies, relying on various methodologies, have emphasized the importance of digital technologies in large-scale protest movements (see, among others, Loader 2008; Earl and Kimport 2011; Bennett and Segerberg

2012; Castells 2015; Manacorda and Tesei 2020).[2] Social media, in particular, has been shown to facilitate coordination, signaling, and mobilization among protesters.[3] As Qin, Strömberg and Wu (2017), we confirm that the close monitoring of social media may help predict where protests are more likely to occur. Moreover, in line with the intuition of Bastos, Mercea and Charpentier (2015), we show that street protests may revitalize online mobilization, thereby creating a feedback loop that helps protest movements stay active longer than initially planned. As such, social media may be instrumental to the return of local politics documented all over the world (Della Porta and Diani 2020; Le Galès 2021).

However, as argued, among others, by Tufekci (2017), social media also dramatically alters the very nature of protest movements. Despite its horizontal nature that facilitates the emergence of leaderless movements, social media is far from a neutral communication technology that would allow equal participation among activists. In particular, social media has long been accused of nurturing ideological segregation (Pariser 2011; Ross Arguedas, Robertson, Fletcher and Nielsen 2022). Even though the real impact of social media on polarization is still an open debate (see, among others, Bail, Argyle, Brown, Bumpus, Chen, Hunzaker, Lee, Mann, Merhout and Volfovsky 2018; Fletcher, Cornia and Nielsen 2020), this paper argues that such mechanisms may be particularly relevant in the case of protest movements, which tend to radicalize over time, partly in response to state repression (Della

[2] In the US, the Tea Party movement in 2009 is arguably the last large-scale protest movement that relied on the support of traditional media, such as local TV or radio talk shows (Madestam, Shoag, Veuger and Yanagizawa-Drott 2013).

[3] Noteworthy examples include Rane and Salem (2012); Acemoglu, Hassan and Tahoun (2018); Clarke and Kocak (2020); Gaffney (2010); Borge-Holthoefer, Rivero, García, Cauhé, Ferrer, Ferrer, Francos, Iniguez, Pérez, Ruiz et al. (2011); González-Bailón, Borge-Holthoefer, Rivero and Moreno (2011); Bursztyn, Cantoni, Yang, Yuchtman and Zhang (2021); Enikolopov, Makarin and Petrova (2020); Jost, Barberá, Bonneau, Langer, Metzger, Nagler, Sterling and Tucker (2018); Fergusson and Molina (2019).

Porta 2018). We show that Yellow vest online communities were subject to both echo chambers and algorithmic filter bubbles that likely contributed to the radicalization of online discussions, in parallel to high levels of violence in the streets. If social media is both a facilitator of protest movements and a pathway to their radicalization, its instrumental role in contemporary protest movements may contribute to explaining why mobilization today can be "very successful in terms of number but tends to be more volatile and intermittent than in the past" (Della Porta 2013).

Our analysis of the Yellow Vests' discussions also exemplifies that social media allows for real-time surveys of the coalitions behind social movements. Access to detailed information about the political preferences of the protesters, their opinion vis-à-vis policy-makers and their media strategies provides researchers with a unique opportunity to revisit longstanding debates on the nature of protest movements (Lipsky 1968; Meyer 2004). In particular, while a large strand of the literature on protest movements has been devoted to measuring their impact on various outcomes,[4] it is now becoming possible to follow protest dynamics over time and space with an unprecedented level of detail. However, these new monitoring possibilities are not without risks, particularly if online conversations allow authoritarian regimes to identify dissenters (Rød and Weidmann 2015; Earl, Maher and Pan 2022).

[4]Examples include Giugni (1998); Gillion (2012); Little, Tucker and LaGatta (2015); Frye and Borisova (2019); Ketchley and El-Rayyes (2021); Reny and Newman (2021); Siegel (2009); Meirowitz and Tucker (2013); Wallace, Zepeda-Millán and Jones-Correa (2014); Branton, Martinez-Ebers, Carey Jr and Matsubayashi (2015); Mazumder (2018); Larson, Nagler, Ronen and Tucker (2019).

2 Context, Data, and Methods

2.1 Brief History

The Yellow vest movement resulted from chance and the social media ecosystem. In May 2018, a motorist, Priscilla Ludosky, created a petition against gasoline taxes on the Change.org platform. Even though the petition had only garnered a few hundred signatures during its first months, it was mentioned in a local newspaper on October 12, 2018. The wife of a truck driver who had been planning a roadblock of the Paris ring road for November 17$^{\text{th}}$ read the article and linked the petition on Facebook. Nine days and thousands of local signatures later, a national newspaper published a new article on the petition and the roadblock project, and signatures skyrocketed nationwide. On October 24, the yellow road security jacket, which every car owner is compelled by law to have in her trunk, was proposed as a rallying sign for angry motorists. The organizers of roadblocks heavily relied on Facebook to spread information, and several dedicated websites were created to list relevant local Facebook groups. On November 17$^{\text{th}}$, hundreds of thousands of protesters blocked hundreds of roads across France.

The movement resorted to more conventional weekly demonstrations in France's main cities as most roadblocks were quickly evacuated. A climax of violence was reached on December 1$^{\text{st}}$ in Paris. The following Saturday, police tanks were mobilized, and 2,000 people were arrested. On December 10$^{\text{th}}$, as a token of peace, President Macron presented a 10-billion-euro plan that significantly bent the government's budgetary policy. In particular, he pledged a €100 per month increase in the minimum wage and excluded charges and taxes on overtime hours in 2019 and any 2018 end-of-year bonuses paid to employees. He further asked for a compilation of lists of grievances (*Cahiers de doléances*, as took place during the French Revolution in 1789) across the country, which was followed by hundreds of town hall meetings meant to allow everyone to voice their concerns through a "Great National Debate" (*Grand Débat National*).

After this response, some roadblocks became permanent campsites, and weekly demonstrations continued for months. However, the number of demonstrators soon became negligible (except in Paris, where some large-scale demonstrations gathering protesters from other parts of France still took place until March 2019). At the same time, the protesters lost public support in the French population and ultimately failed to present a united front for the upcoming elections (the 2019 European Parliament election on May 26$^{\text{th}}$). The movement was still active online in 2022 and organized sporadic protests where Yellow Vests were worn as a badge of honor. As such, this simple piece of garment has become a persistent and divisive icon in the French political landscape.

2.2 Measuring the Intensity of Mobilization

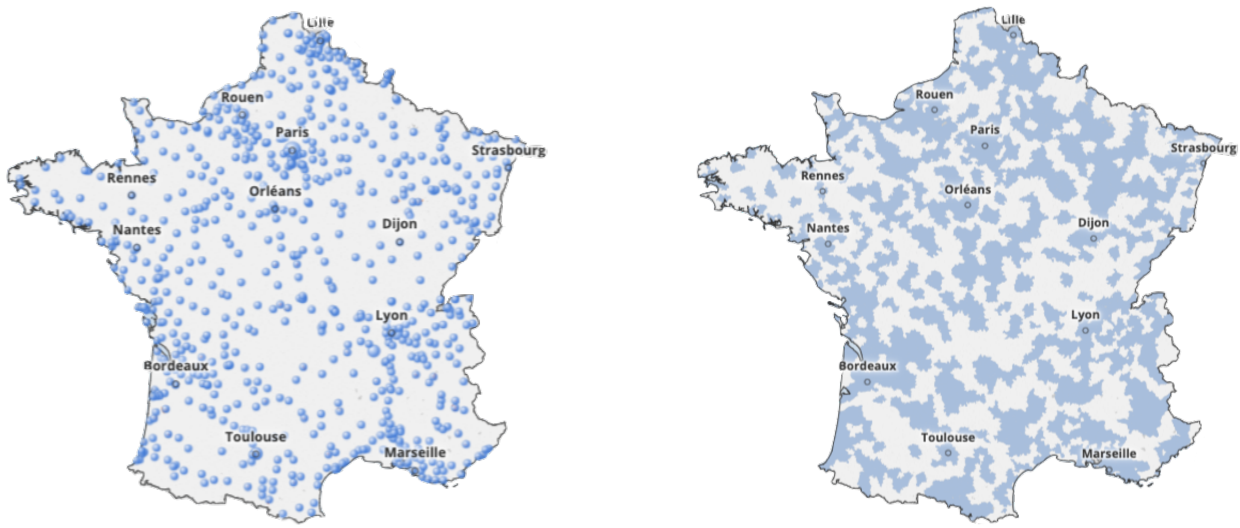
We gathered data on the mobilization using multiple sources (see Appendix A for details). To understand the movement's roots, we retrieved anonymized data on petition signatories from Change.org. The data includes signatories' ZIP codes, allowing us to geolocate them. By October 16, 2019, the petition had garnered 1,247,816 signatures, including 1,043,337 with a valid French ZIP code. We interpret petition signatures as signaling discontent towards the government and a willingness to protest.

To proxy offline mobilization, we collected a map of planned roadblocks on the evening of November 16$^{\text{th}}$. The map was downloaded directly from a website created by the protesters to coordinate demonstrations and roadblocks. This map documented 788 announced roadblocks in metropolitan France (see Panel A in Figure 1), which all pointed to precise road infrastructures (e.g., freeway access ramps, parking lots, but primarily roundabouts) and included specific descriptions of the planned events.[5] Many places were chosen for their

[5]Note that these are declarations of intent to demonstrate. However, as the map was created to coordinate the roadblocks, there was little incentive to falsely declare an intent to demonstrate. Contrary to what happens in autocratic regimes (Clarke and Kocak 2020; Hassan 2021), the French police did not preemptively try to lift the roadblocks on that day.

Figure 1: Blocking Half of France at First Try

A. Geolocation　　　　B. Affected Living Zones



Notes: Panel A displays the geolocation of the 11/17 roadblocks. Panel B displays the living zones with at least one roadblock on 11/17. These living zones gather 49 million people, 77% of the French mainland population.

potential to block traffic and economic activity. To analyze how France was affected by the roadblocks, we use the country's partition in "Living Zones" (hereafter, LZ). They are administrative units defined as the smallest geographical units in which residents can access basic infrastructure and services and conduct a large part of their daily lives. We observe that 551 out of the 1,632 LZs in France experienced a roadblock. They correspond to more than half of the country's population and, as shown in Panel B in Figure 1, to a sizable fraction of the French territory.

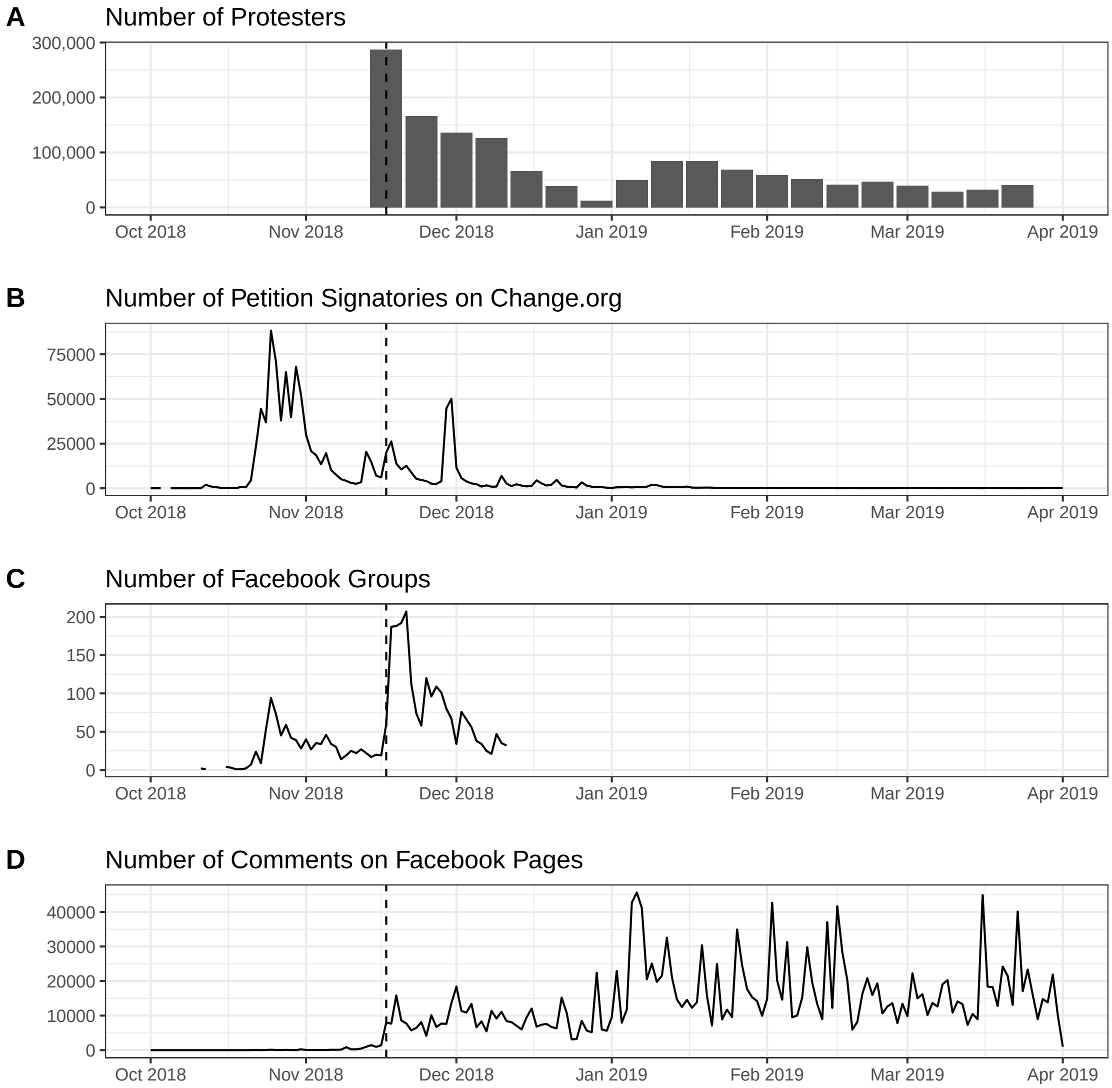
Finally, to document online mobilization, we searched for all public Facebook groups related to the movement. Using the methodology of Gaby and Caren (2012), we compiled a list of the Facebook groups that were still active one month after 11/17 by performing search requests using a large set of keywords linked to the movement. We recorded each group's name, creation date, number of members, and publications. We identified 3,033 groups with a total of over four million members. Over two-thirds of the groups were associated with a geographical area, and more than 40% of the total members belonged to these localized groups. Moreover, only 20% of the posts emanated from national groups, suggesting that localized groups were the most active ones.

Using a similar method, we also identified 617 Facebook pages and used Netvizz (Rieder 2013) to retrieve their content in March 2019: posts, comments, and interactions (such as likes and shares). This corpus features 120,227 posts, 2.1 million comments, 2.8 million sentences, and 21 million interactions. Netvizz did not provide user identifiers associated with each message. To build a panel of Facebook users, we scraped Facebook a second time in January 2022 and collected additional basic user information. This allows us to study the radicalization of Facebook content for a sample of 120,463 users in Section 4.[6]

In Figure 2, we depict the daily time series of the number of petition signatories, the number of Facebook group creations, and the number of comments on Facebook pages. The movement culminated in the streets during the first episode of roadblocks. While the petition was mostly signed before 11/17, there were two distinct episodes of group creation: a small one in the weeks before the roadblocks and a large one immediately afterward. This pattern suggests that Facebook groups were used to organize the roadblocks but also served as virtual meeting places that allowed the movement to continue after an initial mobilization in the streets. The evolution of the intensity of the discussions on dedicated Facebook pages corroborates this hypothesis. Discussions gained importance in January 2019 and, contrary to the weekly number of protesters, remained strongly active during the following months.

