The age of content: How content fuels protests on visual social media

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

July 2025

preliminary draft

Abstract

How does the nature of social media content shape the diffusion of protest? This paper provides causal evidence from the 2023 Nahel riots in France, a nationwide wave of unrest triggered by a police killing and propelled by online mobilization. Using data on 537 protest events and more than 50,000 geolocated Instagram interactions, we construct a municipality-by-day panel of social media exposure. Our identification strategy leverages exogenous variation in pre-riot Instagram connectivity—measured via residualized comment flows on a network seeded on football celebrities' posts—to isolate the diffusion of protest through online ties. We find that digital exposure to protest in connected municipalities significantly increases the probability of local protest, with effects concentrated the following day. Beyond volume, we show that protest diffusion is content-dependent. Pro-movement, emotionally salient, and visually prominent posts are particularly effective at promoting both online engagement and offline action, while exposure to fear-inducing or repressive imagery discourages protest despite increasing attention. Posts with coordination cues (explicit calls to action or protest logistics) have the strongest mobilizing effects. Our findings show that image-based social media accelerate protest diffusion not only by lowering coordination barriers but also by amplifying emotional cues. They reveal how emotional tone, political alignment, and visual structure jointly condition the spread of digital protest.

Keywords: protest, diffusion, social media content, images
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

There is little doubt left that social media facilitate the organization of protest. Their effect has been observed in both democratic and autocratic countries alike (Flückiger & Ludwig, 2025; Casanueva-Artís et al., 2024; Qin et al., 2024; Manacorda & Tesei, 2020; Enikolopov et al., 2020). Their emergence may be responsible for the global increase in the number of protests (Cantoni et al., 2023). The empirical evidence on the role of social media began with the observation that online activity was predictive of off-line mobilization (Acemoglu et al., 2018; Hendel et al., 2017). It was then supplemented by causal evidence that protests spread along pre-existing online social networks (Flückiger & Ludwig, 2025; Manacorda & Tesei, 2020; Enikolopov et al., 2020; Qin et al., 2024).

The apparent univocal evidence that social media contributes to the emergence and diffusion of protests is something of a paradox. On the one hand, their low barriers to entry and reliance on user-generated content allow individuals and grass-roots organizations to bypass the traditional filters of mainstream media and to "self-mediatize" (Castells, 2015; Zhuravskaya et al., 2020). They therefore help potential protesters solve the coordination problem that they face when considering taking to the streets (Enikolopov et al., 2020) and to react faster to repressive forces, reducing the time authorities have to respond (Jost et al., 2018).

On the other hand, social media, just as their name suggests, social media are simply media and are a priori empty pipes that can convey any type of information andprotesters do not have a monopoly on social media usage. Governments can also use social media to influence their citizens and monitor protesters, and repress them. Social media can be used to criticize and hinder protests just as well as to support and organize them. In that respect, whereas Steinert-Threlkeld et al. (2015) observe that an increased use of hashtags related to the coordination of protest preceded the occurrence of protest during the Arab Spring and Qin et al. (2024) document the prevalence of emotional reactions to other protest in China, Enikolopov et al. (2020) report that the information on social media before 2011 was either neutral or positive toward the regime.

Moreover, the content that can be shared on social media has constantly expanded, from the simple text messages studied by Steinert-Threlkeld et al. (2015) and Manacorda & Tesei (2020), to images and videos, which are at the core of networks like Instagram and TikTok and facilitate the exchange of emotional content, which has been found to be powerful motivations to participate in protests (Jasper, 2019; Passarelli & Tabellini, 2017). Yet, whereas the literature has documented the effect of on protest, it has paid little attention to the content shared on those media, thereby overlooking the emotional content of posts and the use of those media against protest. It is therefore key to determine how the content shared on social media affects the spread of protest. This is precisely what this paper does.

More precisely, we study how the content of posts and comments posted on Instagram affected the geographic diffusion in mainland France of the riots that erupted in the early summer of 2023. These riots erupted from June 27 to 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 that would become known as the "Nahel riots" spread across France after a couple of days, affecting 311 municipalities, including small ones that had never experienced a riot before. To study how the riots spread across municipalities, we collected data on the date and location of the 537 riots that took place following the death of Nahel Merzouk. We 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) and the age group to which participants in the riots belonged (Bronner, 2023). To measure the strength of online connections between municipalities, we measure the *a priori* connectedness between two municipalities by the number of comments made by users in a municipality on posts authored by users in the other municipality. We then define the daily exposure of a municipality to the riots through Instagram as the number of riots in connected municipalities weighted by the connectedness of those municipalities. By doing so, we can estimate a panel model where the dependent variable is a dummy set to one if the municipality experienced a riot on a day and the explanatory variable our measure of exposure, controlling for municipality and day fixed effects.

Chiefly, we distinguish the types of content to which municipalities are exposed. In particular, we distinguish content that supports the riots from content that supports the reaction of the police. We also code the level of emotionality of the content and study its effect. We do it for textual content and, most of all, for the image content. This is important as Instagram is image-based and therefore more likely to elicit emotional responses (Casas & Williams, 2019).

We find that exposure of a municipality to riots in other municipalities through Instagram on a given day is predictive of riots in a municipality in the two following days and that the bulk of the effect materializes on the first day that follows exposure, thereby confirming previous evidence (Acemoglu et al., 2018; Steinert-Threlkeld et al., 2015; Qin et al., 2024). The effect is robust and stable across specifications. By weighting online comments by geographic distance to the municipality where the commented picture originates, we show that online proximity goes beyond geographic proximity. We report evidence that the effect is stronger after past exposure to protests in a municipality.

The Nahel riots have several features that make them particularly instructive. The first is that the trigger of the riots is well-identified, the shooting of Nahel Merzouk on June 27, 2023. The second is that, although the riots initially took place in municipalities around Paris, they started affecting more distant and smaller municipalities a couple of days later, before spreading over mainland France. 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).

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 study how social media contributed to spreading riots over an entire sequence of events around what Cantoni et al. (2023) refer to as a sustained protest.

Finally, the Nahel riots share several similarities with riots that took place in other countries, such as the Black Lives Matter movement in the US. Specifically, whereas the police initially claimed that an officer shot Nahel Merzouk in self-defense, two videos showing that Nahel Merzouk had been shot at point-blank went viral on social media, sparking the wave of riots. Also, both Nahel Merzouk and George Floyd belonged to groups that suffer from prejudice, George Floyd being a black American and Nahel Merzouk a French-Algerian. Accordingly, their death sparked suspicion of racism and outrage.

The paper contributes to the literature in several ways. Firstly, it contributes to the literature on the impact of social media on protest by studying riots and social media over an entire sequence of events (Acemoglu et al., 2018; Hendel et al., 2017; Flückiger & Ludwig, 2025; Manacorda & Tesei, 2020; Enikolopov et al., 2020; Qin et al., 2024), as Cantoni et al. (2023) call for, and confirming that social media contribute to spreading protest. Our main contribution to that literature, however, is to show that the effect of social media depends on the content shared. Specifically, content supporting the Nahel riots fanned other riots whereas content opposing the riots or supporting the police reduced their incidence.Moreover, the paper shows the form of the content shared matters. Specifically, it shows that images are particularly effective at prompting reactions both online and offline. In that respect, it contributes to the literature on images (Caprini, 2023). When distinguishing content according to its emotionality, we confirm the intuition that emotions encourage protest (Jasper, 2019; Passarelli & Tabellini, 2017). However, we go beyond that intuition by distinguishing types of emotions. Whereas most emotions encourage online engagement and offline participation in riots, fear acts as a deterrent on offline riots. Finally, in line with the rational model of mobilization, we confirm that content informing participants of the costs and benefits of riots affect the occurrence of riots in the expected direction. Those findings show that rational reasoning coexists with emotional reactions.

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 shows that the impact of social media depends on the type of content shared.. 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. Beginning on June 28, however, the riots spread to other communities, including smaller ones that had never experienced riots before (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).

The majority of protesters were young, with a median age lower than 20 years old.[1] In this age group, the most used social media platform in France is Instagram, an image-centered social media platform that enables users to post, comment, and share pictures and videos publicly or privately (Frommer, 2010) and 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)). [2] We therefore focus on this social media, as it is the most relevant for our analysis.

3 Data

To perform our analysis, we need data on protest, a baseline network of pre-existing digital ties between municipalities, and a series of classifications of protest-related content during the riots. We briefly describe our data sources and core variables below; full details are provided in Appendix C.

Protest Data. Our main protest measure comes from ACLED, which reports geo-referenced protest events with a short description, a list of different actor involved and, for a subsample, the estimated number of participants . We use the description of the events and a large language model (Gemini 2.5 Pro; prompt in Appendix) to classify whether protests support Nahel Merzouk's family or oppose police violence. This yields 537 events across 305 municipalities. As a robustness check, we use data compiled by *Le Monde* (495 events, 311 municipalities), with similar results. Protest activity follows the typical wave pattern observed by Bonnasse-Gahot et al. (2018), peaking on June 29.

[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).

[2] Instagram because it is the most used social media platform in France among people between 16 and 25 years (Diplomeo, 2023)

Figure 1 displays a map with the number of protest events recorded in each municipality across France over the whole wave of protests. To illustrate the evolution of protest activity over time, Appendix Figure 3 provides separate maps for each day of the protest wave, showing the municipalities with protest events on that specific date.

Ex Ante Instagram Network. To identify the causal impact of social media exposure on protest participation, we construct a measure of inter-municipality connectedness *prior* to the onset of the riots. This ex ante network proxies for social pathways along which protest-related content could later diffuse, while remaining exogenous to protest outcomes.

The network is constructed as follows:

1. We begin with the 49 players of the French men's national football team as of June 2023.[3]

2. For each player, we collect all Instagram posts made in the two months before June 27, and scrape up to 50 public comments per post (API-limited). This defines Level 0 interactions.

3. We then recursively gather (i) all posts made by Level 0 commenters within the same time frame – the two months before the death of Nahel –, (ii) all commenters of those posts (Level 1), and repeat the procedure twice more to Level 3. Football players are excluded from subsequent levels and from the construction of the final network.

4. We geolocate all non-celebrity users based on geo-tags, profile bios, recent posts, and ZIP code mentions (see below). When multiple municipalities are referenced, we assign the most frequently mentioned location or default to the largest by population. This procedure geolocates approximately 17.5% of all protest-related Instagram activity

5. We construct a directed network matrix where the exposure from municipality j to i, denoted $\text{ExAnteConnections}_{j \to i}$, is defined as the number of comments made by users in j on posts authored by users in i. Self-links ($i = j$) are excluded. As the numbers of users, posts, and comments in the two municipalities differ, this yields an asymmetric, weighted adjacency matrix capturing directional exposure.

This pre-protest network captures the online attention structure among users plausibly connected to protest behavior, but without reference to the riots themselves. Figure 4 display the spatial intensity of post activity. This measure forms the basis for our variation in subsequent diffusion analysis.

Instagram Protest Content. We collect 1,150 most-liked posts and their comments during this period containing the hashtags `#Nahel` or `#Nael`. Together, these posts generated approximately 2.1 million likes and 80,000 comments, representing a substantial share of total engagement linked

[3]In a robustness check, we go one step further in ensuring political neutrality and exclude four players who posted about Nahel or police violence *a posteriori* finding similar results.

to the protests. For each post, we also record the number of followers of the posting account, which allows us to explore the role of influencer status and potential amplification dynamics. We geolocate both posters and commenters, obtaining approximately 1,000 posts and 961 comment flows with identified origin and destination municipalities. Comment activity spikes on June 29 and mirrors the protest timeline. Figure 2 shows the geographic distribution of comments per municipality of the whole wave of protests. These data serve to characterize the protest discourse and study explicitly the role of social media content on protest diffusion.

To characterize the nature and tone of online protest discourse, we apply a content annotation pipeline using GPT-4o mini (ChatGPT API). Each Instagram post—comprising both textual and visual content—is classified using a detailed prompt constructed according to OpenAI's best-practice guidelines. The prompt first elicits a short description and up to five descriptive keywords. It then asks the model to assess the level of relevance of the post to Nahel riots. [4] We filter out the non relevant posts (around 5%).

Each post is further categorized by its **stance** toward (i) the protests and (ii) the police, using a three-way classification: pro, anti, or neutral/no mention.

Next, we assess **emotional tone** via a subjective "overall emotionality" score (none/low/medium/high), designed to proxy the post's likely affective salience for viewers. We also code the communicative role of the image in the post (main focus, illustration or background or no image present), distinguishing between visually central and incidental content.

