Diffusion of protest through social media: Evidence from Nahel protest in France

Annalí Casanueva-Artís * Pierre-Guillaume Méon† Paul Stephan ‡

June 2024

preliminary draft
–please do not circulate–

Abstract

We study how social media contribute to the geographic diffusion of riots. To do so, we study how the riots that started after 17 year old Nahel Merzouk was shot by the police in Nanterre in the Paris region spread over mainland France in the early summer of 2023. From 27 June to 4 July, violent protests involving several thousand participants took place in 553 municipalities, including municipalities that had never witnessed that type of event. We use the geolocation of posts and comments on Instagram related to the death of Nahel Merzouk to create a measure of exposure of a municipality to the riots of other municipalities not yet affected. Exploiting the panel dimension of social media posts and riots and using the networks of football players and hip-hop artists who posted about the Nahel Merzouk's death, as a source of exogenous variation in the exposure of a municipality to the riots, we find that higher exposure to riots on a day increased the probability of hosting a riot on the following day. We find that the new protests occurred rapidly after exposure. Thus, the higher effect of exposure through social media coincides with an unexpected escalation and geographical reach of the movement that travel effectively through imaged-based social media used by the younger population.

Keywords: protest, diffusion, social media
JEL classification: D74, D83, L82

*Ifo Institute, Big Data Junior Research Group & Center for Industrial Organization and New Technologies
†Université libre de Bruxelles (ULB), CEBRIG & Dulbéa
‡ETH Zurich

1 Introduction

As the emergence of social media coincided with the increase in the number of protests, it is tempting to relate the two phenomena (Cantoni et al., 2023). Social media have two key features that contribute to the appearance of protests: low barriers to entry and reliance on user-generated content (Zhuravskaya et al., 2020). These two features allow single individuals and grass-root organizations to skip the traditional filters of mainstream media and be able to "self-mediatize" (Castells, 2015). Social media allow participants to share information on their discontent and the behavior of the authorities. They also help potential participants solve the central coordination problem that they face when considering taking to the street (Enikolopov et al., 2020). Finally, social media can impede the action of repressive forces as they allow information to travel in real-time, reducing the temporal margin for authorities to react. Since marginalized groups may particularly benefit from this facilitation of communication, social media may act as a "liberation technology" (Manacorda & Tesei, 2020).

So far, the literature has focused on the effect of social media on protest against authoritarian regimes. For example, Acemoglu et al. (2018) observe that expressions of discontent on Twitter were predictive of the number of protesters on Tahrir Square in Cairo between 2011 and 2013. Manacorda & Tesei (2020) document that the spread of mobile phones and the internet in Africa facilitated the organization of mass protests in reaction to economic downturns. Likewise, Enikolopov et al. (2020) report that the penetration of the Russian social network VK contributed to the 2011 wave of protests against electoral fraud. Finally, Qin et al. (2021) show that strikes and protests in China spread through Twitter-like networks. Yet, the US 2020 George Floyd protests and the 2021 Capitol riots, the 2011 London riots in the UK, or the 2018 yellow vests movement in France are reminders that violent riots still exist in mature democracies.

In contrast to previous studies, we therefore study how social networks can contribute to spreading riots across space and over time in a mature democracy. The effect of social media on protest in mature democracies can *a priori* be very different from that in autocratic regimes. The authorities of a democratic country are likely more reluctant to censor or impede this type of communication. They moreover face constitutional checks and balances that constrain the restrictions that they may impose on freedom of expression. Consequently, the fact that information can, theoretically, travel freely may reduce the marginal gain of spreading information through social media instead of other, more traditional ways.

To investigate this issue, we take advantage of riots that took place in France in the summer of 2023 in response to police violence. These riots erupted in France between June 27 and July 4 following the death of 17-year-old Nahel Merzouk who had been shot by the police in Nanterre, in the outskirts of Paris. Starting in the outskirts of Paris, the riots spread across France after a couple of days to affect 553 municipalities, including small ones that had never experienced a riot before.

The 2023 French riots have several features that make them particularly instructive to study the interplay of riots and social media. The first is that they started with a well-identified event, the death of Nahel Merzouk. The trigger of the riots is therefore well-identified, the shooting of a teenager against a backdrop of perceived discrimination and marginalization (Oberti & Maela, 2023).

Secondly, although on the first night following the death of Nahel Merzouk the riots took place in Île-de-France, the region surrounding Paris, they started affecting more distant and smaller municipalities a couple of days later before spreading over mainland France. They therefore give the opportunity to study the role of social media in the geographic diffusion of riots.

The third instructive feature of the 2023 riots is that they occurred over a period of 9 days, from Tuesday 27 June to Wednesday 5 July. We can therefore track the entire wave of riots from its onset to its decay.

Finally, the role of social media was almost instantly highlighted by the press and researchers (Oberti & Maela, 2023; Clairouin et al., 2023), even triggering a public debate about the possibility of shutting down social media platforms during riots (Libération, 2023). Accordingly, the riots give an opportunity to study the role of social media over an entire sequence of events around what Cantoni et al. (2023) refer to as a sustained protest. They therefore allow to study how social media spread riots across space but over time. An important dimension of the role of social media in the reaction to the death of Nahel Merzouk is that visible personalities who are popular in the age group of rioters reacted on their social media. It was the case of several hip-hop artists and members of the French male national football team. Those reactions created an exogenous variation in the exposure to the death of Nahel Merzouk among the followers of those personalities. By collecting data on the location of their followers *before* the event, we obtain a predetermined network along which the news could spread *after* the event allowing us to establish a causal effect of social media.

Moreover, the French 2023 riots share several features with riots that took place in other countries, such as the riots that followed the death of George Floyd which sparked the Black Lives Matter movement in the US. Specifically, after Nahel Merzouk had been shot, the police officially claimed that an officer shot him in self-defense as he was being charged by Nahel Merzouk's car after the latter had refused to comply with a police control. Yet, a few hours after the official report of the police appeared in the media, two videos showing that Nahel Merzouk had been shot at point blank went viral on social media, sparking a week of unrest. Another common feature between the cases of Nahel Merzouk and George Floyd is that they both belonged to groups that suffer from prejudice, George Floyd being a black American and Nahel Merzouk a French-Algerian. Accordingly, their death prompted suspicion of racism and discrimination by the police. In addition, Nanterre, the municipality where Nahel Merzouk lived and was killed, hosts a large disadvantaged population, which eased the identification with Nahel Merzouk of other residents of "banlieues", the term used in the French context to refer to disadvantaged suburban municipalities. The French riots therefore allow to study the entire sequence of reactions to police violence against the member of a disadvantaged group .

With that end in view, we collected data on the date and location of the 553 riots that took place following the death of Nahel Merzouk. We also scraped over the same period the online activity on Instagram, which is the most used social media platform in France among the age group between 16 and 25 years (Diplomeo, 2023). This is important because this is the age group to which participants in the riots belonged (Bronner, 2023). Moreover, Instagram is image-based, as opposed to text-based social networks. It is therefore more likely to elicit emotional responses (Casas & Williams, 2019). Tracking reactions to pictures posted on Instagram is therefore a way to see how emotions contributed to the diffusion of riots. We define the exposure of a municipality to the riots through Instagram as the number of comments of riot-related posts in a municipality on a day. We first define it at the municipality level and estimate a panel model with municipality and day fixed effects where the dependent variable is a dummy set to one if the municipality experienced a riot on that day and the explanatory variable our measure of exposure.

We find that exposure to riots in other municipalities through Instagram on a given day is predictive of riots in a municipality in the following days and that the bulk of the effect materializes on the first day that follows exposure. We find no effect beyond the second day. The effect is robust and stable across specifications. We show that the connection between municipalities that we capture through celebrities and their network on Instagram is more relevant for protest diffusion than other measures of proximity such as the Facebook connectedness index or geographical proximity. However, by weighting online comments by geographic distance to the municipality where the commented picture originates, we show that the effect of online proximity is stronger if municipalities are also geographically close.

The contribution of the paper is therefore fourfold. Firstly, we document how information and communication technology contributes to spreading riots in a mature democracy whereas previous evidence pertained either to illiberal or authoritarian regimes (Manacorda & Tesei, 2020; Enikolopov et al., 2020; Qin et al., 2021) or to the spread of riots but without taking the social media into account (Bonnasse-Gahot et al., 2018).

Secondly, by tracking riots and comments both across space and over time, we can study riots and social media from the angle of sequences of events, as Cantoni et al. (2023) call for. Previous studies either missed the dynamics (Manacorda & Tesei, 2020; Enikolopov et al., 2020; Casanueva et al., 2022), the geography (Acemoglu et al., 2018), or the social media (Bonnasse-Gahot et al., 2018; Cantoni et al., 2023). The only exception is Qin et al. (2021) but they consider China, where the media are closely monitored by the authorities. By doing so, we can show that the effect of exposure to the riots on the probability of observing a riot in a municipality was fast but short-lived. It materializes the day after the exposure and fades away on the following day. We also find that the evolution over time of the marginal effect of exposure to social media on the probability of a riot was non-linear. It was initially small, reached a maximum at the peak of the riot wave, and then faded away. This finding emphasizes that the strength of connections on social media is not reducible to geographic proximity, which may explain why the riots spread to smaller and more distant municipalities than previous waves of riots observed in France (Oberti & Maela, 2023).

Thirdly, we are to our knowledge the first to consider an image-based social media. In this context, the study of image-based social media is particularly important. First because the participants of the riots we study predominantly involved teenagers, who mainly use image-based social media. Second, because images can trigger emotional reactions better than text (Casas & Williams, 2019) and protests can be considered an emotional response to unfair treatment (Passarelli & Tabellini, 2017).

Our final contribution is methodological. We suggest a way to address causality by using a variation in the exposure to the cause of protests driven by prior networks driven by connections to visible personalities, in our case sportsmen and artists. This strategy could easily be applied to other contexts by identifying personalities who are influential among the members of the groups that are studied.

To investigate the spread of protest through social media, the rest of the paper is organized as follows. The next section describes the context of the 2023 riots in France. Section 3 outlines the data set whereas Section 4 describes our empirical strategy. Section 5 reports the baseline results. Section 6 focuses on the dynamics of the relation. Section 7 investigates the heterogeneity of the effect. The last section unsurprisingly concludes.