[6]To protect users' privacy, all users were de-identified. Approximately 30% of pages had been deleted by January 2022. On the remaining pages, we retrieved 46% of the original posts and 18% of the original comments for this second data retrieval (see Appendix Table A.2). To control for selection bias, we extensively compared both datasets. They are similar in terms of their distribution of political language and in terms of the topics discussed. They also display qualitatively similar trends in our three main outcome variables in Section 4.3 (see Appendix A.3.2).

Figure 2: Evolution of Online and Offline Mobilizations



Notes: In Panel A, we show the number of demonstrators reported weekly by the Ministry of the Interior. In Panel B, we plot the daily number of petition signatures. In Panel C, we plot the daily number of new Facebook groups created. Finally, in Panel D, we plot the daily number of messages posted on Facebook pages. The vertical dashed line in all panels corresponds to 11/17.

2.3 Textual Analysis of Facebook Discussions

To analyze discussions on Facebook pages, we rely on NLP methods (see, for an overview, Grimmer and Stewart 2013; Gentzkow, Kelly and Taddy 2019): a topic model, a sentiment analysis, and a political classification of the messages. The technical details of our implementation are provided in Appendix C.

To identify the topics discussed online by the Yellow Vests, we rely on a topic model tailored to analyze short text snippets (Demszky, Garg, Voigt, Zou, Gentzkow, Shapiro and Jurafsky 2019). In our main specification, we allow for 15 different topics, but qualitatively similar results are found with alternative numbers of clusters (see Table C.1). We display the topics obtained in Figure 3. We can group topics into different categories, such as protest organization (A and B), socialization (C and G), and online mobilization (D). Other topics reflect the reasons behind the protests and the political goals the Yellow Vests were trying to achieve (E and F). Finally, five topics refer to antagonistic messages (H, I, J, K, and L) and reflect the protesters' anger toward government officials and their policies.

To measure emotional content in messages, we use a dictionary-based approach that assigns to each sentence a sentiment score ranging from -1 (very negative) to 1 (very positive). Figure 3 splits the 15 topics between those with an average sentiment of messages above zero and those below zero. The five topics we classify as antagonistic are all associated with negative sentiment. Finally, to understand the political stance of messages, we build a measure of partisanship using a supervised learning model based on tweets from parliamentarians.

3 The Online-Offline Feedback Loop

This section shows that online and offline mobilizations reinforce each other. We estimate spatial regressions where we distinguish between mobilization before and after the 11/17 protests. We first document that the 11/17 protests were organized in localities with higher early online mobilization. This is expected – as most roadblocks were organized online – and consistent with previous findings showing the facilitating role of social media in organizing protests. However, the reverse direction of the loop (from street protests to further online mobilization) is novel and warrants deeper investigation. While the time series depicted in Figure 2 does suggest that new Facebook groups were created in the immediate aftermath of

Figure 3: Results of the Topic Model

A. Places [2.9%] B. Towns and hours [4.4%] C. Support [3.9%]

D. Diffusion [4.8%] E. Economic concerns [5.9%] F. Political institutions [7.6%]

G. Food and objects [6%] H. Critiques [6%] I. Insults [4.5%]

J. Violence [5.9%] K. Conspiracy [5.9%] L. Actions [7%]

M. Names [6.6%] N. Foreign Languages [8.4%] O. Other [20%]

Notes: This figure shows wordclouds associated with the fifteen topics we identify in our corpus. The size of words is determined by a term frequency-inverse document frequency (TF-IDF) metric, where each document is the entire collection of sentences associated with a given topic. This metric gives higher scores to words that are (i) more frequent in the corpus and (ii) particularly meaningful for each topic. Wordclouds are boxed inside a rectangle when the average sentiment of messages in the topic is negative. Squared brackets indicate the topic frequency (computed as the share of total messages in the corpus).

the 11/17 protests, this pattern may simply result from the fact that protesters were simultaneously looking for both online and offline ways of expressing their discontent. Therefore, we use an instrumental variables strategy to establish a causal relationship between offline and online mobilizations.

3.1 Empirical Framework

We construct a dataset at the municipality level (indexing observations by m).[7] There are more than 35,000 municipalities in France, and their boundaries date back to the French Revolution.[8] They are the lowest government level, allowing us to gather data on various characteristics. We complement those variables with a novel measure of Facebook penetration (see Appendix A.3.3 for details). Early online mobilization is defined by the vector $M_m^{\text{pre-11/17}}$, which includes the signature rate in the municipality before 11/17 and the number of local Facebook groups before 11/17.

The 11/17 protests are measured by $B_m^{11/17}$, a dummy variable equal to 1 if there was a roadblock in municipality m on 11/17. Finally, later mobilization is measured by $M_m^{\text{post-11/17}}$, which includes the signature rate in the municipality after 11/17, the number of groups created after 11/17, as well as the number of members and publications observed in those later groups, expressed in per-capita terms.

We document the relationship between $M_m^{\text{pre-11/17}}$ and $B_m^{11/17}$ by measuring simple correlations based on the OLS estimation of Equation 1:

$$B_m^{11/17} = M_m^{\text{pre-11/17}}\beta + X_m\gamma + \delta_{LZ(m)} + \varepsilon_m \tag{1}$$

[7]Some variables were only available at higher geographical levels. When relevant, we apportioned them according to municipal population.

[8]Excluding French overseas territories and Corsica from our sample leaves us with 34,434 municipalities.

where X_m is a large set of economic, geographic, demographic, and political controls (see the full list in Table B.1) and $\delta_{LZ(m)}$ is a LZ fixed effect to account for fixed unobserved heterogeneity at a higher spatial level.

Conversely, to gauge the impact of $B_m^{11/17}$ on $M_m^{\text{post-11/17}}$, we use 2SLS estimation. For a set of municipal characteristics Z_m that can serve as instruments of the roadblock probability, we can estimate the following first stage, predicting the probability of a roadblock in a municipality:

$$B_m^{11/17} = \alpha_1 + M_m^{\text{pre}-11/17}\beta_1 + X_m\gamma_1 + \delta_1^{LZ(m)} + Z_m\zeta + \varepsilon_m \tag{2}$$

and the following second stage, regressing a measure of online mobilization after 11/17, M_m^{post}, on the predicted roadblock probability from Equation 2:

$$M_m^{\text{post}} = \alpha_2 + M_m^{\text{pre}}\beta_2 + X_m\gamma_2 + \delta_2^{LZ(m)} + \eta\widehat{B_m^{11/17}} + \epsilon_m, \tag{3}$$

The coefficient η then provides us with the local average treatment effect of an 11/17 roadblock on subsequent online mobilization.

Easy-to-block locations: Roundabouts. To instrument the roadblocks, we leverage the presence of roundabouts in each municipality. The rationale for the relevance of this instrument is that calls for demonstrations urged protesters to block roundabouts. By design, they allow to block several roads at a time and possess a central median strip on which it is convenient to set camp.[9] The identifying assumption is that conditional on observable char-

[9]There always was a conflict in urban planning between favoring policing (such as during Haussmann's renovation of Paris in the XIX$^{\text{th}}$ century (Lefebvre 1968)) and favoring social contact. Car-based urban planning is generally viewed as on the policing side (Davis 1992). In that regard, the occupation of roundabouts by the Yellow Vests represents an ironic turn of events.

acteristics, the distribution of roundabouts only predicts future online mobilization through its impact on roadblocks. The history of roundabouts makes it likely that the conditional distribution of local roundabout density reflects local idiosyncrasies. Roundabouts are partly a French architectural fad, arguably invented in 1906 by the French urban planner Eugène Hénard. France has over sixty-thousand roundabouts (roughly four times more than the United Kingdom). One-third of French municipalities have at least one. While plausible road safety reasons support their use, they can almost always be replaced with traffic lights.

In support of our exclusion restriction, Appendix Table B.2 explores the determinants of the spatial distribution of roundabouts in France. As one can expect, the distribution of roundabouts in France is closely related to the population distribution: it explains more than 40% of the variance in roundabout density. LZ fixed-effects only have little explanatory power, indicating low levels of spatial auto-correlation in roundabout density. Importantly, other controls only explain a residual fraction of the variation in roundabout density after controlling for the population density. Finally, a map of the prediction error of roundabout density after an OLS regression, including our controls, shows a seemingly random distribution (see Appendix Figure B.1).

Assuming the exogeneity of this first instrument, we can leverage a second instrument, which will allow us to test overidentifying restrictions. Indeed, since organizing a roadblock requires significant manpower, protesters had to coordinate to choose roadblock locations. This spatial coordination problem suggests another instrument, which is the mirror image of the first: the density of roundabouts in the other municipalities of the LZ. Because of competition between easy-to-block locations, we expect municipalities surrounded by more roundabouts to be less likely blocked.

3.2 Protests Were Organized Online and Targeted Roundabouts

Table 1 presents OLS estimates of Equations 1 and 2. On top of estimates for β, we show estimates for the coefficients on the Facebook penetration rate. Column (1) displays our results without controlling for early mobilization measures. In line with a vast body of evidence on the role of social media in social movements, we find a significant and positive correlation between Facebook penetration and the occurrence of a roadblock. Column (2) shows that the petition signature rate is positively correlated with the occurrence of a roadblock, which suggests that signature rates may be interpreted as a signal for mobilization potential. More importantly, Column (3) shows that, as expected, early mobilization on Facebook is strongly correlated with the occurrence of a roadblock. In addition, when we include the number of Yellow vest Facebook groups associated with a municipality in the regression, the coefficient associated with the Facebook penetration rate drops by one-third.

Column (4), where we control for both the petition signature rate and the number of groups, shows that the coefficients on the signatures and groups are stable compared to Columns (2) and (3), which suggests that both types of online mobilization are not substitutes for one another. As shown in Column (5), a model that would only control for the existence of those measures would have a predictive power equal to 75% of that of the model with the full set of municipal covariates but without any Yellow-Vest specific controls.

Finally, Column (6) confirms that roundabouts played an essential role in organizing the protests: increasing the density of roundabouts in a municipality by one standard deviation increases the probability of a roadblock by 1.1 p.p. In addition, an increase of one standard deviation in the density of roundabouts in surrounding municipalities decreases the roadblock probability of a municipality by 8.3 p.p. Both variables are statistically significant at the 0.1% level.

3.3 Roadblocks Spurred Further Online Mobilization

Table 2 presents results for the second stage. The Kleibergen-Paap F-statistic equals 25, suggesting that our instruments are reasonably strong. In addition, the high p-values associated with the Hansen J-statistics indicate that we fail to reject the hypothesis that the overidentifying restrictions are valid. Column (1) shows that even though the bulk

Table 1: Predictors of a Roadblock in a Municipality

	(1)	(2)	(3)	(4)	(5)	(6)
Facebook penetration	0.462***	0.476***	0.304***	0.317***		0.301***
	(0.057)	(0.057)	(0.060)	(0.061)		(0.059)
Signatures (pre-17/11)		0.007***		0.006***	0.007***	0.006***
		(0.001)		(0.001)	(0.001)	(0.001)
Nb. groups (pre-17/11)			0.036***	0.035***	0.056***	0.034***
			(0.003)	(0.003)	(0.001)	(0.003)
Roundabouts (municipality)						0.011***
						(0.003)
Roundabouts (LZ)						-0.083***
						(0.017)
Controls	✓	✓	✓	✓		✓
Life zone FE	✓	✓	✓	✓		✓
Observations	34,449	34,449	34,449	34,449	34,475	34,434
Adjusted R-squared	0.20	0.20	0.24	0.24	0.15	0.25

Notes: This table shows OLS estimates for a linear probability model predicting whether a municipality experienced a roadblock or not, as formalized in Equation 1. "Signatures (pre-11/17)" is the municipality's signature rate of the Change.org petition before 11/17; "Nb. Groups (pre-11/17)" is the apportioned number of Facebook groups (from all geographical levels) created before 11/17. These two variables are standardized. We measure Facebook penetration in a municipality as the number of Facebook users who declare to live or come from that municipality, divided by the municipal population. This variable is standardized and divided by 100. The last column represents the first-stage estimates of Equation 3, associated with the second-stage results of Table 2. The two instruments we use are the number of roundabouts per squared kilometer in the municipality and the corresponding average for all other municipalities in the LZ. Both variables are standardized. Standard errors are clustered at the LZ level. *: $p < 0.1$, **: $p < 0.05$, ***: $p < 0.01$.

of petition signatures occurred before 11/17, having a roadblock increases the post-11/17 signature rate by 1.2 standard deviations. This result suggests that protests helped spread information about the Yellow Vests' demands at the end of 2018 when public support for the movement was still high. The previous signatory rate is also correlated with subsequent signatory dynamics.

We also find a strong positive impact of roadblocks on subsequent Facebook activity: a roadblock in a municipality increases the number of new local Facebook groups by 2.9 standard deviations (corresponding to 1.2 additional groups), which translates into an in-

Table 2: Effects of a Roadblock on Post-11/17 Online Mobilization

	(1) Signatures	(2) Groups	(3) Members	(4) Posts
Blockade	1.158***	2.925***	0.214**	0.138**
	(0.253)	(0.692)	(0.090)	(0.065)
Signatures (pre-17/11)	0.599***	-0.001	0.001	-0.002
	(0.022)	(0.006)	(0.008)	(0.007)
Groups (pre-17/11)	-0.036***	0.036	-0.001	-0.001
	(0.010)	(0.036)	(0.007)	(0.006)
Facebook penetration	-0.440	3.243***	5.912**	4.883*
	(0.357)	(0.482)	(2.675)	(2.753)
Controls	✓	✓	✓	✓
Life zone FE	✓	✓	✓	✓
Observations	34,434	34,434	34,434	34,434
Kleibergen-Paap F-stat	24.8	24.8	24.8	24.8
p-value Hansen	0.664	0.538	0.343	0.310

Notes: This table shows 2SLS estimates of the impact of a municipal roadblock on four measures of online mobilization after 11/17: the signature rate of the Change.org petition after 11/17 (column 1), the number of groups created post-11/17 (column 2), the number of members per inhabitant (column 3) and posts per inhabitant (column 4) in these newly created groups. Estimates of the first stage are displayed in column (6) in Table 1. All outcome variables are standardized. We cluster standard errors at the LZ level. *: $p < 0.1$, **: $p < 0.05$, ***: $p < 0.01$.

crease in the number of new members per inhabitant by 0.21 standard deviations, and in the number of posts per inhabitant by 0.14 standard deviations. These three measures of later mobilization on Facebook appear uncorrelated with early online mobilization but are positively correlated with the Facebook penetration rate.

Our results are robust to several specification changes (see Appendix Table B.3). In particular, the estimates are stable if we do not include controls (Panel A) or use only one roundabout instrument instead of two (Panels B and C). These three tests are reassuring regarding the validity of the exclusion restriction. Effects are also reasonably similar if we define location fixed effects and the instrument at the commuting zone ($N = 297$) rather than at the LZ ($N = 1606$) level (Panel D) or if we exclude the Paris region, which stands out along many dimensions (Panel E).

Overall, our results suggest that the 11/17 roadblocks contributed to the expansion of online activity. Roadblocks triggered a new wave of popularity for the Change.org petition and led to a consolidation of the online infrastructure of the Yellow vest movement on Facebook. While street protests subsided, protesters who had met in the streets relied on Facebook to continue their discussions.