Finally, we extract indicators for several **mobilization-related features**. First, we code for the presence of a *call to action*, identifying whether a post urges collective or individual engagement (e.g., "go protest," "contact politicians"); responses are categorized as *main focus*, *present*, or *not present*. Second, we assess whether the post refers to a *specific action*, such as the mention of a protest at a particular time and location or specific advice for mobilization, again using the same three-level coding. Third, we annotate whether the post describes or visually depicts *repression*—that is, police presence, intervention, or violent confrontations with protesters.

For each of these dimensions, the model is asked to provide a brief free-text justification to ensure internal consistency and facilitate validation. Full details and prompt examples, as well as examples for all categories of posts are provided in Appendix C.

Overall, we can construct a municipality-by-day panel consisting of 244,615 observations from June 27 to July 3, 2023, which spans the wave of civil unrest following Nahel Merzouk's death.

Descriptive Statistics Table 9 presents summary statistics for the main variables used in the analysis across municipality-day, municipality-pair, and post levels. At the municipality-day level, protest activity is recorded in approximately 0.2% of observations. Municipalities are associated with an average of 0.003 Nahel-related Instagram posts and 0.007 comments per day. At the municipality-pair level, the average number of observed Instagram connections is 0.00003. This Instagram activity is based on a sub-sample of geo-localized posts and likely captures only a small

[4] Users often post their content using a hashtag that is trending to get more views and engagement despite the content of their post being about something totally orthogonal.

fraction of the total volume of Nahel-related content on the platform and reflects the large number of possible municipality pairs. The average distance between municipalities is 455km.[5]

The content of Instagram posts reveals strong thematic patterns. Among posts talking about Nahel, 69% express clear support, while only 8% express positive sentiment toward the police. Nearly 89% of pro-movement posts place visual content at the forefront, suggesting an emphasis on images over text as a primary vehicle of communication. Emotional content is also prevalent: 66% of pro-movement posts express high emotional intensity, with *anger* (62%) and *sadness* (62%) emerging as the dominant emotions. *Fear* appear in 28% of posts. Other emotions—such as disgust, and surprise—appear less frequently, and no post expresses happiness, consistent with the tone of civil unrest.

Calls to action are present in 48% of pro-movement posts, and about 10% specify concrete instructions, such as locations or times for mobilization (around 21% of posts calling for action call for a specific action). Around 58% of posts reference repression, including violence (19%) and clashes with police (12%).

4 Empirical Strategy

Before introducing the specification of the relation between exposure to protest through Instagram and the propensity to host a protest in our panel of municipalities that we estimate, we first build a measure of exposure to previous protests that depends on online connections and on online connections only.

4.1 A measure of Exposure to protest on social media

To obtain a measure of exposure on day t of municipality i to protest through Instagram, we interact the measure of connectedness of municipality i to each other municipality j at baseline with a dummy variable equal to one if j had a protest on the same day and then sum those interactions over j. In line with Qin et al. (2024), our measure of exposure of municipality i to protest content through Instagram therefore reads:

$$\text{Exposure}_{i,d} = \sum_{i \neq j} (\text{Connections}_{i \to j} \times \text{Protest}_{j,\ d-1})$$

where $\text{Connections}_{i \to j}$ is a fucntion of the number of comments written in municipality i on posts created in municipality j before the death of Nahel and $\text{Protest}_{j,d-1}$ is a dummy variable that turns to one if municipality j hosted a protest on day *d-1*. The rationale for our measure is that users living in municipality i are more likely to learn about a protest that took place in municipality j if

[5] Those descriptive statistics must be interpreted in light of the large number of French municipalities in France. France had 34,955 municipalities in 2022. Moreover, their size distribution is skewed, as 93.7% of them had fewer than 5,000 inhabitants (Direction générale des collectivités locales, 2022). This explains why the number of connections that we capture with our measure is small. In light of the dense network of municipalities, the average distance between the municipalities in our sample is large.

users of municipality i were already connected to municipality j before the onset of the Nahel riots. Accordingly, that measure of exposure to protest is an increasing function of the predetermined connectedness between municipality i to posts created in other municipalities. However, for users in the first municipality to be exposed to protest in other municipalities, the latter must have experienced a protest in the first place. Note that this measure of exposure is directional and non-symmetrical as people from municipality i can comment a lot of post of municipality j while municipality j do not comment posts of municipality i.

We assume that it takes one day for protests to affect exposure. There are two reasons for this. The first is that it takes a little time for a topic to gain momentum. The second is that the Nahel riots typically occurred in the evening or at night, leaving little time for users from another municipality to observe them and react on the same day.[6]

The key feature of the measure of exposure is that interactions between municipalities are measured at baseline, *before* the death of Nahel Merzouk. It is therefore immune to reverse causality that could be due to the fact that the footballers on whom we seeded the network could have attracted followers because of their reaction to Nahel Merzouk's death.

This measure relies on two assumptions. The first is that users in municipalities with a protest react online to the protest on the day of the protest. If they did not, then the exposure to protest of users connected to them in other municipalities would be unaffected and our measure would not capture the relevant variation. The second assumption is that if i is exposed to j at baseline and j had a protest, then users in i will be exposed to the protest-related content of j. In other words, this assumption is that our ex-ante measure of connectedness captures how information travels, including information about protest.

4.2 Isolating online connections

While it leverages rich data on digital interactions, a key challenge of our measure of *ex ante* online connectedness is that it may reflect offline connectedness. In particular, if municipalities with strong Instagram ties are also connected through other channels—such as transportation networks, direct phone calls or private messaging platforms —then our measure of online connected may capture broader ties than ties on social media *per se*. It would therefore prompt us to overestimate the role of social media in the diffusion of riots.

To address this challenge, we use a two-step strategy to isolate the component of Instagram connectivity that is plausibly exogenous to such confounds and net out offline connections. In the first step, we apply to the number of Instagram connections between pairs of municipalities a structural framework inspired by the gravity model commonly applied in the analysis of trade flows, migration patterns, and information diffusion. These models posit that bilateral flows (e.g., trade or migration) between two locations increase with the "mass" (e.g., population) of each location and decrease with geographic distance between them.

In our context, we estimate the following specification:

[6] We provide a series of tests showing that this lag is justified.

$$\frac{\text{Comments}_{i\to j}}{\text{Pop}_i \cdot \text{Pop}_j} = \alpha_0 + \alpha_1 \cdot \frac{1}{\text{Distance}_{ij}} + \nu_i + \delta_j + \varepsilon_{ij}$$

Rather than model the relationship in logarithmic form—which would exclude observations with zero connections—we normalize the connection count by the product of the origin and destination populations, and enter distance in its inverse form to reflect spatial frictions. With this functional form, the marginal effects of distance on social interaction is decreasing. We hypothesize that proximity facilitates interactions most strongly at shorter ranges, with the marginal effect of each additional kilometer declining thereafter. This approach preserves all observations while capturing the notion that proximity facilitates interactions. We also include origin and destination fixed effects to flexibly control for municipality-specific propensities to send or receive digital interactions, thus implicitly absorbing the linear effect of population size in the regression.

In the second step, we retrieve the residuals $\hat{\varepsilon}_{ij}$ from the model, which represent the component of Instagram connections that are exogenous to spatial distance and municipality characteristics and can therefore be interpreted as the part of connections that is specifically online. To return these residuals to the level of actual (non-normalized) connections, we multiply them by the product of the corresponding populations. Finally, we plug that measure of connections in the expression of exposure to protest to construct our residualized exposure variable as:

$$\text{Exposure}_{i,d} = \sum_{i\neq j} \left(\hat{\varepsilon}_{ij} \cdot \text{Pop}_i \cdot \text{Pop}_j\right) \times \text{Protests}_{j,\ d-1}$$

This procedure recovers the component of bilateral Instagram activity that is by construction orthogonal to geographic proximity and population size, and uses it to build a plausibly exogenous measure of exposure to protest content through social media. It allows us to identify the specific causal role of social media connectivity in protest diffusion, net of the spatial and demographic forces that typically structure social ties. We will use this measure for the rest of the paper.

4.3 Baseline specification

To estimate the impact of the exposure to protest of municipality i on the probability that it experiences a protest, we estimate the following two-way fixed effect linear probability model:

$$\begin{aligned}\text{Protests}_{i,d} = {} & \beta_0 + \beta_1 Exposure_{i,d} \\ & + \beta_2 \text{Cumulative protest}_{j,d=0-d-1} \\ & + \eta_i + \delta_d + \varepsilon_{i,d}\end{aligned}$$

where $\text{Protest}_{i,d}$ is set to one if municipality i hosted a protest on day d; $\text{Exposure}_{i,d}$ is the measure of exposure to protest of municipality i defined above; $\text{Cumulative protest}_{j,d=0-d-1}$ is the sum of protest experienced in municipality i since the beginning of the Nahel protest. η_i is a municipality fixed effect, δ_d a day fixed effect, and $\epsilon_{i,j,d}$ the error term. We choose a linear probability model because logit and probit models tend to underestimate the probability of rare events (King & Zeng, 2001). Finally, we estimate the model with ordinary least squares and cluster standard errors at the municipality level to account for the fact that observations of the same municipality on different days are not independent.

The coefficient of interest is β_1, the coefficient of the measure of exposure of municipality i to protest in other municipalities.

The municipality fixed effect controls for constant municipality charateristics that may prompt a municipality to be more likely to host a protest on any day. It captures all the characteristics of the municipality that were constant over the wave of protest, for instance the municipality's political preferences and its demographic, economic, or socioeconomic characteristics.

A key difficulty when studying the diffusion of protest is the reverse Galton problem that would prompt us to attribute to diffusion what may be the result of unit attributes that are spatially clustered (Buhaug & Gleditsch, 2008; Aidt et al., 2022). However, as our period of analysis is limited to six days, most characteristics can be considered as constant. Fundamental characteristics are therefore controlled for by the sets of municipality fixed effects.

By the same token, day fixed effects control for the temporal shocks that may affect all municipalities on the same day. For example, the average effect of a speech of a nationally known politician or the coverage of the protests by national media that can affect the probability of comment and the probability of protest are captured by the day-fixed effects. The two sets of fixed effects therefore capture the unobserved heterogeneity that may have caused a reverse Galton problem or more generally an omitted variable bias. The only variable that changes at the municipality-day level is the exposure to protest in other municipalities.[7]

5 The diffusion of riots along Instagram networks

5.1 Baseline findings on diffusion

This section presents the main empirical evidence on how social media connectivity contributed to the geographic spread of Nahel protests. Table 1 presents the results. All reported coefficients are scaled by a factor of one million for readability.

Columns (1) through (4) report estimates where the dependent variable is the daily number of riots in a municipality. Column (1) presents our baseline specification with municipality and day fixed effects as well as controls for past cumulative protest. The coefficient on lagged exposure—i.e., exposure to protest activity on day $d-1$ via digital ties— is statistically significant. Given a mean of 0.0022 protests per day, this implies that a one-unit increase in exposure corresponds to

[7] We also report the results for alternative combinations of fixed effects structures, including region-day fixed effects. This allows temporal shocks to be different by region and captures temporal shocks that affect all municipalities within one region. For example, the effect of a speech of a local politician or local weather conditions common to the region that affected the likelihood of protest will be captured by region-day fixed effects. Finally, we also 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.

approximately 6% of the daily average—a meaningful magnitude in the context of a rare event given the large number of municipalities in France.[8]

To assess the robustness of our findings to alternative confounding structures, Columns (2) and (3) introduce more saturated fixed effects. Column (2) includes region-by-day fixed effects, while Column (3) introduces department-by-day fixed effects, thereby accounting for region- or department-specific shocks that vary over time—such as local media narratives, coordinated offline mobilization, or weather patterns.

Across these richer specifications, the estimated coefficients remain stable (130.2 and 146.1, respectively) and significant at the 1% level. This consistency supports the identification strategy, suggesting that the estimated exposure effect is not driven by spatially clustered, time-varying unobservables.