2 Background

On 27 June 2023 during a roadside check in Nanterre, a municipality adjacent to Paris, French-Algerian 17-year-old Nahel Merzouk was shot at close range by a police officer at the wheel of the car he was driving. The two passengers of the car, aged 14 and 17, were arrested. Initially, the police officially reported that the car refused a checkpoint and rammed into the police officer who opened fire in self-defense. However, in the hours that followed two videos contradicting the official report of the police, one made by a passer-by and the other shot from his rear-view mirror by the driver of the car that was preceding Nahel Merzouk's, appeared on social media and went viral. The testimonies of the two passengers were also at odds with the official report. On 29 June, the public prosecutor announced the opening of an inquiry for voluntary manslaughter against the police officer who had fired his gun and ordered that he be remanded in custody.

In the meantime, the event sparked a wave of riots pitting youngsters, typically men in their teens or early twenties, against the police in 553 municipalities across France. The police was charged, shot at with fireworks, cars were set on fire, and buildings damaged. The riots were staggered over eight days and nights. Overall, according to the French home office, 5,954 cars were burned, 1,092 buildings were damaged, and 3,462 rioters were arrested and two died. 723 police officers were wounded.

In the first two nights following the death of Nahel Merzouk, the riots took place either directly in Nanterre, where he had been killed, or in municipalities around Paris. However, from June 28 onward, the riots spread to other municipalities, including smaller ones that had never experienced one (Oberti & Maela, 2023). The riots reached a peak around July 1 and then faded away following an unprecedented mobilization of the police. On July 5, the wave was over.

The social media were immediately perceived as playing a key role in the diffusion of riots. First, it is thanks to them that the videos contradicting the official reports of the police were made public. Then, their role in spreading riots and fanning violence was acknowledged by members of the police, the authorities, and the press (Bénézit et al., 2023; Clairouin et al., 2023).

Furthermore, several influential popular artists and star football players quickly reacted. The first was Jules Koundé, French men's national football team's center-back, who denounced "yet another police blunder" on the day of Nahel Merzouk's death (BFMTV, 2023). He was quickly followed by the team's goal-keeper Mike Maignan, and star forward player and captain Kylian Mbappé, whose statement "My France hurts" was widely commented in the media. Their teammates Aurélien Tchouaméni and Paul Pogba followed suit in the following days, as well Achraf Hakimi, a member of Paris Saint-Germain and the Morocco national team (Parisien, 2023b). Hip-hop was not to be outdone as stars of the genre also quickly expressed their outrage and their support for Nahel's Merzouk's family (Parisien, 2023a). Singer Rhoff argued on the social media that a refusal to comply did not allow a police officer to "commit a murder". Kaaris hinted at the role of the social origins of Nahel Merzouk in triggering the reaction of the policeman who had shot him. Jul posted "a thought for Naël (sic) #Nanterre" on Instagram, whereas SCH voiced his support "to Nahel's loved ones and to 'our neighborhoods'". Sadek went as far as announcing that he would donate his next concert fee to Nahel's family(Parisien, 2023a). Those singers are successful and visible. Each one has garnered at least a platinum record and most of them several. Their interventions primed the role of the disadvantaged suburbs where Nahel Merzouk grew. For instance, "neighborhoods" -"*quartiers*" in French- in SCH's post is a synonym of "*banlieues*". However, whereas Rohff, Sadek, and Sadek come from the outskirts of Paris, Jul and SCH come from Marseille, the second largest French municipality, and admittedly capital-city and birthplace of French hip-hop together with Paris, which also has a large disadvantaged population with a migration background. Those sportsmen and artists, with their migration background and their proximity to disadvantaged neighborhoods, eased the identification of rioters with Nahel Merzouk and could easily elicit reactions.

3 Data

For analyzing the diffusion of protests through social media, we construct a municipality-day data set for the period from June 27^{th} to July 3^{th}, 2023, merging data of protest, social media activity, and general municipality characteristics from different sources that we describe below. Table 1 provides descriptive statistics for all the variables we use. We also provide visualizations of the geographic variation of the total number of protests (Figure 1) and the total number of comments of posts talking about Nahel (Figure 2).

Protest data We get all instances of protests and riots related to the shooting of Nahel from the 27th of June until the 3rd of July and their type (i.e. riots with fire, robberies, degradation of public goods, etc) from a compilation produced by the French media outlet "Le Monde". This time frame encompasses all protests related to the death of Nahel that were recorded by main news outlets and each day at least one municipality had a protest. Figure 5 shows the total number of protests per day.

Instagram posts and comments about Nahel We focus our analysis on Instagram because it is the most used social media platform in France among the age group between 16 and 25 years. (Diplomeo, 2023) This population is particularly relevant because participants in protests related to Nahel's death were mainly young [1].

Instagram is an image-centered social media platform that enables users to post, comment, and share pictures and videos publicly or privately (Frommer, 2010). Since Instagram has been shown to be a key source of information for young people in different contexts (see e.g., Gumpo et al. (2020) or Anter & Kümpel (2023)), we analyze Instagram posts about the protests following the death of Nahel. We scrape Instagram posts using the Hashtags *#Nahel* and *#Nael* and geolocalize them. Further, we scrape and geolocalize the comments of each Instagram post. We ended up with 412 geo-localized posts and 3,146 geo-localized comments from 850 municipalities. We show in Figure 5 the number of comments per day across France.

Instagram network at baseline To isolate the causal effect of exposure to protest content on social media, we need to gather data on the online proximity between municipalities before the shooting of Nahel. This a priori network should be independent of the probability of protest but, at the same time, capture users who could be potential future protesters. To do so, we scrape Instagram data on a network of people, using as starting points pop culture celebrities that could be of interest to young men from disadvantaged backgrounds (i.e. potential protesters), in particular football or rap. To make the list of celebrities even closer to the set of Instagram users we are interested in, we choose celebrities that ended up reacting to Nahel's death, but were not expressing political opinions before. We took this list of celebrities from Wikipedia and we cross-checked it with a media article from *Le Parisien* that also has a list of the celebrities reacting to Nahel's death.

After establishing this list of celebrities, we iteratively extract a network using Instagram posts and comments: first, we list the posts from the celebrities posted before in the two months before the death of Nahel, and extract comments from those posts, again limiting to comments posted prior to Nahel's death (due to limitations of the Instagram API, we only get 50 comments per post, while celebrity posts typically receive many more comments). We obtain a list of users, and we iterate the process on this list of level 1 users, obtaining level 2 and (after one more iteration) level 3 users. Finally, we geolocate all the obtained users (ignoring the initial celebrities), and approximate the Instagram network connection by which municipalities comment on other municipalities posts.

[1] Several Media sources have remarked that the great majority of protesters were teenagers, some as young as 13 years old, or young adults, with a median age of 17 years old. See for example Le Parisien or Bronner (2023).

We show the geographical distribution of post and comments linked to this network in Figures 3 and 4.

Facebook Social Connectedness Index As an alternative measure of social media proximity, we use Facebook's Social Connectedness Index (Bailey et al., 2018). This index, available at the *département* level in France, measures how likely it is that users from two departments are connected by Facebook friendship links: the index is the +normalized number of friendship between two départements divided by the Facebook users of the two départements, and with noise added for anonymization. We use this index as an approximation of the connectivity between municipalities. To keep it comparable with our other Instagram proximity measure, we multiply it by the population of both municipalities.

3.1 Geo-coding of posts and comments

A key component of analyzing the diffusion of protests through social media is determining the geographical location of the users posting and commenting about the protests. To geo-localize the posts and comments on Instagram, we solely use publicly available data. We screen the geo-tag of the post, the public biography of each user as well as the geo-tags and description of their latest 10 public posts for names of French municipalities with more than ten thousand inhabitants. In case of multiple mentions of municipalities, we choose the most often appearing municipalities. If multiple municipalities are mentioned equally often, we pick the municipality with the largest population. Additionally, we also screen the available information of the user for a ZIP code. It is not unusual among certain population groups, particularly in communities related to rap to publicly display the residence ZIP code. All together, we geolocalize 412 of the 2,347 posts (around 17.5%) and 3,146 of the 18,283 of the comments (around 17.2%) of the Instagram activity related to Nahel. For the Instagram activity around the 14 celebrities that may be liked and followed by potential rioters, among the 24,434 users collected, 4,377 are located (17.9%).

4 Empirical Strategy

Our empirical strategy assesses whether spatial and temporal variation in the level of exposure to previous protests in other locations through Instagram affected the probability of a protest in a municipality. We use an instrumental variable approach to isolate the effect of exposure to protest activity though social media on the onset of new protest. We focus on municipalities with population greater than ten thousand and consider the level of exposure of each municipality to protest on social media, as measured by the number of comments written by users from the municipality on posts talking about Nahel from a municipality with previous protests. We instrument the comments of posts about Nahel by a mesure of *a priori* connectivity measure through Instagram using people commenting on posts of relevant celebrities. We will first explain our fixed effects structure and then the instrumental variable approach.

4.1 Two-way fixed effects for social media exposure to protests

The first step to study the diffusion of protests through social media consists in constructing a measure of exposure to protests on social media. As described above, we focus on engagement on Instagram, since it is the most used social media platform among 16 to 25-year-old in France, which made up the majority of participants in the protests following the death of Nahel.

To obtain an accurate measure of exposure, we would ideally observe and geolocalize the users who view a post talking about a certain protest in another municipality. Unfortunately, our data does not allow us to know the users who viewed a specific post. However, we can observe the users that commented on a post and we can use publicly available information to geolocalize the author of the comment. We hypothesize that the higher the engagement with posts about protests on a given day, the more likely it is that a municipality will experience a protest in the following days. We describe the construction of our measure of engagement in detail below.