4 The Rise of Online Radicalism

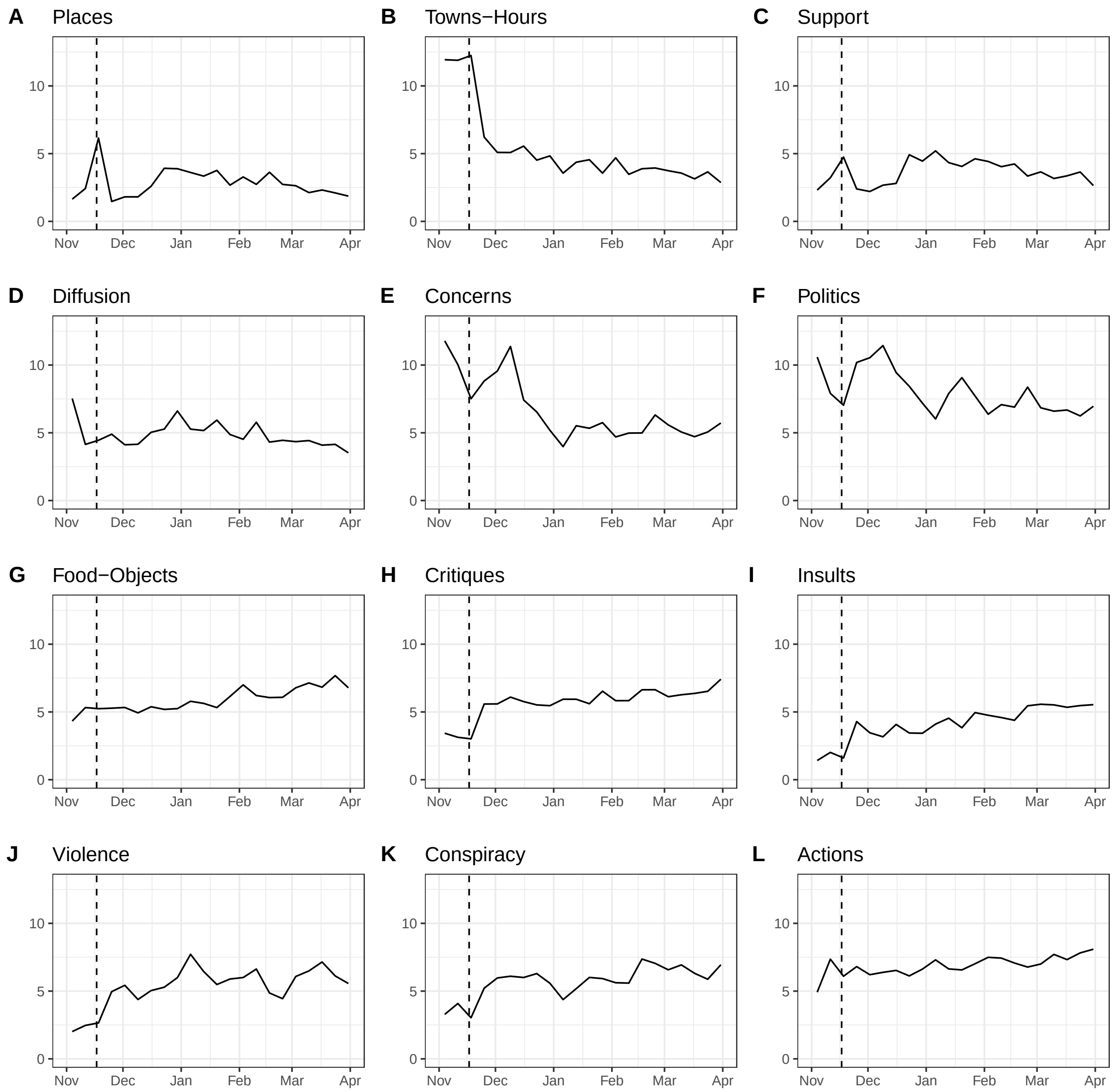
In this section, we document the evolution of online discussions among protesters through the textual analysis of the Yellow vest Facebook pages. Our data allows us to follow the content of Facebook pages between the end of October 2018 and the beginning of April 2019. We first document the evolution of online discussions. We then use the distinction between posts and comments to study the role of Facebook's recommender algorithm in structuring online discussions. Finally, as we observe the messages of individual members over time, we study how the composition of the population of discussants changed and whether some individuals radicalized over the period.

4.1 Online Discussions Became More Radical Over Time

Figure 4 shows that the share of messages associated with political or economic concerns declined while messages of violence, conspiracy theories, and insults became more widespread. Overall, the share of messages associated with antagonistic content (topics H to L) increased by 15 p.p. between November 2018 and March 2019. Other classifications of Facebook messages reflect similar trends over the period: the share of messages classified as negative (resp. associated with a far-right or far-left party) increased by 8 p.p. (resp. 6 p.p.).

Altogether, the content of online discussions became more antagonistic, negative, and polarized. We define these three concurrent dynamics as a "rise of online radicalism". Several

Figure 4: Topic Shares in Facebook Discussions Over Time



Notes: This figure shows weekly shares of the twelve topics of interest shown in Figure 3. For all topics, the vertical dashed line corresponds to 11/17. The share of messages associated with violence is below 2.5% in early November and is consistently above 5% after December 10.

non-exclusive reasons may explain this pattern. In particular, one may think of external reasons: on the positive side, the movement succeeded in having the government revert its policy so that economic concerns were less in need to be discussed. On the negative side, some protests were quite violent and met with police repression. In this context, it is no wonder that topics related to violence or insults rose to prominence.

In what follows, we describe two pieces of quantitative evidence suggesting that part of this increase in online radicalism may be due to specific online processes. First, we show that discussants were disproportionately exposed to radical content by Facebook. Second, we show that new, more radical discussants progressively replaced the more moderate ones.

4.2 Facebook's Algorithm Increased Exposure to Radical Content

A first way to assess how radical content spread on the Yellow vest Facebook pages is to look at the structure of online discussions, which involve an initial post and its associated comments. To assess whether Facebook played an active role in the type of content that the Yellow Vests were exposed to online, we focus on the organization of discussions by Facebook's platform.

While Facebook displays posts chronologically on Facebook pages, it does not deal with their associated comments similarly. Instead, undisclosed algorithms rank comments by what the platform calls "relevance." Since our dataset contains information on the ordering of comments shown to users at the time of the scrape, we can assess whether our radicalization measures are correlated with the recommendations of Facebook's algorithm. To that end, we regress the rank of each comment in our text corpus on our measures of radicalism. Since posts vary a lot in their content and the number of comments they generate, we control for post fixed effects.[10] We also control for a measure of the rank of the comment based on the time when the comment was posted.

Results are displayed in Table 3. Measures of rank based on Facebook suggestions and time of posting are positively correlated. However, the estimates are quite far from unity, which shows that Facebook strongly alters the original ordering of comments. In particular,

[10] Out of our original corpus of 120,227 posts, we focus on the 35,828 with at least two comments. Estimates are qualitatively similar but slightly higher if we do not include post fixed effects.

Table 3: Comments' Rank and Radical Content

	Rank of the Comment (in log)				Comment is Among First Four			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Antagonistic Topic	-0.095***			-0.081***	0.004***			0.003***
	(0.005)			(0.004)	(0.000)			(0.000)
Extreme Parties		-0.028***		-0.017***		0.002***		0.002***
		(0.003)		(0.003)		(0.000)		(0.000)
Negative Sentiment			-0.085***	-0.065***			0.004***	0.003***
			(0.005)	(0.004)			(0.000)	(0.000)
Chronological Order	0.127***	0.126***	0.126***	0.127***	0.014***	0.014***	0.014***	0.014***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.003)	(0.003)	(0.003)	(0.003)
Post Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
R-Squared	0.81	0.81	0.81	0.81	0.48	0.48	0.48	0.48

Notes: This table shows estimates of OLS regressions at the sentence level (N=1,881,976). We restrict the text corpus to comments (and exclude original posts). Some comments are made of several sentences, but results are similar if we restrict the sample to single-sentence comments (61% of the sample). In Columns (1) to (4), the dependent variable is the (log) rank of the comment suggested by Facebook at the time of the scrape. In Columns (5) to (8), the dependent variable is a dummy variable equal to 1 if the comment is among the first four comments suggested by Facebook at the time of the scrape. "Antagonistic Topic" is a dummy variable equal to 1 if the sentence is classified as belonging to an antagonistic topic. "Extreme Parties" is a dummy variable equal to 1 if the sentence is attributed to an extreme party. "Negative Sentiment" is a dummy variable equal to 1 if the sentence is associated with a negative sentiment value. "Chronological Order" is defined as the counterpart of the dependent variable, based on chronological order: the (log) rank of the comment based on chronological order in Columns (1) to (4), and a dummy variable equal to 1 if the comment was among the first four to be posted in Columns (5) to (8). All specifications include a post fixed effect. The values of the R-Squared change for the fourth decimal. In all regressions, we cluster standard errors at the post level. *: p< 0.01, **: p<0.05, ***: p<0.1.

comments associated with our radicalization measures are more likely to be found higher on the list. For example, comments associated with antagonistic topics are displayed at a rank between 8 and 10% higher than other comments. The same patterns appear if we focus on the probability of being a "star comment", which we take as one of the first four comments below the post (10% of our estimation sample). Such comments are likely to appear in users' newsfeeds without further clicking and are, therefore, much more likely to be salient and read by users. These results show that a chronological order of comments would have provided discussants with less radical content.

4.3 Radical Discussants Replaced the Moderates

Filter bubbles that push radical comments on the front page are not the only force driving radicalization. Another specificity of online communities is that participation costs are quite low, and participation may be volatile.[11] As such, online discussions are susceptible to being hijacked by a minority of radical users, who drive others away and attract similarly-minded discussants (Sunstein 2017). To test this mechanism, we decompose the rise in online radicalism between two margins. First, moderate users could have progressively left the movement or been replaced by more radical ones. We refer to this depletion effect as the "extensive margin" of radicalization. Alternatively, active users may have radicalized over time. We refer to such individual changes as the "intensive margin" of radicalization.

To assess whether the trends we observe are more likely to reflect shifts in the intensive or extensive margin of radicalization, we exploit the panel dimension of the data and the fact that we can follow (de-identified) individual Facebook users over time. To identify the intensive margin of radicalization, we can evaluate whether the average user became increasingly likely to post radical messages. To identify the extensive margin, we can evaluate whether the pool of active users becomes increasingly populated with users who (on average) post more radical messages. To this end, we estimate the following fixed-effects equation:

$$Y_s = \delta_{i(s)} + \gamma_{t(s)} + \varepsilon_s, \tag{4}$$

where Y_s is a measure of radicalism of sentence s, $\delta_{i(s)}$ is a fixed effect associated with the user i who posted sentence s, and $\gamma_{t(s)}$ is a fixed effect associated with the month t during which sentence s was posted. Intuitively, $\delta_{i(s)}$ measures user i's propensity to post radical sentences, and $\gamma_{t(s)}$ accounts for the additional propensity of users to post radical sentences during month t.

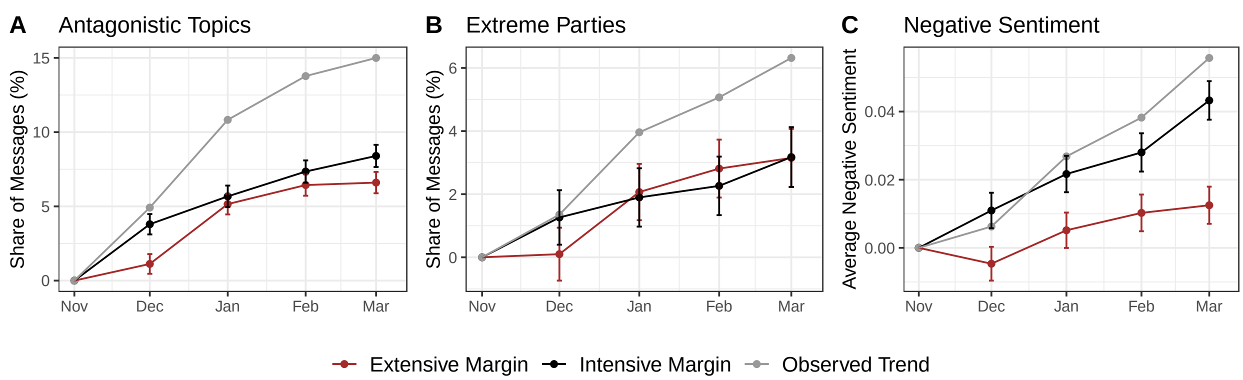
[11] Measuring participation in online communities is notoriously difficult, because of the presence of non-active participants (Malinen 2015).

We can then leverage estimates of user and time fixed effects to decompose the rise of online radicalism into an intensive and extensive margin. Indeed, the average level of radicalism during month t, $\bar{Y}_t$, can be expressed as:

$$\bar{Y}_t = \underbrace{\hat{\mathbb{E}}_t\left[\hat{\delta}_i\right]}_{\text{Extensive margin}} + \underbrace{\hat{\gamma}_t}_{\text{Intensive margin}}, \tag{5}$$

where $\hat{\mathbb{E}}_t\left[\hat{\delta}_i\right] = \sum_i s_{i,t}\hat{\delta}_i$ and $s_{i,t}$ is the share of sentences posted during month t that originated from user i. Hence, the first term of expression 5 corresponds to the average propensity to post radical sentences for users active during the month t. An increase of this term over time means that the share of sentences posted by more radical users increases. An increase in the second term of expression 5 corresponds to an increase in the propensity of any given user to post a radical sentence at a given time.

Figure 5: Extensive and Intensive Margins of Radicalization



Notes: This figure decomposes the increase in online radicalism using Equation 5. Panel A presents estimates for the probability of posting a sentence associated with an antagonistic topic. Panel B presents estimates for the probability of writing a sentence associated with a politically extreme party (i.e., on the far left or the far right). Panel C presents estimates for negative sentiment. We compute standard errors via bootstrap with 200 iterations and plot confidence intervals at the 95% confidence level.

Figure 5 presents a decomposition of our three radicalization measures using Equation 5. In Panel A, the outcome variable is a dummy variable that indicates whether a message was associated with an antagonistic topic. In Panel B, the outcome variable is a dummy variable that indicates whether a message was associated with an extreme political party. In Panel C, the outcome variable is the negative sentiment score associated with a sentence, taking values between -1 and 1. For all three dependent variables, our decomposition exercise suggests that both margins contributed to the radicalization of Facebook content. In addition, both margins played a quantitatively similar role in two out of three measures, although the impact of the extensive margin seems slightly delayed compared to that of the intensive margin.

We interpret this finding on the role of the extensive margin of radicalization as supporting evidence that participation in online communities is quite volatile and prone to being taken over by more radical discussants over time. However, since we lack information on offline protesters, we cannot tell whether this pattern is specific to online mobilization. In addition, our linear decomposition does not allow us to study the interplay between both margins. For example, the radicalization of other discussants may have led to the departure of moderate discussants, or, conversely, remaining discussants may have faced lower moderation from fellow discussants over time. Further understanding the interplay between online and offline sources of radicalization would require having access to individual newsfeeds, combined with detailed accounts of protesting activity.

4.4 Did Radicalization Lead to Demobilization?

The online radicalization of a movement does not necessarily hinder its persistence. For instance, the homogenization of protesters on Facebook may have helped the Yellow Vests consolidate a base of loyal supporters who shared similar views and kept mobilizing over more extended periods. However, the significant increase in antagonistic topics and partisanship prevented the movement from appealing to a broader base of supporters. Thus, the offline mobilization persisted, but with marginal numbers of protesters. Between January and April 2019, polls indicate that the movement lost its overall population support, particularly among centrist voters (see Appendix Figure A.9). At the same time, the number of organized protests remained steady, but those events attracted fewer and fewer participants, even according to the Yellow Vests' own monitoring system (see Appendix Figure A.1).