Column (4) incorporates lag and lead terms to examine the timing of the exposure effect. The results reveal a sharp temporal pattern: the effect is concentrated at lag $d-1$ (coefficient: 132.7; p < 0.01), with no statistically significant effects at lags $d-2$, $d-3$, or at the contemporaneous day d. This aligns with the qualitative timeline of events: protests typically occurred in the evening or night, while social media reactions unfolded with a one-day delay. The absence of a same-day effect reinforces our empirical design, which assumes that the realization of protest in one municipality influences digital exposure elsewhere only with a lag, and not instantaneously and diminishes concerns about possible reverse causality or presence of anticipatory effects. This time structure is particularly striking because it resembles the one reported by Qin et al. (2024). It therefore underlines that although protest waves may unfold gradually in some contexts (Cantoni et al., 2023), they can spread fast elsewhere, especially when protests are triggered by perceptions of injustice rather than strategic considerations (Passarelli & Tabellini, 2017). In our case, the platform we analyze—Instagram— is image-based, and visual content is known to elicit stronger and more immediate emotional responses than text (Casas & Williams, 2019). [9]

In Column (5), we examine the extensive margin of protest diffusion by using a binary dependent variable that equals one if a municipality experienced any protest on a given day. The estimated effect remains statistically significant (coefficient: 76.3; $p < 0.05$), confirming that digital exposure not only increases protest frequency but also triggers new protest onset.

Taken together, these results provide compelling evidence that protest behavior propagated along pre-existing digital ties. The social media induced diffusion occurs fast and is not explained by concurrent regional shocks or spatial confounding.

[8] It is important to keep in mind that overall just over 300 municipalities hosted a riot during the period of analysis over almost 35 000 total municipalities in France.

[9] Figure 5 gives extra insights in the time dimension of the diffusion of riots along social media. It reports the estimated marginal effect of exposure to riots when we allow it to change every day. The main finding is that those marginal effects were positive and statistically significant over the whole wave of riots. It also provides suggestive evidence that the marginal effect of online exposure followed a U-shaped evolution and was accordingly larger at the beginning and the end of the wave. However, that comparison must be taken with care as point estimates are not statistically distinguishable.

5.2 Robustness checks

To assess the robustness of our results and rule out alternative explanations, we perform a battery of complementary checks, which we detail in Appendix B. These tests examine the sensitivity of our findings to data choices, specification changes, placebo conditions, and alternative definitions of treatment and exposure. In particular, we replicate our analysis using a different data source for protest events (daily newspaper *Le Monde* instead of ACLED), account for more granular time and space heterogeneity by allowing day fixed effects to vary by region and department, and include municipality-level linear trends to absorb location-specific dynamics over time.

We also conduct a set of placebo and alternative treatment tests. These include verifying that future protests (at $d+1$) do not predict current-day activity, and ensuring that results are not driven by municipalities like Paris or Nanterre. To rule out endogeneity in our network of exposure, we exclude football players who publicly commented on Nahel Merzouk. We also examine robustness within a richer dyadic setting using municipality-pair fixed effects. Across all variations, our results remain stable and statistically significant, reinforcing confidence in our identification strategy and main findings.

Finally, we give a first look at a channel of transmission from exposure to protest to protest by using posts about Nahel Merzouk rather than protests as the treatment variable. The outcome of that regression shows a positive and statistically significant effect of exposure of a municipality to comments in other municipalities on the propensity of the exposed municipality to host a protest. This provide additional evidence that the mediazation of protest on social media contributed to spreading the Nahel riots.

6 The role of content

While prior sections established that protest activity spread through pre-existing Instagram networks, they treated exposure in the same way regardless of the content shared between two municipalities. Yet, not all protest-related content is likely to have the same mobilizing effect. Some content may inspire people to participate or demoralize them whereas other types of content more directly informs on the costs and benefits of participation (Aidt et al., 2022). Emotional tone and expressions of support to the riots fall in the first category. Calls to action, or depictions of repression fall in the second. Both may each influence how users respond—both online and offline.

To investigate how content affects the effect of exposure, we first classify the posts to which a municipality was exposed in turn along multiple annotated dimensions, specifically political slant, the type of image, emotionality, and information on the costs and benefits of mobilization. We then re-specify our empirical framework to interact the same residualized measure of Instagram connectivity with the number of protest-related posts of each type about Nahel created on day d, rather than protest events themselves. This allows us to isolate how variation in content—rather than simply protest occurrence—drives comment activity and protest incidence.

6.1 Positions on the riots and the police

The Nahel riots generated a lot of supportive comments online, but they also generated a backlash. Likewise, the reaction of the police caused both support and opposition. As comments supporting or opposing the Nahel riots or the police likely have different effects, we begin our analysis of content heterogeneity by examining the stance of a post toward the riots or the police—affects both online engagement and offline mobilization.

Using our annotated corpus of Instagram posts, we classify each post as either supportive or not supportive of the riots, and separately, as either supportive or not supportive of the police. Figures 6 and 6 illustrate how the distribution of post stances evolved over the wave of protests. In the early phase of the riots, most posts were supportive of the movement, but the share of neutral or opposing content increased slightly toward the end of the period. A similar dynamic can be seen in views toward the police: initially, anti-police sentiment dominated, but was gradually replaced by neutral and, to a lesser extent, pro-police content.

We then interact these content types with pre-existing Instagram connectivity to assess how online support for the riots or the police, or the lack thereof, conditioned the effect of posts on the exposed municipality.

Columns (1)–(3) of Table 2 report the effect of exposure to the different types of content on comment activity. Column (1) shows that overall exposure to Nahel-related posts significantly increases the number of comments received in exposed municipalities, indicating that content spreads actively along the network.

In Column (2)-(3), we disaggregate exposure by stance. Exposure to pro-movement content leads to a large and statistically significant increase in comment activity (column 2), while non–pro-movement content (i.e., neutral or negative toward the movement) triggers significantly fewer comments. This asymmetric response suggests that content supportive of the movement generates stronger digital engagement. Similarly, in Column (3), we observe that exposure to pro-police content is associated with a sharp drop in comment activity, whereas exposure to other, i.e., non–pro-police, content increases it. In all cases the coefficients are statistically significant at the one-percent level.

Columns (4)–(6) turn to the main outcome of interest: protest activity. Column (4) confirms that total exposure to Nahel-related posts significantly increases the number of protests. Column (5) shows that exposure to pro-movement content significantly increases protest activity, while exposure to non–pro-movement content reduces it. In Column (6), the effect of pro-police content on the probability to host a riot is negative and significant, while other content leads to a higher probability of having a riot. The pattern mirrors the engagement results in Columns (2) and (3), and underscores that the tone of content—not just its existence—shapes the diffusion of riots.

6.2 Images

Visual content is known to carry affective power, reduce cognitive load, and drive disproportionate attention—especially in digital environments. Images may act as interpretive frames, condensing complex narratives into emotionally legible signals. Accordingly, visually salient news images systematically bias readers' beliefs and behavioral intentions in politically charged settings(Caprini, 2023). In the context of protests, images can vividly convey emotional cues, group identity, and urgency in ways that text alone may not.

To isolate the effect of visual prominence, we restrict the analysis to posts supportive of the movement, and compare posts in which the image was classified as the main communicative focus with those where the image was not. All regressions control for exposure to non–pro-movement content to net out the baseline ideological contrast. Table 3 presents the results.

Column (1) shows that both types of pro-movement posts—those where the image is the focal point and those where it is not—are strongly associated with increased comment activity in connected municipalities. Their effect is statistically significant at the one-percent level. Column (2) turns to the main outcome of interest: protest activity. Here, only pro-movement posts with a prominent visual focus are significantly associated with an increase in riots, while posts without prominent images have a statistically insignificant effect.

Our results suggest that pro-riot posts with visually prominent content are particularly effective in spurring riots, while not necessarily disproportionally increasing online engagement in the form of comments. The first result is in line with the evidence that visual cues can slant the perception of audiences (Caprini, 2023).

6.3 Emotions

Emotions are powerful motivations to participate in protests (Jasper, 2019; Passarelli & Tabellini, 2017). To gauge their effect on the spread of protest, we build on Ekman and Friesen's (1971) framework of six basic emotions—anger, disgust, fear, happiness, sadness, and surprise—and investigate whether the emotional content embedded in pro-movement posts conditions their ability to propagate through digital networks and catalyze offline mobilization.[10]

Table 4 presents the results separately for digital engagement (Panel A) and protest activity (Panel B), across the full set of emotions. We begin by examining whether posts annotated as highly emotional—regardless of the specific emotion— are more likely to generate engagement or riots than more neutral posts (column 1). We then disaggregate by individual emotion types to assess whether distinct emotions have differential effects (columns 2-7). Our regressions include three mutually exclusive exposure variables: (i) pro-movement posts with a focal emotion, (ii) pro-movement posts with no high emotion or a different emotion, and (iii) non–pro-movement posts. As such, the coefficients capture the marginal effect of exposure to each category, holding exposure to the others constant.

Panel A of Table 4 shows that exposure to pro-movement posts with high emotional intensity is strongly associated with increased comment activity. This holds across several individual emotions: anger, sadness, and disgust all significantly boost online engagement compared to municipalities with less or no exposure to those specific emotional tones. Somewhat surprisingly, pro-movement

[10]Given the nature of the unrest, no posts in our sample were classified as triggering happiness.

posts that lack high emotion or express another emotion are also consistently associated with high levels of engagement, and in most cases, their coefficients are larger than those for the focal emotion category. Pro movement posts that spark anger are less effective in creating online engagement than pro-movement posts not sparking anger (coefficient 36.4 vs. 17.6, and statistically distinguishable), suggesting that less intensely emotional content may sometimes generate more online interaction. Across specifications, exposure to posts not in favor of the movement decreases online engagement.

Panel B of Table 4 explores whether emotional content translates into offline mobilization. Column 1 shows that exposure to pro-movement posts with overall highly emotional content increases protest incidence relative to municipalities with less or no exposure to that content. The effect is statistically significant at the one-percent level. Exposure to pro-protest content less emotionally charged is also positively associated to protest, at the five-percent level, at a potentially lower magnitude but the coefficients are not statistically significant different from each other.

When disaggregating by emotion, exposure to posts that lack anger, disgust, or surprise appears more strongly associated with protest than exposure to posts that express those emotions. In contrast, exposure to post with sad content are effective in inciting protest. Across specifications exposure to content that is not in favor of the movement is associated with decreased protest activity. The case of fear is especially instructive: despite some elevated comment activity, exposure to fear-inducing content is associated with a large and negative effect on protest occurrence. It therefore seems that fear deters participation in riots, possibly by increasing the perception of the cost of participation or paralyzing potential participants (Jasper, 2019).

Together, these results suggest that determinants of online engagement can be very different from the ones of offline action. While exposure to less emotionally charged posts is generally more effective in creating online engagement to posts about Nahel, this is not true for offline activities. The effect of exposure to emotionally charged content on protest activity depends on the type of emotion. Fear may backfire by raising perceived costs of participation, while sadness may promote solidarity and shared grievance.

This cautions against the view that emotional intensity unequivocally drives engagement and mobilization. Instead, our findings show that some emotions encourage mobilization, others, fear in particular, deter it.

6.4 The usual suspects: Coordination and Repression

Rational agents will weigh the costs and benefits of mobilization. By coordinating to take to the streets together, they can lower the cost of participation (Passarelli & Tabellini, 2017). Accordingly, successful protest requires collective action—individuals must believe others will join and know when, where, and how to act. Conversely, repression raises the perceived costs of participation, potentially deterring mobilization, although it may also provoke backlash and fuel further dissent. Social media platforms mediate both processes—facilitating the spread of logistical information and visibility of protest, while also amplifying images of repression and violence. In what follows, we examine how exposure to content reflecting these two forces—mobilizing appeals and depictions of repression—differentially shapes online engagement and protest diffusion.

6.4.1 Coordination Cues

A central insight from the literature on collective action is that successful protest hinges not only on individual grievances but on the ability to coordinate behavior among participants: individuals are more likely to join when they believe others will do the same and when they have clear information about when and where to mobilize (Cantoni et al., 2019). Social media platforms are thought to lower coordination costs by enabling the rapid dissemination of mobilizing messages, visibility of shared intent, and logistical details (Acemoglu et al., 2018; Enikolopov et al., 2020; Tufekci, 2017; Steinert-Threlkeld, 2017). We explore this mechanism by explicitly analyzing the content of the posts to which potential protesters are exposed to, again focusing on posts that supported the Nahel riots. We categorize pro-movement posts into those that contain calls to action, those that specify concrete protest guidance (i.e., mention specific places or times to reunite or give practical advice on how to protest), and those that remain supportive but do not invite participation. We hypothesis that posts that do not merely express support but also urge action or provide concrete details—such as time, place, or tactics—can facilitate protest participation by signaling mutual intent, resolving strategic uncertainty, and lowering the informational threshold for joining.

Table 5 presents results for both online engagement (columns 1–2) and protest activity (columns 3–4). All categories are mutually exclusive and exhaustive among our sample of posts referring to Nahel Merzouk, meaning coefficients reflect the marginal association of exposure to each type of post, holding constant exposure to the others.