For each day, we consider all Instagram posts regarding Nahel from all municipalities j that had a protest on any previous day (we exclude the posts from municipalities that did not have a protest in the days before). Then we take all the comments on those posts from municipalities i (excluding comments from a municipality to itself, i.e. where $i = j$, as we are interested in the geographical diffusion). The sum of all these comments is our basic measure of exposure as described in the next equation:

$$\text{Measure of exposure}_{i,d} = \sum_{i \neq j} \text{Comments}_{i \to j,d}$$

where $\text{Comments}_{i \to j,d}$ are comments about Nahel on day d in municipality i in reply to posts from a municipality j that experienced a protest.

Once we have estimated the measure of exposure to a protest on social media, we investigate the impact of this measure on the probability of observing a protest in the following days. We take advantage of the panel structure of our data to compute a two-way fixed effect linear probability model with municipality and region-day fixed effects. We estimate the following regression, clustering standard errors at the municipality level:

$$\text{Protest}_{i,d} = \beta_0 + \beta_1 \text{Measure of exposure}_{i,d-1} + \alpha_i + \gamma_{rd} \tag{1}$$

The use of this fixed effect structure has several advantages. First, it captures the unobserved endogeneity related to the characteristics of the municipality that are time-invariant. All characteristics regarding the root municipality determinants of political preferences and the demographic, economic, or social characteristics will be captured by municipality-fixed effects. For example, urban municipalities could be more likely to organize protests and be more engaged in social media. Alternatively, left-wing municipalities could be more prone to protest and thus comment more and at the same time fear police repression less for protest as their institutions could mitigate or prevent a strong police intervention. All this possible endogeneity will be captured by municipality-fixed effects.

Given our short period of analysis, the municipality fixed effects capture a big part of the variation that could raise endogeneity concerns because they can capture the effect of characteristics that vary relatively fast but remain constant over a six-day period. For instance, in municipalities with a tourism-based economy and seasonal jobs, unemployment rates, particularly among young population, could impact both the available time to engage with social media content and the opportunity cost of protesting and could change rapidly between seasons but are unlikely to vary substantially in a six day period between the end of June and beginning of July.

Second, it captures the temporal shocks that affect all municipalities. For example, the average effect of a speech of a nationally known politician that can affect the probability of comment and the probability of protest will be captured by the day-fixed effects. We allow these temporal shocks to be different by region. This will capture temporal shocks that affect all municipalities within one region. For example, the effect of a speech of a politician known just in the region or specific weather conditions common to the region that made protest less likely will be captured by region-day fixed effects.

In addition to this structure of fixed effects, we also show the results for alternative combinations of fixed effects structures. Finally, we include linear municipality trends to control for trends that can be related to both the treatment and the outcome. For example, a linear trend that can be related to both the treatment and the outcome that could increase both the number of comments to posts about protest and protest itself.

4.2 Using baseline social media exposure as instrument

To isolate the effect of exposure to protests while addressing potential endogeneity concerns we employ an instrumental variable approach. As an instrument, we construct a network of social media exposure *at baseline* between municipalities that is relevant for the demographic that participated in the protest but is independent of the protests themselves. As described in the data section, this network is formed through interactions (through commenting) on a sample of posts, collected starting from posts by national hip-hop and football celebrities.

To capture the network we are interested in (ie. potential protesters), we start from a set of celebrities that could be of interest to young men from disadvantaged backgrounds. To make the list of celebrities even closer to the set of Instagram users we are interested in, we choose celebrities that ended up reacting to Nahel's death. Importantly, the celebrities we choose do not have a particular political leaning and are not known to express political opinions systematically, which makes it less likely that they attract users because of their political opinions. Also importantly, we consider the network *before* the death of Nahel, alleviating the concern that the celebrities could attract users because of their reaction to Nahel's death. This strategy is able to give us the "relevant" network (i.e. captures users that *a priori* could protest) while being exogenous to the treatment (because the list of celebrities we consider were not politically aligned *a priori*, and because we take the network *before* the death of Nahel).

We compute, for each pair of city, a directed Instagram influence called $\textit{Instagram Connections}_{j \to i}$, defined as the number of comments from users located in city i on posts from users located in city j. Then, we define our instrument as the sum of the influences of all municipalities that had protest at $d-1$ or before:

$$\text{Instrument}_{i,d} = \sum_{j \neq i} \text{Instagram Connections}_{j \to i} \times \text{Protest Before}_{j,d-1} \tag{2}$$

where Protest $\text{Before}_{j,d-1}$ is a dummy equal to one if municipality j had a protest on day $d-1$ or any previous day. Our measure of connectivity do not vary over time but the number of municipalities which connections are considered does as the number of municipalities that hosted a protest before day d changes over time.

Thus, our first stage writes as follows:

$$\begin{aligned} \text{Measure of exposure}_{i,d-1} = \pi_0 + \pi_1 \sum_{j \neq i} \text{Instagram Connections}_{j \to i} \times \text{Protest Before}_{j,d-1} \\ + \delta_i + \lambda_r d + \nu_{id} \end{aligned} \tag{3}$$

The results of our first stage, present in panel B of Table 7, show, as expected, that higher *a priori* connection to municipalities with protest in the previous days predict the number of comments to post about Nahel. The different fixed effects structures have F statistic close to or above 10. In particular, our preferred specification (columnn 5), has a F-statistic of more than 37, way above the commonly used threshold of 10 (Stock & Yogo, 2005).

The exogeneity assumption is ensured by solely using *pre-protest* interactions to construct the network and by including only celebrities that were not politically vocal prior to the protests. Lastly, the instrument fulfills the exclusion criterion as we expect the a priori network to have no direct effect on the likelihood to protest since it is principally apolitical and all protest-related interactions are captured by our main network of interest.

In the second stage of our 2SLS strategy, we use the fitted values of the exposure to protest as explanatory variable of the likelihood to protest. Our second stage, thus, writes as follows:

$$\text{Protest}_{id} = \beta_0 + \beta_1 \widehat{\text{Measure of exposure}}_{i,d-1} + \alpha_i + \gamma_r d + u_{id} \tag{4}$$

5 Main Results

The main results are presented in Table 7. Panel A show the second stage results corresponding to the empirical strategy described above. Panel B displays the results of the first stage, panel C the reduced form and panel D the OLS without instrumenting the measure of exposure. Different columns represent different fixed-effects structures. Column 1 shows the results including municipality FE (fixed effects), column 2 includes day FE, column 3 region-day FE, column 4 includes municipality and day FE and finally, column 5 includes municipality and region-day fixed effects and represents our preferred specification that we will use for the rest of the analysis.

All the estimates across the different fixed effects structures are positive and statistically significant at the 10% level. Importantly, all second stage results (panel A) show relatively similar magnitudes showing that our instrument is ortogonal to municipality characteristicts.

Interpreting the magnitudes, an increase of 1 comment to a post of a municipality that had a previous protest increases the probability of having a protest the day after by around 3.5 percentage points (pp). Alternatively, an increase in 1% of comments increases the likelihood of protest by 0.04%. Focusing on the estimates of the reduced form (panel C), an additional connection to the network of relevant celebrities, increases the probability of protest by around 0.12 pp.

The OLS results (panel D) are smaller in magnitude than the 2SLS (around half). This could happen either because the OLS is biased downwards or because the LATE is different than the ATE. In our case, the LATE captures the effect for the subpart of the population that is most likely to riot because we explicitly chose the celebrities that could interest young men from disadvantage background (i.e. potential rioters).

Overall, the results show that a higher level of exposure to protests in other municipalities through Instagram increases the probability of having a protest and that this increase is not due to the intrinsic characteristics of the municipalities or temporal shocks increasing both Instagram connections and riots.

5.1 Robustness checks

In our main table of results, we already show that our estimates are robust to different fixed effects structures. In this section, we perform additional empirical tests to assess the robustness of our estimates. We present the results in Table 3. Panel A presents the results for the second stage, panel B presents the first stage, panel C the reduced form and panel D the non-instrumented OLS results. Each column show the result fordifferent robustness checks that we describe below. The first column show the baseline identification for comparison.

Excluding Nanterre The shooting of Nahel Merzouk occurred in Nanterre. Therefore, a lot of posts and comments refer to this location without the post or the comments actually coming from Nanterre. Additionally, since the first protest occurred in the same municipality where Nahel was shot, studying protests occurring in Nanterre will not inform us about the geographical diffusion of protest. We exclude Nanterre from the sample, both as posting and as commenting municipality in column 2. The estimates remain positive and significant and the coefficients more than triple in magnitude.

Excluding Paris To alleviate a possible concern of our results being driven by outlier observations, we exclude Paris, the capital of France, the most populated city of France and the cities where most of the celebrities we consider to construct our proximity between municipalities though Instagram measure perform their activity. Results (presented in column 3) remain robust with a magnitude almost double the magnitude of our baseline specification. We present the results excluding both Nanterre and Paris in column 4 with robust results and higher magnitude.

Including municipality-trends We include in column 5 linear trends that can vary by municipality. This controls by the local escalation of concern about police brutality that can increase both on-line activity around content related to Nahel and the probability of protest. Results remain robust and increase in magnitude moderately but the instrument becomes less strong and the F-stat drops to slightly bellow 10.

Allowing for spatial correlation Observations can be correlated spatially which may lead to a wrong estimation of the standard error. We allow the spatial correlation of our observation within certain spatial windows using Conley standard errors (Conley, 1999). We consider four different spatial windows: 25km, 50km, 100km, and 200km within which we allow observations to be correlated. Results are presented in columns 6-9 of panel A of Table 3 and show that the results are robust to spatial correlation.[2]

6 Mechanism

7 Dynamics of the diffusion

7.1 Speed of diffusion

Protest can spread fast or be preceded by a long build-up (Cantoni et al., 2023). In our case, we expect the diffusion to occur rather fast for several reasons. First, because we analyze a social media platform that is image-based and images trigger emotions more strongly and easily than text (Casas & Williams, 2019). When emotions are stronger, the willingness to protest can increase, particularly if protests occur as an emotional reaction to unfair treatment (Passarelli & Tabellini, 2017) rather than after a strategic desition. Second, because the protests that we consider were not authorized by authorities and did not show signs of formal organization efforts. In this context, an immediate reaction is required to increase the probability of an effective action before an official repression.

[2] For power-computational constraints, we are only able to present the results for the strategy focusing on municipalities but we plan to present the results for municipality pairs in an updated version of the paper.

