Breaking the Echo Chamber: Social Media Networks and Political Conflict

Luis Menéndez *, Daniel Montolio†
Hannes Mueller ‡, Francesco Slataper§

August 15, 2025

Abstract

This article exploits data from a political conflict between language groups to show how political events can rapidly redefine how these groups interact on social media. Leveraging a unique dataset of 26 million retweets by 120 000 Catalan- and Spanish-speaking Twitter users, we estimate individual exposure to tweets with a network-based model. We then compare two shocks in the same region and year: the Barcelona terror attack and the Catalan independence referendum of 2017. The referendum, and related police violence, triggered a symmetric jump in retweeting across groups. The terror attack, by contrast, did not lead to a similar realignment.

Keywords: Social Media, Echo-chambers, Political Conflict, Retweet Behavior, Social Networks, Ethno-linguistic Conflict, Polarization
JEL Codes: D74, C55, C45

*IAE-CSIC,UAB, BSE. luisigmenendez@gmail.com
†UB & IEB. montolio@ub.edu
‡IAE-CSIC, BSE, CEPR. h.mueller.uni@gmail.com
§UNC. slata@unc.edu
¶Daniel Montolio acknowledges financial support from the Spanish Ministry of Science, Innovation and Universities under the project PID2022-140661OB-I00. Mueller acknowledges support provided by the Severo Ochoa Programme for Centres of Excellence in R&D (Barcelona School of Economics CEX2024-001476-S), funded by MCIN/AEI/10.13039/501100011033 and AGAUR-Generalitat de Catalunya (2021-SGR-416). We thank audiences at the IAE, UAB, the Barcelona Supercomputing Centre and the CEPR Paris Symposium and the meeting of the Catalan Economics Association. All remaining errors are our own.

Introduction

The media plays a crucial role in shaping politics, and the recent literature highlights the growing influence of social media in this process (Besley and Burgess (2002), Zhuravskaya et al. (2020)). A particularly concerning aspect of social media is its tendency to foster echo chambers, environments where individuals are primarily exposed to information that reinforces their existing beliefs. Echo chambers can lead to diverging information environments across different segments of the population. This phenomenon is especially relevant in contexts of political conflict, where fragmented information may contribute to the malfunctioning of political institutions and the escalation of violence (Levy and Razin (2019), Müller and Schwarz (2020), Blattman (2023)). However, existing research offers mixed findings on whether social media increases or decreases the exposure to opposing views.[1]

Building on this literature, we argue that the effects of social media on political conflict are highly context-dependent and shaped by the nature of offline events. In particular, we focus on how political events influence online behavior and, in turn, the diffusion of information within and across network communities. We analyze communication in the weeks before and after two widely discussed events that occurred in Spain in 2017: the terrorist attack on August 17 in Barcelona and the Catalan independence referendum, marked by police violence, on October 1. In our analysis we exploit the fact that both events took place in the same region (Catalonia) within a short period of time and received extensive coverage in both traditional and social media across the country.

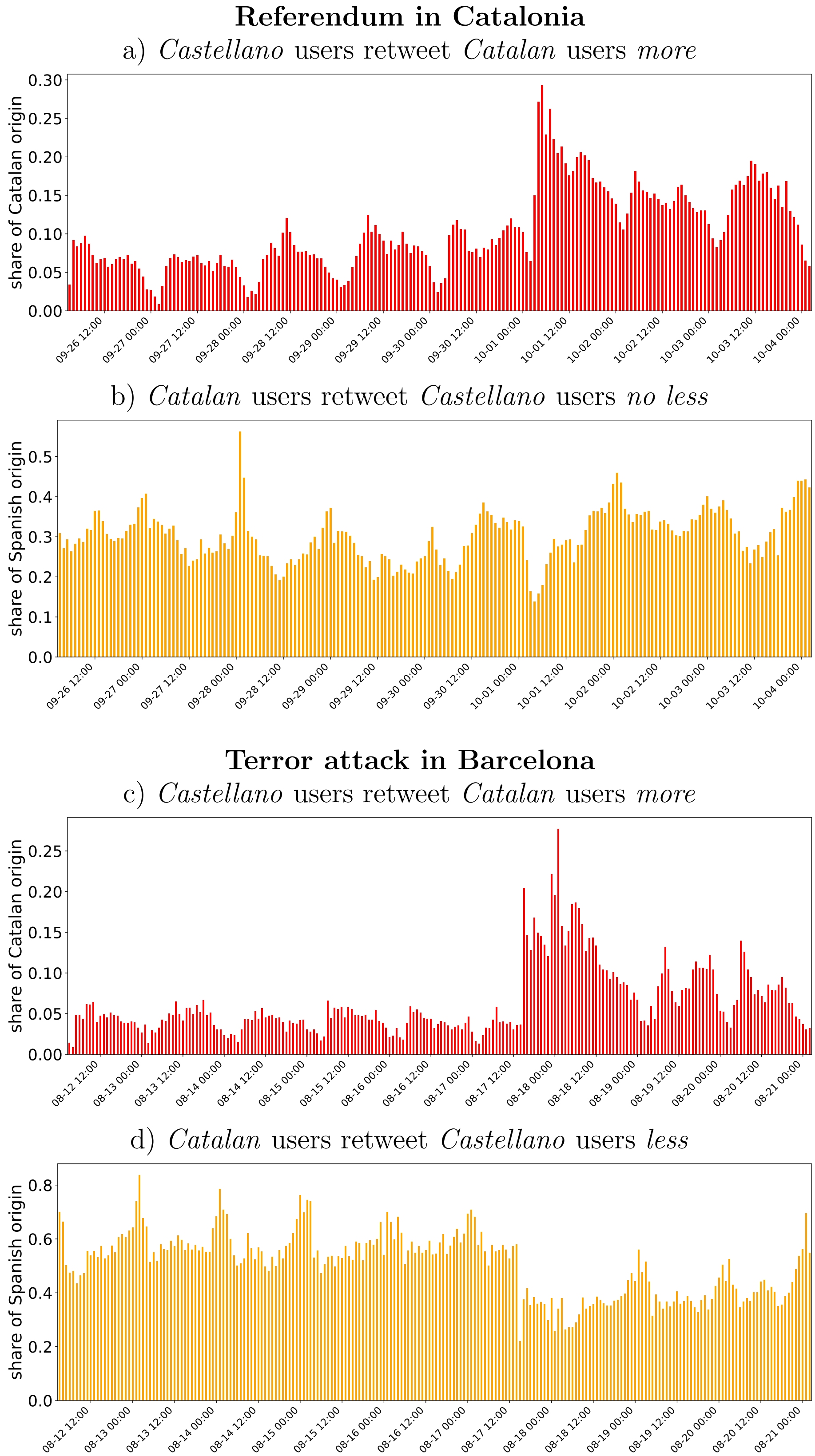
A key aspect for our study is Catalan bilingualism. Catalan users can choose to engage online in either Catalan or Spanish (hereafter referred to as Castellano). Drawing on a unique corpus of nearly 120,000 Twitter (now X) users and 26 million retweets, we document the existence of these two online language communities. Using the terrorist attack as benchmark, we then examine how events following the referendum reshaped individual user behavior on Twitter.

Figure 1 offers a first overview of our dataset. Panels a) and b) show aggregate cross-sharing between language groups in the period of interest, a key metric reflecting the main focus of our study. Red bars show the retweeting of Catalan users by Castellano users normalized by total Castellano retweeting. Orange bars show the reverse, retweeting of Castellano users by Catalan users normalized by total Catalan retweeting. For comparison, panels (c) and (d) illustrate intergroup retweet rate around August 17, 2017, the day of the terror attack.[2]

[1]See, for example, Gentzkow and Shapiro (2011), Boxell et al. (2017), Allcott et al. (2020), Barberá (2020), Levy (2021), Bowen et al. (2023) and Allcott et al. (2024).

[2]Figures covering the entire sample period are available in Appendix Figures B3 and B4.

Figure 1: Retweets across language groups



Notes: Graphs show the share of retweets for users in two different language groups. The red figures show the share of tweets coming from Catalan users in all tweets retweeted by Castellano users over time. The orange figures show the share of tweets coming from Castellano users retweeted by Catalan users over time.

We observe a sharp increase in Castellano retweeting Catalans in the wake of both events. However, the response by the Catalan network to Castellano accounts differs considerably. Figure 1 b) shows a slight upward shift on the 30th of September, in the politically charged atmosphere leading up to the referendum, and is only interrupted by a slump around midnight. Figure 1 d), on the other hand, shows a sharp drop of cross-sharing following the terror attack.

Our research approaches the data in Figure 1 through the conceptual distinction of *exposure* and *retweet rates*. We conceptualize exposure, a latent feature in our data, as the visibility of content. Exposure is a function of user activity, the supply of content, network structure and retweet rates of the members of each language community. If retweet rates across groups increase, each community is *ceteris paribus* more exposed to the other group's tweets, making further cross-sharing more likely. Behavioral changes at the individual level therefore induce changes in segregation of the two language communities. Studying retweet rates is complicated by these latent, dynamic changes in exposure. For example, it is possible that the sharp increase in Castellano retweeting Catalans in Figure 1 is driven by changes in the supply of Catalan tweets leading to an increase in exposure and not by an "opening up" of the Castellano language community.

Our main finding is that both language communities changed their retweet rates significantly as the events surrounding the referendum unfolded. Surprisingly, the response was symmetric — each group increased the relative retweet rates for messages from the other group. However, the pattern was completely different in the days after the terror attack, where our analysis suggests no or short-lived effects. This discrepancy suggests that at least part of the decrease in retweeting illustrated in Figure 1 panel d) is a simple "crowding out" effect, where the surge in supply of Catalan messages displaced those from Castellano users. This is in line with similar effects seen in traditional media, where high volumes of newsworthy content during major events crowd out other reporting (see, e.g. Eisensee and Strömberg (2007); Durante and Zhuravskaya (2018)).

However, this raises the question why the two language communities opened up in response to an event that might have reinforced echo chambers. It is particularly striking that Catalan users, the primary victims of police brutality, amplified messages originating in the Castellano language group rather than reinforcing messages from their own group. To study this, we analyze the content of the messages that broke the Catalan echo chamber most effectively. We find that Catalan users leveraged sympathetic messages from Castellano voices declaring outrage over the violence. This can be interpreted as a kind of *"Nixon goes to China"* phenomenon: those least expected to support a cause can lend it exceptional credibility when they do (Cukierman and Tommasi (1998)).

In doing so, the Catalan response echoes a broader principle in the study of protest and political conflict: non-violent resistance can be especially effective when it provokes state overreach. Documenting such repression and amplifying outsider outrage — often termed *political jiu-jitsu* — can strengthen the protesters' position.[3] The police crackdown on voters in Catalonia served as a vivid instance of this dynamic. Viral tweets from Catalan users condemning police brutality circulated widely among Castellano audiences and made information from Catalan accounts reach deeper into the Castellano-speaking network. Declarations of outrage over the police violence by Castellano users reached deeply into the Catalan network.

Our findings speak to broader theoretical debates on information diffusion and polarization in networks. Exposure to cross-cutting viewpoints can mitigate polarization or exacerbate it depending on structural conditions and how users react to new information (Golub and Jackson (2012); Dandekar et al. (2013); Bramoullé et al. (2014)). In our case, network permeability and shared outrage to police violence enabled temporary convergence across group lines. Our approach — simulating exposure via network-based models — offers a tractable and robust way to track how external shocks reconfigure communication patterns and could be adapted for other cases where group identities intersect with political events.

Finally, our results contribute to the literature on echo chambers, online polarization and ethno-linguistic conflict. The role of social media for these phenomena is likely highly context-specific. The literature has found, for example, that social media networks can be used to coordinate protest (Enikolopov et al., 2020) or motivate hate crimes (Müller and Schwarz, 2020). Catalonia in 2017 offers a unique chance in this regard as two important events of different character took place within a short time span. The different behavior of Twitter users under the two events shows that social media networks adapt to what is being discussed. Our results on the 1st of October show that important events can restructure online spaces, prompting users to prioritize ideological commonality, for example, over group identity. This is particularly important in ethnically or linguistically polarized societies which have been shown to be more prone to armed conflict (Montalvo and Reynal-Querol (2005); Esteban et al. (2012); Michalopoulos and Papaioannou (2016)) and dysfunctional politics (Alesina et al. (1999), Posner (2005), Burgess et al. (2015)).

[3] See Sharp (1973) on this mechanism; Chatterjee (1974) provides a strategic model of its logic and Muñoz and Anduiza (2019) show, for the Spanish anti-austerity movement known as the 15-M or *indignados* that emerged during the financial crisis, that exposure to violent repression increases popular support for social movements by galvanizing bystanders, who might otherwise remain passive, into active sympathy and engagement.

Events can be used to amplify group identity and mobilize for conflict (Fearon and Laitin (2000); Sen (2006); Racek et al. (2024)) and this is the context for the findings in Müller and Schwarz (2020). However, it is also possible that events unify across ethno-linguistic lines (Depetris-Chauvin et al. (2020)) and we would expect social media to play a unifying, conflict-reducing role in these circumstances. Our main contribution in this regard is that our data allows us to observe changes in individual behavior over time and link it to specific content. This allows us to track the effect of a unifying event through revealed preferences and contrast this effect with an event that was not unifying across linguistic lines.

1 Study Background

1.1 The Independence Vote and Police Violence

Before introducing our statistical model, we describe the real-world context behind the online interactions we observe. Our research examines a recent, intense phase of a longstanding conflict.

Catalans have long harbored aspirations for greater autonomy and, in some periods, outright independence from Spain. In more recent times, the independence movement gained momentum from 2010 onwards. On June 9 2017, the Catalan Government, led by President Carles Puigdemont, officially announced an independence referendum for Catalonia. The Spanish Constitutional Court immediately (September 7) suspended the law, and forbade both the Catalan government and any of its municipalities to participate, finance or otherwise support to the referendum. As soon as the sentence was published, the Spanish government directed Catalan municipalities and public officers to publicly declare their intentions (and thus their allegiance) within 48 hours. This ultimatum backfired, as the vast majority explicitly defied the Court's order. The Catalan government then proceeded to organize the voting.