5 Conclusion

Large protest movements are now a combination of online and offline mobilizations. Many have noted that social media favors the emergence of protests by lowering coordination costs and making it easier to signal discontent. Our study confirms that social media and online protests likely increase the number of street protesters at the start of a protest movement. Moreover, we provide novel evidence that real-life demonstrations may also intensify subsequent online engagement, thus prolonging protest movements' lifespan. However, these persistent online communities are subject to radicalization, and we provide several pieces of evidence suggesting that the social media infrastructure itself may drive part of this radicalization process. Put together, our findings highlight one core tension of hybrid social movements: on the one hand, social media allow online communities to arise from everywhere, even remote locations such as rural areas or suburban fringes; on the other hand, a strong dependency on a leaderless social media infrastructure presents risks of radicalization and may dampen the ability to structure long-lasting, effective political campaigns.

References

Acemoglu, Daron, Tarek A Hassan, and Ahmed Tahoun, "The Power of the Street: Evidence From Egypt's Arab Spring," *Review of Financial Studies*, 2018, *31* (1), 1–42.

Algan, Yann, Elizabeth Beasley, Daniel Cohen, and Martial Foucault, *Les Origines Du Populisme*, La République des Idées. Seuil, 2019.

Arguedas, A Ross, Craig Robertson, Richard Fletcher, and Rasmus Nielsen, "Echo Chambers, Filter Bubbles, and Polarisation: A Literature Review," 2022.

Bail, Christopher A., Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, Haohan Chen, M. B. Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky, "Exposure to opposing views on social media can increase political polarization," *Proceedings of the National Academy of Sciences of the United States of America*, 2018, *115* (37), 9216–9221.

Bastos, Marco T., Dan Mercea, and Arthur Charpentier, "Tents, Tweets, and Events: The Interplay Between Ongoing Protests and Social Media," *Journal of Communication*, 2015, *65* (2), 320–350.

Bendali, Zakaria and Aldo Rubert, "Les Sciences Sociales en Gilet Jaune," *Politix*, 2020, *4*, 177–215.

Bennett, W Lance and Alexandra Segerberg, "The Logic of Connective Action: Digital Media and the Personalization of Contentious Politics," *Information, Communication & Society*, 2012, *15* (5), 739–768.

Borge-Holthoefer, Javier, Alejandro Rivero, Iñigo García, Elisa Cauhé, Alfredo Ferrer, Darío Ferrer, David Francos, David Iniguez, María Pilar Pérez, Gonzalo Ruiz et al., "Structural and Dynamical Patterns on Online Social Networks: The Spanish May 15th Movement as a Case Study," *PloS One*, 2011, *6* (8), e23883.

Boyer, Pierre C., Thomas Delemotte, Germain Gauthier, Vincent Rollet, and Benoît Schmutz, "The Origins of the Gilets Jaunes Movement," *Revue Économique*, 2020, *71* (1), 109–138.

Branton, Regina, Valerie Martinez-Ebers, Tony E Carey Jr, and Tetsuya Matsubayashi, "Social Protest and Policy Attitudes: The Case of the 2006 Immigrant Rallies," *American Journal of Political Science*, 2015, *59* (2), 390–402.

Bursztyn, Leonardo, Davide Cantoni, David Yang, Noam Yuchtman, and Jane Zhang, "Persistent Political Engagement: Social Interactions and the Dynamics of Protest Movements," *American Economic Review: Insights*, 2021, *3* (2), 233–50.

Castells, Manuel, *Networks of Outrage and Hope: Social Movements in the Internet Age*, Wiley, 2015.

Clarke, Killian and Korhan Kocak, "Launching revolution: Social media and the Egyptian uprising's first movers," *British Journal of Political Science*, 2020, *50* (3), 1025–1045.

Cointet, Jean-Philippe, Pedro Ramaciotti Morales, Dominique Cardon, Caterina Froio, Andrei Mogoutov, Benjamin Ooghe, and Guillaume Plique, "De Quelle(s) Couleur(s) Sont Les Gilets Jaunes? Plonger Des Posts Facebook Dans Un Espace Idéologique Latent," *Statistique Et Société*, 2021, *9* (1-2), 79–107.

Davis, Mike, *City of Quartz: Excavating the Future in Los Angeles*, Vintage Books, 1992.

Della Porta, Donatella, *Can Democracy Be Saved?*, Polity Press, 2013.

_ , "Radicalization: A Relational Perspective," *Annual Review of Political Science*, 2018, *21*, 461–474.

Della Porta, Donatella and Mario Diani, *Social Movements: An Introduction*, Wiley-Blackwell, 2020.

Demszky, Dorottya, Nikhil Garg, Rob Voigt, James Zou, Matthew Gentzkow, Jesse Shapiro, and Dan Jurafsky, "Analyzing Polarization in Social Media: Method and Application to Tweets on 21 Mass Shootings," in "Proceedings of NAACL-HLT" 2019, pp. 2970–3005.

Earl, Jennifer and Katrina Kimport, *Digitally Enabled Social Change: Activism in the Internet Age*, Mit Press, 2011.

_ , Thomas V. Maher, and Jennifer Pan, "The digital repression of social movements, protest, and activism: A synthetic review," *Science Advances*, 2022, *8* (10), eabl8198.

Enikolopov, Ruben, Alexey Makarin, and Maria Petrova, "Social Media and Protest Participation: Evidence From Russia," *Econometrica*, 2020, *88* (4), 1479–1514.

Fergusson, Leopoldo and Carlos Molina, "Facebook Causes Protests," *Documento CEDE*, 2019, (41).

Fletcher, Richard, Alessio Cornia, and Rasmus Kleis Nielsen, "How Polarized Are Online and Offline News Audiences? A Comparative Analysis of Twelve Countries," *The International Journal of Press/Politics*, 2020, *25* (2), 169–195.

Frye, Timothy and Ekaterina Borisova, "Elections, Protest, and Trust in Government: A Natural Experiment From Russia," *Journal of Politics*, 2019, *81* (3), 820–832.

Gaby, Sarah and Neal Caren, "Occupy Online: How Cute Old Men and Malcolm X Recruited 400,000 US Users to OWS on Facebook," *Social Movement Studies*, 2012, *11* (3-4), 367–374.

Gaffney, Devin, "#iranElection: Quantifying Online Activism," in "In Proceedings of the Web Science Conference" Citeseer 2010.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy, "Text as Data," *Journal of Economic Literature*, 2019, *57* (3), 535–74.

Gillion, Daniel Q, "Protest and Congressional Behavior: Assessing Racial and Ethnic Minority Protests in the District," *Journal of Politics*, 2012, *74* (4), 950–962.

Giugni, Marco G, "Was It Worth the Effort? The Outcomes and Consequences of Social Movements," *Annual Review of Sociology*, 1998, *24* (1), 371–393.

González-Bailón, Sandra, Javier Borge-Holthoefer, Alejandro Rivero, and Yamir Moreno, "The Dynamics of Protest Recruitment Through an Online Network," *Scientific Reports*, 2011, *1* (1), 1–7.

Grimmer, Justin and Brandon M Stewart, "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts," *Political Analysis*, 2013, *21* (3), 267–297.

Hassan, Mai, "Coordinated Dis-Coordination," *American Political Science Review*, 2021, pp. 1–15.

Jost, John T, Pablo Barberá, Richard Bonneau, Melanie Langer, Megan Metzger, Jonathan Nagler, Joanna Sterling, and Joshua A Tucker, "How Social Media Facilitates Political Protest: Information, Motivation, and Social Networks," *Political Psychology*, 2018, *39*, 85–118.

Ketchley, Neil and Thoraya El-Rayyes, "Unpopular Protest: Mass Mobilization and Attitudes to Democracy in Post-Mubarak Egypt," *Journal of Politics*, 2021, *83* (1), 291–305.

Larson, Jennifer M, Jonathan Nagler, Jonathan Ronen, and Joshua A Tucker, "Social Networks and Protest Participation: Evidence From 130 Million Twitter Users," *American Journal of Political Science*, 2019, *63* (3), 690–705.

Le Galès, Patrick, "The Rise of Local Politics: A Global Review," *Annual Review of Political Science*, 2021, *24* (1), 345–363.

Lefebvre, Henri, *Le Droit À La Ville*, Anthropos, 1968.

Lipsky, Michael, "Protest as a Political Resource," *American Political Science Review*, 1968, *62* (4), 1144–1158.

Little, Andrew T, Joshua A Tucker, and Tom LaGatta, "Elections, Protest, and Alternation of Power," *Journal of Politics*, 2015, *77* (4), 1142–1156.

Loader, Brian D, "Social Movements and New Media," *Sociology Compass*, 2008, *2* (6), 1920–1933.

Madestam, Andreas, Daniel Shoag, Stan Veuger, and David Yanagizawa-Drott, "Do Political Protests Matter? Evidence From the Tea Party Movement," *Quarterly Journal of Economics*, 2013, *128* (4), 1633–1685.

Malinen, Sanna, "Understanding user participation in online communities: A systematic literature review of empirical studies," *Computers in Human Behavior*, 2015, *46*, 228–238.

Manacorda, Marco and Andrea Tesei, "Liberation Technology: Mobile Phones and Political Mobilization in Africa," *Econometrica*, 2020, *88* (2), 533–567.

Mazumder, Soumyajit, "The Persistent Effect of US Cvil Rights Protests on Political Attitudes," *American Journal of Political Science*, 2018, *62* (4), 922–935.

Meirowitz, Adam and Joshua A Tucker, "People Power or a One-Shot Deal? A Dynamic Model of Protest," *American Journal of Political Science*, 2013, *57* (2), 478–490.

Meyer, David S., "Protest and Political Opportunities," *Annual Review of Sociology*, 2004, *30* (1), 125–145.

Pariser, Eli, *The Filter Bubble: What the Internet Is Hiding*, Princeton university press, 2011.

Qin, Bei, David Strömberg, and Yanhui Wu, "Why Does China Allow Freer Social Media? Protests Versus Surveillance and Propaganda," *Journal of Economic Perspectives*, 2017, *31* (1), 117–40.

Rane, Halim and Sumra Salem, "Social Media, Social Movements and the Diffusion of Ideas in the Arab Uprisings," *Journal of International Communication*, 2012, *18* (1), 97–111.

Reny, Tyler T and Benjamin J Newman, "The Opinion-Mobilizing Effect of Social Protest Against Police Violence: Evidence From the 2020 George Floyd Protests," *American Political Science Review*, 2021, *115* (4), 1499–1507.

Rieder, Bernhard, "Studying Facebook via Data Extraction: The Netvizz Application," in "Proceedings of the 5th annual ACM web science conference" ACM 2013, pp. 346–355.

Rød, Espen Geelmuyden and Nils B Weidmann, "Empowering activists or autocrats? The Internet in authoritarian regimes," *Journal of Peace Research*, 2015, *52* (3), 338–351.

Sebbah, Brigitte, Natacha Souillard, Laurent Thiong-Kay, and Nikos Smyrnaios, "Les Gilets Jaunes, Des Cadrages Médiatiques Aux Paroles Citoyennes," *Rapport De Recherche Préliminaire*, 2018.

Siegel, David A, "Social Networks and Collective Action," *American Journal of Political Science*, 2009, *53* (1), 122–138.

Steinert-Threlkeld, Zachary C, "Spontaneous collective action: Peripheral mobilization during the Arab Spring," *American Political Science Review*, 2017, *111* (2), 379–403.

_ **, Delia Mocanu, Alessandro Vespignani, and James Fowler**, "Online social networks and offline protest," *EPJ Data Science*, 2015, *4* (1), 1–9.

Sunstein, Cass R, *# Republic: Divided Democracy in the Age of Social Media*, Princeton university press, 2017.

Tufekci, Zeynep, *Twitter and tear gas: the power and fragility of networked protest*, New Haven; London: Yale University Press, 2017.

Wallace, Sophia J, Chris Zepeda-Millán, and Michael Jones-Correa, "Spatial and Temporal Proximity: Examining the Effects of Protests on Political Attitudes," *American Journal of Political Science*, 2014, *58* (2), 433–448.

Zhuravskaya, Ekaterina, Maria Petrova, and Ruben Enikolopov, "Political Effects of the Internet and Social Media," *Annual Review of Economics*, 2020, *12*, 415–438.

Supplementary Material
(for Online Publication)

Contents

A Data Sources

A.1 Street Protests

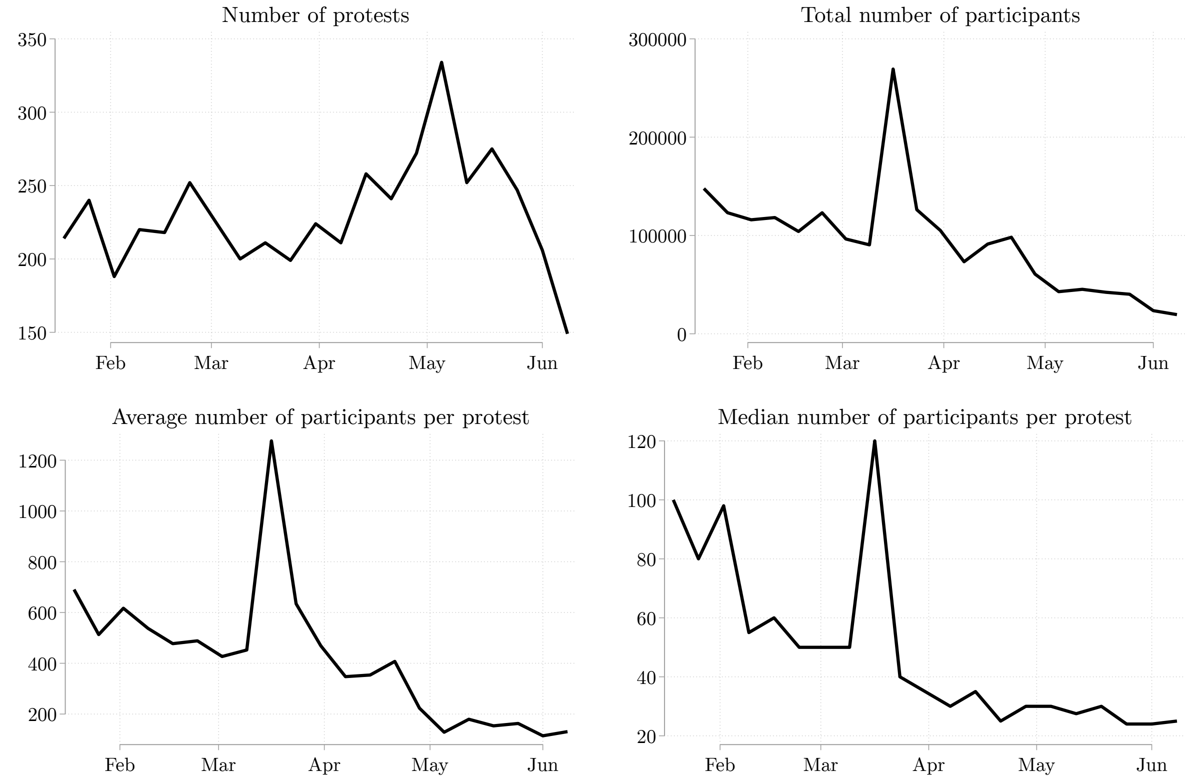
A website (www.blocage17novembre.fr) was created to coordinate the mobilization. It provided a map of the organized blockades, updated in real-time. As of November 16, the map documented 788 geolocated blockades. We use this map to document the offline mobilization of the Yellow Vests.

Starting from January 19th, 2019 (the seventh week of the Yellow Vest movement), a group of Yellow Vests, called *Le Nombre Jaune* ("the Yellow Number") started to collect statistics about the number of participants to Yellow Vest demonstrations across the country. Each week, they published a dataset containing a list of Yellow Vest demonstrations that took place on that week's Saturday, along with the estimated number of demonstrators that participated in each protest. To build these datasets, members of Le Nombre Jaune relied on articles from local newspapers, videos published online, as well as reports from demonstrators. We show in Figure A.1 measures of the intensity of offline Yellow Vest activity in 2019, taken from the Nombre Jaune datasets.