To test that possibility, we consider several lags of the explanatory variable. We investigate the velocity of diffusion by looking at the effect of comments on previous days. If the diffusion happens fast then we would expect the effect of comments several days before to be less important in explaining protest than comments the day just before a possible occurrence. Results are presented in Table 4. Panel A presents the results for the second stage of the 2SLS, panel B the first stage, panel C the reduced form and panel D the un-instrumented OLS. Different columns show different lag windows for comments. As hypothesized, comments on $d-1$ are the stronger predictor of protest, and magnitudes diminish and become statistically insignificant when considering comments on $d-2$ and at $d-3$, suggesting that the diffusion occurs rather fast.

7.2 Effect over time

So far, we have assumed a constant effect over time of comments on the probability of a riot. Yet, the spread of riots followed a wave as Figure 5 shows. Bonnasse-Gahot et al. (2018) observed the same phenomenon during the 2005 riots. To understand the temporal dynamics of the effect of exposure to protests on social media, we consider the daily effects of comments on Instagram. Figure 4 shows the effect of a comment written in a municipality on the previous day on the probability of protesting for each day of our period of analysis. Sub-figure (a) shows the estimates for the second stage of 2SLS; sub-figure (b) shows the results for the reduced form and sub-figure (c) the results of the un instrumented two way fixed effect OLS. Dots represent point estimates and lines confidence intervals.

The effect is significant only on the days of high protesting activity and is insignificant on the first and the last two days of the analyzed data frame. These results are consistent between the different sub-figures. These temporal dynamics of the effect are consistent with the following explanation: the first two days, protests took place mainly in Nanterre and in places that would usually protest police violence incidents. The protest in these places is not dependent on social media networks (we can consider them always takers). However, starting on June 29th, the protest spread across France in smaller towns that had not seen this type of protest before. Absent social media, residents of those municipalities may not have sympathized with rioters in larger municipalities. We can interpret these results as showing that this phase of the protest was much more driven by social media content. Finally, the number of protests decreased, which matches the decrease in effect. In our case, the usual *pic* [3] on the evolution of a movement (Cantoni et al., 2023) (which in this context also correspond to an unexpected geographical escalation) would be, at least in part, explained by the diffusion through social media.

[3] One key feature of riots documented by recent literature is that they spread geographically, although rioters themselves do not travel. Focusing on the 2005 riots in France, Bonnasse-Gahot et al. (2018) observe that their diffusion across municipalities could be accurately described by a SIR model, usually used to study epidemics. Likewise, studying 1.2 million protest events across 218 countries between 1980 and 2020, Cantoni et al. (2023) observe that protests tend to spread across municipalities over a few weeks before they reach a peak and decay.

7.3 Considering different networks

In Table 5, we consider the effect of other types of connectivity measures between municipalities to understand the relative importance of connections through an imaged-based social media mainly used by young people (i.e. Instagram). We perform the same reduced form analysis as in our main empirical strategy but we change the treatment to consider first another measure of online proximity among municipalities but based on Facebook and a measure of basic offline proximity: distance. Column 1 show our reduced form baseline specification of comparison, column 2 the results with the Facebook Social Connectedness index described in the data section, column 3 show the results using geographical proximity and column 4 runs a horse race between all three measures. All measures are standardize to be able to compare the magnitudes.

Both looking at individual results or at the estimated of the variables considered simultaneiously, we see that the network connection captured by the network around celebrities, explain better the diffusion of protest and predict better the probability of protesting than other measures of social connectedness. In the horse race (column 4) , the only significant result is the Instagram connectivity measure. Online connection through Facebook are significant (column 2) with a lower magnitude but become insignificant when we consider all variables in the same regression, likely because of the correlation between both measures. Geographical proximity by its own, does not seem to play a significant enough role to have a significant coefficient. However, as we will detail in the next section, the effect of online proximity is even more important in explaining protest diffusion in geographically close municipalities.

8 Heterogeneity of the effect

8.1 Heterogeneity according to distance

Identical levels of exposure to protests occurring in different municipalities may have different effects. For example, people may be more prone to react to posts of people who are geographically closer to them. We consider this possible heterogeneity by giving a greater weight to the comments of posts of closer municipalities compared to posts from far away municipalities. In particular, we define a level of influence that decreases monotonically with distance but that does not tend to infinity at very close distances. The next two equations describe precisely our weighting strategy.

$$\text{Weighted measure of exposure}_{i,d} = \sum \text{Comments}_{i \to j,d} \times \text{Influence}_{i,j}$$

where $\sum \text{Comments}_{i \to j,d}$ is our previous measure of exposure and $\text{Influence}_{i,j}$ is the weight defined as follow:

$$\text{Influence}_{i,j} = \frac{threshold}{threshold + \text{DistanceGeo}_{i,j}}$$

For any given value of the threshold, the comments at distance zero are weighted one.[4] The

[4]Note that there will not be comments from a post of a municipality at zero distance because we eliminate all comments

higher the threshold, the slower the weights decrease with distance. For example, with a 50km threshold, comments from a protest of a municipality at 50 km will weigh half as much as posts of municipalities at 0 km, while comments of a protest at 100 km will be weighted 1/3. With a threshold of 100km, comments from a protest at 50km will be weighted 2/3, and comments on a protest at 100 km will be weighted 1/2.

Results are presented in Table 6. Panel A presents the results for the second stage of the 2SLS; panel B shows the first stage; panel c the reduced form and panel D the un-instrumented OLS. Column 1 is the baseline specification for comparison. Columns 2-5 change the *threshold* of the *influence* measure described above going to the threshold that weighs the different distances more equally (i.e. closer to the baseline specification) in column 2 to the threshold that puts relatively more weight to closer municipalities in column 5. Magnitudes increase steadily as higher weights are given to close cities. This suggests that people are more likely to protest in their own municipality if they are exposed, through social media, to protests occurring in neighboring regions than if the protest happened in municipalities further away. The effect of online exposure to protest is then higher with geographical proximity.

9 Conclusion

In this paper, we analyze the diffusion of riots through image-based social media after 17-year-old Nahel Merzouk was shot by the police in France. To do so, we leverage networks based on posts from Instagram – the social media platform most commonly used among the young population – that contain mentions of Nahel and the comments to these posts. We instrument the Instagram activity around Nahel by a baseline measure of online social connectedness using the Instagram network of celebrities followed by young men from disadvantaged networks (i.e. potential rioters).

Using an *a priori* network of social media exposure as instrument, we find that higher exposure to past riots through image-based social media increased the probability of a new riot. We investigate the dynamics of the diffusion and find that exposure to past riots has rapid effects on new mobilization. The effect is strongest on the day following exposure to a riot and then declines in magnitude before becoming insignificant on the third day. Second, we look at the heterogeneity of the effect and find that the effect is driven by the three days in the middle of our studied period. Finally, we show that the connection between municipalities that we capture through celebrities and their network on Instagram is more relevant for protest diffusion than other measures of proximity such as the Facebook connectedness index or geographical proximity. However, the effect of online proximity is stronger if municipalities are also geographically close.

Overall, the results align with the view that protest can trigger information cascades, allowing individuals to dynamically update beliefs and leading to an unexpected momentum of social movements (Kuran, 1989, 1991; Lohmann, 2000; González, 2020; Bursztyn et al., 2021). We show that these cascades travel effectively through specific networks on social media and diffuse protests.

that commented on a post from the same municipality.

References

Acemoglu, D., Hassan, T. A., & Tahoun, A. (2018). The power of the street: Evidence from egypt's arab spring. *Review of Financial Studies*, 31(1), 1–42.

Anter, L. & Kümpel, A. S. (2023). Young adults' information needs, use, and understanding in the context of instagram: A multi-method study. *Digital Journalism*.

Bailey, M., Cao, R., Kuchler, T., Stroebel, J., & Wong, A. (2018). Social connectedness: Measurement, determinants, and effects. *Journal of Economic Perspectives*, 32(3), 259–80.

BFMTV (2023). Equipe de france: Maignan, mbappé, koundé... la colère des bleus après la mort d'un jeune tué par un tir de police à nanterre.

Bonnasse-Gahot, L., Berestycki, H., Depuiset, M.-A., Gordon, M. B., Roché, S., Rodriguez, N., & Nadal, J.-P. (2018). Epidemiological modelling of the 2005 french riots: a spreading wave and the role of contagion. *Scientific reports*, 8(1), 107.

Bronner, L. (2023). Les leçons d'émeutes sans précédent. *Le Monde*.

Bursztyn, L., Cantoni, D., Yang, D. Y., Yuchtman, N., & Zhang, Y. J. (2021). Persistent political engagement: Social interactions and the dynamics of protest movements. *American Economic Review: Insights*, 3(2), 233–250.

Bénézit, J., Lenoir, A., Richardot, R., Rof, G., Schittly, R., & Traullé, F. (2023). Emeutes après la mort de nahel m. : jeunes, survoltés et hyperconnectés, des émeutiers au profil complexe. *Le Monde*.

Cantoni, D., Kao, A., Yang, D. Y., & Yuchtman, N. (2023). *Protests*. Technical report, National Bureau of Economic Research.

Casanueva, A., Avetian, V., Sardoschau, S., & Saxena, K. (2022). Social media and the broadening of social movements: Evidence from black lives matter. *Available at SSRN 3831819*.

Casas, A. & Williams, N. W. (2019). Images that matter: Online protests and the mobilizing role of pictures. *Political Research Quarterly*, 72(2), 360–375.

Castells, M. (2015). *Networks of outrage and hope: Social movements in the Internet age*. John Wiley & Sons.

Clairouin, O., Laurent, S., & Reynaud, F. (2023). Mort de nahel m. : avec les réseaux sociaux, des émeutes en direct. *Le Monde*.

Conley, T. G. (1999). Gmm estimation with cross sectional dependence. *Journal of econometrics*, 92(1), 1–45.

Diplomeo, B. (2023). Sondage resaux sociaux jeunes 2023.

Enikolopov, R., Makarin, A., & Petrova, M. (2020). Social media and protest participation: Evidence from russia. *Econometrica*, 88(4), 1479–1514.

Frommer, D. (2010). What is instagram? *Business Insider*.