In columns 1-2, we find that exposure to pro-movement posts with calls to action is associated with significantly increased comment activity. However, the effect is comparable to or smaller than that of posts without any call to action, which also yield strong engagement. When dis-aggregating posts containing a call for action from those containing specific information on tactics, places or times, we see again that exposure to both posts containing non-specific call for action and no call for action at all significantly increase the number of protests (column 2). Exposure to posts providing specific instructions—such as protest locations or times—shows a positive but statistically imprecise effect. Calls to action may carry more strategic than expressive value and might thus not elicit immediate commenting behavior.

Columns 3-4 confirm this and reveals a clearer pattern: coordination cues strongly predict protest mobilization. Exposure to pro-movement posts containing any type of call to action is more effective in increasing protests than exposure to post without a call (column 3). Posts offering specific action guidance—such as protest time and place—are particularly effective, with a coefficient of 30.89 (statistically different from posts with a non-specific call, no call at all or posts not supporting the movement), the largest observed across all categories. Despite being less effective, non-specific calls also significantly increase protest activity while posts with no mobilizing message are associated with lower protest rates, and even negative effects when isolated (column 4). These findings are consistent with the coordination hypothesis: content that lowers the perceived cost of participation facilitates protest diffusion, even if it does not provoke more online discourse. This distinction between digital engagement and real-world action highlights that mobilizing content operates through clarity and logistical relevance, not just salience.

Overall, these results support the view that social media enables protest not merely by spreading ideas, but by offering coordination infrastructure—especially when messages include direct appeals or actionable details. While symbolic support may encourage online debates, calls to action are what mobilize.

6.4.2 Repression, Violence, and the Discouragement of Protest

On the one hand, repression may produce fear-based deterrence (Davenport, 2007), reduce turnout by increasing perceived personal costs (Young, 2019), or signal the high price of dissent. On the other hand, highly visible repression can sometimes generate outrage-based mobilization (Francisco, 1996), particularly when perceived as illegitimate or disproportionate (King et al., 2013). Whether protest grows or shrinks in the face of repression is therefore ambiguous. Here, we examine whether exposure to protest-related Instagram posts depicting repression, clashes between protesters, participants and police officers, or broader violence affects comment activity and protest outcomes.

Columns 1-3 of Table 6 report regressions taking comments as their dependnet variable. Column 1 shows that posts depicting repression and those that do not are both associated with a significant increase in comment activity. Column 2 distinguishes clashes with the police and other types of shown repression and shows that only the former statistically significantly affects comments whereas the latter is statistically insignificant. Posts describing violence more broadly and those not showing violence (column 3) also lead to online interaction.

Columns 4-6 take the occurrence of a riot as dependent variable. Here, posts depicting repression, clashes, or generalized violence are all significantly associated with reduced protest activity. Exposure to content showing clashes with police officers are the most effective in reducing protest activity, suggesting that content portraying protest risks or confrontations may serve a deterring function, discouraging mobilization. This interpretation aligns with findings from Young (2019), who documents that repression during the Ferguson protests in the U.S. discouraged turnout in the most affected areas. In contrast, exposure to pro-movement posts not featuring repression or violence is consistently associated with increased protest incidence across columns.

These findings echo the deterrent effect observed in our analysis of emotional tone, where exposure to fear-related content significantly reduced protest participation. Repression imagery and fear-inducing emotional cues appear to reduce mobilization by amplifying perceived risks and suppressing collective efficacy.

Taken together, these results underscore an important asymmetry: repression-related content attracts attention but suppresses action, consistent with prior work showing that violent or repressive imagery increases visibility and circulation on social media, even when it does not necessarily translate into mobilization (King et al., 2013). While it may amplify digital engagement, it appears to discourage physical protest, likely by shifting perceived costs or signaling increased state vigilance.

7 Conclusion

This paper provides causal evidence that social media—and specifically, an image-based platform like Instagram—play a central role in the geographic diffusion of protest. Studying the 2023 Nahel riots in France, we show that protest behavior spread along pre-existing online connections, with exposure to protest in connected municipalities significantly increasing the likelihood of protest locally. The diffusion is rapid, materializing within a day, and remains robust across specifications including rich fixed effects structures, and passes placebo tests.

Crucially, we move beyond exposure volume and demonstrate that protest diffusion is content-dependent. Posts that support the movement, emotionally resonant, visually salient, or contain calls to action disproportionately drive both online engagement and offline mobilization. Conversely, exposure to repression, violence, or fear-inducing content dampens protest activity, even as it increases digital interaction. This asymmetry underscores the distinction between expressive and mobilizing content, and highlights how emotional and strategic cues interact to shape collective action.

Our findings contribute to a growing literature on the micro-mechanisms of protest contagion, providing novel evidence on how affective and visual elements condition the mobilizing power of social media. They also point to a broader theoretical insight: social media are not intrinsically favorable to protest. Their content can either move people to act or deter them from acting.

These results have important implications for scholars of political communication, protest dynamics, and digital governance. As image-based platforms continue to shape the landscape of contentious politics, understanding how content features translate into real-world outcomes will be vital to anticipating the trajectories of social movements—and the counterforces that aim to deter them.

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.

Aidt, T., Leon-Ablan, G., & Satchell, M. (2022). The social dynamics of collective action: evidence from the diffusion of the swing riots, 1830–1831. *Journal of Politics*, 84(1), 209–225.

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

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*.

Buhaug, H. & Gleditsch, K. S. (2008). Contagion or confusion? why conflicts cluster in space. *International Studies Quarterly*, 52(2), 215–233.

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.

Cantoni, D., Yang, D. Y., Yuchtman, N., & Zhang, Y. J. (2019). Protests as strategic games: experimental evidence from hong kong's antiauthoritarian movement. *The Quarterly Journal of Economics*, 134(2), 1021–1077.

Caprini, G. (2023). Visual bias.

Casanueva-Artís, A., Avetian, V., Sardoschau, S., & Saxena, K. (2024). Late adoption and collective action: Social media expansion and the diffusion of black lives matter.

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*.

Davenport, C. (2007). *State Repression and the Domestic Democratic Peace*. Cambridge University Press.

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

Direction générale des collectivités locales (2022). Les chiffres-clés des collectivités locales.

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

Flückiger, M. & Ludwig, M. (2025). The structure of online social networks and social movements: Evidence from the black lives matter protests. *Journal of Public Economics*, 246, 105373.

Francisco, R. A. (1996). Coercion and protest: An empirical test in two democratic states. *American Journal of Political Science*, (pp. 1179–1204).

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

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.

Hendel, I., Lach, S., & Spiegel, Y. (2017). Consumers' activism: the cottage cheese boycott. *RAND Journal of Economics*, 48(4), 972–1003.

Jasper, J. M. (2019). *The emotions of protest*. University of Chicago Press.

Jost, J. T., Barberá, P., Bonneau, R., Langer, M., Metzger, M., Nagler, J., Sterling, J., & Tucker, J. A. (2018). How social media facilitates political protest: Information, motivation, and social networks. *Political psychology*, 39, 85–118.

King, G., Pan, J., & Roberts, M. E. (2013). How censorship in china allows government criticism but silences collective expression. *American political science Review*, 107(2), 326–343.

King, G. & Zeng, L. (2001). Logistic regression in rare events data. *Political analysis*, 9(2), 137–163.

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

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 (Le) (2023). 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. (2024). Social media and collective action in china. *Econometrica*, 92(6), 1993–2026.

Steinert-Threlkeld, Z. C. (2017). Spontaneous collective action: Peripheral mobilization during the arab spring. *American Political Science Review*, 111(2), 379–403.

Steinert-Threlkeld, Z. C., Mocanu, D., Vespignani, A., & Fowler, J. (2015). Online social networks and offline protest. *EPJ Data Science*, 4(1), 1–9.

Tufekci, Z. (2017). *Twitter and Tear Gas: The Power and Fragility of Networked Protest.* Yale University Press.

Young, L. E. (2019). The psychology of state repression: Fear and dissent decisions in zimbabwe. *American Political Science Review*, 113(1), 140–155.

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

8 Figures and Tables

Figure 1. Protests

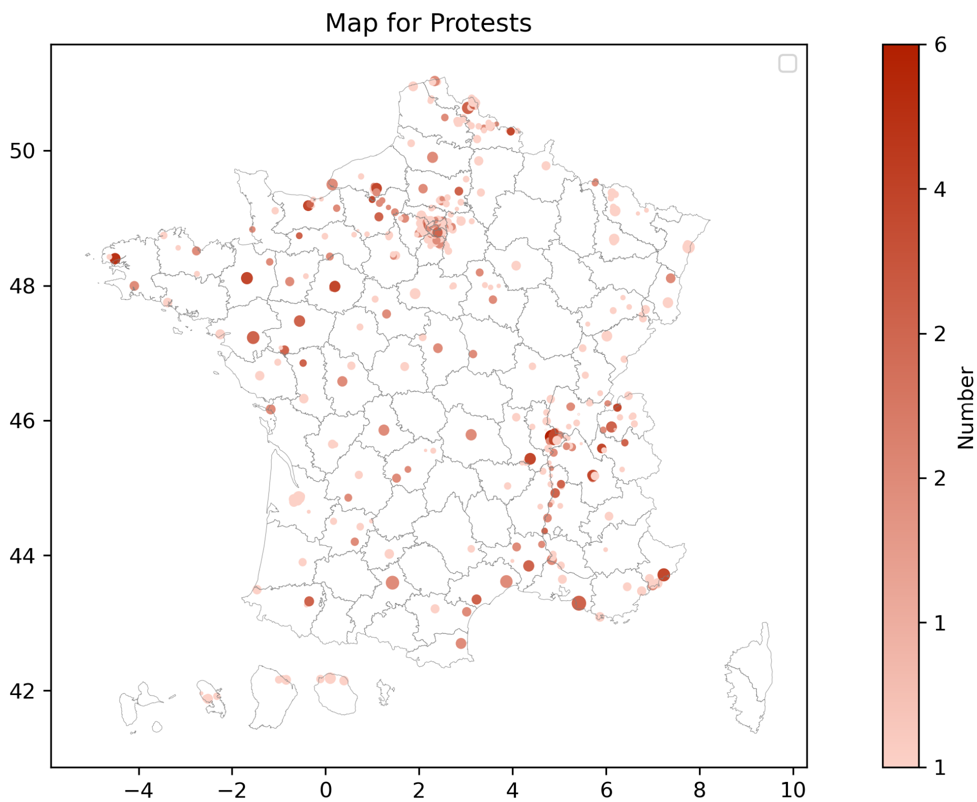


Figure 2. Comments

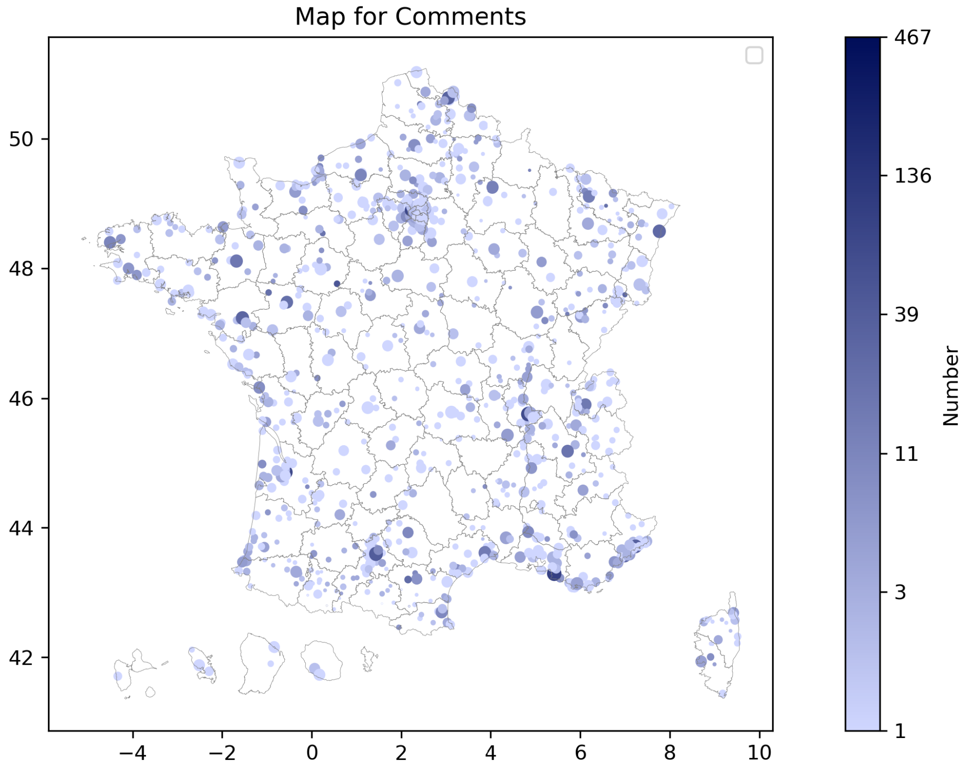


Table 1. **Diffusion of protest through social media.**

	Protests at d				Has protest
	(1)	(2)	(3)	(4)	(5)
Exposure to Protests $d-1$	130.9***	130.2***	146.1***	132.7***	137.5***
	(33.03)	(32.99)	(49.52)	(42.51)	(35.75)
Exposure to Protests $d-2$				0.618	
				(51.63)	
Exposure to Protests $d-3$				10.58	
				(8.402)	
Exposure to Protests d				-13.63	
				(42.94)	
Cumulative protests	Y	Y	Y	Y	Y
Municipality FE	Y	Y	Y	Y	Y
Day FE	Y				Y
Region-day FE		Y			
Department-day FE			Y		
Observations	244,615	244,615	244,608	244,615	244,615
R-squared	0.448	0.450	0.454	0.448	0.435
Mean dep. var.	0.00220	0.00220	0.00217	0.00220	0.00212