On the night of September 30, 2017, the eve of the referendum, hundreds of citizens gathered at designated polling stations, including schools and community centers, to ensure these locations remained accessible for voting. Social media and particularly Twitter, emerged as an arena for competing narratives, rapidly polarizing users into opposing camps. The online discourse was intensified by the strategic use of trolls and bots, which amplified extremist messages, coordinated artificial campaigns and contributed significantly to online polarization.[4]

[4]Twitter identified and removed numerous accounts that were suspected of being involved in coordinated disinformation campaigns and these accounts, and their activity, is therefore not included in our dataset. We made use of Twitter's "Information Operations Archive" and found no presence of trolls in our dataset.

The referendum day itself was characterized by significant unrest and violence. Early in the morning, Spanish police forces began operations to shut down polling stations and confiscate voting materials. Despite these efforts, many polling stations managed to open. Throughout the day, numerous instances of police violence were reported, further galvanizing the pro-independence movement and drawing widespread condemnation from international observers. Images and videos of the clashes were widely shared on social media, drawing condemnation from human rights organizations, foreign governments, and international media. The European Union and the United Nations called for dialogue and restraint, emphasizing the need for a peaceful resolution to the conflict.[5]

1.2 The Terror Attack

As a placebo test, we compare online behavior on October 1, 2017, with that observed two months earlier, following the terrorist attack in Barcelona. On August 17, 2017, Barcelona suffered a terrorist attack when a van deliberately drove into pedestrians along La Rambla, a popular tourist area, killing 14 people and injuring over 130. The Islamic State (ISIS) claimed responsibility for the attack. The incident drew international media attention and solidarity, marking one of Spain's deadliest terrorist attacks since the 2004 Madrid train bombings.

1.3 Opinion Poll Data

Figure 2 provides context for these events using data from the *Centro de Investigaciones Sociológicas* (CIS) barometer, which conducts quarterly surveys of about 4,000 respondents. The survey records each respondent's location and asks their opinion of the central government with the question "In general, how would you rate the job the PP (People's Party) government is doing: very good, good, fair, bad, or very bad?" We collected all quarters from 2017 up to the last quarter before the change in government in April 2018. Both in Catalonia and in the rest of Spain, approval of the central government declined after October 1st. However, the drop was particularly pronounced among Catalan respondents, where the average rating fell by 0.17 points.

It is important to keep in mind that this was the context where the social media communication we analyze took place. Figure 2 illustrates how the events between August to October

This is consistent with Twitter's policy which removed any trace of these tweets including all the interactions with them. Although no longer accessible from the official X webpage, datasets for the Catalonian foreign influence tweets can be found at the online repository https://github.com/simabasel/cib-data/tree/master.

[5]Some of the most popular videos about police violence shared in Twitter can be seen at https://catalanrepression.github.io/ and https://spanishpolice.github.io/.

Figure 2: Survey valuation on Central Government

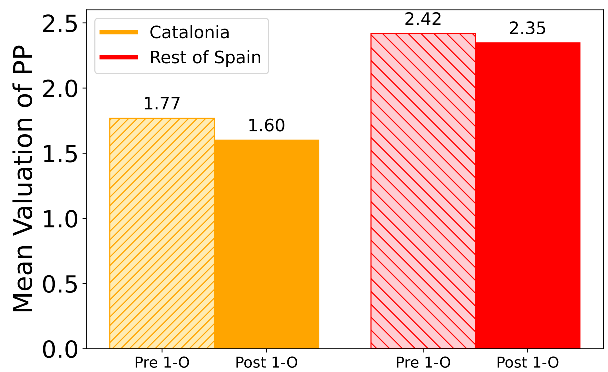


Figure notes: Bars show the average rating of the Central Government (right-wing People's Party, PP) from the CIS barometer, for two periods: before October 1, 2017, and from October 1, 2017, until the PP left power in 2018. Scores range from 0 (very negative) to 5 (very positive), and are shown separately for Spanish respondents (excluding Catalans) and Catalan respondents. Light-colored, dashed bars represent the average rating for January, April, and July 2017; solid bars represent the average rating for October 2017, January 2018, and April 2018.

represented a public relations' disaster for the Rajoy government (People's Party), alienating Catalans in particular.

2 Statistical Model

2.1 Model Set-up

The goal of our statistical model is to interpret the patterns we observe in the data, for example those illustrated by Figure 1. More precisely, our aim is to understand whether the events around the 1st of October led to changing retweet behavior between the Castellano- and Catalan-speaking groups. Our initial hypothesis was that stronger echo chambers would emerge as a result of heightened political tensions, as messages within each language groups would be shared more fervently and out-group messages ignored.

The key decision we want to study at the individual level is, therefore, whether to retweet a message or not.[6] Retweeting simultaneously validates (and thus, legitimizes) and amplifies a message, making it a central indicator of segregation or polarization of online communities.

Testing our hypothesis presents a significant challenge, as users tend to be far more present

[6]Note that we can distinguish between comments and retweets. Our focus on retweets is motivated by the fact that they provide an unambiguous expression of support, whereas comments may deliver mockery, criticism or any other sentiment (Boyd et al., 2010).

and active during periods of intense political debate, such as the one under consideration.[7] Our statistical model therefore aims to separate *exposure*, the visibility or availability of different types of messages, and the *retweet rate*, the behavioral likelihood of resharing them. This distinction is essential. If we manage to isolate exposure empirically, we can identify the effect of content on the observed changes in retweet behavior, controlling for the fluctuations in overall user activity or online mobilization.

We index tweets (messages) by m and users by i. Each message m has a content and an origin user, denoted $o(m)$. Each interaction is also defined by the linguistic community of the origin user $lan(o(m))$ and of the retweeting user $lan(i)$, where $lan(o(m)), lan(i) \in \{esp, cat\}$ indicate Castellano (esp) and Catalan (cat) language groups, respectively. Denote the retweet decision of message m by individual i at time t as $rt_{m,i,t} \in \{0, 1\}$. Define exposure of individual i to message m at time t as $Z_{m,i,t} \in \{0, 1\}$.

Formally, we can then write that retweets are driven by latent exposure to a message and the period-specific retweet rate given this exposure as:

$$\begin{aligned} \Pr\big(rt_{m,i,t} = 1\big) = {} & \Pr(Z_{m,i,t} = 1) \, \Pr\big(rt_{m,i,t} = 1 \mid Z_{m,i,t} = 1\big) \\ & + \Pr(Z_{m,i,t} = 0) \, \Pr\big(rt_{m,i,t} = 1 \mid Z_{m,i,t} = 0\big). \end{aligned} \tag{1}$$

Note that while the resulting retweets, the 1s on the left-hand-side in equation 1, are observable, the latent factors on the right-hand-side are not. The goal of our structural model is to derive parameter estimates capturing the likelihood of a user i, exposed to message m in period t, retweeting it:

$$\Pr\big(rt_{m,i,t} = 1 \mid Z_{m,i,t} = 1\big) \quad = \quad \alpha_{m,i,t}.$$

Our goal is to track how this retweet behavior across language groups evolves over time.[8] For instance, we want to understand how often Catalan users $lan(i) = cat$ retweet content originating from Castellano users $lan(o(m)) = esp$ and how this changes with the events we study. With a slight abuse of notation, we define four sets of of time-varying parameters: $\alpha_{esp,esp,t}$, $\alpha_{esp,cat,t}$, $\alpha_{cat,esp,t}$ and $\alpha_{cat,cat,t}$ to quantify inter-group retweet rates.

[7] See Appendix Figure B2 showing the total number of hourly tweets between July and November 2017.

[8] Intuitively, one would assume that retweets are impossible for users that are not exposed to content, $Z_{m,i,t} = 0$. However, depending on the model of exposure we empirically observe retweets from unexposed users. For our main empirical results we will focus on exposed users in this case, with robustness checks conducted on the sample of unexposed users.

This allows us to focus on retweets within and across language groups. Total retweets of messages from origin-language $g \in \{cat, esp\}$ by user i in hour t are defined as

$$RT_{g,i,t} := \sum_{m:\, lan(o(m))=g} rt_{m,i,t}$$

and the retweets of messages from origin-language g from all users in language group $h \in \{cat, esp\}$ are defined as

$$RT_{g,h,t} := \sum_{i:lan(i)=h} RT_{g,i,t}.$$

These are observable in the data but driven by both aggregate exposure and group-specific retweet rates. Given that $rt_{m,i,t}$ is observable only when $rt_{m,i,t} = 1$, modeling exposure is essentially an exercise in modeling the decisions not to retweet. The more $rt_{m,i,t} = 0$ we generate from our models of exposure, the lower will be our estimates of $\alpha_{m,i,t}$.

An obvious reason for why the exposure to content would change over time, for example, is that the rate at which tweets by a language group are produced varies over time. If the Catalan language group doubles its production of tweets, the network will *ceteris paribus* be exposed to twice as many messages with Catalan user origin, $lan(o(m)) = cat$. Another important factor for exposure is activity. If users spend more time online scrolling through tweets, they are exposed to more content. Finally, the idea of echo chambers is precisely that exposure to content is distributed very unequally across the user network. Network characteristics and the local activity of a network can therefore matter as well.

To tackle these issues, we propose two simple models of exposure. First, as a simple benchmark, we model exposure uniformly and directly linked to observables - every active user is equally exposed to all active messages. However, this model is obviously not able to study phenomenon such as echo chambers. We, therefore, model the impact of the social network explicitly at the individual level by making the assumption that the retweets in the network of an individual are an important indicator of exposure. In this view, messages travel through the user network by being retweeted by connected users.

2.1.1 Uniform Exposure

We first assume that all active users in our network are equally exposed to all active tweets. Non-active users never retweet. Retweets are independently distributed but retweet rates can differ systematically across time and language groups. Call the set of active users $\mathcal{I}_t$ and the active tweets $\mathcal{M}_t$. We define these through the hourly activity. To be specific, we call a user active if they retweet at least once within an hour

$$\mathcal{I}_t = \left\{ i \middle| \sum_m rt_{m,i,t} > 0 \right\}.$$

and messages are active if they are retweeted at least once

$$\mathcal{M}_t = \left\{ m \middle| \sum_i rt_{m,i,t} > 0 \right\}.$$

In order to derive useful measures of exposure we need to distinguish active messages and users by language group. We denote, for instance, the exposure to Catalan messages as

$$\mathcal{M}_{cat,t} = \left\{ m \middle| \sum_i rt_{m,i,t} > 0 \wedge lan(o(m)) = cat \right\}$$

and the set of active Castellano users as

$$\mathcal{I}_{esp,t} = \left\{ i \middle| \sum_m rt_{m,i,t} > 0 \wedge lan(i) = esp \right\}.$$

Under the assumption of uniform exposure it is straightforward to derive exposure and retweet rates. Assume, for example, that we want to analyze an active Castellano user, $lan(i) = esp$, and how much this user i retweets Catalan content over time. We would then build the sum above as

$$E[RT_{cat,i,t}] = |\mathcal{M}_{cat,t}| \times \alpha_{cat,esp,t}.$$

Denote the total number of active Castellano users as $|\mathcal{I}_{esp,t}|$ the total number of retweets by language group $lan(i) = esp$ of messages originating in language group $lan(o(m)) = cat$ is then simply given by

$$E[RT_{cat,esp,t}] = |\mathcal{M}_{cat,t}| \times \alpha_{cat,esp,t} \times |\mathcal{I}_{esp,t}|$$

where the formulation of the parameters $\alpha_{lan(o(m)),lan(i),t}$ as time-varying should be regarded as a short-cut of modeling the changing nature of content in m and factors like the social desirability to retweeting different types of content in the Castellano language group. Retweet rates can then be estimated as:

$$\widehat{\alpha_{cat,esp,t}} = \frac{RT_{cat,esp,t}}{|\mathcal{M}_{cat,t}| \times |\mathcal{I}_{esp,t}|}.$$

But content draws users onto social media and, hence, we observe different levels of activity by users even if we would hold the users constant. For example, it could be that the events such as the police violence and the terror attack attracted users onto the Twitter app mechanically increasing retweet rates of both Catalan users and Castellano users. This time-varying level of activity is baked into our estimates of $\alpha_{m,i,t}$. A simple way to counter this in the aggregate data is to normalize the retweet rate across language groups by the within retweet rate. Formally, we define the relative retweet rate of Castellano users of content from Catalan users as:

$$\widehat{\alpha^{rel}_{esp,t}} = \frac{\widehat{\alpha_{cat,esp,t}}}{\widehat{\alpha_{esp,esp,t}}} = \frac{RT_{cat,esp,t} \times |\mathcal{M}_{esp,t}|}{RT_{esp,esp,t} \times |\mathcal{M}_{cat,t}|} \tag{2}$$

where the number of active Castellano users $|\mathcal{I}_{esp,t}|$ cancels out. Similarly, we write the relative retweet rate of Catalan users of content from Castellano users as

$$\widehat{\alpha^{rel}_{cat,t}} = \frac{\widehat{\alpha_{esp,cat,t}}}{\widehat{\alpha_{cat,cat,t}}} = \frac{RT_{esp,cat,t} \times |\mathcal{M}_{cat,t}|}{RT_{cat,cat,t} \times |\mathcal{M}_{esp,t}|} \tag{3}$$

where the number of active Catalan users $|\mathcal{I}_{cat,t}|$ cancels out. As levels of user activity vary dramatically, we will always report these relative retweet rates. Note, that this controls for general levels of activity of the language groups on the outlet side, i.e. we implicitly control for how active users are in retweeting and focus on how much they retweet one type of content compared to other content.

The assumption of uniform exposure is not particularly realistic and by design would only allow us to study echo chambers by language groups. This is a problem for the estimation of retweet rates if there are time-varying components that lead to language group-specific variation in exposure. Therefore, we propose a second method to capture exposure at the individual level through the user network.

2.1.2 Network Exposure

We maintain the focus on active users, $i \in \mathcal{I}_t$, and only model the behavior of those exposed to content. However, now we change the definition of who is exposed to derive individual variation in exposure to different messages as these messages are shared in the social network of a user.