González, F. (2020). Collective action in networks: Evidence from the chilean student movement. *Journal of Public Economics*, 188, 104220.

Gumpo, C. I. V., Chuchu, T., Maziriri, E. T., Madinga, N. W., & Maziriri, E. (2020). Examining the usage of instagram as a source of information for young consumers when determining tourist destinations. *SA Journal of Information Management*, 22.

Kuran, T. (1989). Sparks and prairie fires: A theory of unanticipated political revolution. *Public choice*, 61(1), 41–74.

Kuran, T. (1991). The east european revolution of 1989: is it surprising that we were surprised? *The American Economic Review*, 81(2), 121–125.

Libération, 2023. Mort de nahel : après les émeutes, les réseaux sociaux sous pression du gouvernement.

Lohmann, S. (2000). Collective action cascades: An informational rationale for the power in numbers. *Journal of Economic Surveys*, 14(5), 655–684.

Manacorda, M. & Tesei, A. (2020). Liberation technology: Mobile phones and political mobilization in africa. *Econometrica*, 88(2), 533–567.

Oberti, M. & Maela, G. L. G. (2023). Analyse comparée et socio-territoriale des émeutes de 2023 en france. *Mimeo, Sciences Po*.

Parisien, L. (2023a). Nahel, 17 ans, tué par un policier : Jul, sch, sadek... des stars du rap réagissent après le drame. *Le Parisien*.

Parisien, L. (2023b). Nahel aurait pu être mon petit frère : la lettre ouverte de tchouaméni après la mort de l'adolescent. *Le Parisien*.

Passarelli, F. & Tabellini, G. (2017). Emotions and political unrest. *Journal of Political Economy*, 125(3), 903–946.

Qin, B., Strömberg, D., & Wu, Y. (2021). Social media and collective action in china.

Stock, J. & Yogo, M. (2005). *Testing for Weak Instruments in Linear IV Regression*, (pp. 80–108). Cambridge University Press: New York.

Zhuravskaya, E., Petrova, M., & Enikolopov, R. (2020). Political effects of the internet and social media. *Annual review of economics*, 12, 415–438.

10 Figures and Tables

Figure 1. Protests

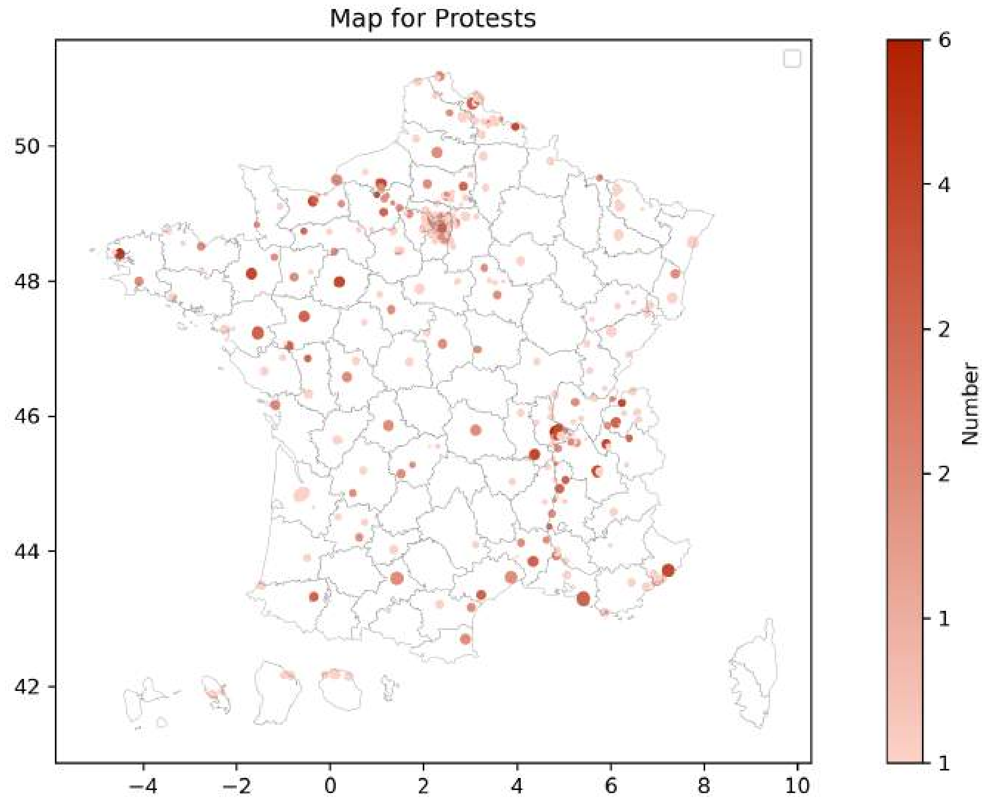


Figure 2. Comments

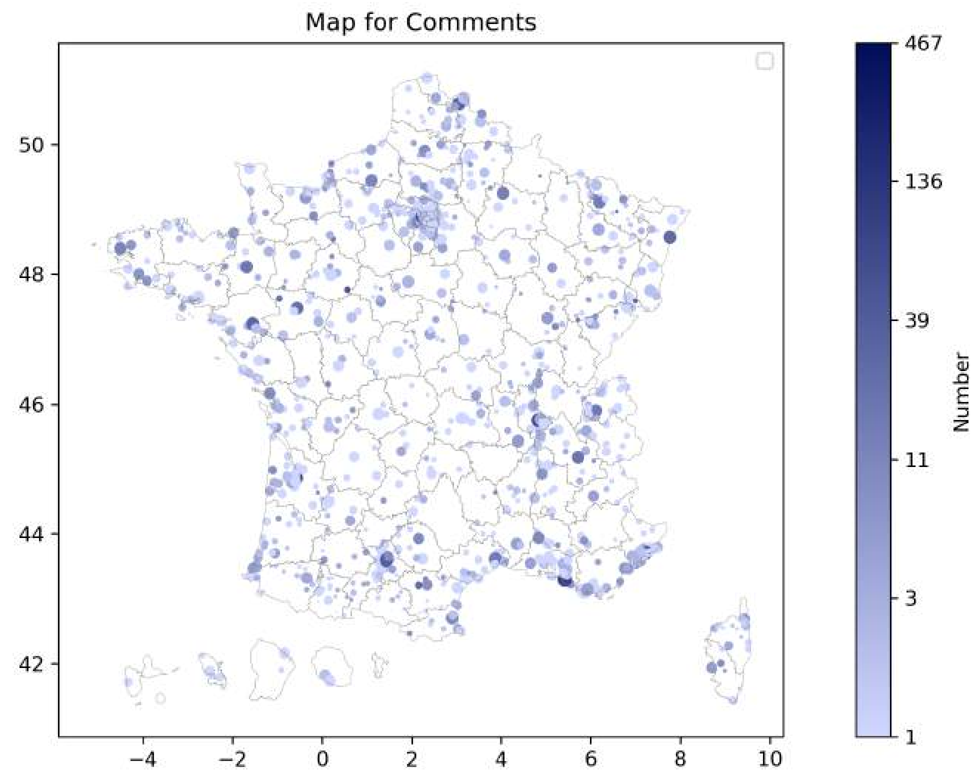


Figure 3. Posts in network at baseline from celebrity posts 2 months before Nahel's death