A.2 Change.org Petition

Change.org generously gave us access to an anonymized list of the signatories of the petition which launched the Yellow Vests movement. Each observation is associated with the date of signature and the ZIP code of the signatory. We restrict the data to signatures in mainland France and with a valid ZIP code. Using the ZIP code, we are able to associate each signature with a municipality, and therefore compute the signature rate in each municipality by dividing the number of signatures in each municipality by its population. In some instances, a ZIP code is associated with several municipalities. In these cases, we allocate signatures associated to this ZIP code across relevant municipalities proportionally to population. In Figure A.6, we map the distribution of signature rates over France.

Figure A.1: Measures of offline Yellow Vest activity from Le Nombre Jaune



Notes: This figure describes the intensity of Yellow Vest protests in the first part of 2019, as reported by Le Nombre Jaune. On March 16th, 2019, the Yellow Vest protests were organized jointly with a demonstration for climate awareness ("marche pour le climat"), and the numbers from Le Nombre Jaune for that date include participants to both events, explaining the observed spike in mobilization.

A.3 Facebook Activity

The main websites coordinating demonstrations listed local Facebook groups.[12] To document online mobilization, we looked for public Facebook groups and pages related to the movement. Due to the limitations of the Facebook API, we had to look for groups and pages manually, between December 12 and December 15, 2018 for groups and between March 21 and March 23, 2019 for pages. We used Netvizz to retrieve content between April 2 and April 10, 2019. Note that Netvizz did not allow us to retrieve actual discussions happening on Facebook groups. We use a keyword search approach to find Facebook groups and pages, performing requests on Facebook's search engine and manually retrieving results.

[12] First blocage17novembre.fr, then gilets-jaunes.com and giletsjaunes-coordination.fr.

These searches were performed using temporary sessions in order to minimize bias induced by Facebook's algorithm.

For groups, our aim was to retrieve as many groups linked to the Yellow Vests as possible. To this end, we started by searching for the keywords "gilet jaune" and "hausse carburant", on their own and associated with the the codes and names of the départements and of the former and current regions, as well as the names of all municipalities with more than 10,000 inhabitants.[13] Then, we performed further searches with the keywords "hausse taxes", "blocage", "colere" and "17 novembre", associated with the names of the French départements, the names of the former and current regions, and the same list of municipalities as before. Finally, we performed searches for the following keywords: "gillet jaune", "gilets jaune", "manif 17 novembre", "manif 24 novembre", "manif 1 decembre", "manif 8 decembre", "macron 17 novembre", "macron 24 novembre", "macron 1 decembre", "macron 8 decembre","blocus 17 novembre", "blocus 24 novembre", "blocus 1 decembre", "blocus 8 decembre", "blocage 17 novembre", "blocage 24 novembre", "blocage 1 decembre", "blocage 8 decembre".[14]

For pages, as our aim was not to retrieve the universe of active Yellow Vests communities but simply a sample of messages large enough to perform text analysis, we relied on a smaller number of searches, searching for the keywords "gilet jaune" and "blocage hausse carburant" on their own or associated with the codes and names of the départements as well as a list of

[13]Restricting the keywords used to these large municipalities is necessary as the number of municipalities in France is very high. It might introduce a bias towards groups associated to denser areas. Fortunately, this bias is reduced by a characteristic of Facebook's algorithm: when searching for groups and pages associated with a municipality on the platform, Facebook also retrieves results associated to nearby municipalities.

[14]We reviewed all the search results manually to only keep the groups clearly associated with the mouvement.

Table A.1: Characteristics of Facebook groups

Targeted Audience	Groups	Members	Publications
General	502 (63%)	2,372,217	255,131
Region	164 (81%)	244,930	135,857
County	717 (81%)	507,729	320,263
Municipality	1,638 (65%)	983,057	742,036
Total	3,033 (70%)	4,109,325	1,453,878

Notes: In the first column of this table, we show the number of Facebook groups for each geographic focus. We infer the group's targeted audience from its name. In parentheses, we indicate the share of the number of groups created after 11/17. Other columns show the total number of members and the total number of publications (this number is right-censored by Facebook at 10,000 publications per group). The last line ("Total") includes 12 "foreign" groups, 11 of which were created after 11/17, including 1,392 members and associated with 591 publications.

the largest cities.[15]

A.3.1 Yellow Vests Groups

For each group, we recorded the group's name, creation date, number of members, and number of publications. We eventually identified 3,033 groups in total, with over four million members. Over two-thirds of the groups were associated with a geographical area, and more than 40% of the total members belonged to these localized groups. Moreover, only 20% of the posts emanated from national groups, suggesting that localized groups were the most active type. Table A.1 presents descriptive statistics on the dataset. Figure A.7 displays the spatial distribution of these groups before (Panel A) and after (Panel B) 11/17.

[15] The complete list of further keywords used is the following: paris; marseille; lyon; toulouse; nice; nantes; strasbourg; montpellier; bordeaux; lille; rennes; reims; le havre; saint etienne; toulon; grenoble; dijon; angers; villeurbanne; le mans; nimes; aix en provence; brest; clermont ferrand; limoges; tours.

A.3.2 Yellow Vests Pages

We identified 617 Facebook pages and used Netvizz to retrieve their content (Rieder 2013): posts, comments, and interactions (such as likes and shares).[16] This corpus features over 121,000 posts, 2.1 million comments, and 21 million interactions. Since Netvizz did not provide user ids associated with scraped content, we scraped Facebook again in January 2022 and collected additional basic user information. To protect users' privacy, all user ids were de-identified. Approximately 30% of pages had been deleted by January 2022. On the remaining pages, we retrieved 46% of the original posts and 18% of the original comments for this second data retrieval (see Table A.2). Both datasets appear similar both in terms of their distribution of political preferences and in terms of the topics discussed. They also display qualitatively similar trends in our three main outcome variables in Section 4.3, though the second dataset generally displays larger increases in radical attitudes.

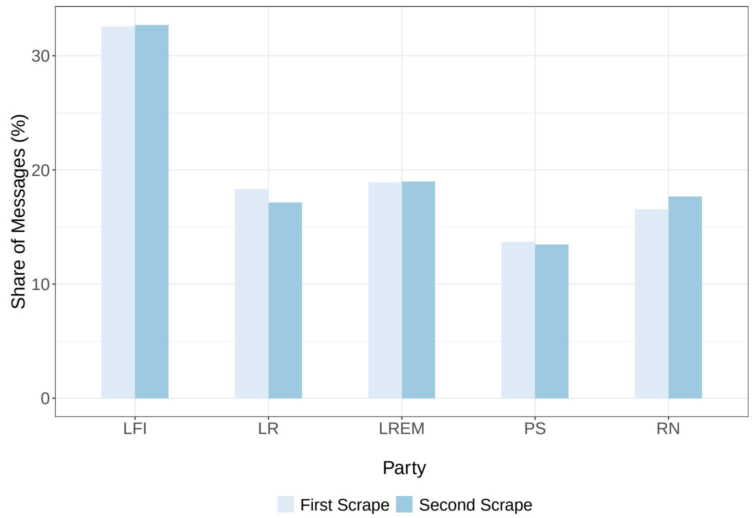
Table A.2: Comparison Between the Two Data Collections on Facebook Pages

Data Collection	Pages	Posts	Comments	Sentences	Users
First	617	120,242	1,936,921	2,860,427	NA
Second	411	56,062	352,733	706,182	120,463

Notes: This table presents simple count metrics to compare the datasets resulting from our two data collections on Facebook pages.

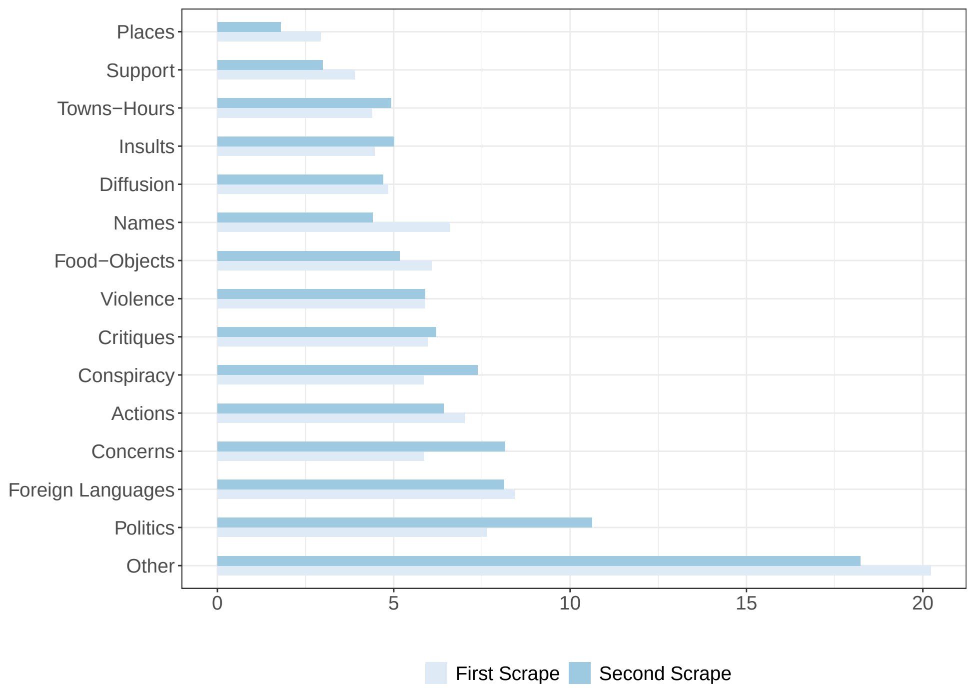
[16]Netvizz is no longer available since the 21st August 2019.

Figure A.2: Political Attitudes for Each Data Collection



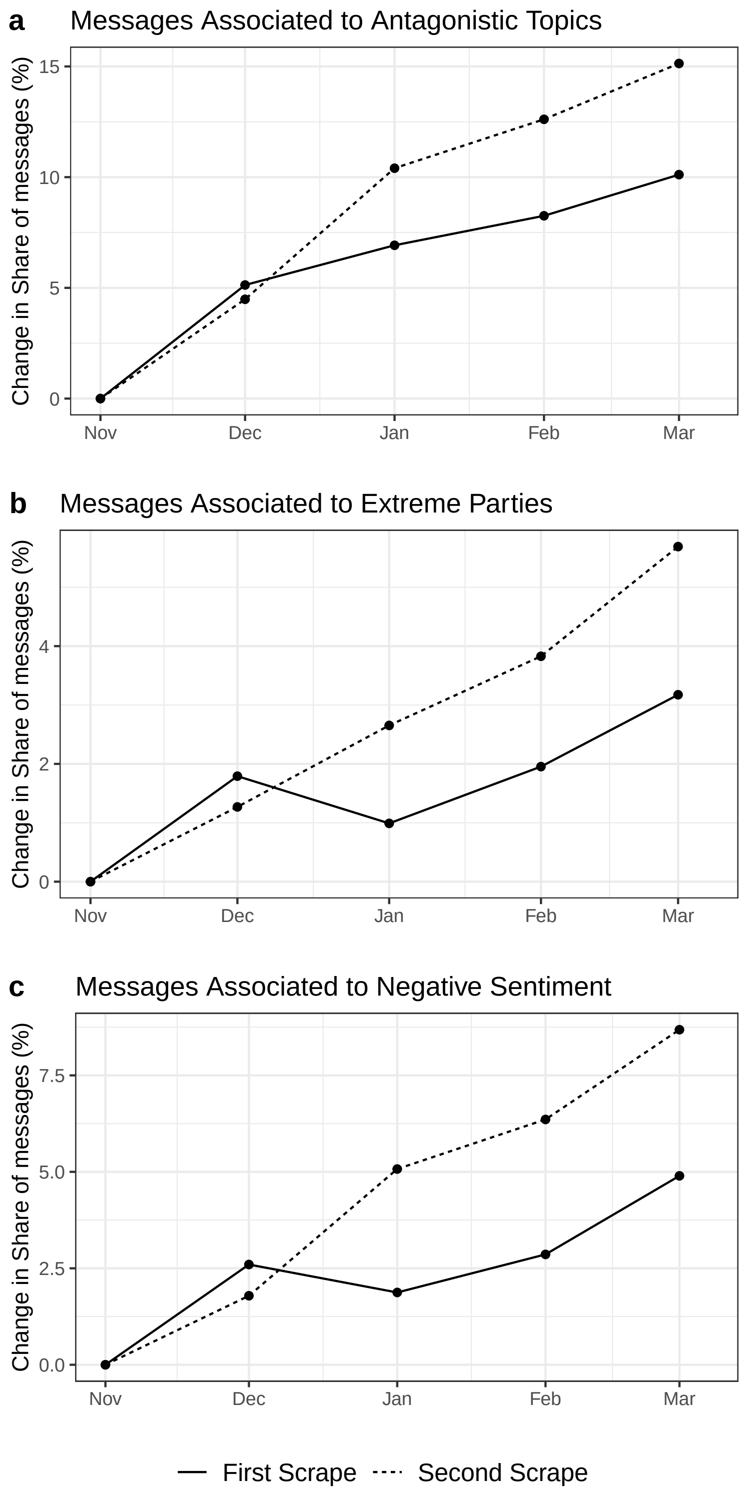
Notes: This figure compares the predicted political leaning of sentences for the first (in light blue) and second (in dark blue) data collection. We assign a political leaning to each sentence in our corpus based on the probability of it being pronounced by a given party according to our supervised learning model.

Figure A.3: Topic Shares for Each Data Collection



Notes: This figure compares the share of messages assigned to each topic for our first (in light blue) and second (in dark blue) data collection on Facebook pages.

Figure A.4: Evolution of Outcomes for Each Data Collection



Notes: This figure compares observed trends in radical attitudes for our first (solid line) and second (dashed line) data collection on Facebook pages. Panel A presents changes in the share of sentences associated with an antagonistic topic. Panel B presents changes in the share of sentences associated with a politically extreme party (i.e., on the far left or the far right). Panel C presents changes in the share of sentences associated with negative sentiment.

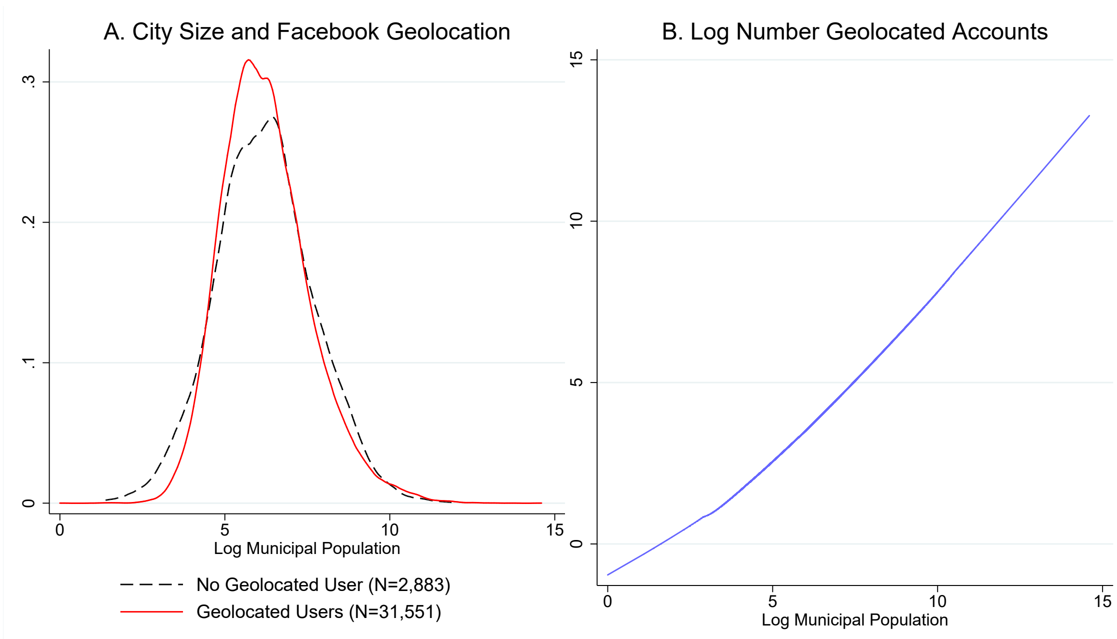
A.3.3 Facebook Penetration Rates

To measure Facebook penetration within France, we leverage one of the largest data leaks in the history of Facebook. In 2019, a massive dataset of Facebook users was made publicly available.[17] For half a billion Facebook accounts (including 30 million French accounts), the dataset contains its creation date, the name of the user, marital status, self-declared location, and phone number. The hackers compiled a close-to-comprehensive list of Facebook public profiles by searching through Facebook ID numbers in ascending order. Survey data from 2018 shows that about 60% of all French adults have a Facebook account, which is consistent with the number of French accounts leaked as the adult population of the country is slightly above 50 million (see here for details). It is, to this date, the largest publicly available dataset of Facebook users with location information.