Call the network of user i, A_i. This is assumed to be exogenous and fixed. We say that a user j is in this network if $j \in A_i$. All messages which are retweeted by the network of user i are then the set

$$\mathcal{M}_{A_i,t} = \left\{ m \middle| \sum_{j \in A_i} rt_{m,j,t} > 0 \right\}$$

and we notate exposure of user i to messages originating from, for example, Catalan users as

$$\mathcal{M}_{cat,A_i,t} \quad = \quad \left\{ m \middle| \sum_{j \in A_i} rt_{m,j,t} > 0 \wedge lan(o(m)) = cat \right\}.$$

This formulation distinguishes between the language group where a message originates, $lan(o(m))$, and the users that retweet the message from the network of user i, $j \in A_i$. We model how the social network functions as the channel through which the user is exposed to messages originating in the two language groups. If users in the network A_i like some content more than other content, they will be more likely to retweet these messages and form a better communication channel between $o(m)$ and i.

In Section 3 we discuss the simulation of exposure to Catalan content, $\mathcal{M}_{cat,A_i,t}$, and Castellano content, $\mathcal{M}_{esp,A_i,t}$. This allows us to calculate the number of messages an individual user is exposed to as the size of the sets $\mathcal{M}_{cat,A_i,t}$ and $\mathcal{M}_{esp,A_i,t}$. Building sums over this individual exposure over all users from a language group then delivers total exposure. Total exposure of Castellano users to Catalan content, for example, is defined as $\sum_{i \in \mathcal{I}_{esp,t}} |\mathcal{M}_{cat,A_i,t}|$.

Our estimate of retweet rates of Castellano users of messages originating from Catalan users is then

$$\widehat{\alpha_{cat,esp,t}} = \frac{RT_{cat,esp,t}}{\sum_{i \in \mathcal{I}_{esp,t}} |\mathcal{M}_{cat,A_i,t}|}$$

where the denominator is total exposure of Castellano users to content originating from Catalan users. The estimate of the retweet rate of Catalan users of content from Castellano users is

$$\widehat{\alpha_{esp,cat,t}} = \frac{RT_{esp,cat,t}}{\sum_{i \in \mathcal{I}_{cat,t}} |\mathcal{M}_{esp,A_i,t}|}.$$

If language groups are segregated to some degree then network exposure will generate relatively low exposure across groups in the first few minutes of a tweet. Catalan users will, for example, be exposed to small sets of active Castellano messages, $\mathcal{M}_{esp,A_i,t}$, through their network, A_i. But, as Castellano messages enter the Catalan user network more Catalans will have users in their network retweeting messages with Castellano origin leading to larger sets $\mathcal{M}_{cat,A_i,t}$ and more active Catalan users, $\mathcal{I}_{cat,t}$.

As in the model with uniform exposure we then define the relative retweet rate of Castellano users of Catalan content as

$$\widehat{\alpha_{esp,t}^{rel}} = \frac{RT_{cat,esp,t} \sum_{i \in \mathcal{I}_{esp,t}} |\mathcal{M}_{esp,A_i,t}|}{RT_{esp,esp,t} \sum_{i \in \mathcal{I}_{esp,t}} |\mathcal{M}_{cat,A_i,t}|} \tag{4}$$

where now the number of active Castellano users, $|\mathcal{I}_{esp,t}|$, does not cancel out as every user is exposed to different messages through their network. The relative retweet rate of Catalan users of content from Castellano users as

$$\widehat{\alpha_{cat,t}^{rel}} = \frac{RT_{esp,cat,t} \sum_{i \in \mathcal{I}_{cat,t}} |\mathcal{M}_{cat,A_i,t}|}{RT_{cat,cat,t} \sum_{i \in \mathcal{I}_{cat,t}} |\mathcal{M}_{esp,A_i,t}|}. \tag{5}$$

It is important to highlight two challenges to identification implied by our models of exposure. In the network model we assume that two users are connected if they tend to retweet the same tweets. We then treat the resulting network structure as fixed over time. In other words, we use the observed pattern of who retweets what—across our entire five-month sample—as an indicator of the latent social structure underlying communication. Holding connections fixed allows us to study both the dynamics of exposure and retweet behavior rather than modeling the strategic formation of links.[9] To sharpen this focus on the change in retweet behavior over time, we produce a version of retweet rate estimates in which we estimate retweet rates by hour directly at the individual level using regression analysis. Regressions have the advantage that we can control for user fixed effects. This allows us to focus on changes in behavior that cannot be driven by the fixed network structure we estimate from the entire sample.

The second challenge to identification is that we approximate exposure for each user but do not know the Twitter algorithm. This is a common situation in social media studies. To advance on the issue we simulate exposure hour-by-hour using a model in which exposure and retweeting happen simultaneously. However, the simultaneity takes away any hope for causal identification. It is purely a modeling device to make sense of the data. We will return to this issue in the conclusions as it gives a direction for future research.

[9]We leave endogenous network formation to future work, as discussed in De Paula (2020). Link formation will show up in our estimates as changing retweet rates where a user suddenly reacts differently to content she is exposed to.

3 Implementation

In what follows we discuss our dataset and how we implemented the simulation of exposure in this data. We start by discussing the Twitter data we rely on. Our method proceeds as follows: 1) we define user language groups using the tweet language; 2) we define the user network using joint retweet behavior, and 3) we simulate exposure for our two models of uniform and network exposure. Note, the model with uniform exposure can be estimated directly from the raw tweet data once user language groups have been defined. Finally, we introduce a formal test for break points which allows us to economize on figures in the various robustness checks that we run.

3.1 Data

We initially conducted live monitoring of the entire Spanish Twitter-sphere in the year 2017 using Twitter's API access at the time. The resulting dataset was not deemed of high quality enough due to several outages on important days like the 1st of October 2017. However, the data gave us a list of active user accounts in this period. Given the development of the Twitter API in the following years, we were able to hydrate the full available tweet history for all our user ids making use of Twitter API 2.0 under the research access. This gave us a comprehensive data of Twitter activity in Spanish spanning 01-07-2017 to 31-10-2017 from 155,949 user accounts.

To further complement this data, we obtained user names for the most prominent Spanish politicians and incorporated their Twitter activity around that period (appendix Table B3). In this way, we added additional 84 accounts with their full activity. Overall, we obtained 53,022,160 retweets. Due to capacity constraints, we reduce our data including only users that were present both in the periods [2017-07-01 - 2017-10-01] and [2017-10-01 - 2017-11-30] and only those that have more than 5 retweets.

3.2 Defining Language Groups

We classify users into Catalan and Castellano groups based on the language of their retweets, applying a fixed threshold to distinguish the groups. Specifically, a user is labeled as a *Catalan* user if the proportion of their retweets in Catalan exceeds this cutoff. This means that some users that identify themselves as Catalan, perhaps even as independentists, are classified as Castellano if they exclusively retweet in Spanish. On the other hand, Castellano users who do not tweet (speak) in Catalan but have many Catalan contacts and retweet media content from them with a Catalan description will be classified as Catalan users.

It can also be that the Catalan users in our data do not live in Catalonia or that Castellano users do. To decide on the optimal cut-off to define language groups, we used the locations posted by the users in the dedicated tag to check whether they considered themselves from Catalonia, the rest of Spain, or abroad. The threshold for language group identification is then chosen to maximize the out-of-sample accuracy in predicting location for the users who filled out this tag.[10] This ensures consistency between linguistic and geographical affinity in our simulation, as the two are strongly entwined and jointly constitute the regional identity that we see at play here. As a result, we arrive at a cut-off of 4% - users with a higher share of retweets in Catalan are categorized as Catalan.

For the rest of the analysis we exclude the *other* language category (7,022 users representing 6% of the total users) and focus only on the Castellano and Catalan communities. We also focus on the days around the two events. This leads to a total of 116,956 users and 25,860,287 retweets.

A large part of the user network is shown in Figure 3. Each dot represents the entire tweet history of a user in our sample period. Red dots are Castellano users and orange dots are Catalan users. Users in the figure appear closer together if they retweet the same tweets (higher a_{ij}) and if they retweet in the same language.

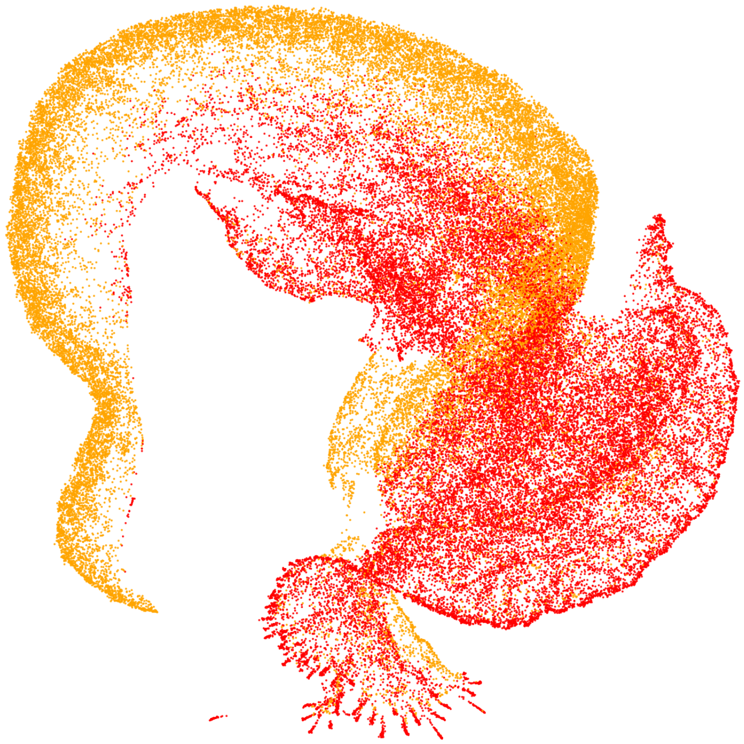
There is significant mixing of what we call Catalan and Castellano users. At the same time, we can clearly identify user language groups in the user networks, i.e. many purely red and orange neighborhoods. This means that there are very clearly defined retweet networks driven by language and joint retweet behavior in which retweeting is concentrated. In other words, our way to define language groups picks up something about segregated retweeting behavior in defined communities: the echo chambers. We will demonstrate in our results section that the distance between dots in Figure 3 even captures elements of the underlying political conflict of Catalan independence and attitudes towards the referendum.

3.3 Building the Network

Consider a static network approach in which we have a set N of users that produce a set of M tweets overall. Denote by X the $N \times M$ matrix that indicates whether a user retweets a tweet in our sample period. Each entry a_{ij} of $A \equiv XX'$, represents the number of retweets that users i and j have in common and total retweet activity for $i = j$.

[10]In Appendix B.1 we describe the location matching methodology together with the results on the optimal cut-off, shown in Figure B1.

Figure 3: User network



Notes: Each dot represents the entire tweet history of a user in our sample period. Users are colored according to our language cutoff rule. Red dots are Castellano users and orange dots are Catalan users. Users appear closer together if they retweet the same tweets (higher a_{ij}, see next subsection) and if the language of their retweets is more similar. More specifically, we use a Uniform Manifold Approximation and Projection (UMAP) of summary statistics of the joint retweet matrix A, the number of retweets in Catalan and Castellano and the number of network members from the two language groups and feed this data.

We model the network by assuming that the lower triangular matrix represents a network of connections between users.[11] We chose an ad-hoc way of defining a connection through the counts a_{ij} and ignore the intensity of links beyond this. Define the social network as $\tilde{A}$ with entries $\tilde{a}_{ii} = 0$ and, for $i \neq j$:

$$\tilde{a}_{ij} = \begin{cases} 1 & if \quad a_{ij} \geq 10 \\ 0 & otherwise. \end{cases}$$

Note that the network is defined through retweet behavior on the entire period 01-07-2017 to 31-10-2017.[12] We choose 10 joint retweets in this entire period and not a lower number to

[11]Note, that being in a *network* here can mean that two users follow the same users or regularly visit the same news webpages and see the same tweets there and then get redirected to Twitter when they retweet the content.

[12]Following the release of the Twitter algorithm (March 31, 2023), the recommendation system has been shown to rely both in in-network and out-network connections to personalize the feed. Specifically, the algorithm assigns half of its weight to content from users that are directly followed and the other half to

simplify the analysis. Still, the final code took six months to simulate the exposure for two weeks of data using this model and so a lower threshold would have been prohibitive for the hardware we had available.[13] A higher threshold would be possible but we tried to include as many weak links as possible as we hypothesized that these to be important drivers of the massive changes we observe. Thus, our social network is represented with $\{N, \tilde{A}\}$ with links defined by users that have retweeted together a minimum of 10 times.

3.4 Simulating Dynamic Exposure

The static network assumes fixed connections over the whole sample period. To model dynamic exposure we use this static network and keep constructing an hour-by-hour account of the three days around our events of interest. For every user-retweet, we keep track of whether other users in the network also retweeted the same tweet in the same hour. The network structure thus allows us to operationalize exposure: any tweet produced or shared by a user gets spread this way in the fixed user network. One mechanism would be that the network of the user will be more likely to see the tweet in their Twitter feed. This allows us to estimate how exposed users (and unexposed users) react to language specific content.

Consider a tweet m produced by some origin $o(m)$ and a vector of who retweeted it in period t: $a_{m,t}$. We simulate exposure as follows: $\tilde{A} \cdot a_{m,t}$ is a vector of size N with entries that represents how many users in a user's network retweeted m in period t. On the operational side, the model can be implemented despite the inherent complexity because exposure can be simulated through simple matrix multiplication. The fact that these are vector operations allows us to simulate billions of exposures.

3.5 Identifying Breaking Points

In order to identify trend changes in retweet rates, we use a simple t-test on the difference in means for the relative α we estimate. Specifically, we let a window of size W run over the dataset to test for a mean change before and after period T in the middle of the window.[14] For a given window size ($W = 8$ hours in our results below), we iteratively partition the data into all observations up to the split date and from the split date onward. We then perform a two-sided t-test to compare the means of these two segments.

out-of-network content that has received engagement, such as retweets, from users in the direct network. This design highlights the importance of interactions within a user's network in determining exposure. See this post for an annotated version of the algorithm.

[13] We run our codes at the University of Barcelona Faculty of Economics Windows Server with 256Gb RAM and 32 cores. The pipeline plate diagram for data generation is shown in Appendix Figure A1.