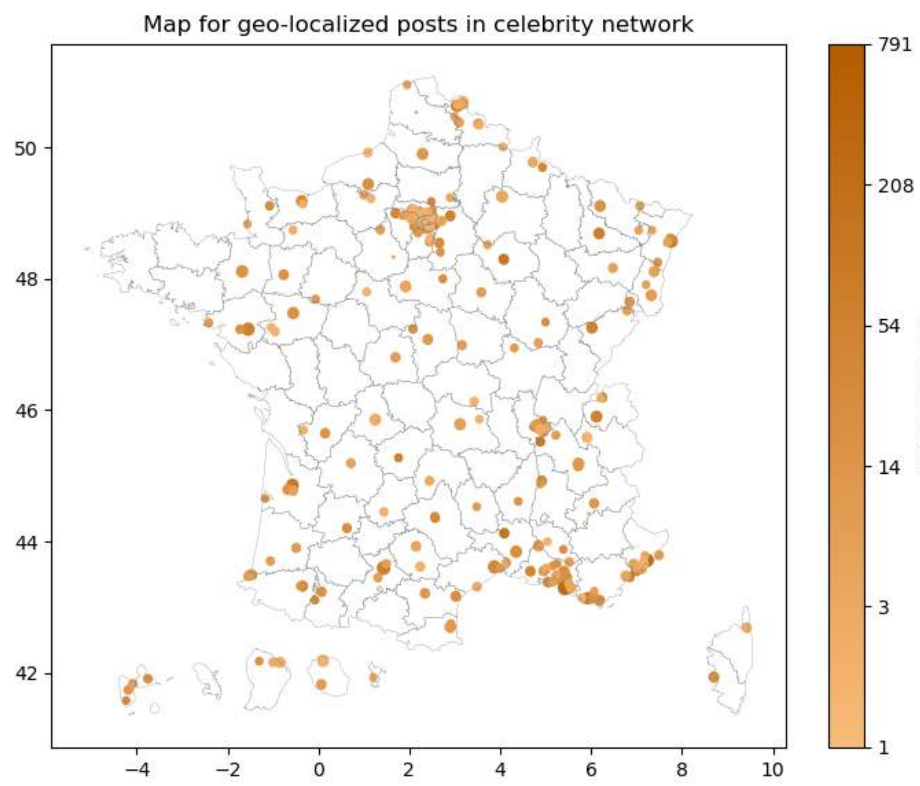


Figure 4. Comments in network at baseline from celebrity posts

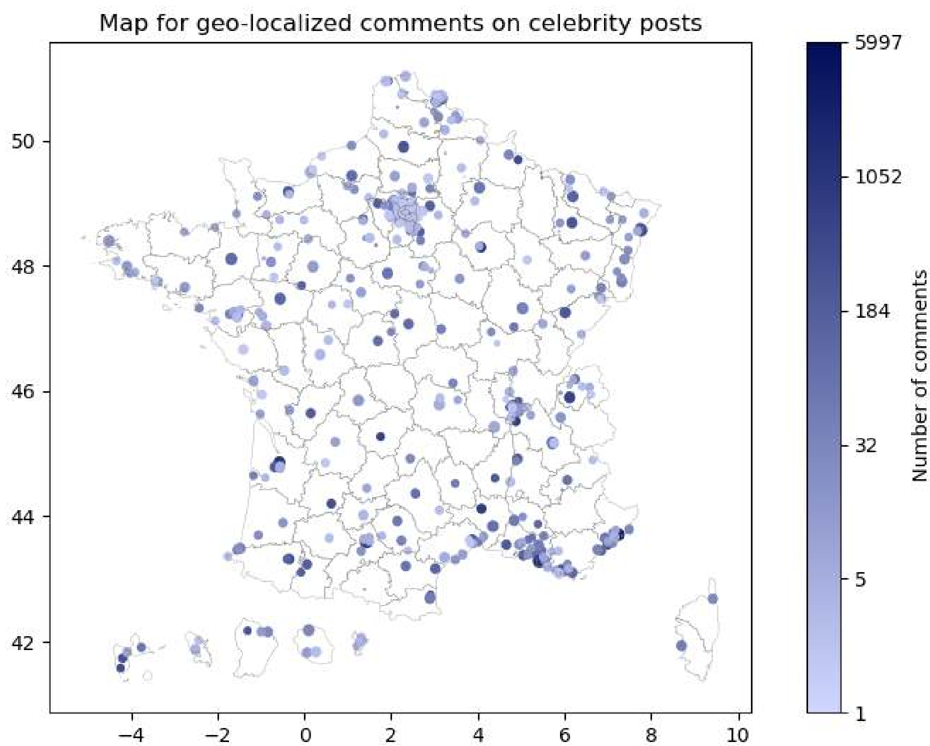


Figure 5. Protests and comments per day

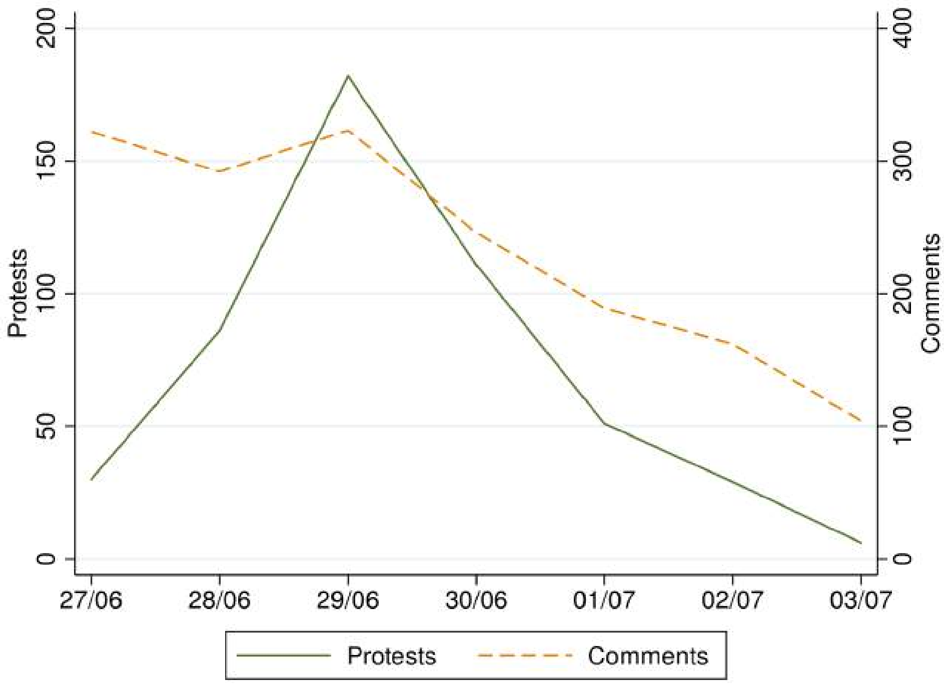
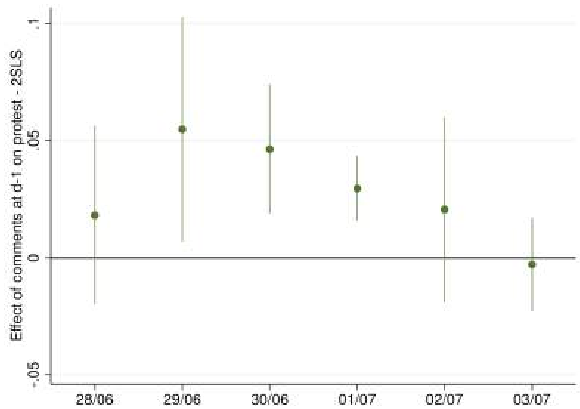
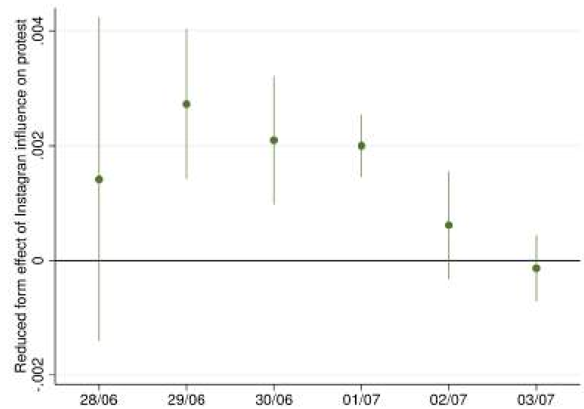