We combine string pattern matching techniques with human supervision to link self-declared locations of users to French administrative data. By doing so, we match 10.8 million accounts to a municipality. Although it is not easy to verify the representativeness of this geolocated subset, we geolocate a positive number of users in over 90% of the municipalities. As shown in Panel A of Figure A.5, municipalities with no geolocated account are fairly similar in size to the other municipalities. As shown in Panel B of Figure A.5, the log-log relationship between the total population and the number of geolocated users is almost linear, with a slope of 1. Finally, we define Facebook penetration as the number of accounts per inhabitant. Panel B in Figure A.8 shows the spatial distribution of this variable, which displays much more heterogeneity within regions than between regions.

[17] *Source:* https://www.bbc.com/news/technology-56772772. Note that leaked data has been recently used in research articles on tax evasion (Alstadsæter, Johannesen and Zucman 2019). We only use this data to construct municipality-level statistics with sufficient statistical anonymity.

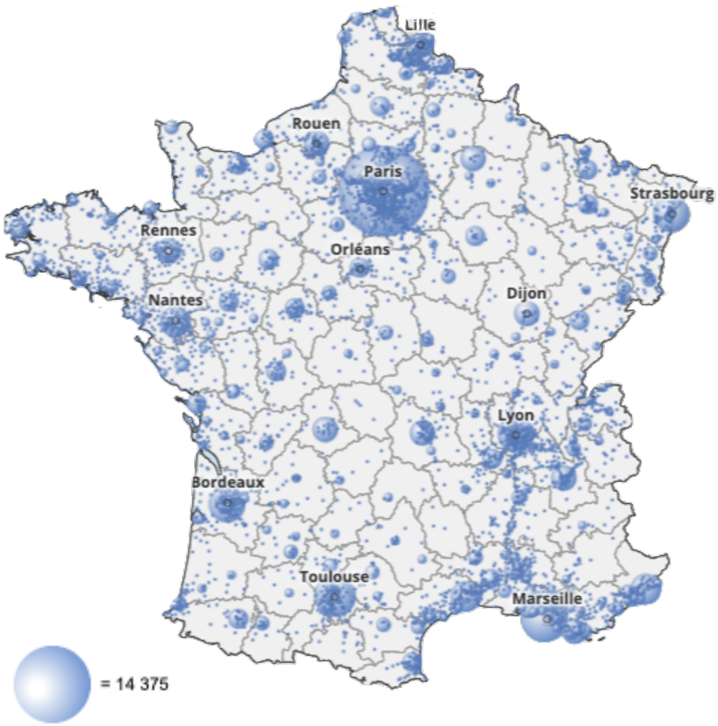
Figure A.5: Geolocation of Facebook Accounts



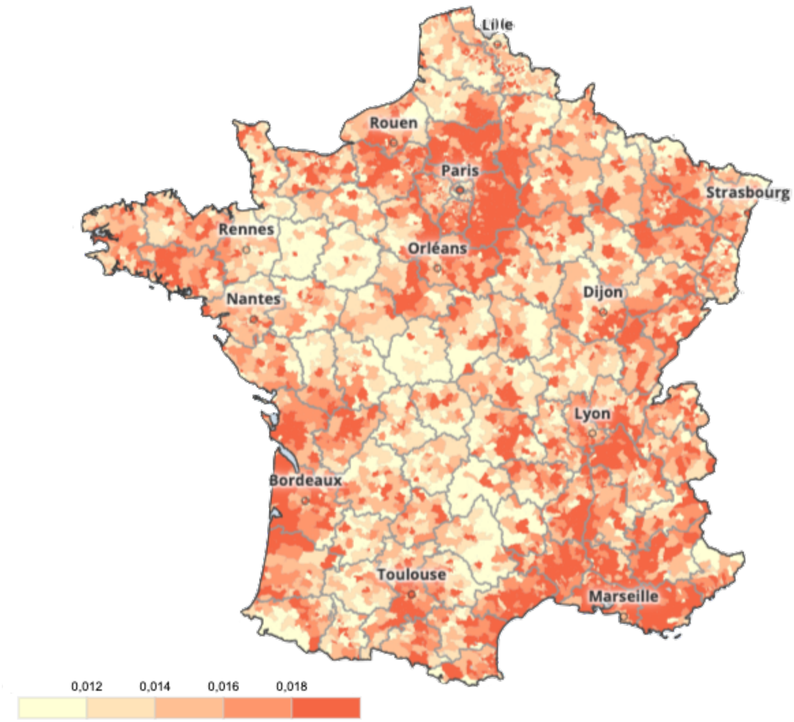
Notes: Panel A shows the distribution of municipal population for municipalities with at least one geolocated account and municipalities without. Panel B shows a Lowess regression of the (log) number of geolocated municipal accounts and the (log) population.

Figure A.6: Signature Rate of the Change.org Petition by Municipality

A. Absolute value



B. Per inhabitant

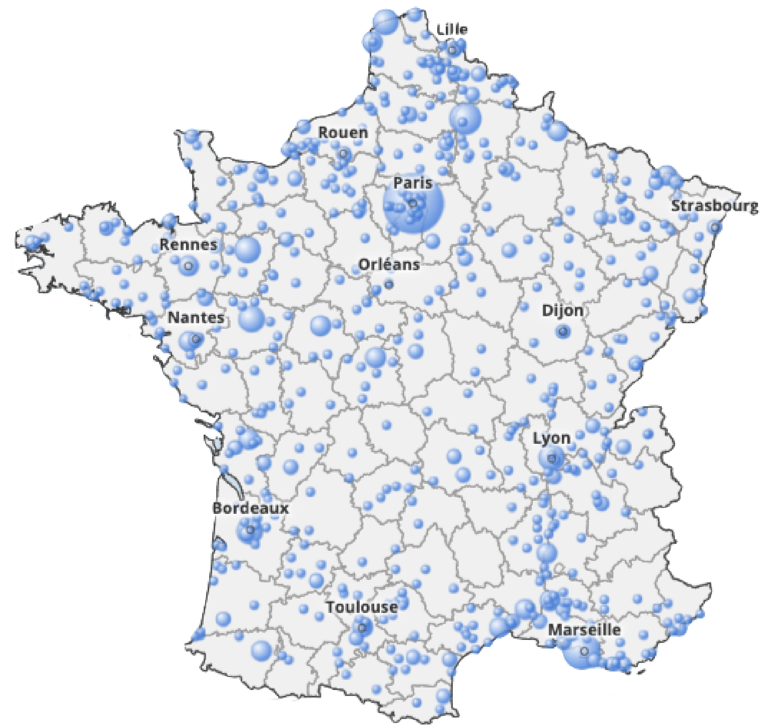


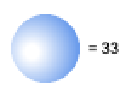
Notes: Figure A displays the number of signature per municipality. Figure B displays the signature rate (signature per inhabitant) by municipality.

Figure A.7: Number of Local Groups per Municipality.

A. Before 17/11

B. After 17/11



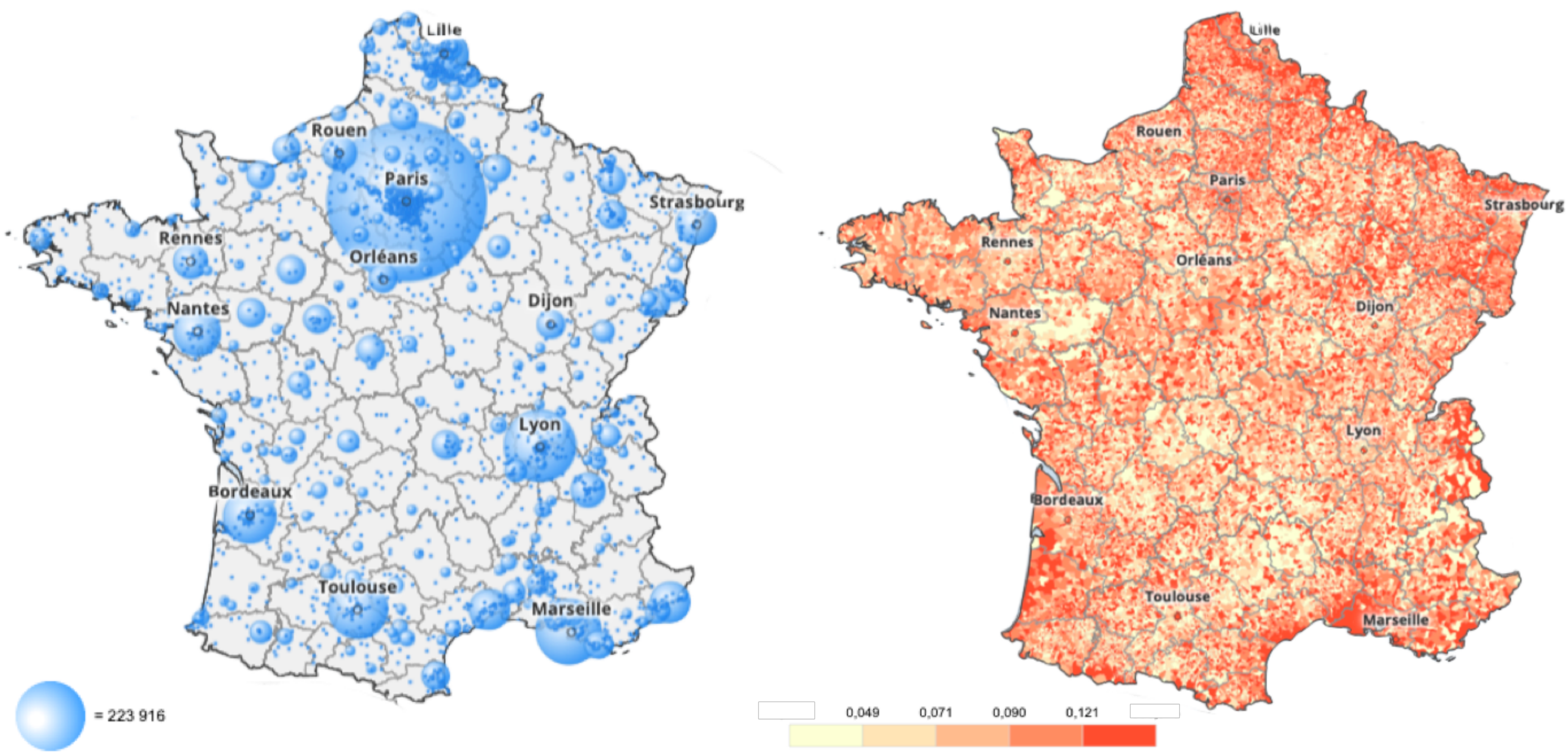


Notes: The two figures display the number of Yellow Vests local groups per municipality. Figure A corresponds to group creation before the First Act of blockades, while Figure B corresponds to group creation after the First Act.

Figure A.8: Spatial Distribution of Facebook Users

A. Absolute value

B. Per inhabitants



Notes: These two maps display the spatial distribution of Facebook users (who declare either to live in or to come from the location) in absolute values at the municipality level. The map on the left-hand side displays the values in proportion of circle radius, while the map on the right-hand side displays the value in terms of color intensity.

A.4 Tweets of Politicians

We build a dataset of tweets by politicians who belonged to the lower chamber of the French Parliament (the *Assemblée Nationale*) between 2017 and 2022. We consider the five largest French political parties: Rassemblement National (RN), Les Républicains (LR), La République en Marche (LREM), le Parti Socialiste (PS) and La France Insoumise (LFI). Politicians use Twitter to speak to their constituents directly. Thus, tweets are closer to daily social media messages than parliamentary speeches. They provide a natural, labeled dataset to train a machine learning classifier of party affiliation based on written text. We then use our classifier to infer online protesters' political partisanship based on their Facebook messages. The complete list of politicians at the Assemblée Nationale is available here. The dataset of French politicians on Twitter is available here. We retrieve the 3200 last tweets of each politician via the Twitter API.

A.5 Administrative Data

We construct a wide set of local controls. The set of municipal controls included in our regressions may be grouped as follows:

Geography includes the population of the municipality, its density, the distance to the closest city with over 20,000 inhabitants and 100,000 inhabitants, whether the municipality was classified as urban in 2015, and whether it switched from rural to urban between 1999 and 2015.

Source: Census (RP, complementary exploitation), 2016, INSEE.

Transport includes the shares of the employed population commuting by car and public transportation, as well as the median commuting distance.

Source: Census 2016, INSEE. Déclarations Annuelles de Données Sociales (DADS), 2015, INSEE.

Economy includes the local unemployment rate, the fraction of employees with a non-permanent contract, mean income, and population immigrant share.

Source: Census 2016, INSEE. DADS, 2015, INSEE.

Occupation includes the share of the different *catégories socio-professionelles* defined by INSEE: executive, independent, middle-management, employee, manual worker and agriculture.

Source: Census 2016, INSEE.

Age includes the shares of the population in the following groups: 18-24 y.o.; 25-39 y.o.; 40-64 y.o.; over 65 y.o.

Source: Census 2016, INSEE.

Education includes the shares of the population without a high-school diploma, and with a university degree.

Source: Census 2016, INSEE.

Vote includes the vote share for the five major candidates in the 2017 presidential election (Macron, Le Pen, Fillon, Mélenchon, Hamon), as well as the share of abstention.

Source: Ministry of the Interior.

Signature is the local signature rate of the Change.org petition before 11/17.

Source: Change.org.

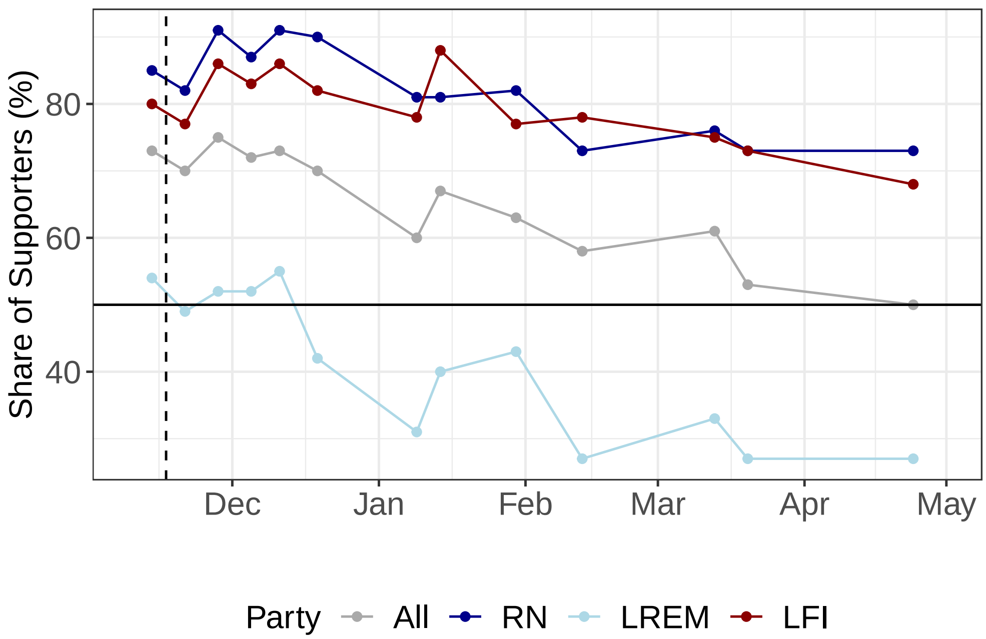
LZ is a set of 1,606 dummies for Living Zones.