Note: Effect of the diffusion of protest through Instagram. OLS estimates with two-way fixed effects. Sample is all municipalities in France over the period 27 June to 3 July 2023. Exposure to protest is the interaction of the residualized measure of Instagram connections with the presence of protest the day before as defined in 4. Columns 1-3 vary the set of fixed effects: column 1 include municipality and day FE, column 2 municipality and region-day FE and column 3 municipality and department-day FE. Our preferred specification is column 1. Column 4 include different lags of the treatment. Column 5 changes the outcome from count of protests to presence of protest and thus estimate a linear probability model. All columns include cumulative past protest as control. Standard errors are clustered at municipality level *p<0.1; **p<0.05; ***p<0.01.

Table 2. **Effect of posts supporting or opposing the Nahel protests and the police**

	Comments on Nahel posts			Protests d		
	Baseline (1)	Movement (2)	Police (3)	Baseline (4)	Movement (5)	Police (6)
Exposure to Posts d	6.264***			3.687***		
	(2.135)			(0.932)		
Exposure to Positive posts d		25.80***	-108.8***		7.189***	-5.435***
		(2.703)	(25.32)		(0.893)	(1.558)
Exposure to Other posts d		-37.14***	11.70***		-4.092***	4.168***
		(3.437)	(2.050)		(1.168)	(0.947)
Cumulative protests	Y	Y	Y	Y	Y	Y
Day FE	Y	Y	Y	Y	Y	Y
Municipality FE	Y	Y	Y	Y	Y	Y
Observations	244,615	244,615	244,615	244,615	244,615	244,615
R-squared	0.675	0.675	0.675	0.446	0.447	0.446
Mean dep. var.	0.00733	0.00733	0.00733	0.00220	0.00220	0.00220

Note: Effect of exposure to social media post content with different support for the Nahel protests and for the police on commenting activity and protest incidence. OLS estimates with two-way fixed effects. Sample includes all municipalities in France over the period 27 June to 3 July 2023. Exposure to posts is defined as the interaction of the residualized measure of Instagram connections with post characteristics on a given day. Columns 1–3 report the effect on the number of comments on Nahel-related posts; columns 4–6 report the effect on protest counts. Column 1 and 4 include overall exposure to Nahel-related posts. Columns 2 and 5 differentiate exposure by content sympathetic to the movement; columns 3 and 6 by content supportive of the police. All regressions control for cumulative protests, municipality and day fixed effects. Standard errors are clustered at the municipality level. *p<0.1; **p<0.05; ***p<0.01.

Table 3. **The role of visual prominence**

	Comments on Nahel posts (1)	Protests d (2)
Exposure to Pro-mov. where Img. main focus	21.06***	7.633***
	(5.177)	(2.122)
Exposure to Pro-mov. where Img. not main focus	46.46***	5.253
	(11.27)	(4.646)
Exposure to Not pro-mov.	-34.11***	-4.376***
	(4.672)	(0.562)
Cumulative protests	Y	Y
Day FE	Y	Y
Municipality FE	Y	Y
Observations	244,615	244,615
R-squared	0.675	0.447
Mean dep. var.	0.00733	0.00220

Note: Effect of exposure to protest-related Instagram content emphasizing visual elements on commenting behavior and protest activity. OLS estimates with two-way fixed effects. Sample covers all municipalities in France during the period 27 June to 3 July 2023. Exposure is defined via the interaction of the residualized Instagram connection metric with the presence of posts categorized by whether they support the movement and the visual prominence of the image, as explained in 4. Column 1 reports effects on the number of comments on Nahel-related posts, and column 2 on the count of protests. Distinction is made between pro-movement posts where the image is the main focus, pro-movement posts with text emphasis, and content not supporting the movement. All specifications include cumulative past protest controls, municipality and day fixed effects. Standard errors are clustered at the municipality level. *p<0.1; **p<0.05; ***p<0.01.

Note: Clustered at municipality level

Table 4. **The role of emotions**

	overall emotionality (1)	anger (2)	disgust (3)	fear (4)	sadness (6)	surprise (7)
Panel A: Comments on Nahel posts						
Exposure to Pro-mov. posts with high emotion	19.18***	17.64***	391.5***	400.3	19.39***	-3.805e+06
	(4.106)	(3.341)	(83.33)	(517.3)	(3.156)	(3.230e+06)
Exposure to Pro-mov. posts with no high emotion or other emotions	40.30***	36.42***	16.76***	2.786	43.96***	25.96***
	(4.012)	(3.600)	(4.571)	(30.49)	(3.297)	(2.712)
Exposure to Not pro-mov. posts	-41.70***	-36.37***	-24.90***	-14.66	-46.76***	-36.38***
	(2.763)	(3.539)	(4.912)	(29.74)	(3.181)	(2.963)
R-squared	0.675	0.675	0.675	0.675	0.675	0.678
Mean dep. var.	0.00733	0.00733	0.00733	0.00733	0.00733	0.00733
Panel B: Protests at d						
Exposure to Pro-mov. posts with high emotion	8.555***	3.213	4.201	-159.0***	8.116***	-69,939
	(2.037)	(2.531)	(51.06)	(10.92)	(1.781)	(152,732)
Exposure to Pro-mov. posts with no high emotion or other emotions	4.198**	12.36***	7.263***	17.40***	4.563**	7.192***
	(1.819)	(1.423)	(2.133)	(0.703)	(1.844)	(0.894)
Exposure to Not pro-mov. posts	-3.150	-3.721***	-4.192***	-14.07***	-2.702	-4.079***
	(1.929)	(1.021)	(0.767)	(1.629)	(2.496)	(1.169)
R-squared	0.447	0.447	0.447	0.447	0.447	0.447
Mean dep. var.	0.00220	0.00220	0.00220	0.00220	0.00220	0.00220
Cumulative protests	Y	Y	Y	Y	Y	Y
Day FE	Y	Y	Y	Y	Y	Y
Municipality FE	Y	Y	Y	Y	Y	Y
Observations	244,615	244,615	244,615	244,615	244,615	244,615

Note: Effect of emotional content in Instagram posts on commenting behavior and protest occurrence. OLS estimates with two-way fixed effects. The sample includes all municipalities in France during the period 27 June to 3 July 2023. Exposure to posts is constructed by interacting the residualized Instagram connection measure with the presence of pro-movement and non-pro-movement posts, classified by expressed emotions as detailed in 4. Panel A shows results on the number of comments on Nahel-related posts; Panel B shows effects on the number of protests. Each column isolates a different dominant emotion—anger, disgust, fear, happiness, sadness, and surprise—as well as the overall effect. "High emotion" refers to posts where a given emotion is classified as high by our LLM analysis. All specifications include controls for cumulative past protests, day and municipality fixed effects. Standard errors are clustered at the municipality level. *p<0.1; **p<0.05; ***p<0.01.

Table 5. **Coordination: call to action**

	Comments on Nahel posts		Protests d	
	(1)	(2)	(3)	(4)
Exposure to Pro mov. with call to action	24.20***		13.64***	
	(8.141)		(1.802)	
Exposure to Pro mov. with specific action		26.64		30.89***
		(33.03)		(5.180)
Exposure to Pro mov. with con-specific call		23.84***		11.04***
		(7.304)		(1.475)
Exposure to Pro mov. without call to action	27.61***	27.18***	-0.140	-3.240***
	(4.392)	(7.608)	(0.940)	(1.045)
Exposure to Not pro-mov.	-37.84***	-37.33***	-1.249	2.405
	(4.872)	(7.082)	(1.482)	(2.107)
Cumulative protests	Y	Y	Y	Y
Day FE	Y	Y	Y	Y
Municipality FE	Y	Y	Y	Y
Observations	244,615	244,615	244,615	244,615
R-squared	0.675	0.675	0.447	0.447
Mean dep. var.	0.00733	0.00733	0.00220	0.00220

Note: Effect of exposure to calls to action in Instagram posts on online engagement and protest activity. OLS estimates with two-way fixed effects. The sample covers all municipalities in France over the period 27 June to 3 July 2023. Exposure is constructed by interacting the residualized Instagram connection measure with the presence of protest-related content, distinguishing between pro-movement and not pro-movement posts, and whether the post includes a call to action. Column 1 and 2 report effects on comments on Nahel-related posts while columns 3 and 4 report effects on protest incidence. Columns 1 and 3 examine general calls to action, while column 2 and 4 differentiates between specific and non-specific calls. "Specific action" refers to posts that suggest concrete behaviors (e.g., attending a protest in a specific place at a specific time); "non-specific call" includes general encouragement to act. Posts with no calls to action are also separately identified. All regressions include controls for cumulative past protests, municipality and day fixed effects. Standard errors are clustered at the municipality level. *p<0.1; **p<0.05; ***p<0.01.

Table 6. **Repression**

	Comments on Nahel posts			Protests at d		
	(1)	(2)	(3)	(4)	(5)	(6)
Exposure to Pro-mov. with repression	23.78***			-8.951***		
	(3.046)			(2.880)		
Exposure to Pro-mov. with Clashes		30.10			-31.03***	
		(39.90)			(4.572)	
Exposure to Pro-mov. with other repression		22.08*			-3.257	
		(12.06)			(3.236)	
Exposure to Pro-mov. with violence			17.23**			-14.25***
			(7.331)			(3.707)
Exposure to Pro-mov. with other	30.34***	30.72***	27.24***	15.69***	15.50***	10.80***
	(6.726)	(5.339)	(3.117)	(1.373)	(1.377)	(0.560)
Exposure to Not pro-mov.	-37.69***	-38.18***	-34.90***	-1.395	-0.0174	1.509***
	(3.290)	(4.486)	(3.810)	(1.076)	(1.166)	(0.515)
Cumulative protests	Y	Y	Y	Y	Y	Y
Municipality FE	Y	Y	Y	Y	Y	Y
Day FE	Y	Y	Y	Y	Y	Y
Observations	244,615	244,615	244,615	244,615	244,615	244,615
R-squared	0.675	0.675	0.675	0.447	0.448	0.447
Mean dep. var.	0.00733	0.00733	0.00733	0.00220	0.00220	0.00220

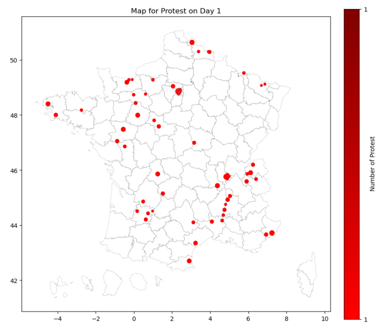
Note: Effect of exposure to repression-related content in Instagram posts on online engagement and protest activity. OLS estimates with two-way fixed effects. The sample covers all municipalities in France over the period 27 June to 3 July 2023. Exposure is constructed by interacting the residualized Instagram connection measure with the presence of protest-related content, distinguishing between pro-movement and not pro-movement posts, and whether the post refers to repressive actions. Columns 1 to 3 report effects on comments on Nahel-related posts, while columns 4 to 6 report effects on protest incidence. Column 1 and 4 examine general repression mentions; columns 2 and 5 further differentiate between "clashes" with police and "other repression"; columns 3 and 6 focus on references to "violence." The category "other" refers to posts not mentioning repression. All regressions include controls for cumulative past protests, municipality and day fixed effects. Standard errors are clustered at the municipality level. *p<0.1; **p<0.05; ***p<0.01.