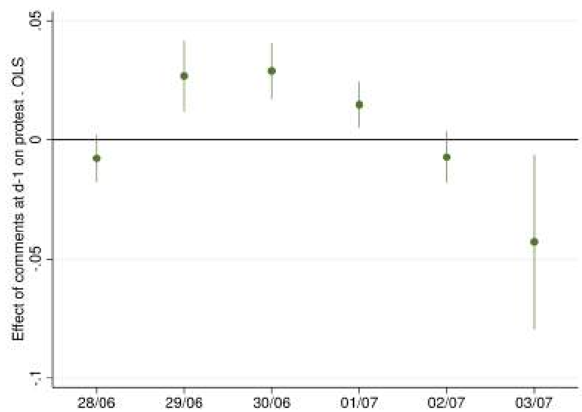
Figure 6. Effect per day



(a) 2SLS

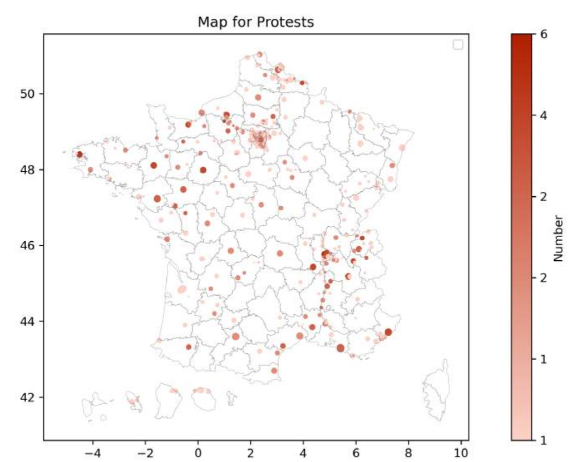


(b) Reduced form

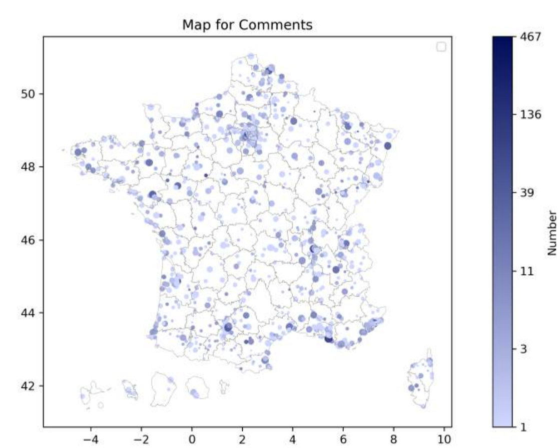


(c) OLS

Figure 7. Total Protests and Comments

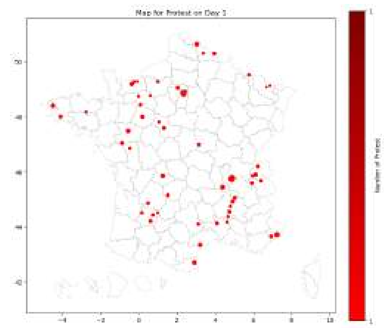


(a) Protests

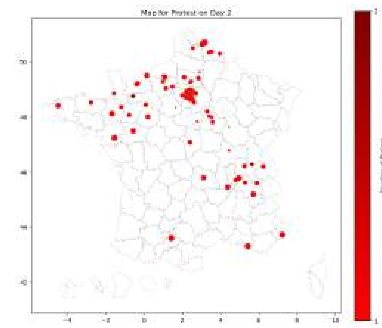


(b) Comments

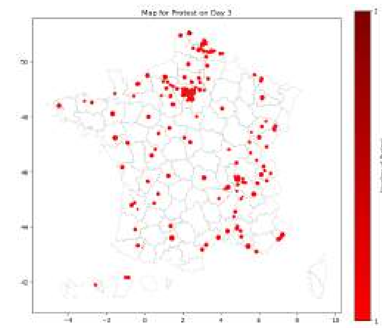
Figure 8. Daily plots of Protests



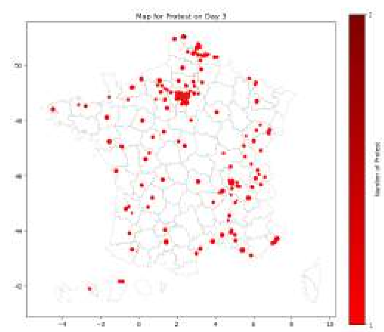
(a) Protests Day 1



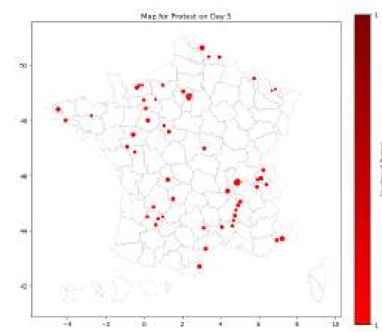
(b) Protests Day 2



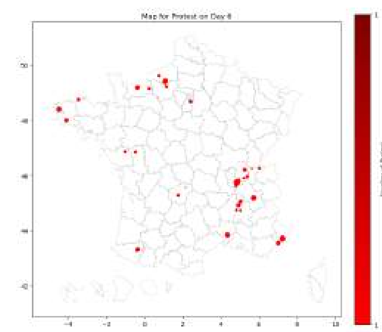
(c) Protests Day 3



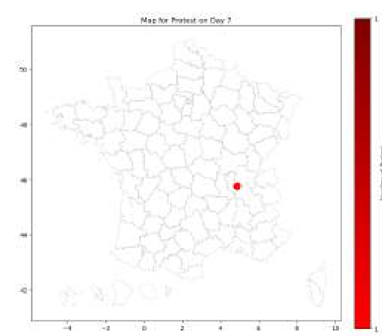
(d) Protests Day 4



(e) Protests Day 5

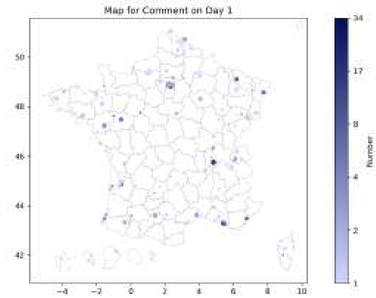


(f) Protests Day 6

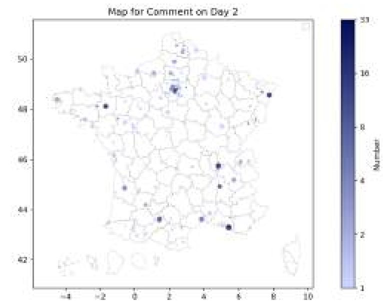


(g) Protests Day 7

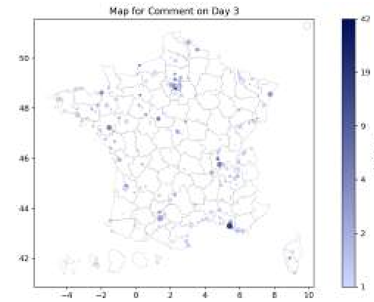
Figure 9. Daily plots of Comments



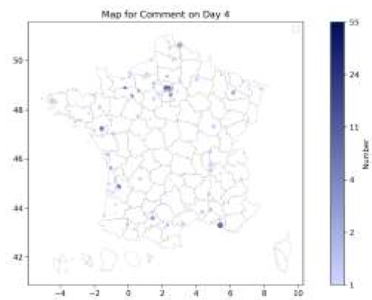
(a) Comments Day 1



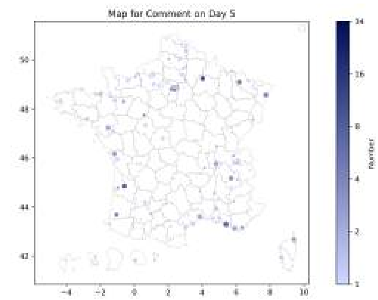
(b) Comments Day 2



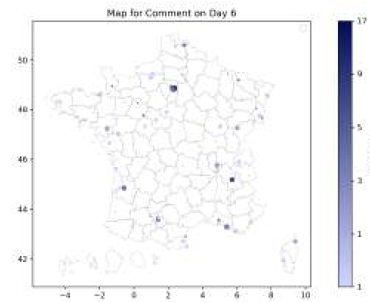
(c) Comments Day 3



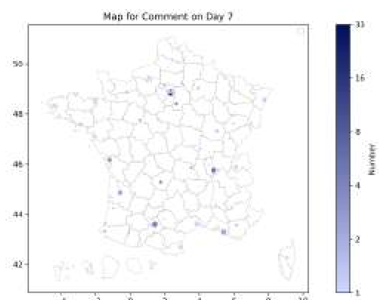
(d) Comments Day 4



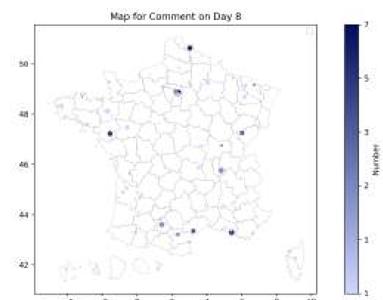
(e) Comments Day 5



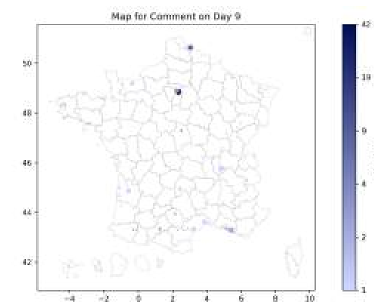
(f) Comments Day 6



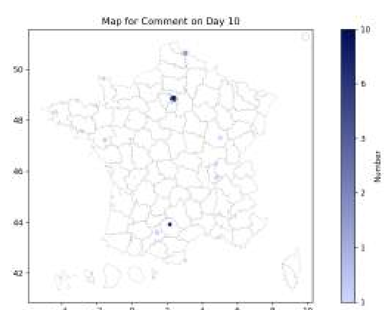
(g) Comments Day 7



(h) Comments Day 8



(i) Comments Day 9



(j) Comments Day 10

Table 1. **Descriptive statistics**

	count	mean	sd	min	max
Panel A: Municipality dataset					
Protest	7294	0.0679	0.2542	0.0000	1.0000
Comments $d-1$	7294	0.0846	1.0097	0.0000	47.0000
Instagram influence	7294	1.9917	17.8150	0.0000	601.0747
Influence comments $d-1$, 100 km	7294	0.0362	0.7680	0.0000	41.5557
Influence comments $d-1$, 50 km	7294	0.0289	0.6937	0.0000	37.7683
Influence comments $d-1$, 30 km	7294	0.0224	0.5868	0.0000	32.0106
Influence comments $d-1$, 15 km	7294	0.0178	0.4883	0.0000	26.6310
Cumul protest $d-1$	7294	0.2471	0.6164	0.0000	5.0000
Cumul comments $d-2$	7294	0.2183	3.0370	0.0000	164.0000

Note: Descriptive statistics of the main variables used in the empirical analysis. In panel A presents the unit of observation is a municipality and in panel B it is a municipality pair.

Table 2. **Diffusion of protest through social media.**

	Protest				
	(1)	(2)	(3)	(4)	(5)
Panel A: 2SLS					
Comments $d-1$	0.0346***	0.0362*	0.0372*	0.0327***	0.0336***
	(0.0101)	(0.0188)	(0.0192)	(0.00895)	(0.00659)
F first stage	36.90	54.50	54.87	36.95	37.06
Panel B: First stage					
Instagram connections × Protest before	0.0369***	0.0463***	0.0464***	0.0370***	0.0371***
	(0.00608)	(0.00628)	(0.00626)	(0.00609)	(0.00609)
Panel C: Reduced Form					
Instagram connections × Protest before	0.00128***	0.00168**	0.00173**	0.00121***	0.00124***
	(0.000264)	(0.000669)	(0.000677)	(0.000240)	(0.000218)
Panel D: OLS					
Comments $d-1$	0.0176***	0.0248***	0.0255***	0.0154**	0.0165**
	(0.00602)	(0.00701)	(0.00705)	(0.00621)	(0.00683)
Comment muni. FE	Y			Y	Y
Day FE		Y		Y	
Region-day FE			Y		Y
Observations	7,294	7,294	7,294	7,294	7,294

Note: Effect of social media comments on posts from municipalities with protests on protests. Each observation is a municipality-day unit. The sample is restricted to municipalities with population greater than 10,000. Time horizon is June 27th 2023 until July 3th 2023. Comments are measured as the comments posted from municipality m posted at $d-1$ on posts from any municipality that had a protest at $d-1$ or before. The outcome is the number of protests that occurred on day d in municipality m. Panel A presents the 2SLS estimate. Comments are instrumented by the sum of the Instagram connections to municipalities that had protest at $d-1$ or before, measured on a sample of Instagram posts. Panel B presents the first stage, panel C the reduced form regression, and panel D the OLS estimate. Different columns present different fixed effect structures. Column 1 includes municipality fixed effects. Column 2 includes day fixed effects. Column 3 includes region-day fixed effects. Column 4 combines municipality fixed effects with day fixed effects, and column 5 with region-day fixed effects. Standard errors (in parenthesis) are clustered by municipality. Kleibergen-Paap rk Wald F statistic provided. *p<0.1; **p<0.05; ***p<0.01.