Source: INSEE.

A.6 Polls

The polling institute ELABE conducted several surveys between November 2018 and April 2019 for the news Channel BFM TV. Figure A.9 reports their results on the evolution of public support for the Yellow Vests movement.

Figure A.9: Evolution of the Support for the Yellow Vests



Notes: This figure plots the share of the population who declared they were supportive or sympathetic to the Yellow Vests movement over time. The vertical dashed line corresponds to 11/17. ELABE, the survey institute from which we collected data, conducted polls on 11/14/2018, 11/21/2018, 11/28/2018, 12/5/2018, 12/11/2018, 12/19/2018, 1/9/2019, 1/14/2019, 2/13/2019, 3/13/2019, 3/20/2019, and 4/24/2019.

B Supplement for "The Online-Offline Feedback Loop"

B.1 Control variables and early mobilization

Table B.1: Variance Decomposition: Yellow Vests Movement (pre-17/11)

	Signatures		Nb. Groups		Blockade	
	(1)	(2)	(3)	(4)	(5)	(6)
Explained variance $\hat{y}$	**8.43**	**37.05**	**34.61**	**38.17**	**17.94**	**23.48**
Fixed effects (LZ)		**22.55**		**1.41**		**1.00**
Density		0.10	0.82	1.35	0.36	0.27
Population	2.90	5.84	21.53	20.75	5.51	8.01
Population squared	0.00	0.00	3.94	7.18	0.63	0.57
Pop. spline 50/75th percentile	0.12	0.26	0.01	0.00	0.00	0.01
Pop. spline 75/100th percentile	2.92	5.90	7.28	5.27	9.18	11.34
Population measures	**6.01**	**12.10**	**33.59**	**34.55**	**15.68**	**20.19**
Dist. to closest mid size city	0.05	0.55	0.02	0.04	0.01	0.04
Dist. to closest large city	0.08	0.17	0.01	0.06	0.01	0.01
Urban municipality	0.10	0.29	0.24	0.53	0.87	0.88
Urbanized since 1999	0.00	0.03	0.01	0.02	0.02	0.02
Geography	**0.23**	**1.04**	**0.28**	**0.65**	**0.91**	**0.94**
Share commuting by car	0.02	0.06	0.04	0.08	0.06	0.07
Share commuting by public transp	0.12	0.11	0.19	0.27	0.14	0.10
Median commuting distance	0.02	0.00	0.00	0.00	0.02	0.02
Commuting	**0.17**	**0.17**	**0.23**	**0.35**	**0.22**	**0.19**
Average wage income	0.24	0.02	0.00	0.00	0.01	0.00
Share in CDI	0.02	0.04	0.01	0.02	0.07	0.07
Unemployment rate	0.09	0.05	0.11	0.22	0.23	0.20
Share retail workers	0.00	0.00	0.00	0.00	0.00	0.01
Share executives	0.04	0.04	0.00	0.01	0.02	0.02
Share intermediate workers	0.10	0.12	0.01	0.03	0.02	0.01
Share clerical workers	0.06	0.07	0.01	0.01	0.02	0.01
Share blue collar	0.03	0.03	0.00	0.00	0.00	0.00
Labor market	**0.57**	**0.36**	**0.14**	**0.29**	**0.37**	**0.33**
Share 18 to 24 y.o.	0.01	0.01	0.13	0.27	0.20	0.23
Share 25 to 39 y.o.	0.04	0.02	0.01	0.02	0.03	0.03
Share 40 to 64 y.o.	0.01	0.01	0.03	0.03	0.04	0.03
Share over 65 y.o.	0.37	0.30	0.00	0.00	0.00	0.01
Age groups	**0.43**	**0.33**	**0.16**	**0.33**	**0.27**	**0.29**
Share with HS degree	0.03	0.00	0.00	0.00	0.01	0.01
Share with college degree	0.10	0.02	0.01	0.02	0.01	0.00
Education	**0.13**	**0.03**	**0.01**	**0.02**	**0.02**	**0.01**
Fillon vote	0.14	0.19	0.00	0.00	0.01	0.00
Hamon vote	0.00	0.00	0.01	0.01	0.01	0.01
Le Pen vote	0.55	0.13	0.00	0.03	0.02	0.02
Macron vote	0.02	0.01	0.00	0.00	0.03	0.03
Far left vote	0.04	0.01	0.01	0.01	0.01	0.01
Abstention	0.00	0.00	0.03	0.12	0.26	0.27
2017 election votes	**0.75**	**0.34**	**0.05**	**0.17**	**0.33**	**0.34**
Share roads with reduced speed	0.01	0.00	0.02	0.05	0.06	0.06
Share of diesel vehicles	0.12	0.09	0.04	0.19	0.05	0.10
Motorists	**0.13**	**0.09**	**0.06**	**0.25**	**0.11**	**0.16**
Facebook penetration	**0.01**	**0.03**	**0.09**	**0.16**	**0.04**	**0.03**

Notes: Following Shorrocks (1982), this table displays the factor contribution of each explanatory variable on the signature rate per inhabitant before 11/17 (Columns (1) and (2)), the number of Facebook groups before 11/17 (Columns (3) and (4)), and a dummy variable for the existence of a blockade on 11/17 (Columns (5) and (6)).

B.2 Details on the roundabout instrument

Table B.2: Variance Decomposition: Roundabouts

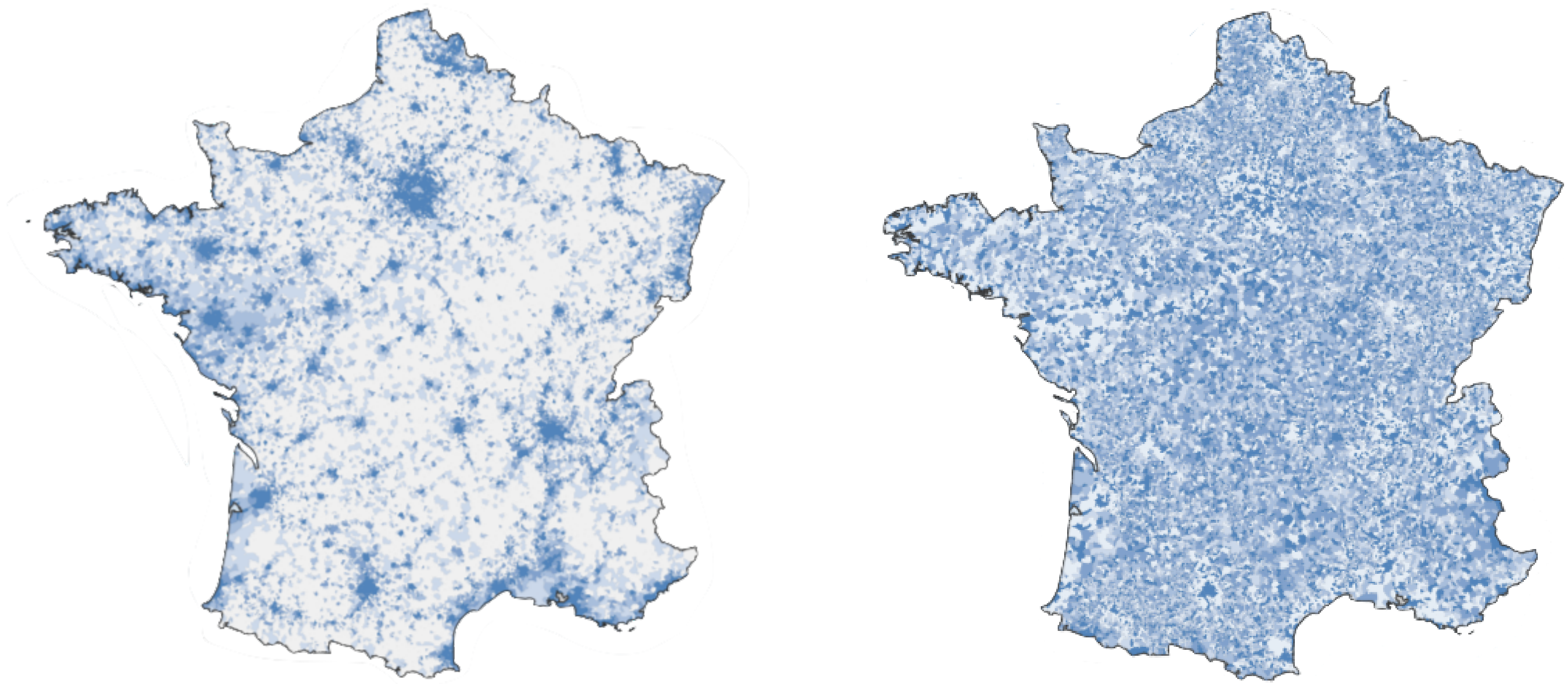
	Roundabout density	
	(1)	(2)
Explained variance $\hat{y}$	**45.33**	**51.75**
Fixed effects (LZ)		**3.18**
Density	1.69	2.30
Population	18.81	18.41
Population squared	0.04	0.06
Pop. spline 50/75th percentile	0.06	0.04
Pop. spline 75/100th percentile	19.85	19.80
Population measures	**40.44**	**40.61**
Dist. to closest mid size city	0.32	0.87
Dist. to closest large city	0.11	0.12
Urban municipality	2.09	3.35
Urbanized since 1999	0.03	0.04
Geography	**2.55**	**4.38**
Share commuting by car	0.02	0.06
Share commuting by public transp	0.33	0.54
Median commuting distance	0.04	0.05
Commuting	**0.39**	**0.65**
Average wage income	0.04	0.14
Share in CDI	0.25	0.38
Unemployment rate	0.07	0.12
Share retail workers	0.03	0.04
Share executives	0.03	0.04
Share intermediate workers	0.02	0.03
Share clerical workers	0.04	0.05
Share blue collar	0.01	0.01
Labor market	**0.50**	**0.81**
Share 18 to 24 y.o.	0.32	0.49
Share 25 to 39 y.o.	0.04	0.08
Share 40 to 64 y.o.	0.09	0.10
Share over 65 y.o.	0.01	0.04
Age groups	**0.46**	**0.72**
Share with HS degree	0.00	0.00
Share with college degree	0.05	0.03
Education	**0.05**	**0.03**
Fillon vote	0.00	0.00
Hamon vote	0.03	0.03
Le Pen vote	0.07	0.10
Macron vote	0.21	0.18
Far left vote	0.02	0.02
Abstention	0.10	0.22
Motorists	**0.11**	**0.24**
Share roads with reduced speed	0.04	0.10
Share of diesel vehicles	0.46	0.72
Motorists	**0.50**	**0.82**
Facebook penetration	**0.00**	**0.00**

Notes: Following Shorrocks (1982), this table displays the factor contribution of each explanatory variable on the the density of roundabouts in the municipality (Columns (1) and (2)), and on the density of roundabouts in the other municipalities of the LZ (Columns (3) and (4)).

Figure B.1: Roundabout density

A. Roundabouts by squared kilometer

B. Residuals



Notes: Panel A shows the density of roundabouts in mainland France, with darker colors corresponding to higher density. Panel B shows the residual density of roundabouts after controlling for the set of controls described in Section 3. Color intensity corresponds to quantile thresholds.

B.3 Robustness of the 2SLS results

Table B.3: Impact of Blockades on Post-17/11 Online Mobilization: Alternative Specifications

	(1) Signatures	(2) Groups	(3) Members	(4) Posts
Panel A: Without controls				
Blockade	2.740***	5.874***	0.155***	0.104***
	(0.283)	(0.490)	(0.0513)	(0.0385)
Kleibergen-Paap F-stat	30.6	30.6	30.6	30.6
p-value Hansen	0.001	0.022	0.036	0.029
Panel B: Only municipal instrument				
Blockade	1.261***	3.289***	0.287*	0.198*
	(0.345)	(0.868)	(0.151)	(0.110)
Kleibergen-Paap F-stat	14.7	14.7	14.7	14.7
Panel C: Only LZ instrument				
Blockade	1.086***	2.673***	0.163**	0.0971*
	(0.304)	(0.842)	(0.0740)	(0.0566)
Kleibergen-Paap F-stat	37.0	37.0	37.0	37.0
Panel D: Commuting zone instead of LZ				
Blockade	0.632**	3.430***	0.289***	0.186**
	(0.255)	(0.896)	(0.0967)	(0.0745)
Kleibergen-Paap F-stat	12.1	12.1	12.1	12.1
p-value Hansen	0.099	0.868	0.131	0.140
Panel E: Excluding Paris region				
Blockade	0.840***	2.992***	0.395***	0.287**
	(0.309)	(0.901)	(0.144)	(0.121)
Kleibergen-Paap F-stat	18.9	18.9	18.9	18.9
p-value Hansen	0.656	0.635	0.067	0.065

Notes: This table shows estimates corresponding to variations of the regressions of Table 2. Panel A shows results for the 2SLS estimation of Table 2 when we do not include any municipal control nor LZ fixed effects. In Panel B (resp. C), we show 2SLS results using the roundabout density of the municipality as an instrument (resp., the density of roundabouts in other municipalities of the LZ) only. In Panel D, we control for commuting-zone fixed effects instead of LZ fixed effects and instead of considering the density of roundabouts in other municipalities of the LZ as a second instrument, we use density of roundabouts in other municipalities of the commuting zone. In all panels but Panel E, the number of observations is 34,434. In all regressions, we cluster standard errors at the LZ level (except in Panel F, where we cluster standard errors at the commuting zone level). *: p< 0.01, **: p<0.05, ***: p<0.1.

C Supplement for "The Rise of Online Radicalism"

C.1 Text Pre-processing

We process all text corpora in the same way. We remove emojis, links, accents, punctuation, social media notifications (e.g., "Yellow Vests changed their profile picture"), and stopwords from the corpus. We also lowercase the text and lemmatize words. We keep hashtags and user mentions but drop all tokens which occur less than ten times in the Facebook corpus.[18] This leaves us with approximately 40,000 unique tokens in the corpus. Most documents in our corpora are short text snippets (e.g., a phrase or a sentence). Some are longer and span over multiple sentences (e.g., Facebook posts). To keep all documents comparable, we work with unigrams at the sentence level.