[14] Zussman and Zussman (2006) study break points in asset prices in a similar way.

We record the absolute value of each t-statistic across the series, and we identify the maximum absolute value among them which represents the point where the potential structural break is strongest. Results are robust to varying window sizes.

4 Results

This section presents our main results for relative retweet rates (α) using the methodology outlined in Sections 2 and 3. We start with results for the referendum and then contrast these with the terror attack. We also discuss robustness to the exposure model and to the introduction of individual fixed effects.

Section 5 will discuss the channels for the changes we observe. We show how specific content was responsible for the observed results and introduce a political dimension in addition to the language communities to study ideological realignment.

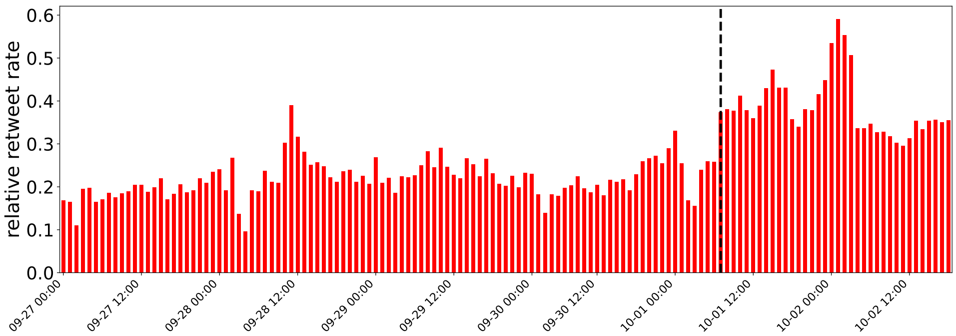
4.1 Main Results

Figure 4 illustrates the changes in retweet behavior during the referendum period, employing the uniform exposure model. The relative retweet rates calculated using Eq. (2) for Castellano users (panel a, red) and Eq. (3) for Catalan users (panel b, orange), reveal significant shifts. Notably, Castellano users substantially altered their retweeting patterns toward Catalan users beginning on the morning of October 1st, coinciding with the unfolding events. Table 1 reports the results of the rolling t-test. The break point is found to be on the 1st of October at 7am. This is indicated as a dashed line in Figure 4 a).

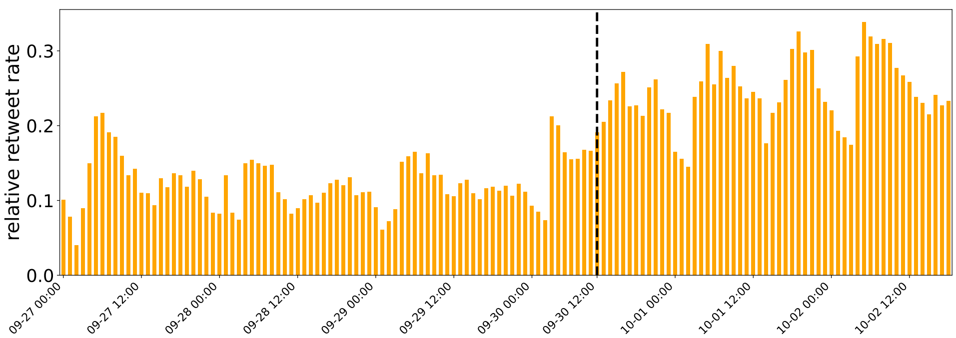
Prior to the referendum and the ensuing police violence, the relative retweet rates stood around 0.2, suggesting that Castellano users were retweeting Catalan users only about one-fifth as frequently as their own language group. However, following the referendum, this rate increased significantly to a range between 0.4 and 0.5. Thus, the Castellano community became notably more receptive to Catalan-origin tweets, more than doubling their cross-language engagement.

Figure 4: Relative retweet rates (uniform exposure) - 1st of October

a) *Castellano* users retweet *Catalan* users more



b) *Catalan* users retweet *Castellano* users more



Notes: Figures show the relative α values calculated as in Eq. (2) and Eq. (3). The red figure shows the share of tweets coming from *Catalan* users retweeted by *Castellano* users over time. The orange figure shows the share of tweets coming from *Castellano* users retweeted by *Catalan* users over time.

Figure 4 b) shows that the reverse was also true. We detect a break point around noon on the 30th of September. Catalan users more than doubled their relative retweet rate of Castellano users from around 10 percent to over 20 percent. Again, this is quite a dramatic change in behavior with the Catalans doubling their retweet rates of Castellano content. The figure also shows that these changes are persistent, i.e. they last until the end of the sample period.

It is tempting to regard the low relative Catalan retweet rates as an aversion to Castellano content but the relative levels in retweet rates across language groups are not easily comparable. The sheer size of the Castellano user group means that more content is produced and shared and this is not controlled for through the uniform exposure model. Still, it is interesting to note that relative rates are far below one for both groups indicating some aggregate "closeness" of both groups suggesting echo chamber dynamics along language lines.

Figure 5 and Table 2 shows results using the network model. We find break points similar to the uniform exposure model. The t-test in Table 2 finds the break points at 9am on the

Table 1: Mean break in relative retweet rates (uniform exposure) - 1st of October

	Castellano users		Catalan users	
	date/hour	t-stat	date/hour	t-stat
uniform exposure (mean alphas)	10-01h07	15.86	09-30h12	17.78

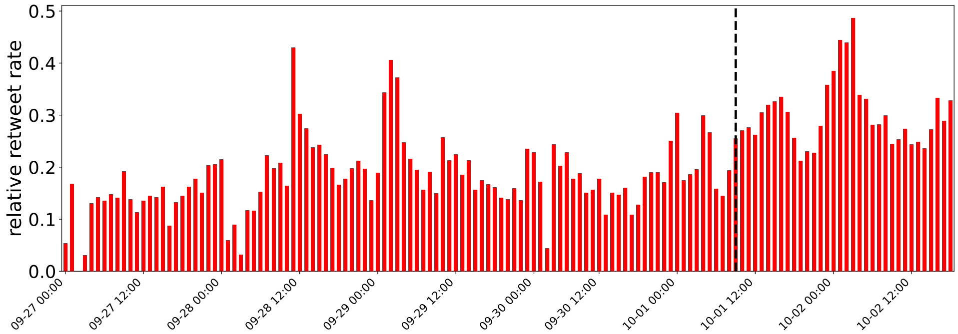
Note: Table shows result of t-test on break in means for relative retweet rates in the period 09-27 at midnight and 10-02 at midnight using the whole sample. The column "date" reports the date with the maximum t-statistic in this period and the column "t-stat" reports the t-statistics. Under "Castellano users" we report test results for relative retweet rates of *Castellano* users retweeting *Catalan* users relative to retweet rates of *Castellano* users retweeting *Castellano* users. Under "Catalan users" we report test results for relative retweet rates of *Catalan* users retweeting *Castellano* users relative to retweet rates of *Catalan* users retweeting *Catalan* users.

1st of October and at 11am on the 30th of September. Both user groups again more than doubled their relative retweet rates across groups. For Castellano users retweeting Catalans we again find a significant increase of relative rates (from around 0.2 to 0.35). The relative retweet rate of Catalan users of Castellano user content is now estimated to be much higher in levels and more noisy with a surge in the early morning hours of the 29th of September - where retweet activity is extremely low overall. But the surge is still strongly significant and meaningful from around 60 percent on the 30th of September to 120 percent in the afternoon of the 1st of October.

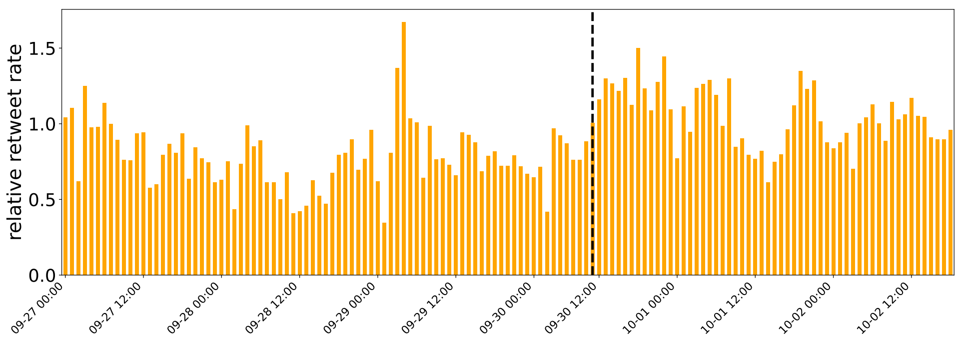
The level shift in Figure 5 b) when compared to Figure 4 b) indicates that we obtain a very different image of openness for the Catalan user group through the network model. At the height of the crisis of the 1st of October, Catalans were equally likely to retweet content coming from Catalan and Castellano users when they were exposed to it. This finding is driven by a more realistic model of exposure in which not all Catalan users are exposed to all Castellano content equally. The level of the relative retweet rates of the Castellano users of Catalan content in Figure 5 a) is very similar to the uniform model in Figure 4 a).

Figure 5: Relative retweet rates (network exposure) - 1st of October

a) *Castellano* users retweet *Catalan* users more



b) *Catalan* users retweet *Castellano* users more



Notes: Figures show the relative α values calculated as in Eq. (4) and Eq. (5) conditional on exposed users. The red figure shows the share of tweets coming from *Catalan* users retweeted by *Castellano* users over time. The orange figure shows the share of tweets coming from *Castellano* users retweeted by *Catalan* users over time.

Table 2: Mean break in relative retweet rates (network exposure) - 1st of October

	Castellano users		Catalan users	
	date/hour	t-stat	date/hour	t-stat
network exposure (mean alphas)	10-01h09	8.89	09-30h11	7.13

Note: Table shows result of t-test on break in means for relative retweet rates in the period 09-27 at midnight and 10-02 at midnight conditioning on exposed users. The column "date" reports the date with the maximum t-statistic in this period and the column "t-stat" reports the t-statistics. Under "Castellano users" we report test results for relative retweet rates of *Castellano* users retweeting *Catalan* users relative to retweet rates of *Castellano* users retweeting *Castellano* users. Under "Catalan users" we report test results for relative retweet rates of *Catalan* users retweeting *Castellano* users relative to retweet rates of *Catalan* users retweeting *Catalan* users.

It is tempting to regard the dramatic change in retweet rates across groups as a feature of an extremely newsworthy event taking place in Catalonia that somehow reconfigured communication in the retweet network. This could at least explain an increase in relative retweet rates of Catalan content by Castellano users. To test this idea, we study the Terror Attack which took place on the afternoon (14:56 UTC) of the 17th of August 2017 in Barcelona. We expect relative retweet rates to increase in the aftermath of the attack.

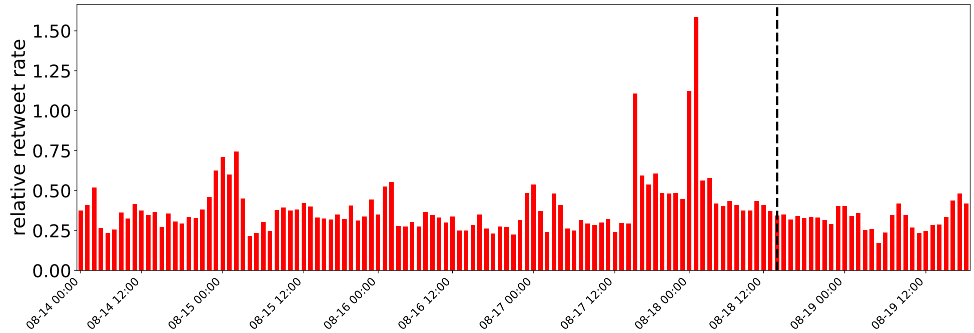
Figure 6 and Table 3 present the results for the terror attack event using the uniform exposure model. Here, Castellano users briefly showed an increase in retweeting Catalan content immediately following the attack. However, unlike the referendum scenario, this spike was temporary and did not result in a sustained shift in retweet behavior. Thus, the impact of the terror attack on cross-language engagement was short-lived rather than lasting. In fact, the mean break method indicates the most significant shift is downward in the aftermath of the terror attack.[15] We find null-results for Catalans. There is no discernible shift in Catalans retweeting Castellano users.

This underlines the role of events as drivers of different behavior. While the newsworthiness of the terror attack drove increased tweet production from Catalan origin, it only lead to an increase in retweet rates in the Castellano part of the network after the event for those exposed. The dramatic increase of Castellano users retweeting Catalan users appears as much more a supply effect when decomposed into exposure and retweet rates in both the uniform and network model. In addition, there is no evidence for increasing retweet rates from Catalan users to Castellano users. This is important as it suggests the increased retweet rates after the 30th of September are due to specific incentives to retweet Castellano messages in the context of the referendum and police violence.

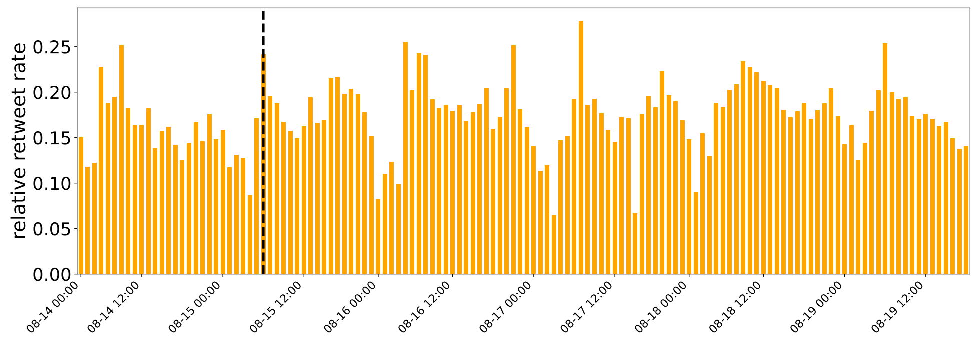
[15]The method does suggest an upward shift at the time of the terror attack as well but it is smaller. The break is consistent under the network exposure model, shown in Appendix Figure B5 and Table B1.