Table 3. **Robustness checks**

		excl.	excl.	excl.	Municipality trends +		Conley standard errors			
	baseline	Nanterre	Paris	both	day FE	region-day FE	25km	50 km	100 km	200 km
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: 2SLS										
Comments $d-1$	0.0336***	0.132**	0.0569**	0.150**	0.0568**	0.0586**	0.0336**	0.0336*	0.0336*	0.0336*
	(0.00659)	(0.0569)	(0.0284)	(0.0641)	(0.0249)	(0.0242)	(0.0169)	(0.0173)	(0.0173)	(0.0179)
F first stage	37.06	9.643	4.036	7.455	8.918	9.037				
Panel B: First stage										
Instagram × Protest $d-1$	0.0371***	0.0230***	0.0566**	0.0215***	0.0847***	0.0850***	0.0371**	0.0371**	0.0371**	0.0371**
	(0.00609)	(0.00742)	(0.0282)	(0.00786)	(0.0284)	(0.0283)	(0.0149)	(0.0149)	(0.0149)	(0.0149)
Panel C: Reduced form										
Instagram × Protest $d-1$	0.00124***	0.00305**	0.00322**	0.00322**	0.00481***	0.00498***	0.00124**	0.00124**	0.00124**	0.00124**
	(0.000218)	(0.00130)	(0.00136)	(0.00136)	(0.00110)	(0.00103)	(0.000531)	(0.000539)	(0.000538)	(0.000583)
Panel D: OLS										
Comments $d-1$	0.0165**	0.0328	0.0175***	0.0415**	0.0187***	0.0200***	0.0165***	0.0165***	0.0165***	0.0165***
	(0.00683)	(0.0213)	(0.00655)	(0.0191)	(0.00583)	(0.00607)	(0.00564)	(0.00533)	(0.00544)	(0.00489)
Comment muni FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Region-day FE	Y	Y	Y	Y		Y	Y	Y	Y	Y
Day FE					Y					
Municipality trend					Y	Y				
Observations	7,294	7,294	7,294	7,294	7,294	7,294	7,294	7,294	7,294	7,294

Note: Effect of social media comments on posts from municipalities with protests on protests. Dependent variable is the number of protests on day d. Each observation is a municipality-day unit. The sample is restricted to municipalities with population greater than 10,000. Time horizon is June 27th 2023 until July 3th 2023. Comments are measured as the comments posted from municipality m posted at $d-1$ on posts from any municipality that had a protest at $d-1$ or before. The outcome is the number of protests that occurred on day d in municipality m. Panel A presents the 2SLS estimate. Comments are instrumented by the sum of the Instagram connections to municipalities that had protest at $d-1$ or before, measured on a sample of Instagram posts. Panel B presents the first stage, panel C the reduced form regression and panel D the OLS regression. Different columns present different robustness checks. Column 1 shows the baseline results for comparison. Column 2 removes comments to and from Nanterre, column 3 to and from Paris, and column 4 to and from both. Columns 5 and 6 include municipality-specific trends. In column 5, we replace the usual region-day fixed effect by a day fixed-effect, while the usual region-day fixed effect is kept in column 6. Columns 7 to 10 use clustering that accounts for spatial correlation, varying the distance at which correlation is taken to vanish from 25 km to 200 km. Standard errors (in parenthesis) are clustered by municipality except in columns 7-10. Kleibergen-Paap rk Wald F statistic provided when available. *p<0.1; **p<0.05; ***p<0.01.

Table 4. **Lagged comments**

Outcome:	Protest at d		
	Comments and instagram at		
	$d-1$ (baseline)	$d-2$	$d-3$
	(1)	(2)	(3)
Panel A: 2SLS			
Lagged comments	0.0336***	0.00946	-0.00354
	(0.00659)	(0.00575)	(0.00628)
F first stage	37.06	31.62	37.79
Panel B: First stage			
Instagram × Lagged protest	0.0371***	0.0489***	0.0577***
	(0.00609)	(0.00870)	(0.00939)
Panel C: Reduced form			
Instagram × Lagged protest	0.00124***	0.000463	-0.000205
	(0.000218)	(0.000336)	(0.000348)
Panel D: OLS			
Lagged comments	0.0165**	0.0137***	-0.00330
	(0.00683)	(0.00479)	(0.00734)
Observations	7,294	7,294	7,294
Muni. FE	Y	Y	Y
Region-Day FE	Y	Y	Y

Note: Effect of social media comments on posts from municipalities with protests on protests. D Each observation is a municipality-day unit. The sample is restricted to municipalities with population greater than 10,000. Time horizon is June 27th 2023 until July 3th 2023. Comments are measured as the comments posted from municipality m posted at $d-1$ ($d-2$ in column 3, $d-3$ in column 3) on posts from any municipality that had a protest at $d-1$ (resp. $d-2$, $d-3$) or before. The outcome is the number of protests that occurred on day d in municipality m. Panel A presents the 2SLS estimate. Comments are instrumented by the sum of the Instagram connections to municipalities that had protest at $d-1$ or before (resp. $d-2$, $d-3$), measured on a sample of Instagram posts. Panel B presents the first stage, and panel C the reduced form regression. Different columns present different fixed effect structures. Column 1 includes municipality fixed effects. Column 2 includes day fixed effects. Column 3 includes region-day fixed effects. Column 4 combines municipality fixed effects with day fixed effects, and column 5 with region-day fixed effects. Standard errors (in parenthesis) are clustered by municipality. Kleibergen-Paap rk Wald F statistic provided. *p<0.1; **p<0.05; ***p<0.01.

Table 5. **Diffusion of protest through social media, distance.**

	Protests			
	(1)	(2)	(3)	(4)
Instagram influence	0.0222***			0.0430*
	(0.00388)			(0.0256)
Social connectivity		0.0190**		0.00543
		(0.00907)		(0.0331)
Distance			0.0151	-0.0419
			(0.0146)	(0.0380)
Observations	7,294	7,280	7,294	7,280
R-squared	0.332	0.332	0.331	0.333
Municipality FE	Yes	Yes	Yes	Yes
Region-day FE	Yes	Yes	Yes	Yes

Note: Effect of connections to municipalities with protests on local protests. Dependent variable is the number of protests on day d. Each observation is a municipality-day unit. The sample is restricted to municipalities with population greater than 10,000. Time horizon is June 27th 2023 until July 3th 2023. Three types of influences are measured: instagram influence are the number of instagram connections measured in the Instagram comments sample to municipalities with protests at $d-1$ or before. Similarly, social connectivity is the sum of the (unnormalized) Facebook Social Connectedness Index to municipalities with protests at $d-1$ or before. Finally, distance is the sum of the population of nearby municipalities that had a protest before, multiplied by a weight decreasing linearly from 0 to 200 km. All models include region-day fixed effects and municipality fixed effects. Standard errors (in parenthesis) are clustered by municipality. Kleibergen-Paap rk Wald F statistic provided. *p<0.1; **p<0.05; ***p<0.01.

Table 6. **Differential effect according to distance. Changing the weights**

	Protest				
	(1)	(2)	(3)	(4)	(5)
Panel A: 2SLS					
Influence comments $d-1$	0.0336***	0.0403***	0.0429***	0.0480***	0.0552***
	(0.00659)	(0.00915)	(0.00924)	(0.00990)	(0.0115)
F first stage	37.06	36.16	41.79	45.39	43.59
Panel B: First stage					
Influence instagram	0.0371***	0.116***	0.163***	0.221***	0.262***
	(0.00609)	(0.0192)	(0.0252)	(0.0327)	(0.0397)
Panel C: Reduced form					
Influence instagram	0.00124***	0.00466***	0.00699***	0.0106***	0.0144***
	(0.000218)	(0.000687)	(0.00104)	(0.00172)	(0.00272)
Panel D: OLS					
Influence comments $d-1$	0.0165**	0.0259***	0.0287***	0.0341***	0.0412***
	(0.00683)	(0.00323)	(0.00306)	(0.00329)	(0.00379)
Observations	7,294	7,294	7,294	7,294	7,294
Influence	1	$100/(100+d_{geo})$	$50/(50+d_{geo})$	$25/(25+d_{geo})$	$15/(15+d_{geo})$
Municipality FE	Yes	Yes	Yes	Yes	Yes
Region-day FE	Yes	Yes	Yes	Yes	Yes

Note: Effect of social media comments on posts from municipalities with protests on protests. Dependent variable is the number of protests on day d. Each observation is a municipality-day unit. The sample is restricted to municipalities with population greater than 10,000. Time horizon is June 27th 2023 until July 3th 2023. Comments are measured as the comments posted from municipality m posted at $d-1$ on posts from any municipality that had a protest at $d-1$ or before, weighted by a distance-dependent weight. The outcome is the number of protests that occurred on day d in municipality m. Panel A presents the 2SLS estimate. Comments are instrumented by the sum of the Instagram connections to municipalities that had protest at $d-1$ or before, measured on a sample of Instagram posts, weighted by the same distance-dependent weight as comments. Panel B presents the first stage, panel C the reduced form regression, and panel D the OLS
All models include region-day fixed effects and municipality fixed effects. Standard errors (in parenthesis) are clustered by municipality. Kleibergen-Paap rk Wald F statistic provided. *p<0.1; **p<0.05; ***p<0.01.

Table 7. **Diffusion of protest through social media. - Restricting comments to Protest $d-1$**

	Protest				
	(1)	(2)	(3)	(4)	(5)
Panel A: 2SLS					
Comments $d-1$	0.0807*	0.0769*	0.0789*	0.0568**	0.0581**
	(0.0447)	(0.0445)	(0.0458)	(0.0276)	(0.0276)
F first stage	5.052	11.46	11.59	4.994	5.043
Panel B: First stage					
Instagram connections × Protest yesterday	0.0542**	0.0505***	0.0507***	0.0545**	0.0548**
	(0.0241)	(0.0149)	(0.0149)	(0.0244)	(0.0244)
Panel C: Reduced Form					
Instagram connections × Protest yesterday	0.00438***	0.00389***	0.00400***	0.00309***	0.00318***
	(0.00119)	(0.00128)	(0.00131)	(0.000776)	(0.000750)
Panel D: OLS					
Comments $d-1$	0.0248***	0.0359***	0.0366***	0.0171**	0.0181**
	(0.00914)	(0.0130)	(0.0130)	(0.00682)	(0.00706)
Comment muni. FE	Y			Y	Y
Day FE		Y		Y	
Region-day FE			Y		Y
Observations	7,294	7,294	7,294	7,294	7,294

Note: Effect of social media comments on posts from municipalities with protests on protests. Each observation is a municipality-day unit. The sample is restricted to municipalities with population greater than 10,000. Time horizon is June 27th 2023 until July 3th 2023. Comments are measured as the comments posted from municipality m posted at $d-1$ on posts from any municipality that had a protest at $d-1$ or before. The outcome is the number of protests that occurred on day d in municipality m. Panel A presents the 2SLS estimate. Comments are instrumented by the sum of the Instagram connections to municipalities that had protest at $d-1$ or before, measured on a sample of Instagram posts. Panel B presents the first stage, panel C the reduced form regression, and panel D the OLS estimate. Different columns present different fixed effect structures. Column 1 includes municipality fixed effects. Column 2 includes day fixed effects. Column 3 includes region-day fixed effects. Column 4 combines municipality fixed effects with day fixed effects, and column 5 with region-day fixed effects. Standard errors (in parenthesis) are clustered by municipality. Kleibergen-Paap rk Wald F statistic provided. *p<0.1; **p<0.05; ***p<0.01.