Note: All municipalities (no restriction by population), ACLED data, data: day-municipality observations, Clustered at municipality level

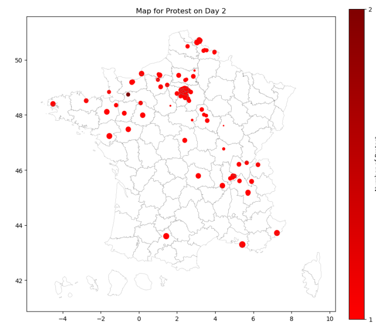
Online Appendix

Appendix A: Appendix: Additional results

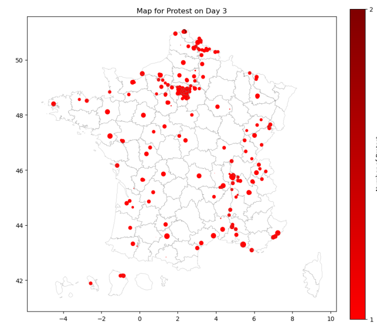
Figure 3. Daily plots of Protests



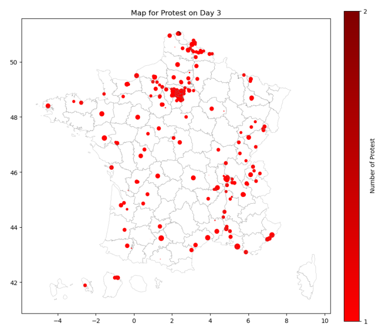
(a) Protests Day 1



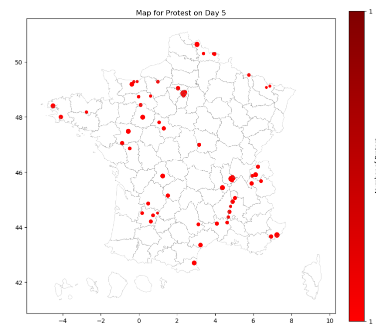
(b) Protests Day 2



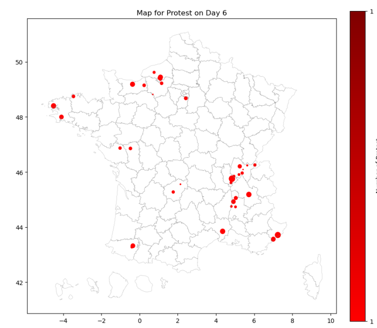
(c) Protests Day 3



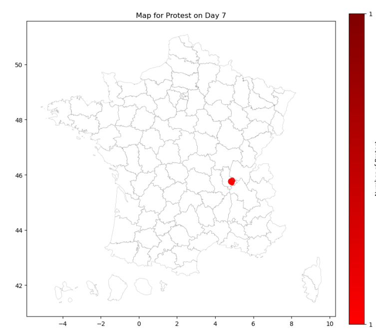
(d) Protests Day 4



(e) Protests Day 5



(f) Protests Day 6



(g) Protests Day 7

Figure 4. Posts in network at baseline from celebrity posts 2 months before Nahel's death