Figure 6: Relative retweet rates (uniform exposure) - Terror attack

a) *Castellano* users retweet *Catalan* users more



b) *Catalan* users retweet *Castellano* users no more



Notes: Figures show the relative α values calculated as in Eq. (2) and Eq. (3). The red figure shows the share of tweets coming from *Catalan* users retweeted by *Castellano* users over time. The orange figure shows the share of tweets coming from *Castellano* users retweeted by *Catalan* users over time. Dashed line represents the date with maximum absolute t-statistic from the break test.

Table 3: Mean break in relative retweet rates (uniform exposure) - Terror attack

	Castellano users		Catalan users	
	date/hour	t-stat	date/hour	t-stat
uniform exposure (mean alphas)	08-18h14	-1.99	08-15h06	2.05

Note: Table shows result of t-test on break in means for relative retweet rates in the period 08-10 and 08-24 at midnight using the whole sample. The column "date" reports the date with the maximum t-statistic in this period and the column "t-stat" reports the t-statistics. Under "Castellano accounts" we report test results for relative retweet rates of *Castellano* users retweeting *Catalan* users relative to retweet rates of *Castellano* users retweeting *Castellano* users. Under "Catalan accounts" we report test results for relative retweet rates of *Catalan* users retweeting *Castellano* users relative to retweet rates of *Catalan* users retweeting *Catalan* users.

Before we move on, it is important to discuss the interpretation of our findings. In the uniform exposure model we assume that all users are exposed to all active content and can react to it. In the network model we simulate exposure on a network that we define through joint retweet behavior, i.e. on a network that is endogenous to the behavior we want to study. The network model, therefore, captures the reorientation of users to new message origins through the changes in retweet rates we estimate. In other words, our definition of the network allows us to model the breaking of language group echo chambers explicitly as changes in retweet rates holding the network fixed.

Following this idea, we show in the robustness section that our network results are robust to user fixed effects, i.e. the same users suddenly started to retweet different users within their own network. The fixed effects control for the network position, and we are able to study changes in retweet rates *within* user. This lends credibility to our interpretation of our main results as a changing openness to content from the other language group starting with the referendum and the police violence.

4.2 Additional Results and Robustness Checks

In this section, we describe additional results that place our main results in a better context and test the robustness of our findings. We start with an analysis of exposure in the network model.

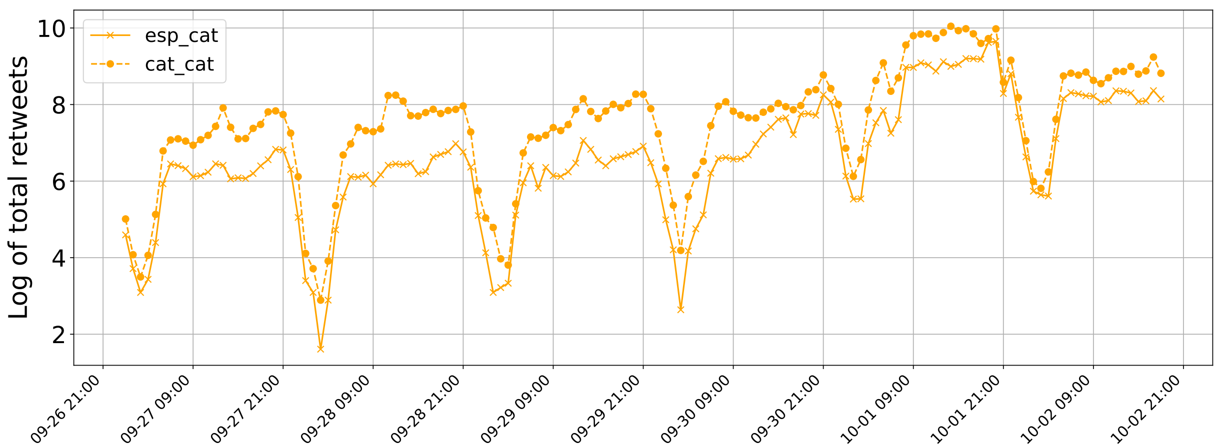
4.2.1 Network Exposure over Time

Changes in retweet behavior induced by changes in content led to dramatic shifts in exposure as measured by the network model. Figures 7 and 8 show the log total number of Catalan and Castellano users exposed by language of origin, respectively. We now use this variation to gauge the relative size of the time variation absorbed by the exposure terms in Eq. (2) and Eq. (3).

We can clearly observe the dramatic increase of exposure in all origin-destination combinations. Figure 7 shows how exposure of Catalan users to tweets from Castellano users approaches the same level as the exposure to tweets from Catalan users on the evening of the 30th of September. The difference decreases and does not recover completely afterwards. While the pattern is not as clear for Castellano users (Figure 8), we also observe a persistent decrease in the log difference of exposure. Relative exposure across language groups increases considerably after the 30th of September.

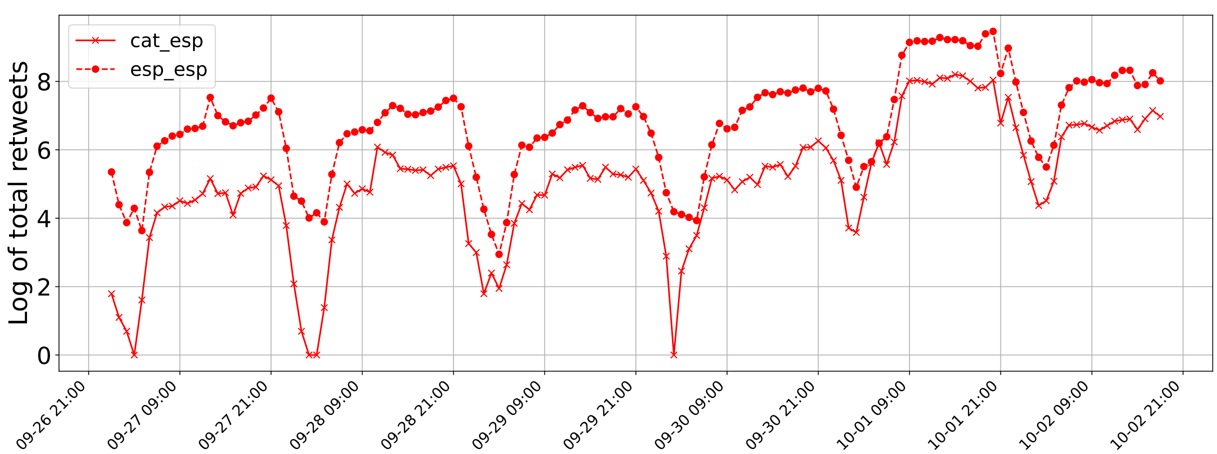
We can use the network model to create counterfactual scenarios simulating overall retweets

Figure 7: Catalan exposure by language of origin - 1st of October



Notes: The figure shows the log total of tweets of exposed Catalan users by language of origin.

Figure 8: Castellano exposure by language of origin - 1st of October



Notes: The figure shows the log total of tweets of exposed Castellano users by language of origin.

under alternative exposure numbers. In Appendix B.4 we calculate the overall cross-language retweets in the 24 hours after the break points identified in Table 2 in a counterfactual scenario where exposure follows the same pattern as in 24 hours before the break points. Overall, retweets across would have been 80 percent lower for retweets by Catalan users of Castellano content and 74 percent lower for Castellano users retweeting Catalan content. Within-group exposure also increased dramatically but the counterfactuals show a decrease of "only" 64 and 67 percent if we assume the exposure of the previous days.

In this way, the network model allows us to illustrate how the changes in retweet behavior led to a completely different experience for Twitter users. The counterfactual predictions demonstrate the dramatic increase in visibility the two language groups gained of each other. It also lends support to the idea of analyzing relative retweet rates as dramatically shifting exposures led to a saturation of users.

4.2.2 Regressions at the User Level

Our exposure models allow us to simulate users who are exposed to a message but do *not* retweet it. At the aggregate level, these simulated non-retweets are simply counted and included in our calculations of total exposure across language groups. Additionally, by simulating exposure at the individual level for a subsample, we can perform a robustness check using individual fixed effects.

The use of user fixed effects is an important robustness check in the face of the endogeneity problem inherent in the way we model uniform and network exposure. Connections in the network exposure in particular are defined through joint retweeting in the three months of our dataset. The fixed effects control for the network position of users in the period and we are therefore able to study changes in retweet rates *within* user as they open up to new type of user origins.

We run regressions on a sample of 15,000 users four days before and two days after the main event for the uniform and network exposure models. We filter exposed individuals and expand the dataset by millions of rows in which a user is exposed to a message but does not retweet it. For our 15,000 users this leads to a dataset with 1.5 billion observations in the uniform exposure model and close to 0.1 billion observations in the network exposure model.

To obtain estimates of exposure, we then run regressions at the user/retweet level with message origin as main explanatory variables. The key advantage to moving to the individual level is that we can run regressions with user fixed effects. This filters out all selection effects and only tracks the reorientation of behavior *within* users over time as events unfolded. The regression we run is:

$$\begin{aligned} rt_{imt} = &\sum_{\tau} \beta^{\tau}_{cat,cat} \left(d_t(\tau) \times cat\_cat_{imt}\right) + \sum_{\tau} \beta^{\tau}_{cat,esp} \left(d_t(\tau) \times cat\_esp_{imt}\right) + \\ &\sum_{\tau} \beta^{\tau}_{esp,esp} \left(d_t(\tau) \times esp\_esp_{imt}\right) + \sum_{\tau} \beta^{\tau}_{esp,cat} \left(d_t(\tau) \times esp\_cat_{imt}\right) + \gamma_i + \epsilon_{imt} \end{aligned} \tag{6}$$

where $d_t(\tau)$ are hourly time dummies and $lan(o(m))\_lan(i)_{imt}$ are interaction of dummy variables indicating the dyadic language relationships between the original user and the destination user for a given message m at time t. These exposure dummies are given by the respective model. In the uniform exposure model we assume that all active users are automatically exposed to all messages that are active at time t. In the network model, active users are only exposed if their direct network retweeted a message in the same hour t. We always show results with and without individual fixed effects γ_i. Analogous to Eq. (3) and Eq. (2), we then calculate the ratios $\frac{\hat{\beta}^{\tau}_{cat,esp}}{\hat{\beta}^{\tau}_{cat,cat}}$ and $\frac{\hat{\beta}^{\tau}_{esp,cat}}{\hat{\beta}^{\tau}_{esp,esp}}$.

Table 4: Robustness to Individual Fixed Effects - 1st of October

	Castellano users		Catalan users		Fixed effects	N
	date/hour	t-stat	date/hour	t-stat		
uniform exposure	10-01h07	10.07	9-30h12	6.80		1,592,419,876
uniform exposure	10-01h07	9.15	9-30h15	5.77	✓	1,592,419,876
network exposure	10-01h09	6.20	9-30h11	4.69		97,934,738
network exposure	10-01h09	4.63	9-30h11	3.81	✓	97,934,738

Note: Table shows result of t-test on break in means for relative retweet rates in the period 09-29 at midnight and 10-02 at midnight for the regression analysis. The column "date" reports the date with the maximum, absolute t-statistic in this period and the column "t-stat" reports the t-statistics. Under "Castellano users" we report test results for estimated coefficients *Castellano-Catalan* relative to *Castellano-Castellano* users. Under "Catalan users" we report test results for estimated coefficients *Catalan-Castellano* relative to *Catalan-Catalan* users. Column (3) marks results that have been obtained under individual fixed affects. The last column reports the total number of observations.

Table 4 summarizes results from regressions using both the uniform exposure and network exposure models, conducted with and without individual fixed effects. The identified break points without fixed effects align closely with those observed using the complete sample, supporting the robustness and generalizability of these findings. Results with fixed effects are remarkably robust - only the cutoff for Catalan users retweeting Castellano users under uniform exposure changes by three hours. This robustness underscores that we are able to capture changes in user behavior that were correlated with changes in content exposure and not changes in self-selection onto Twitter. We can interpret our main results as a changing openness to content from the other language group starting with the referendum and the police violence.

5 Evidence on the Drivers of Changing Retweet Rates

We have thus far presented evidence on a sudden change in retweet behavior in our two language communities in which both groups open to each other in the wake of the political events around the 1st of October 2017 but only in one direction (if at all) with the terror attack in Barcelona on the 17th of August of the same year. In this section we analyze the channels driving this change. To better understand the dramatic shifts in communication, we first conduct an analysis of the most retweeted content around the 1st of October. We also provide evidence for the strengthening of retweets along ideological lines.

5.1 Content Analysis

What is the content that drove this relative opening to the other language group? In Appendix Tables B5 to B8 we report the most reshared messages with Castellano and Catalan origin for the 30th of September and 1st of October.

Two, similarly phrased, tweets in these tables stand out and symbolize the mechanisms underlying the opening of language groups on the 1st of October. The two tweets came from two different journalists. While both were working for Spanish National TV (RTVE), one was Catalan and is classified as a Catalan speaker in our dataset, whereas the other was Spanish and classified accordingly (Castellano). The message written by the Catalan user reads as follows:

> *"Shame, that's what I feel as a RTVE journalist seeing the biased treatment that the news is giving. This is manipulation."*[16]

The message of her Castellano colleague, with very similar wording, was the most retweeted message by Catalan users.[17] We observe other similar tweets in our dataset as a group of RTVE journalists decided to coordinate and create messages that were retweeted broadly, transcending language group boundaries.

We also examine the extent to which these top tweets spread across the language communities. Figure 9 shows a network graph representation for tweets from Appendix Tables B5 to B8. We contrast a tweet that was retweeted at a high rate by the Castellano language community in Figure 9 a) with a tweet that resonated much less in Figure 9 b). The network graphs are generated using the full network information so that the position of users in the graph is somewhat informative about their general retweet behavior. Castellano users close to the Catalan language group tend to retweet Catalan content at a higher rate.

The figures, therefore, capture our idea of exposure extremely well. The further away from the yellow/red interface a tweet travels the "deeper" it penetrates a language group and the more exposure that group has to the content of the tweet. Figure 9 a) shows the spread of a condemnation of the police violence by a Catalan twitter account who shared a picture of a voting ballot with a "no" vote, i.e. against independence, that had the following sentence

[16] All tweets have been translated into English using ChatGPT.