C.2 Topic Model

The standard approach for topic modeling in the text as data literature is to rely on Latent Dirichlet Allocation (LDA) models. LDA models documents as a distribution over multiple topics. Though this is often a reasonable assumption, it is implausible in the case of short text snippets (such as sentences) which often refer to only one topic (Yan, Guo, Lan and Cheng 2013). For this reason, standard topic models are known to perform poorly on such short texts. As an alternative, we build a custom topic model in the spirit of Demszky et al. (2019). First, we produce word embeddings for the corpus and represent each sentence as a vector in the embedding space. We train a Word2Vec model using Gensim's implementation, with moving windows of eight tokens and ten iterations of training. We build sentence embeddings as the weighted average of the constituent word vectors, where the weights are smoothed inverse term frequencies (to assign higher weights to rare/distinctive words) (Arora, Liang and Ma 2017). The resulting embedding space allows for a low-dimensional representation

[18]The frequency threshold does not influence results, but allows us to remove many uncommon spelling mistakes and other idiosyncrasies related to social media data.

of text, in which phrases which appear in similar contexts are located close to one another. Second, we group sentence vectors together into a small set of clusters. The goal is to have different clusters for different topics in the text. We rely on the K-Means algorithm. We train the algorithm on 100,000 randomly drawn sentences and predict clusters for the rest of the corpus. We use the ten closest words to the cluster centroids to manually label topics.[19] We choose to work with 15 topics for our main results. However, since the number of topics is a hyperparameter in our topic model, we also present resulting topics when specifying 5, 10, and 20 clusters (see Table C.1). To further inspect the topic model, we present the closest phrase to the centroid of each topic below. These phrases may be understood as the most representative text snippet for each topic. We present the pre-processed (as opposed to raw) phrases.

Critiques *visiblement representer peuple francais devenir lamentable attitude mepris*

Insults *sale batard hont francais macron bouffon macron batard degage fumier*

Diffusion *vouloir publier information verifier site diffuser savoir etre derriere info*

Towns-Hours *samedi 5 janvier rdv 10h place verdun marche rdv 18h zenith pau partir convoi tarbes depart 18h30 max 19h co voiturage voir place*

Conspiracy *souverainiste racisme fascisme etre frontal pensee correct tourner nation occidental homme blanc judeo chretien etre utilise arme psychologique mediatique tres puissant hegemonie moral ideologique pouvoir perdurer peuple europeen culpabiliser gauche systematiquement instrumentaliser ad horreur second guerre mondial discrediter national lui meme homme blanc nom jamais devoye*

[19] We also considered alternative labeling options, such as term frequency-inverse cluster frequency, which yield similar results.

Concerns *2000 euro concerne restaurer service public disparu poste hopital maternite ecole instauration revenu minimum lieu aide diffus demander complexe limitation salaire 10 smic augmentation salaire meme proportion gros salaire reprise dette banque france banque prive limitation montant demander maison retraite ecole vraiment gratuite fourniture activite livre gratuit lieu donner aide servir chose detail complet utilisation impot blocage tipp salaire elu 4 smic fin privilege egalite transparence fonds*

Actions : *malheureusement laisse choix vouloir change aller falloir arreter pacifiste attendre roi rigoler voir faire defoncer tomber nuit*

Foreign Languages *marie jo laziah*

Names *rajoute prenom chaine rose annick patricia nelly angel sophia mary didier gabrielle maya pierre fanny magali ludivine isabelle nicole nathan marie patricia jeannine serge josiane eric marie fleur rose laly severine emilie delphine nanou ophelie yohann laurer nanou aya magdalena aurelie angele chantal fanny carine brigitte yael sylvie virginie dominique rachel frederic audrey benjamin marie jeanne phil laurence rachel jeremy annie patricia agnes nini*

Violence *france ordre pouvoir continuer agresser impunite civil etre legitime defense cas attaque voir rue tv journaliste faire photo etre blesser flashball coup venir porter plainte ordre justement*

Other *oui faire accord jean michel*

Politics *faire site internet permettre inscrire revendication monde pouvoir proposer soutenir d lier etre veritable logique fin possibilite revendiquer systeme constitution battre revolte revolutionnaire systeme place deja logique pre institution etre legitimer adhesion populaire*

Support *bonjour lilly cur courage etre fille formidable faire gros bisou*

Places *79 44 85 16 13 80 06 01 53 36 69 bcp 17*

Food-Objects *jamais faire greve vie etre fan kro merguez pis odeur pouilleux sentir pisse odeur pneu cramer*

C.3 Sentiment Analysis

To measure emotional content in Facebook messages, we use a dictionary-based approach that assigns to a sentence a sentiment score ranging from -1 (very negative) to 1 (very positive). For each sentence, the sentiment score is obtained as the average of the sentiment scores of its constituent words. We rely on the VADER (Valence Aware Dictionary for Sentiment Reasoning) library for our main results. We present five of the most negative and five of the most positive sentences according to the VADER sentiment analyzer.[20]

Examples of the most positive sentences:

honneur gilet jaune

mdr

bravo

mercii jeune meilleur facon aider progres meilleur monde

bravo gabin media honnete souhaite reussite merite equipe bravo gj

Examples of the most negative sentences:

macron demission

macron cabanon castananer enfer

florence menteur

bande pourriture batard

castaner assassin degage voleur menteur

Our measure of sentiment could vary depending on the dictionary used. As a robustness check, we rely on French TextBlob as an alternative dictionary for word sentiment. We find that the VADER dictionary's density has larger tails as it tends to classify more sentences to the extremes

[20]Sentences can be long and with many repetitions. For readability, we remove sequences of repeated tokens.

Table C.1: Results of the Topic Model for Alternative Numbers of Clusters

Panel A: Results of the Topic Model for 5 clusters

Associated words
04, nimes, arras, nime, 77, narbonne, albi, chambery, 47, orleans
pouvoir, etre, consequent, favoriser, necessaire, n, global, politique, specifique, constitue
merde, connard, salopard, pourriture, encule, putain, hont, honte, batard, ordure
gabin, live, sympa, app, brancher, stp, ramous, cool, stabilisateur, coupure
laziah, misfortune, #noussommesgiletsjaune, dellacherie, exhort, substituons, sansone, pajalo, victory, naeim

Panel B: Results of the Topic Model for 10 clusters

Associated words
etre, n, peuple, meme, politique, faiblesse, nefaste, veritable, gouvernement, destructeur
annuel, beneficiaire, compenser, bonus, salaire, taxation, production, exoneration, delocalisation, embauche
cr, flic, flics, policier, gazer, projectile, charger, manifestant, matraque, gendarme
zappe, zapper, tpmp, humoriste, fakenew, interviewe, conversation, cnew, interviewer, bfmtv
orlane, magdalena, grilo, correia, gourdon, leal, caudrelier, malaury, macedo, khaye
connard, merde, encule, bouffon, conard, pd, salope, enculer, fdp, batard
adhesion, charte, valider, definir, modalite, eventuel, prealable, specifique, necessaire, proposer
04, nimes, arras, albi, nime, royan, 77, narbonne, chambery, 47
courage, courag, bravo, felicitation, formidable, bisou, bisous, genial, soutien, continuation
sansone, dutie, facilitate, soldiers, auv, weier, unterstutzen, #jiletsjaune, ausbeutung, seem

Panel C: Results of the Topic Model for 20 clusters

Associated words
beneficiaire, compenser, salaire, bonus, annuel, exoneration, plafonner, taxation, embauche, reduction
omo, #noussommesgiletsjaune, laziah, houpette, nooooon, jeoffrey, chab, limitatif, exhort, cageot
aller, faire, voir, la, etre, oui, vraiment, merde, savoir, meme
englos, royan, sisteron, pontivy, arras, seclin, hendaye, douai, roanne, albi
twitter, diffuse, info, publier, fb, diffuser, relater, page, interview, information
adhesion, structuration, proposer, proposition, definir, charte, structurer, concertation, revendication, necessaire
maud, johanna, gomes, anai, melanie, gregory, rudy, armand, melissa, mathias
bisous, courage, felicitation, courag, bisou, bravo, formidable, soutien, genial, coucou
asservissement, domination, peuple, deposseder, destructeur, gouvernance, oppression, politique, veritable, appauvrissement
recours, illegal, sanction, infraction, poursuite, condamnation, delit, penal, abusif, commettre
41, 52, 58, 47, 38, 61, 69, 37, 46, 82
canette, chaussette, bouteille, cendrier, plastique, peintur, toilette, saucisson, scotch, brosse
cr, flic, flics, frapper, tabasser, matraquer, policier, gazer, matraque, tabasse
mafieux, imposteur, larbin, escroc, acolyte, magouilleur, maffieux, corrompu, dictateur, sbire
kassav, akiyo, diritti, sempr, dittaturer, etait, popolo, quando, anch, infami
stupide, pathetique, affliger, pitoyable, malsain, stupidite, abject, irrespectueux, insultant, grossier
15h, 17h30, 16h30, 10h, 14h00, 11h, gare, 8h30, 18h, 18h30
laziah, #noussommesgiletsjaune, gourdon, misfortune, orlane, grilo, victory, duquesnoy, dellacherie, macedo
#jiletsjaune, created, soldiers, #assembleenationale, #coletesamarelo, #parisprotest, dutie, unterstutzen, #france3, sansone
connard, encule, batard, salope, fdp, merde, conard, enculer, pd, salopard

Notes: This table shows the clusters defined by our the topic model when requesting alternative numbers of topics (5, 10, and 20). For each topic, we report the closest words to the cluster centroid (measured by cosine similarity).

of the sentiment spectrum. Nonetheless, both measures suggest an increase in average negative sentiment between November 2018 and March 2019. Figure C.1 decomposes the increase in average negative sentiment (as measured by TextBlob) using the method outlined in Section 4.3. Results are qualitatively similar to the main text results.

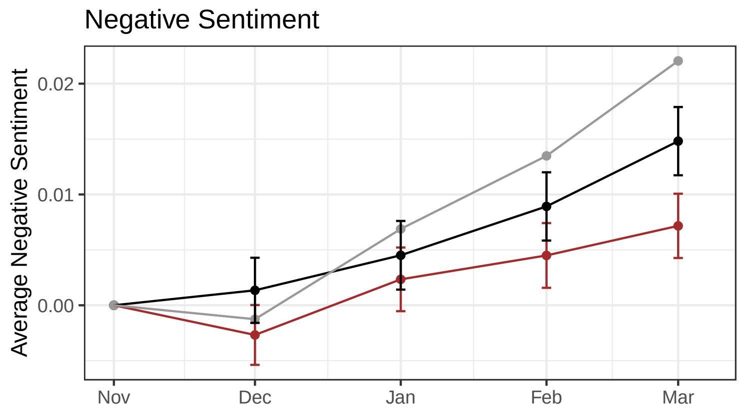


Figure C.1: Margins for Negative Sentiment Using TextBlob

Notes: This figure decomposes the increase in average negative sentiment using the method outlined in Section 4.3. We compute sentiment scores based on the TextBlob dictionary. Results are qualitatively similar to the main text results.

C.4 Political Partisanship Model

Our principal classification method is multinomial logistic regression. Given the large size of the vocabulary, we further penalize the regression with the L1-norm (Lasso) to force some coefficients to zero (Friedman, Hastie, Tibshirani et al. 2001). We consider the five largest French political parties: le Rassemblement National (RN), les Républicains (LR), la République en Marche (LREM), le Parti Socialiste (PS) and la France Insoumise (LFI). We parametrize the probability that a text snippet $\mathbf{x}$ is from party k as:

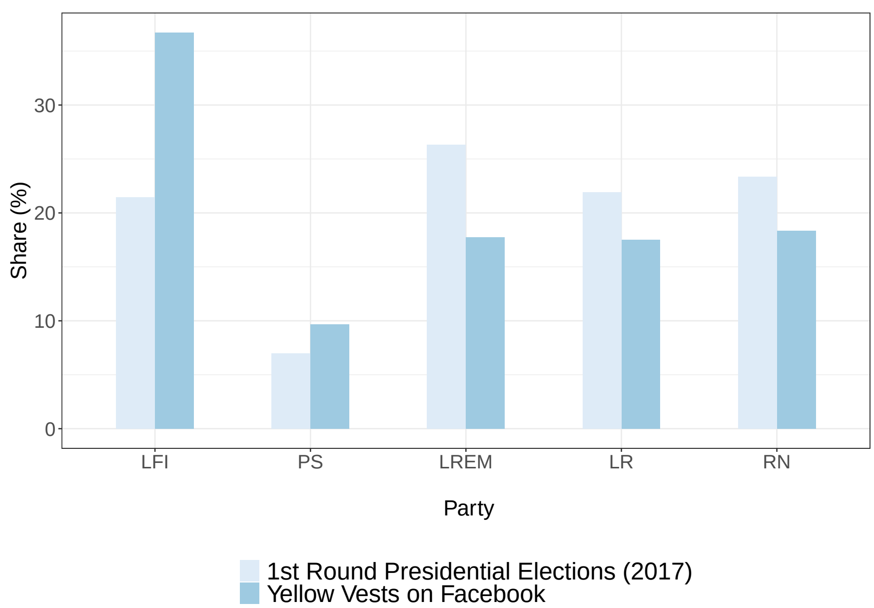
$$\mathrm{P}\big(\text{party} = k \,|\, \mathbf{x}\big) = \frac{\exp(\mathbf{w_k} \cdot \mathbf{x} + b_k)}{\sum_j \exp(\mathbf{w_j} \cdot \mathbf{x} + b_j)}$$

in which $\mathbf{w_k}$ are specific coefficients to be estimated for party k. Given the large size of the vocabulary, we further penalize the multinomial logistic regression with the L1-norm (Lasso) to force some coefficients to zero. As some unigrams are not informative of political partisanship, the penalization mitigates over-fitting of the training set by shrinking coefficients.

To validate the model, we shuffle the corpus and split it into 80% training data and 20% test data. We build the classifier in the training set and evaluate its performance in the test set. The model has an accuracy score of 55.5%. A random guess would correctly infer the author's party 20% of the time. Our model thus assigns the correct party to a text snippet between two and three times more often than a guess at random would. For comparison, Peterson and Spirling (2018) predict party affiliation with an accuracy between 60 and 80% for two parties. In this case, a guess at random would get the label right 50% of the time.

Results are presented in Figure C.2.

Figure C.2: Predicted Political Leaning of the Yellow Vests



Notes: This figure compares the predicted political leaning of the active Yellow Vests users on Facebook (in dark blue) to the scores of each party at the first round of the presidential elections (in light blue). Vote shares at the elections are modified so as to sum up to a hundred (there were other smaller parties that we exclude from the analysis). We assign a political leaning to each Facebook user in our corpus based on the average probability of her sentences being pronounced by a given party according to our classifier.

References

Alstadsæter, Annette, Niels Johannesen, and Gabriel Zucman, "Tax Evasion and Inequality," *American Economic Review*, June 2019, *109* (6), 2073–2103.

Arora, Sanjeev, Yingyu Liang, and Tengyu Ma, "A simple but tough-to-beat baseline for sentence embeddings," *Conference paper at ICLR 2017*, 2017.

Demszky, Dorottya, Nikhil Garg, Rob Voigt, James Zou, Matthew Gentzkow, Jesse Shapiro, and Dan Jurafsky, "Analyzing Polarization in Social Media: Method and Application to Tweets on 21 Mass Shootings," in "Proceedings of NAACL-HLT" 2019, pp. 2970–3005.

Friedman, Jerome, Trevor Hastie, Robert Tibshirani et al., *The Elements of Statistical Learning*, Vol. 1, Springer series in statistics New York, 2001.

Peterson, Andrew and Arthur Spirling, "Classification Accuracy as a Substantive Quantity of Interest: Measuring Polarization in Westminster Systems," *Political Analysis*, 2018, *26* (1), 120–128.

Rieder, Bernhard, "Studying Facebook via Data Extraction: The Netvizz Application," in "Proceedings of the 5th annual ACM web science conference" ACM 2013, pp. 346–355.

Shorrocks, Anthony F, "Inequality Decomposition by Factor Components," *Econometrica*, 1982, pp. 193–211.

Yan, Xiaohui, Jiafeng Guo, Yanyan Lan, and Xueqi Cheng, "A Biterm Topic Model for Short Texts," in "Proceedings of the 22nd international conference on World Wide Web" 2013, pp. 1445–1456.