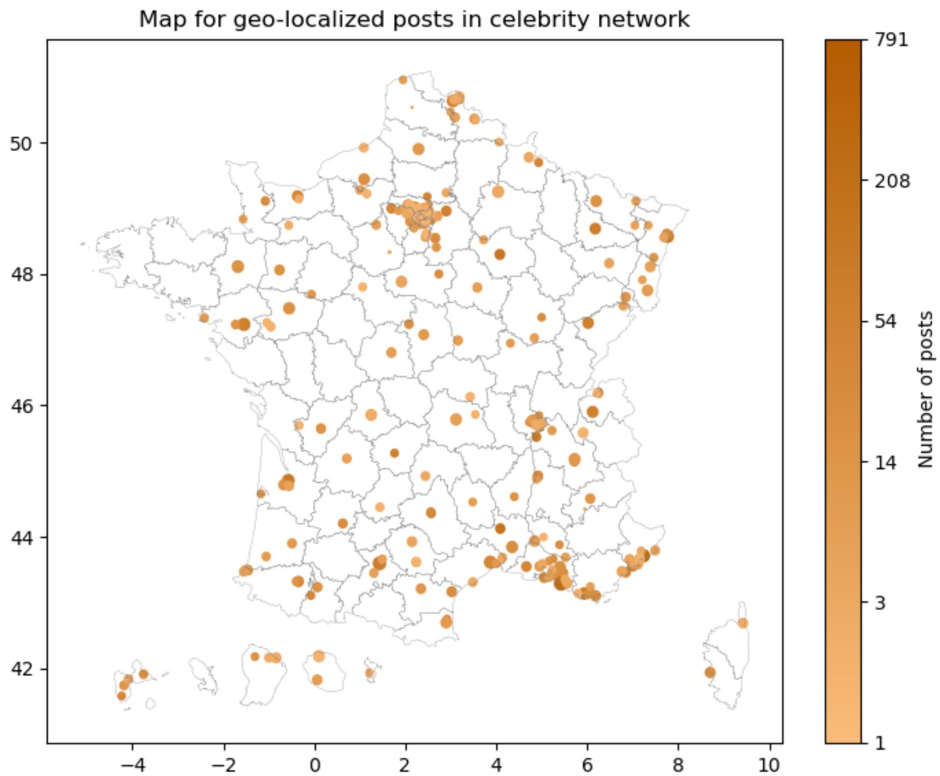


Figure 5. Daily protest and effect of diffusion per day

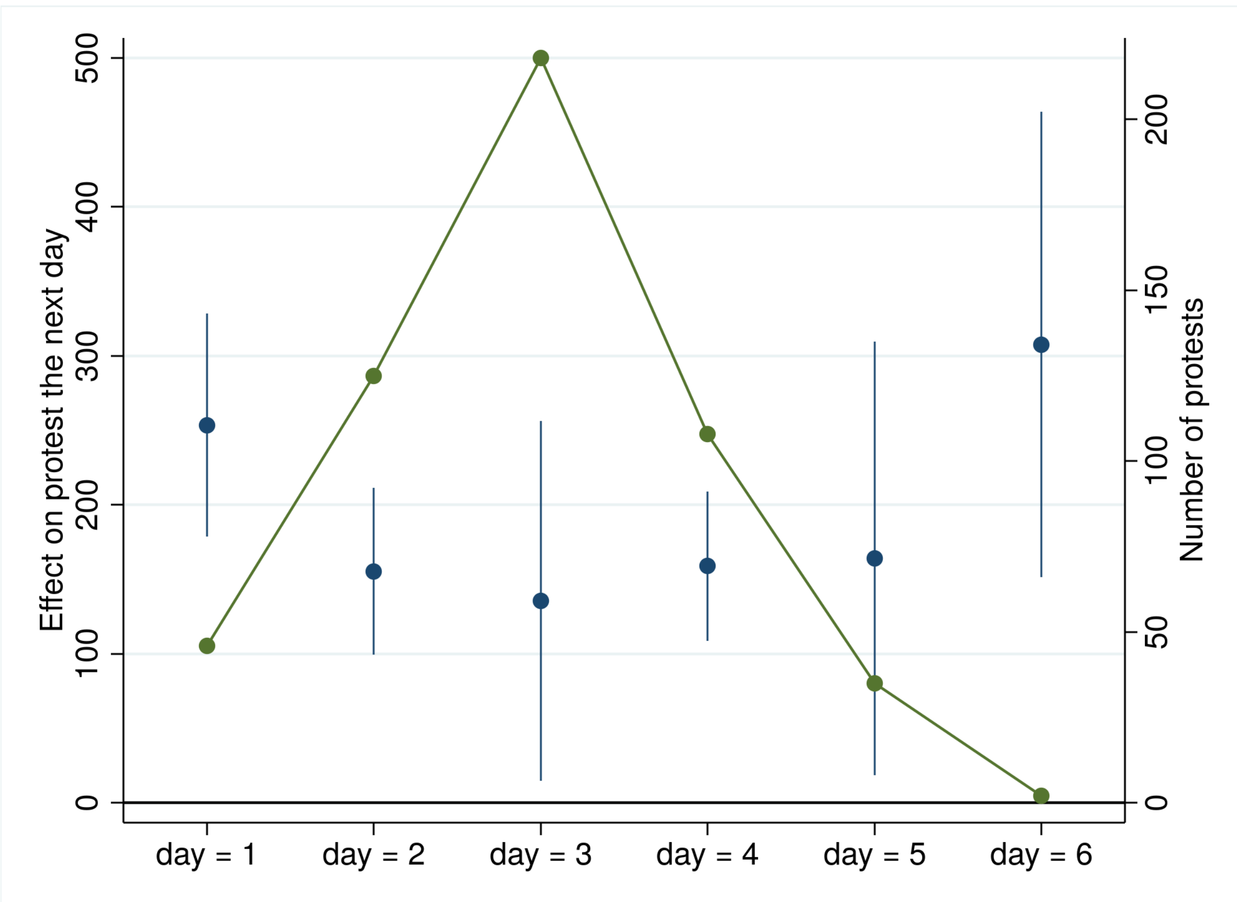


Figure 6. Evolution of opinions on movement

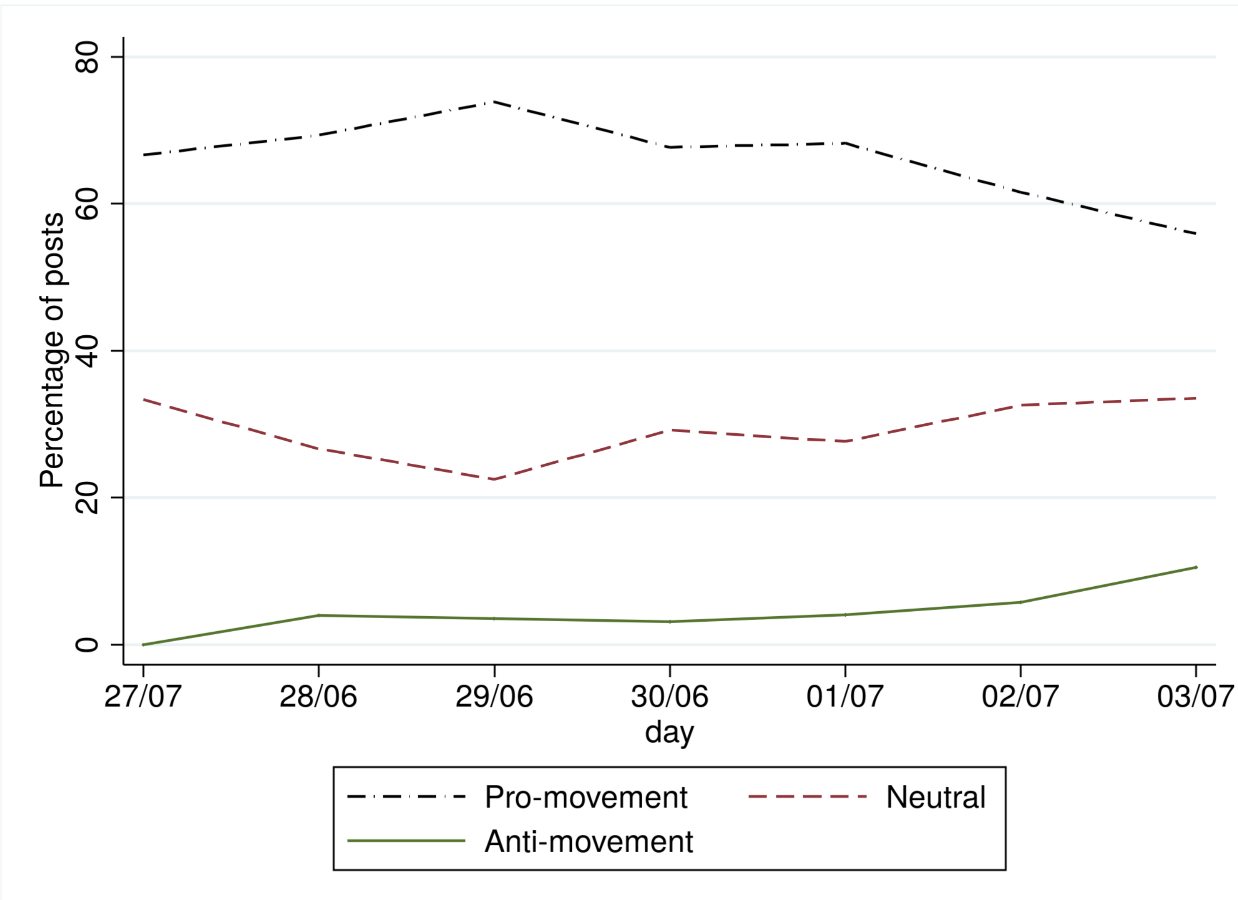
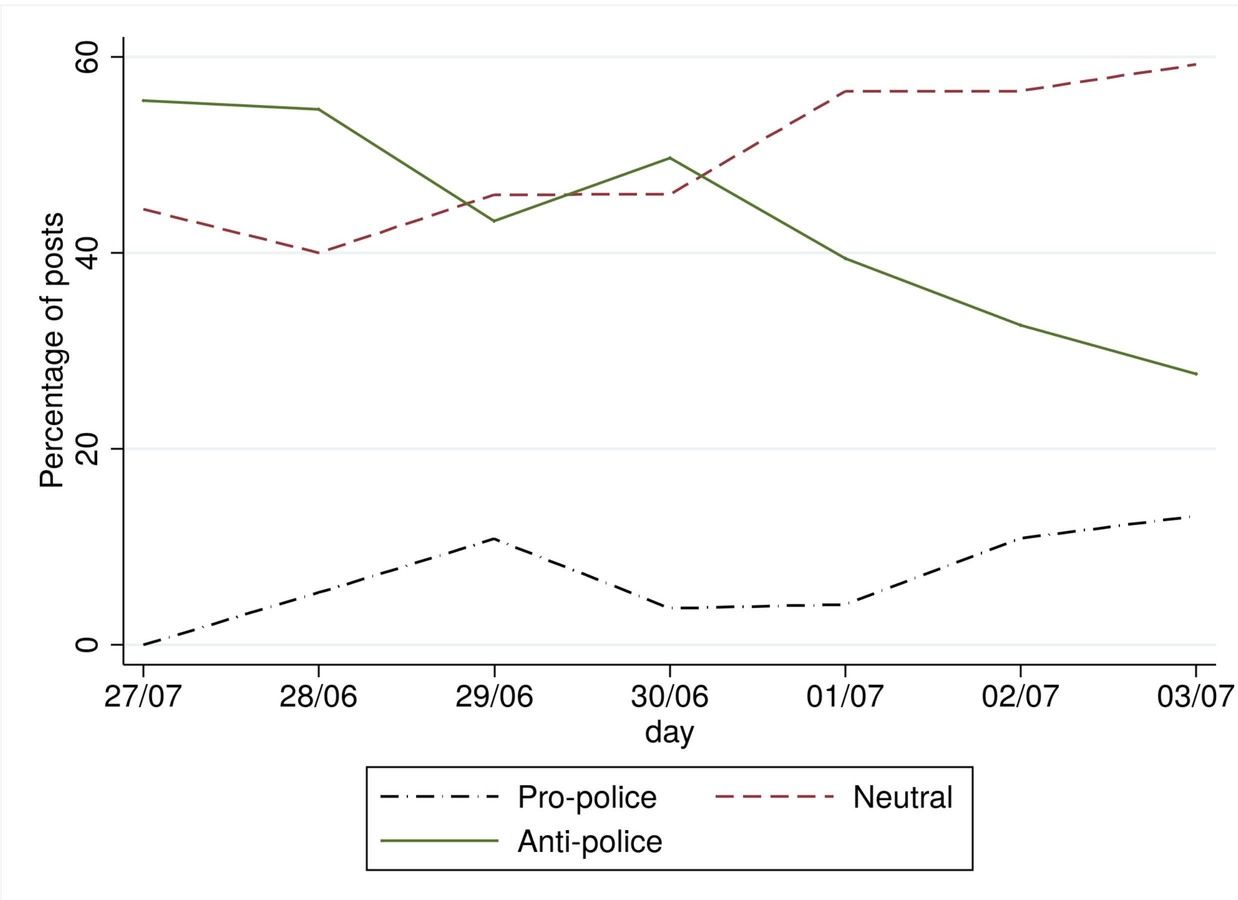


Figure 7. Evolution of opinions of police



Appendix B: Appendix: Robustness Checks

We present here the set of robustness checks discussed in the main text, each designed to test the validity and stability of our empirical strategy. These robustness checks are reported in Tables 7 and 8 and summarized below.

Alternative data source for protests We replicate our main analysis using protest data from *Le Monde* instead of ACLED. Despite differences in event coverage and method, the core findings remain consistent in sign and significance, suggesting that our results are not specific to one source.

Full Instagram network (non-residualized) In our baseline, we residualize the Instagram connectivity measure using a gravity model. As a robustness check, we use the raw follower network without residualization. The results remain similar, suggesting that connections in our Instagram network are largely independent from other connections.

Region- and department-specific day fixed effects To control for regional time-varying shocks (e.g., localized news cycles or administrative actions), we allow day fixed effects to vary by region and department. Results are stable, indicating that the effects are not driven by unobserved time-varying factors at intermediate spatial levels.

Municipality trends To further account for potential unobserved heterogeneity, we include municipality-specific linear time trends in our specification. These trends help control for gradual, location-specific shifts that might affect both online activity and the likelihood of protest. For instance, a municipality may experience a progressive rise in public concern over policing, or increased mobilization through local networks, independent of any single protest event. By absorbing these slow-moving factors, municipality linear trends ensure that our estimated effects are not confounded by underlying temporal dynamics specific to individual locations.

Alternative treatment: posts about Nahel Instead of actual protests, we use a shift in online posting about Nahel as a treatment proxy. The consistency of effects using this alternative definition supports the idea that our mechanism captures broader shifts in protest-related discourse.

Placebo: future protests ($d+1$) We test whether protest events at day $d+1$ predict outcomes on day d, which would cast doubt on our identification. The absence of significant effects provides reassurance against reverse causality or anticipatory behavior.

Excluding vocal football players Our measure of *ex ante* connectedness between municipalities is based on Instagram follower networks of professional football players. While the majority of these players are not publicly political, a few members of the French men's national team reacted publicly to the death of Nahel, including statements or posts that expressed solidarity or political views. This raises a concern about the exogeneity of our exposure network: followers of these particular players may have engaged with them not only for athletic interest, but also for their political stance, especially in the aftermath of Nahel's death. If this were the case, the connectedness metric could partially reflect ideological alignment or selective attention to protest-related content, rather than neutral network structure.

To address this concern, we exclude all posts and comments that were collected starting from the accounts of the four players who publicly expressed views about the event: Kylian Mbappé, Jules Koundé, Aurélien Tchouaméni, and Paul Pogba.[11] This restriction removes approximately 4 out of the 49 footballers in our dataset. As reported in column 6 of Table **??**, the exclusion of these potentially influential nodes has a negligible effect on the results. Estimates remain similar in magnitude and significance to our main specification, suggesting that our network measure captures general patterns of social connectedness rather than being driven by politically salient figures.

Excluding Nanterre and Paris We address concerns that our results may be disproportionately driven by a few highly salient municipalities—specifically Nanterre and Paris. Nanterre is the location where the shooting of Nahel Merzouk occurred, making it symbolically central to the protests and frequently referenced in posts and comments, regardless of users' actual location. This could create measurement noise in our geographic assignment of online activity. To mitigate this concern, we exclude Nanterre entirely from the sample, both as the origin of posts and as a commenting municipality. As shown in column 3 of Table 8, the estimated effects remain positive, statistically significant, and of similar magnitude.

Paris also presents a distinct concern due to its size, centrality in media and political discourse, and its disproportionate weight in our measure of social media connectedness, especially given that many of the football players in our dataset are active in Paris-based clubs. To rule out that our findings are driven by idiosyncratic dynamics in the capital, we exclude Paris in a separate specification. Finally, we exclude both Nanterre and Paris simultaneously. The results remain robust under these exclusion criteria, indicating that our main effects are not an artifact of either city's influence.

Excluding football players who talked about Nahel Some French men national team selection football players who we use to measure *ex ante* connectedness between municipalities reacted publicly to the death of Nahel. This may raise the concern that some of the people interested in their Instagram content followed them because of their broader political views. If this were the case, it would question the exogeneity of the network to the case of Nahel. Although this concern is alleviated by the fact that the players that we consider were, in general, little politicized, we address it by excluding all posts of these politically vocal football players. This excluded 4 out of 49 players.[12] The results are very similar to those of the main specification.

[11]See Parisien (Le) (2023).

[12]These players were Kylian Mbappé, Jules Koundé, Aurélien Tchouaméni, and Paul Pogba. See Parisien (Le) (2023).

Table 7. **Diffusion of protest through social media: robustness checks**

	Protests d							
	baseline	Le Monde	full network	region-day FE	department-day FE	trends	posts	$d+1$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Exposure to Protests $d-1$	130.9***	130.7***	106.6**	130.2***	146.1***	136.6***		
	(33.03)	(33.61)	(52.24)	(32.99)	(49.52)	(40.78)		
Exposure to Posts $d-1$							2.450***	
							(0.596)	
Exposure to Protests $d+1$								-23.42
								(16.12)
Cumulative protests	Y	Y	Y	Y	Y	Y	Y	Y
Day FE	Y	Y	Y		Y	Y	Y	Y
Region-day FE				Y				
Department-day FE					Y			
Municipality FE	Y	Y	Y	Y	Y	Y	Y	Y
Municipality trends						Y		
Observations	244,615	244,615	244,615	244,615	244,608	244,615	244,615	244,615
R-squared	0.448	0.448	0.380	0.450	0.454	0.579	0.446	0.445
Mean dep. var	0.00220	0.00220	0.00235	0.00220	0.00217	0.00220	0.00220	0.00220

Note: Robustness checks for the effect of protest diffusion through Instagram on protest activity. Two-way fixed effects estimation. The sample includes all municipalities in France over the period 27 June to 3 July 2023. Exposure is constructed by interacting the residualized Instagram connection measure with the presence of protests on the previous day, unless otherwise noted. Column 1 reproduces the baseline specification from Table 1. Column 2 uses Le Monde newspaper coverage for the data on protest. Column 3 uses the full Instagram network before taking the residualization of the gravity model. Columns 4 and 5 include region-day and department-day fixed effects, respectively. Column 6 adds municipality-specific linear time trends. Column 7 replaces protest exposure with post exposure. Column 8 includes a placebo test using future protest exposure ($d+1$). All regressions include controls for cumulative past protests and municipality fixed effects. Day fixed effects are included in all columns except 4 (region-day FE) and 5 (department-day FE). Standard errors are clustered at the municipality level. *p<0.1; **p<0.05; ***p<0.01.

Table 8. **Diffusion of protest through social media: robustness checks**

	Protests d			
	baseline	excl. Paris	excl. Nanterre	excl. both
	(1)	(2)	(3)	(4)
Exposure to Protests $d-1$	130.9***	147.0***	130.8***	162.9**
	(33.03)	(49.98)	(45.87)	(80.29)
Cumulative protests	Y	Y	Y	Y
Municipality FE	Y	Y	Y	Y
Observations	244,615	244,608	244,608	244,601
R-squared	0.448	0.443	0.439	0.433
Mean dep. var	0.00220	0.00217	0.00217	0.00214

Note: Additional robustness checks for the effect of protest diffusion through Instagram on protest activity. Two-way fixed effects estimation. The sample includes all municipalities in France over the period 27 June to 3 July 2023. Exposure is constructed by interacting the residualized Instagram connection measure with the presence of protests on the previous day. Column 1 reproduces the baseline specification from Table 1. Column 2 excludes Paris, the capital and largest city. Column 3 excludes Nanterre, where Nahel was killed. Column 4 excludes both Paris and Nanterre. All regressions include controls for cumulative past protests and municipality and day fixed effects. Standard errors are clustered at the municipality level. *p<0.1; **p<0.05; ***p<0.01.

Appendix C: Appendix: Data

This appendix provides a detailed account of our data sources, construction of key variables, classification methods, and geolocation strategy. Our final dataset is a municipality-by-day panel for the period June 27 to July 3, 2023. The unit of analysis is the municipality-day, with around around 35,000 municipalities and 250,000 municipality-day observations. We provide descriptive statistics of our variables in Table 9.

C.1 Protest Data

We compile protest data from two complementary sources: (i) the Armed Conflict Location & Event Data Project (ACLED) and (ii) a protest database curated by *Le Monde*.