[17] *"As a RTVE journalist, I feel ashamed of the treatment RTVE gave to 1-O. Again, public television at the service of the government #SOS."*

written on the ballot paper:

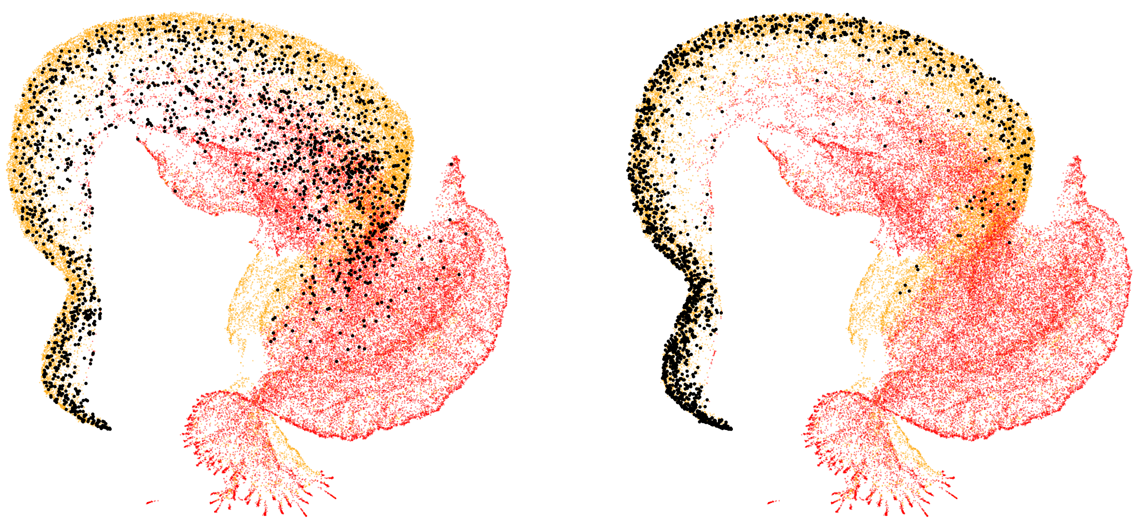
> *"This referendum is illegitimate but I cannot stay at home while they beat up my people."*

This message was retweeted deeply into the Castellano user network, probably because it made explicit the separation of the issue of the referendum and the opposition to the police violence.

Figure 9: Tweet examples of Catalan users

a) Tweet example 1st of October: Catalan user condemns police violence

b) Tweet example 30th of September: jab by Catalan politician



Notes. The English translation of the tweet shown in panel a is: *"This ballot was in my ballot box, at Rafael de Casanovas school in Llefià. Look what your batons achieve #pride #shame `https://t.co/ToAIZc6CKg`"*. The picture shows a ballot with a "no" vote. The handwritten text reads: *"This referendum is illegitimate but I cannot stay home while they beat up my people."* The English translation of the tweet shown in panel b is: *"This photograph is the most eloquent speech to explain what is at stake in tomorrow's referendum `https://t.co/apmCThjSGw`"*. The picture shows unionist protesters trying to take down a sign *"more democracy"* written in Catalan.

Figure 9 b) shows the retweets for a tweet by the Catalan president Carles Puigdemont on the the 30th of September. The tweet showed unionist protesters trying to take down a sign *more democracy* in Catalonia. While the tweet was heavily retweeted, even by Castellano users, it did not penetrate the Castellano network in the same way. In the Appendix Figure B6 we show two similar cases for Castellano users spreading more and less into the Catalan user network.[18]

From the other messages in Tables B5 to B8 it is clear that the surge in Catalan users'

[18] The two users are Catalans which suggests some mis-classification is possible where Catalans chose to communicate in Spanish.

retweeting of Castellano content on the 30th of September reflects a response to the escalating political crisis surrounding the Catalan independence referendum. A significant portion of the most retweeted messages involved real-time updates on police deployments and security measures, suggesting that Catalan users actively sought to monitor law enforcement activities.[19]

Additionally, the retweeted content prominently featured expressions of support for the referendum as a democratic right and critiques of right-wing demonstrations. Messages highlighting police actions, symbolic resistance, and perceived threats to democratic norms originating in Castellano users resonated with Catalan users. By amplifying such messages, they not only accessed critical information but also reinforced a narrative of resistance against a right-wing, repressive central state.

Our hypothesis from this analysis is that Twitter communication on the 30th of September and 1st of October reflects the formation of a new political coalition of left-wing users across language groups, united by the outrage at the police violence, commanded by a right-wing central government, aggrieved by the illegal vote. The linguistic echo chambers broke down as the conflict shifted away from language identity, making the Castellano/Catalan divide a less relevant marker. We test this hypothesis in the next section.

5.2 Methodology for Classifying User Ideology

In order to test the hypothesis of re-alignment, we classify users into a left-right wing political spectrum and check whether we find heterogeneous responses when language communities are interacted with the inferred political affiliations.

Previous work has inferred political ideology from Twitter follower relations (Barberá, 2015). We use a simple approach to position users on a left-right ideological spectrum based on the politicians they follow, considering only direct links. First, we collect Twitter usernames of the most prominent Spanish and Catalan politicians and match them with their party affiliations (Table B3). Next, we retrieve all Twitter followers of these politicians and match them with our sample users. This strategy reveals that 60% of our sample (or 68,603 users) follow at least one politician.

We asked ChatGPT 4.0 to evaluate each political party on a left-right spectrum with values in $\{-1, -0.5, 0, 0.5, 1\}$. The scores assigned to each party are shown in Table B4. Finally, we computed an ideology score for each user by averaging the scores of the politicians they follow. We will call users "left-wing" if they have a negative score and vice-versa.

[19] This aligns with the sense of urgency and threat among Catalans that we, as locals, can attest to.

Figure B7 illustrates the estimated density of ideology scores across our entire sample, which is skewed towards the left ideology: 65% of Catalans are left-wing compared to 51% for the Castellano users (Figure B8). This is consistent with survey estimates in the same time period.[20]

5.3 Heterogeneous Changes of Retweet Rates by Ideology

We present the results for the relative alphas as outlined in Eq. (3) and Eq. (2), under the uniform exposure assumption, but now derive heterogeneous effects by conditioning on the ideology of the origin users.[21] This way, we test whether results of increased sharing across language groups were driven by content from a specific part of the political spectrum.

Panels a) and b) of Figure 10 show the relative alphas for Catalan users resharing left-wing and right-wing Castellano users; respectively. Catalan users move from relative α's of 30 percent to nearly 70 percent after the police violence. Figure b) shows that this trend break is not present for content originating from right-wing Castellano users. Mean break results reported in Table 5 confirm the significant, upward break on the midnight of the 30th of October for the left-wing origins and the absence of any upward trend break for the right-wing origins.

A similar pattern is shown in Figures 11 a) and b) for Castellano users retweeting left-wing and right-wing Catalans. Castellano users moved from relative α's of 20 percent to around 40 percent on the 1st of October. The absence of break for the Castellano reposting right-wing Catalans confirms the cross-linguistic realignment along ideological lines by left-wing users.

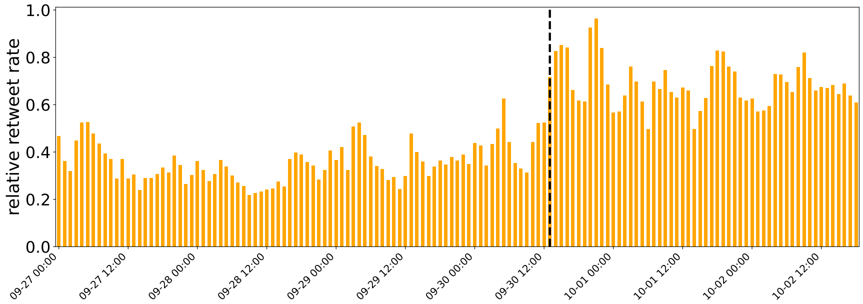
In summary, for those users where we can track political leaning, we find that the changes in communication across language groups are mainly driven by content originating from left-wing accounts in both language groups. Relative retweet rates for content originating from left-wing users shoot up with the events of the 1st of October. There are no equivalent changes in the retweet rates of right-wing content.

[20]We calculated the mean ideology using survey data from various years from CIS for the month of June and July finding always distributions skewed to the left. Scores are not directly comparable since the questions differ across survey weights (intention vs last vote) and the parties considered also vary.

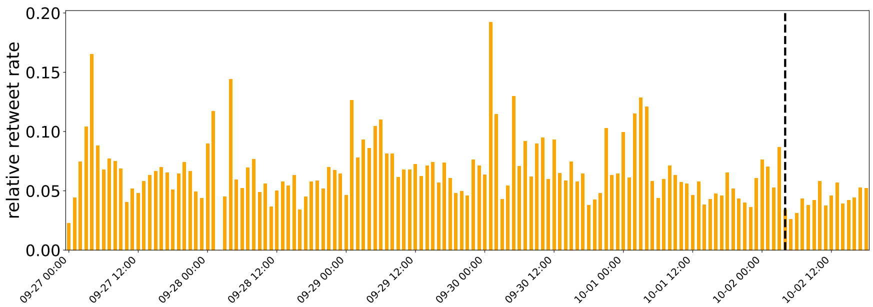
[21]We focus on the uniform exposure model for computational reasons. The equivalent network results have not been generated.

Figure 10: Relative retweet rates (uniform exposure) of Catalan users conditional on ideology - 1st of October

a) *Catalan* users retweet left-wing *Castellano* users more



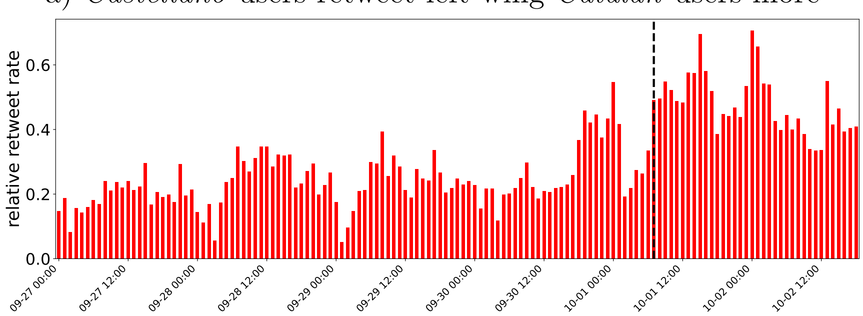
b) *Catalan* users retweet right-wing *Castellano* users no more



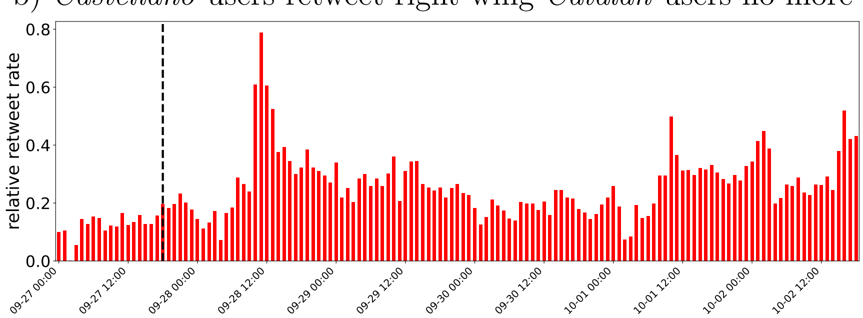
Notes: Figures show relative α values calculated as in Eq. (2) and Eq. (3). The top (bottom) panel shows the share of tweets from left-wing (right-wing) *Castellano* users retweeted by *Catalan* users over time.

Figure 11: Relative retweet rates (uniform exposure) of Castellano users conditional on ideology - 1st of October

a) *Castellano* users retweet left-wing *Catalan* users more



b) *Castellano* users retweet right-wing *Catalan* users no more



Notes: Figures show relative α values calculated as in Eq. (2) and Eq. (3). The top (bottom) panel shows the share of tweets from left-wing (right-wing) *Catalan* users retweeted by *Castellano* users over time.

Table 5: Mean break in relative retweet rates (uniform exposure) - 1st October

	Catalan users (left origin)		Catalan users (right origin)	
	date/hour	t-stat	date/hour	t-stat
uniform exposure (mean alphas)	09-30h13	10.75	10-02h04	-3.04
	Castellano users (left origin)		Castellano users (right origin)	
uniform exposure (mean alphas)	10-01h07	5.61	09-27h18	1.665

Note: Table shows result of t-test on break in means for relative retweet rates in the period 09-27 at midnight and 10-02 at midnight. The column "date" reports the date with the maximum t-statistic in this period and the column "t-stat" reports the t-statistics. Under "Castellano users (left origin)" we report test results for relative retweet rates of all *Catalan* users retweeting *left Castellano* users relative to retweet rates of all *Catalan* users retweeting all *Catalan* users. Under "Catalan users (right origin)" we report test results for relative retweet rates of *Catalan* users retweeting *right Castellano* users relative to retweet rates of *Catalan* users retweeting *Catalan* users. Under "Castellano users (left origin)" we report test results for relative retweet rates of all *Castellano* users retweeting *left Catalan* users relative to retweet rates of all *Castellano* users retweeting all *Castellano* users. Under "Castellano users (right origin)" we report test results for relative retweet rates of *Castellano* users retweeting *right Catalan* users relative to retweet rates of *Castellano* users retweeting *Castellano* users.

6 Conclusions

Our paper analyzes online communication in the context of political events in Catalonia. We use a unique dataset of retweets from over 120,000 users spanning the months surrounding the referendum of independence held on the 1st of October 2017. Our analysis focuses on changes in cross-linguistic retweeting behavior, distinguishing between exposure to and engagement with content. This allows us to isolate how key political events reshape social media networks, breaking down language-based divisions and fostering new ideological coalitions.

We interpret our findings through the lens of the emergence of a pan-linguistic political coalition in Spain during this turbulent period. The referendum and, more significantly, the state's response involving police violence were associated with a shift from linguistic-based segmentation towards ideological realignment. This is important because the shift in alignments, assisted by the PP's corruption troubles, laid the groundwork for Pedro Sánchez's (Socialist Party) minority government in 2018. The absence of a similar pattern regarding the terror attack demonstrate that it is the special character of these political events that led to the reorientation of online networks.