ACLED. The ACLED dataset is a systematic collection of real-time political disorder data. It provides geo-referenced events, including latitude, longitude, event type, and a short textual description. We restrict attention to protests in France between June 27 and July 3, 2023. ACLED includes both peaceful and violent events. To isolate those related to Nahel Merzouk, we use a custom prompt on the Gemini 2.5 Pro large language model (LLM), which reads the event notes and classifies whether the protest (i) explicitly references Nahel or his death, or (ii) expresses discontent with police violence.

The model classifies 537 protest events in 305 municipalities as relevant to our analysis. Each event is geocoded and matched to its INSEE municipality code via spatial join using GIS shape files. We verify a random subsample of the LLM classifications manually to ensure consistency and accuracy. Details of the prompt and examples are available upon request.

Le Monde. The French newspaper *Le Monde* compiled a list of riots and protests during the same period. This dataset includes event type (e.g., property destruction, arson, looting), location, and time. It serves as an independent source for robustness checks. We identify 495 protest events in 311 municipalities. We verify that results are not driven by choice of protest dataset.

C.2 Instagram Data: Connectedness Network

We construct a proxy for online connectedness between municipalities prior to the protest wave. The objective is to measure the latent social media exposure between municipalities that may act as a channel for protest diffusion, while avoiding post-treatment bias.

Seed Accounts. We begin with the list of 49 football players in the French men's national team as of June 2023. These individuals are public figures with wide social media reach, particularly among younger male users and people from disadvantage background (the specific socio-demographic group that protested after the death of Nahel).

Scraping Procedure. For each player, we retrieve Instagram posts made in the two months prior to June 27. For each post, we collect the first 50 publicly available comments (API-limited). This forms Level 0. We then extract:

- **Level 1:** All posts made by Level 0 commenters in the same two-month window, and the commenters on those posts.
- **Level 2:** Posts made by Level 1 commenters, and their own commenters.
- **Level 3:** Posts made by Level 2 commenters, and their own commenters.

This iterative scraping stops at Level 3. The resulting dataset includes a rich network of post to comments relation.

Constructing the Matrix. We geolocate each user (detailed below) and define a directed weighted matrix $\text{ExAnteConnections}_{j \to i}$ as the number of comments from users in municipality j on posts authored by users in i, excluding within-municipality interactions. The matrix is asymmetric, capturing directional exposure rather than mutual engagement. This structure is well-suited to measure potential influence or information flow.

Assumptions. We assume that pre-existing comment patterns on Instagram approximate the network of potential exposure during the protest period. The use of celebrity-seeded networks ensures partial exogeneity, while still targeting the demographic most likely to engage in or be exposed to protest behavior.

C.3 Instagram Data: Protest Content

In addition to the baseline network, we collect Instagram activity explicitly referencing the Nahel protests. We scrape all public posts and comments that contain the hashtags `#Nahel` and `#Nael` from June 27 to July 3. These hashtags capture the protest discourse with high precision, though we acknowledge they may underrepresent subtler or less explicit forms of political expression.

Content Volume. We collect approximately 1,000 posts and 961 comments with successful geolocation of both the commenter and the poster. Figures 2 and **??** visualize temporal and spatial variation. Notably, comment volume peaks on June 29, contemporaneous with peak protest activity.

Engagement Structure. For each comment, we identify both the origin (commenter's municipality) and the target (post's author municipality). This allows us to construct a similar directional interaction structure as in the baseline network, but focused solely on protest-related discourse.

C.4 Geolocation Strategy

Assigning users to municipalities is central to our identification strategy. We geolocate users using only publicly available information. Our geolocation algorithm follows a hierarchical rule set:

1. Check the geo-tag metadata of the user's post.
2. Parse user biography text for municipality names (restricted to French municipalities with population >10,000).
3. Examine geo-tags and captions of the user's 10 most recent posts.
4. Identify explicit ZIP codes in bios or captions.

When multiple municipalities are found, we assign the most frequently occurring one. If multiple appear with equal frequency, we default to the most populous. This procedure yields a geolocation success rate of approximately 17.5% for protest-related activity, and a higher rate for baseline network data given more user metadata.

Table 9. **Descriptive statistics**

	observations	mean	sd	min	max
Municipality-day level data					
Protests	244615	.0021953	.0486022	0	4
At least one protest	244615	.0021217	.0460131	0	1
Nahel posts	244615	.0031233	.3325984	0	76
Comments on Nahel posts	244615	.007334	.2590149	0	53
Exposure to Protests d	244615	4.78e-07	.0000439	-1.47e-06	.0119045
Exposure to Posts	244615	9.35e-06	.0012569	-.000034	.3145283
Population	244615	1964.455	15153.01	0	2162598
Municipality-pair level data					
Distance (km)	1221048192	455.155	640.99	0.413913	13456.05
Connections	1221048192	0.000032	0.197677	0.000000	4312.715
Residual connections	1221048192	0.000030	0.197236	-1.628827	4299.763
Post-level data					
Positive towards movement	774	.6899225	.4628242	0	1
Positive towards police	774	.0826873	.2755871	0	1
Image main focus if pro mov.	534	.8876404	.3161044	0	1
High emotion if pro mov.: overall	534	.6573034	.4750561	0	1
High emotion if pro mov.: anger	534	.6198502	.4858786	0	1
High emotion if pro mov.: disgust	534	.0018727	.0432742	0	1
High emotion if pro mov.: fear	534	.0280899	.1653846	0	1
High emotion if pro mov.: happiness	534	0	0	0	0
High emotion if pro mov.: sadness	534	.6179775	.4863376	0	1
High emotion if pro mov.: surprise	534	.0018727	.0432742	0	1
Call to action if pro mov.	534	.4831461	.5001844	0	1
Specific action if pro mov.	534	.1011236	.301775	0	1
Repression if pro mov.	534	.5786517	.4942382	0	1
Clashes if pro mov.	534	.1235955	.3294281	0	1
Violence if pro mov.	534	.1947566	.3963845	0	1

Note: Descriptive statistics for the main variables used in the analysis. The table reports means, standard deviations, and minimum/maximum values across three levels of analysis. The first panel presents statistics at the municipality-day level for 244,615 observations, including protest incidence (as count and binary), the number of Nahel-related Instagram posts and comments, two measures of exposure (to protests and to posts), and population size. The second panel describes data at the municipality-pair level (over 1.2 billion observations), including geographical distance, Instagram connections, and the residualized measure of connections used in the empirical strategy. The third panel reports statistics at the post level, including indicators for whether content is positive toward the movement or police, whether the image is the main focus, presence of high emotional content (overall and by emotion type), presence and specificity of calls to action, and references to repression, clashes, or violence.

C.5 Post examples

Post Examples by Stance Category

Pro-Protest

29 juin 2023, Nanterre. Mounia Merzouk organise une marche blanche en la mémoire de son fils Nahel, tué à l'âge de dix-sept ans d'une balle dans le thorax. Plusieurs milliers de personnes se réunissent et défilent alors dans les rues de la ville.
`#nanterre #nahel #marcheblanche #paris #france`

Anti-Protest

Rien ne justifie la mort d'un jeune de 17 ans. Ce qui est arrivé à #Nael est inexplicable et inexcusable. [...] Mais depuis plusieurs jours, les scènes de violences indéfendables se multiplient [...] Ces attaques inqualifiables contre les institutions, contre les élus, contre les policiers, contre les civils qui n'ont rien demandé, contre nos concitoyens, nos enfants, sont autant inacceptable.

Pro-Police

`#Nael #nanterre #emeutes`

Anti-Police

`#montreuil #montreuilcity #montreuilstreetart #montreuilgraffiti #streetartmontreuil #montreuil93 #graffiti #instaphoto #instalife #city #ville #cityphotography #montreuilcityzoo #france #iledefrance #paris #nahel`

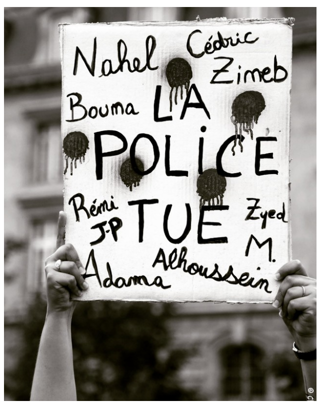
Post Examples: Image Focus

Main Focus

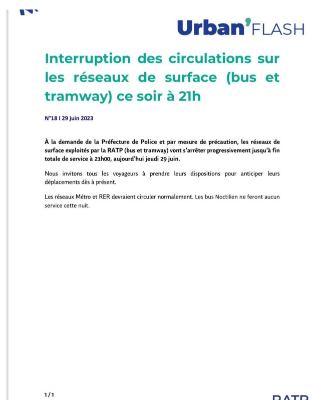
Background Image

No Image

Post Examples: Emotional Intensity



High Emotional Intensity


Urban'FLASH

Interruption des circulations sur les réseaux de surface (bus et tramway) ce soir à 21h

N°18 I 29 juin 2023

À la demande de la Préfecture de Police et par mesure de précaution, les réseaux de surface exploités par la RATP (bus et tramway) vont s'arrêter progressivement jusqu'à fin totale de service à 21h00, aujourd'hui jeudi 29 juin.

Nous invitons tous les voyageurs à prendre leurs dispositions pour anticiper leurs déplacements dès à présent.

Les réseaux Métro et RER devraient circuler normalement. Les bus Noctilien ne feront aucun service cette nuit.

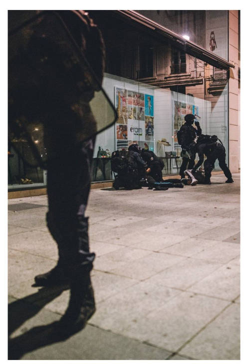
1/1

Low Emotional Intensity

Post Examples: Coordination and Repression

Call to Action

Specific Action



Repression

C.6 Prompts

Prompt for Classifying ACLED protest

I will give you a list of description of events along with an identifier. Please answer in CSV format with the following information: identifier, whether this is a protest or riot against police brutality following the death of Nahel M, whether this is a protest against the protests and riots following the death of Nahel M.

Prompt for Describing Image Content

You are an expert system analyzing Instagram posts related to protests following the death of Nahel Merzouk in June 2023. Each post includes an image, its caption, and any visible text within the image. Take your time to carefully and holistically analyze all available information—image content, embedded text, and caption—to extract meaningful, structured insights suitable for regression and data analysis. Your output should be a Python dictionary with the following structure:
=== VISUAL DESCRIPTION ===
"image_summary": A 1-2 sentence description. "keywords": A list of 5 relevant keywords.
=== PROTEST FRAMING ===
Each of the following should be a dictionary with: "value": one of the listed options, "justification": short reasoning based on image and caption.
Fields: "protest_stance": ["pro-protest", "anti-protest", "neutral/not mentioned"], "police_stance": ["pro-police", "anti-police", "neutral/not mentioned"], "view_on_nahel": ["favorable", "unfavorable", "neutral/not mentioned"], }.
Ensure the final output is a valid Python dictionary ready to be appended to a pandas DataFrame. Take your time to evaluate the post carefully. === POST TYPE === "relevance": the relevance of the post to the topic of the protest following the death of Nahel, police reaction, systemic racism, etc. A post is not related if it merely uses a #nahel hashtag but the post is clearly entirely unrelated otherwise. [none/low/medium/high], "language": the language the post is written in. If the post includes multiple languages, list them separated by commas. "image_type": identify the image type, excluding any text that may have been added over the image: photos / political cartoon / drawing / blank image "image_usage": how the image is used: main focus / illustration (for accompanying text in image) / background (if the image is barely visible/doesn't add any information) === EMOTIONS TRIGGERED IN THE VIEWER === "overall_emotionality": How likely the post is to trigger an emotional reaction from the viewer [none/low/medium/high], "overall_emotionality_justification": short reasoning based on image and caption, "anger"/"disgust"/"fear"/"happiness"/"sadness"/surprise": for each of these emotions, whether the post will trigger this emotion, and a strength ([none/low/medium/high]) "emotions_justification": short justification for these emotions === CONTENT === "Call_to_action": whether the post include a call to action (for example "go protest", "contact politicians") - answer with "main focus", "present", "not present", "Specific_action": whether the post mentions a specific action (for example a protest at a given time and location), or gives specific advice for acting - answer with "main focus" "present" "not present" "Repression": does the post or image mention or show police presence, action or repression of protests (main focus/present/not present) "Clashes": does the post or image mention or show clashes with the police (main focus/present/not present) "Violence": does the post or image mention or show violence (main focus/present/not present) "Content_justification": a short justification for the above answers
Ensure the final output is a valid Python dictionary ready to be appended to a pandas DataFrame. Take your time to evaluate the post