The concept of Catalan non-violence provides a useful interpretative frame for understanding this mobilization. In this view, non-violent protests during the referendum put the focus on the coercive actions of the state, fostering broader public sympathy and cross-group solidarity. The ideological backlash against police violence underscores how actions by the state can inadvertently mobilize opposition through non-violent resistance. At times of strong de-democratization movements in part of the world, this is an important lesson.

Our study highlights the potential for political events to reshape online network structures, even in environments often dismissed as echo chambers. Our findings suggest that we need to explicitly model the motivation for online communication to study the role of social media in generating echo chambers. Future research could then try to understand communication dynamics in other contexts marked by ethno-linguistic tensions and varying institutional frameworks. Understanding how political institutions and events mediate online polarization can inform both policy interventions and platform governance in mitigating digital and offline divides.

An important weakness of our data is that we can only approximate exposure. Our approach should be regarded as an attempt of conceptualizing exposure in a way that can be implemented empirically in large datasets. Using the same observational type of data, exposure could be simulated dynamically by increasing the granularity on the time dimension and assuming that past exposure drives retweet behavior. The effect of exposure could be better identified with such a dynamic model. However, network exposure expands extremely fast. One would need to re-compute exposure at least minute by minute and estimate a much larger set of parameters to identify the effect of past exposure on retweeting behavior. We leave this to future research.

References

Alesina, A., Baqir, R., and Easterly, W. (1999). Public goods and ethnic divisions. *Quarterly Journal of Economics*, 114(4):1243–1284.

Allcott, H., Braghieri, L., Eichmeyer, S., and Gentzkow, M. (2020). The welfare effects of social media. *American Economic Review*, 110(3):629–76.

Allcott, H., Gentzkow, M., Mason, W., Tucker, J. A., et al. (2024). The effects of facebook and instagram on the 2020 election: A deactivation experiment. *Proceedings of the National Academy of Sciences*, 121(21).

Barberá, P. (2020). Social media, echo chambers, and political polarization. *Social media and democracy: The state of the field, prospects for reform*, pages 34–55.

Barberá, P. (2015). Birds of the same feather tweet together: Bayesian ideal point estimation using twitter data. *Political Analysis*, 23(1):76–91.

Besley, T. and Burgess, R. (2002). The political economy of government responsiveness: Theory and evidence from india. *The Quarterly Journal of Economics*, 117(4):1415–1451.

Blattman, C. (2023). *Why We Fight: The Roots of War and the Paths to Peace*. Penguin Books Ltd, London.

Bowen, T. R., Dmitriev, D., and Galperti, S. (2023). Learning from Shared News: When Abundant Information Leads to Belief Polarization*. *The Quarterly Journal of Economics*, 138(2):955–1000.

Boxell, L., Gentzkow, M., and Shapiro, J. M. (2017). Is the internet causing political polarization? evidence from demographics. Technical report, National Bureau of Economic Research.

Boyd, D., Golder, S., and Lotan, G. (2010). Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In *Proceedings of the 43rd Hawaii International Conference on System Sciences (HICSS)*, pages 1–10. IEEE.

Bramoullé, Y., Kranton, R. E., and D'Amours, M. (2014). Strategic interaction and networks. *American Economic Review*, 104(3):898–930.

Burgess, R., Jedwab, R., Miguel, E., Morjaria, A., and Padro i Miquel, G. (2015). The value of democracy: Evidence from road building in kenya. *American Economic Review*, 105(6):1817–1851.

Chatterjee, B. B. (1974). Search for an appropriate game model for gandhian satyagraha. *Journal of Peace Research (JPR)*, 11(1):21–29.

Cukierman, A. and Tommasi, M. (1998). When does it take a nixon to go to china? *American Economic Review*, pages 180–197.

Dandekar, P., Goel, A., and Lee, D. T. (2013). Biased assimilation, homophily, and the dynamics of polarization. *Proceedings of the National Academy of Sciences*, 110(15):5791–5796.

De Paula, Á. (2020). Econometric models of network formation. *Annual Review of Economics*, 12(1):775–799.

Depetris-Chauvin, E., Durante, R., and Campante, F. (2020). Building nations through shared experiences: Evidence from african football. *American Economic Review*, 110(5):1572–1602.

Durante, R. and Zhuravskaya, E. (2018). Attack when the world is not watching? u.s. news and the israeli-palestinian conflict. *Journal of Political Economy*, 126(3):1085–1133.

Eisensee, T. and Strömberg, D. (2007). News droughts, news floods, and u.s. disaster relief. *Quarterly Journal of Economics*, 122(2):693–728.

Enikolopov, R., Petrova, M., and Sonin, K. (2020). Social media and protest participation: Evidence from russia. *Econometrica*, 88(4):1473–1512.

Esteban, J., Mayoral, L., and Ray, D. (2012). Ethnicity and conflict: An empirical study. *American Economic Review*, 102(4):1310–1342.

Fearon, J. D. and Laitin, D. D. (2000). Violence and the social construction of ethnic identity. *International Organization*, 54(4):845–877.

Gentzkow, M. and Shapiro, J. M. (2011). Ideological segregation online and offline. *The Quarterly Journal of Economics*, 126(4):1799–1839.

Golub, B. and Jackson, M. O. (2012). How homophily affects the speed of learning and best-response dynamics. *Quarterly Journal of Economics*, 127(3):1287–1338.

Levy, G. and Razin, R. (2019). Echo chambers and their effects on economic and political outcomes. *Annual Review of Economics*, 11:303–328.

Levy, R. (2021). Social media, news consumption, and polarization: Evidence from a field experiment. *American Economic Review*, 111(3):831–870.

Michalopoulos, S. and Papaioannou, E. (2016). The long-run effects of the scramble for africa. *American Economic Review*, 106(7):1802–1848.

Montalvo, J. G. and Reynal-Querol, M. (2005). Ethnic polarization, potential conflict, and civil wars. *American Economic Review*, 95(3):796–816.

Muñoz, J. and Anduiza, E. (2019). 'if a fight starts, watch the crowd': The effect of violence on popular support for social movements. *Journal of Peace Research*, 56(4):485–498.

Müller, K. and Schwarz, C. (2020). Fanning the Flames of Hate: Social Media and Hate Crime. *Journal of the European Economic Association*, 19(4):2131–2167.

Posner, D. N. (2005). *Institutions and Ethnic Politics in Africa*. Cambridge University Press.

Racek, D., Davidson, B. I., Thurner, P. W., Zhu, X. X., and Kauermann, G. (2024). The russian war in ukraine increased ukrainian language use on social media. *Communications Psychology*, 2:1.

Sen, A. (2006). *Identity and Violence: The Illusion of Destiny*. W.W. Norton & Company, New York.

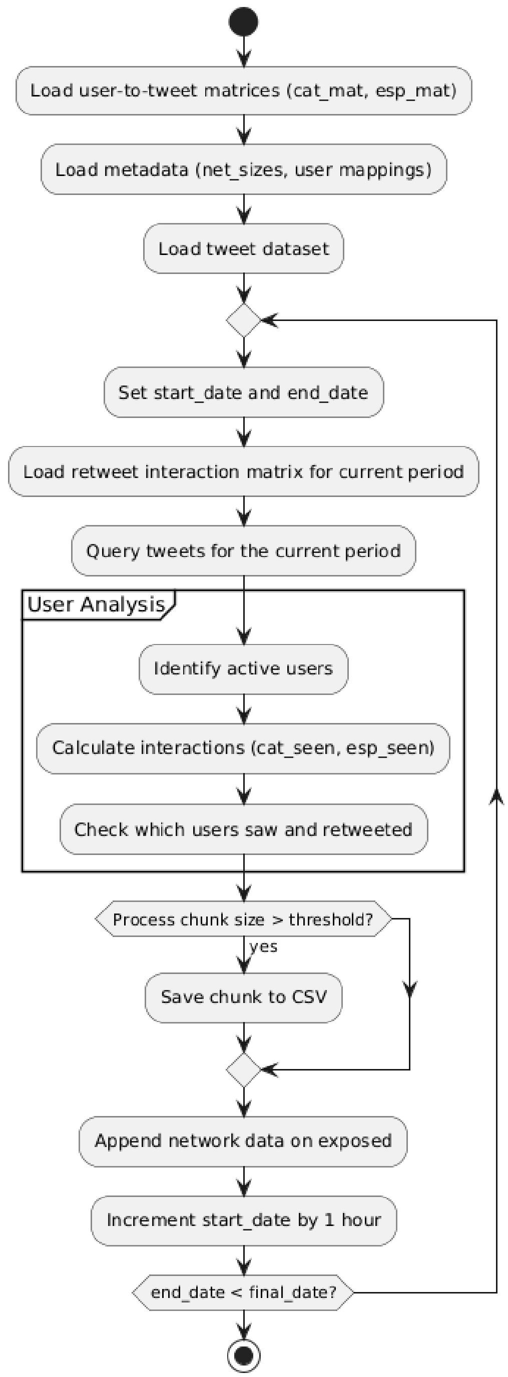
Sharp, G. (1973). *The Politics of Nonviolent Action*. Porter Sargent.

Zhuravskaya, E., Petrova, M., and Enikolopov, R. (2020). Political effects of the internet and social media. *Annual Review of Economics*, 12(1):415–438.

Zussman, A. and Zussman, N. (2006). Assassinations: Evaluating the effectiveness of an israeli counterterrorism policy using stock market data. *Journal of Economic Perspectives*, 20(2):193–206.

A Pipeline

Figure A1: Pipeline network creation

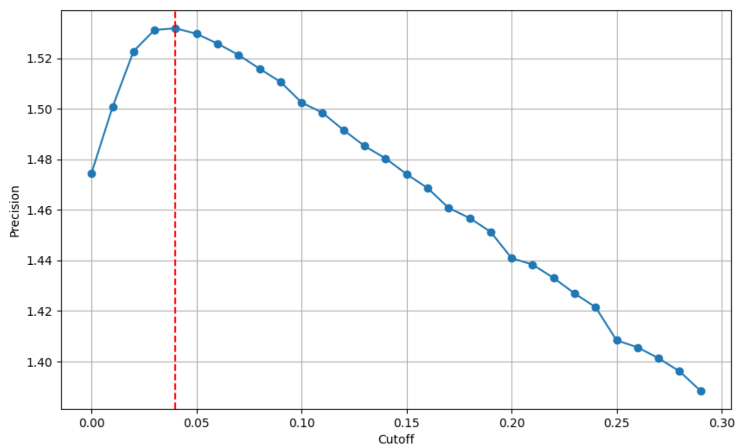


B Additional Material

B.1 Methodology and results cut-off language classification

This code categorizes Twitter user locations into three groups: Catalonia (”cat”), Spain (”esp”), or other (”oth”). Using predefined lists of Catalonia and Spain cities, our algorithm first attempts a direct match to classify each location based on the author’s location description. If unsuccessful, it employs geolocation (using the Nominatim service) for a more precise search. For cases that remain unclassified, the code further breaks down multi-word locations into fragments to check for partial matches. Additionally, it examines user descriptions for place-related keywords. This multi-layered approach improves classification accuracy by handling inconsistencies in user-reported locations.

Figure B1: Sum of Precision for cut-off Thresholds for Catalans and Spanish



B.2 Tweet figures for the sample period

Figure B2: Total tweets July to November 2017

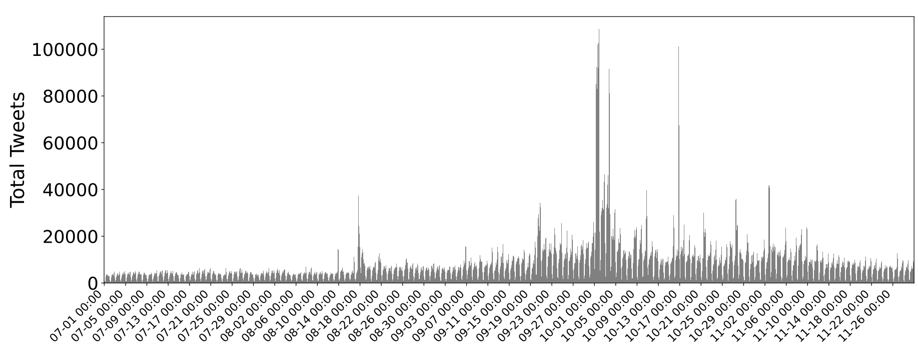


Figure B3: Share of Catalan origin in Castellano retweets

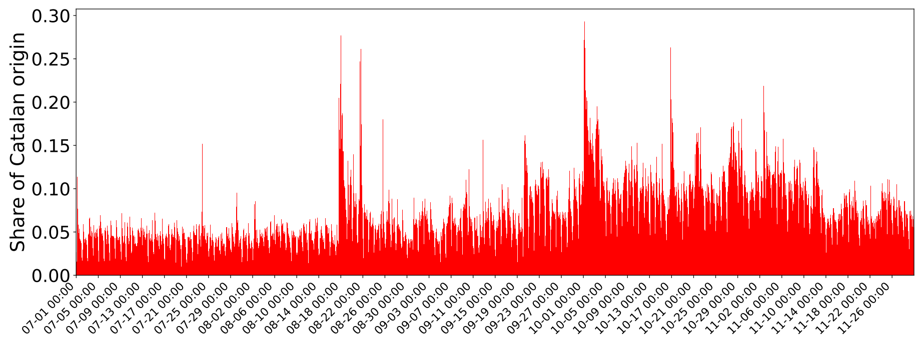
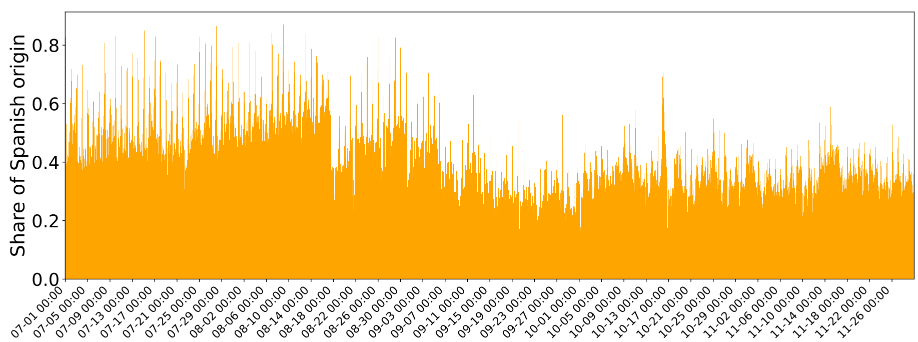
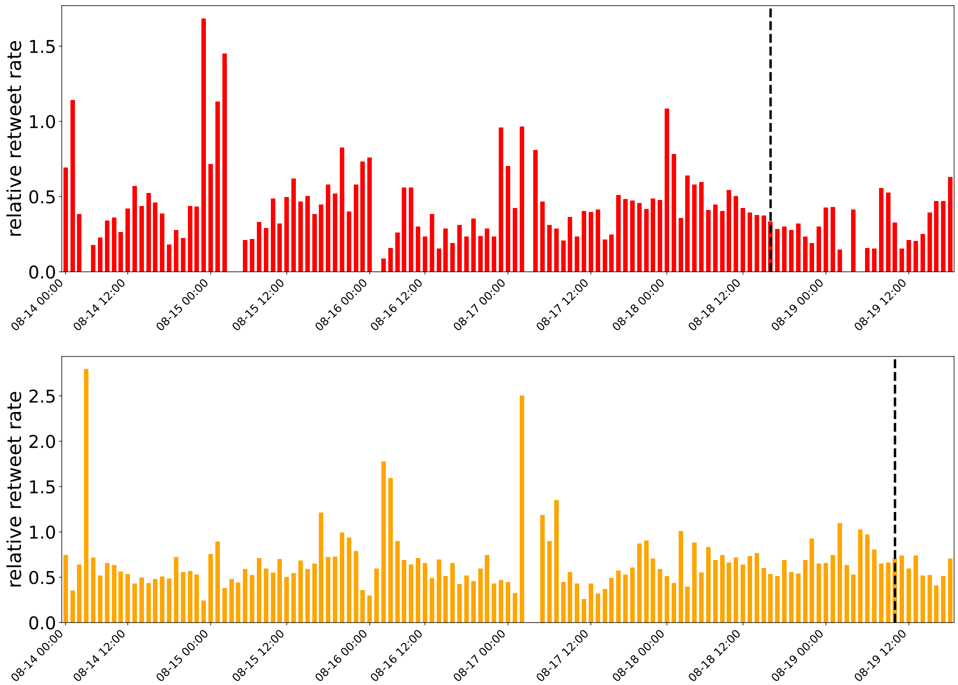


Figure B4: Share of Castellano origin in Catalan retweets



B.3 Network exposure model results for terror attack

Figure B5: Relative retweet rates (network exposure) - Terror attack



Notes: Figures show the relative α values calculated as in Eq. (4) and Eq. (5) conditional on exposed users. The red figure shows the share of tweets coming from *Catalan* users retweeted by *Castellano* users over time. The orange figure shows the share of tweets coming from *Castellano* users retweeted by *Catalan* users over time.

Table B1: Mean break in relative retweet rates (network exposure) - Terror attack

	Castellano users		Catalan users	
	date/hour	t-stat	date/hour	t-stat
network exposure (mean alphas)	08-18h16	-2.33	08-19h10	-0.61

Note: Table shows result of t-test on break in means for relative retweet rates in the period 08-10 and 08-24 at midnight using the network model. The column "date" reports the date with the maximum t-statistic in this period and the column "t-stat" reports the t-statistics. Under "Castellano accounts" we report test results for relative retweet rates of *Castellano* users retweeting *Catalan* users relative to retweet rates of *Castellano* users retweeting *Castellano* users. Under "Catalan accounts" we report test results for relative retweet rates of *Catalan* users retweeting *Castellano* users relative to retweet rates of *Catalan* users retweeting *Catalan* users.

B.4 Back of the envelope exposure counterfactual calculations

In order to quantify the importance of our exposure measure, we present below back of the envelope calculations where we pretend exposure would remain as before the break and predict the total number of tweets by language categories. Specifically, we consider the break-test date (t_{break}) calculated in the relative alphas of section 4.1 (Figure 5) and consider the 24 hours before and after this date. We then calculate the predicted retweets by hour after the break, i.e. for $t = t_{break} + 1, t_{break} + 2, ..., t_{break} + 24$ as

$$\widehat{RT_{esp,cat,t}} = \widehat{\alpha_{esp,cat,t}} \times \sum_{i \in \mathcal{I}_{cat,t}} |\mathcal{M}_{esp,A_i,t}|. \tag{7}$$

and then use the exposure numbers from before the break to calculate a counterfactual number of retweets by hours $t = t_{break} + 1, t_{break} + 2, ..., t_{break} + 24$, i.e. using the formula

$$\widehat{RT^{cf}_{esp,cat,t}} = \widehat{\alpha_{esp,cat,t}} \times \sum_{i \in \mathcal{I}_{cat,t-24}} |\mathcal{M}_{esp,A_i,t-24}|. \tag{8}$$

The numbers in the text then come from comparing the sum over the 24 hours after the break from the actual with the counterfactual retweet number from 24 hours before.

B.5 Evidence on unexposed users

As mentioned above, the exposure model aims to approximate the mixed factors that contribute to asymmetric content exposure across users, providing a more credible scenario than the uniform model. However, a disadvantage of the network model is that it does not strictly impose zero retweeting behavior for unexposed users.

We therefore perform the same test on the retweet rate of users our model classified as *un*exposed (Table B2), and the identified break point remains close to our main result and statistically significant.

B.6 Additional network graphs

We showed the spread of top Catalan users' retweets in Figure 9. Here, we present the analogous graph for the top tweets originating from Castellano users.[1] Figure B6, panel (a), depicts the spread of the tweet:

> *"Those who devised this plan to prevent the referendum may not realize that what they have caused is for Catalonia to leave for good today."*

This tweet was made by the well-known Catalan journalist Jordi Évole. Despite being Catalan, Évole is classified as a Castellano user in our dataset, which aligns with the linguistic

[1] The two users are Catalans but classified as from the Castellano language group as they chose to communicate in Spanish.

Table B2: Mean break in relative retweet rates (unexposed) - 1st of October

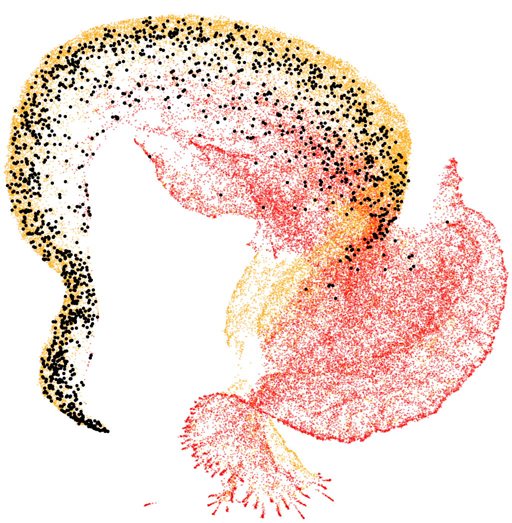
	Castellano accounts		Catalan accounts	
	date/hour	t-stat	date/hour	t-stat
network exposure (mean alphas)	10-01h07	14.18	10-01h05	9.89

Note: Table shows result of t-test on break in means for relative retweet rates in the period 09-27 at midnight and 10-02 at midnight conditioning on unexposed users. The column "date" reports the date with the maximum t-statistic in this period and the column "t-stat" reports the t-statistics. Under "Castellano accounts" we report test results for relative retweet rates of *Castellano* users retweeting *Catalan* users relative to retweet rates of *Castellano* users retweeting *Castellano* users. Under "Catalan accounts" we report test results for relative retweet rates of *Catalan* users retweeting *Castellano* users relative to retweet rates of *Catalan* users retweeting *Catalan* users.

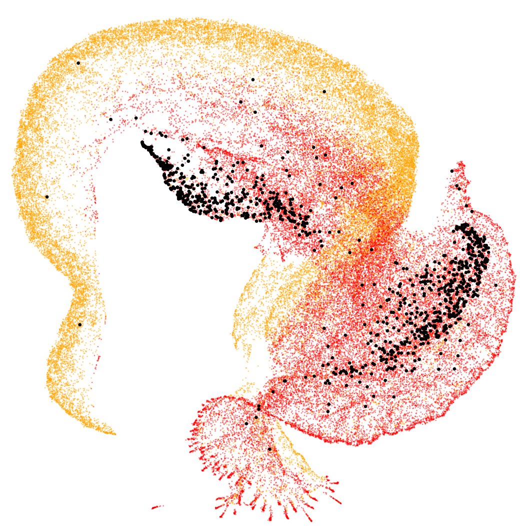
division outlined in this paper, as the majority of his tweets are written in Spanish. Furthermore, Évole is widely recognized for his neutral stance in the Catalan independence conflict, often positioning himself as a mediator advocating for dialogue rather than polarization. His use of Spanish as the primary language of communication is consistent with his broader attempt to bridge divides, reflecting the linguistic categorization used in our analysis.

Figure B6: Tweet examples of Castellano users

a) Tweet example 1st of October: Castellano user condemns policy violence



b) Tweet example 30th of September: Castellano user has two identities



Notes. (a) English translation: *"Those who devised this plan to prevent the referendum may not realize that what they have caused is for Catalonia to leave for good today."* (b) English translation: *"Proud to be who I am: Catalan and Spanish."*

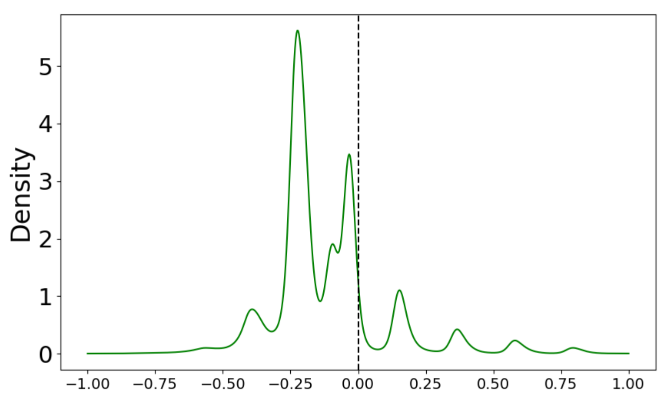
The message clearly penetrated into the Catalan network, as evidenced by its significant presence across Catalan-speaking communities. This contrasts sharply with the top Castellano tweet on September 30th, shown in panel (b), which was posted by Olympic gold medalist Joel González. His tweet featured a picture of himself holding a Spanish flag alongside the caption:

> *"Proud to be who I am: Catalan and Spanish."*

These two examples illustrate the role of linguistic identity as a marker that extends beyond mere regional affiliation. Both Évole and González are Catalan-born but express themselves predominantly in Spanish, reflecting a linguistic choice that transcends the boundaries of Catalan nationality. This mechanism aligns with the concept of the Gandhi Trap, where condemnation of state actions (in this case, the police response during the referendum) serves to solidify oppositional identity, reinforcing linguistic and ideological divides rather than bridging them. The spread of Évole's message among Catalan users, despite his classification as a Castellano speaker, highlights how linguistic expression can both reflect and influence political and social networks, intensifying the symbolic dimensions of the conflict.

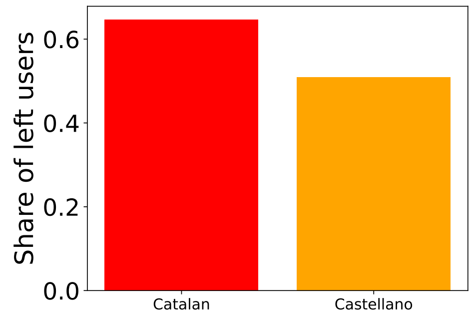
B.7 Additional material and figures for ideology

Figure B7: Estimated density for the ideology of our users.



The figure shows the estimated kernel density for the ideology score assigned to our sample users according to the methodology outlined in Section 5.

Figure B8: Share of left users by language category.



The figure shows the proportion of left-wing Catalan and Castellano users out of the total users that follow at least one of our politicians.

Table B3: Politician Twitter accounts and party affiliations.

Politician	**Party**	**Politician**	**Party**	**Politician**	**Party**
AITOR_ESTEBAN	PNV	AdaColau	EnComúPodem	Adrilastra	PSOE
Albert_Rivera	Cs	BelenHoyo	PP	CarmenMoriyon	ForoAsturias
EduMadina	PSOE	EsterMunoz85	PP	ForcadellCarme	ERC
GFVara	PSOE	GirautaOficial	Cs	GuillermoDiazCs	Cs
IdiazAyuso	PP	Ignacos	PP	InesArrimadas	Cs
IratxeGarper	PSOE	JACS_JaumeACS	JxCat	JRBauza	PP
Jaumeasens	En Comù	JoseMariaMazon	PP	JuanMa_Moreno	PP
KRLS	JxCat	LauraBorras	JxCat	ManuelaCarmena	Podemos
MartinaVelardeG	Podemos	MertxeAizpurua	EHBildu	MiquelBuch	JxCat
Mireia_veca	CUP	MonederoJC	Podemos	NestorRego	BNG
PabloEchenique	Podemos	PabloIglesias	Podemos	QuimTorraiPla	JxCat
Ramon_Espadaler	PSC-PSOE	RevillaMiguelA	PRCantabria	SSumelzo	PSOE
Santi_ABASCAL	VOX	SaraGimnez	Cs	TeresaRodr_	Podemos
Tonicanto1	Ciudadanos	Yolanda_Diaz_	PSOE	_mireiaborras	VOX
agarzon	Podemos	anioramas	CCanaria	arnauramirez	PSC-PSOE
astro_duque	PSOE	cayetanaAT	VOX	eledhmel	VOX
eloysuarezl	PP	felipe_sicilia	PSOE	teresa_perales	PP-PAR
gabrielorriaga	PP	gabrielrufian	ERC	gimenezbarbat	Cs
gmariscalanaya	PP	hermanntertsch	VOX	herrerobono	PP
ierrejon	Podemos	joanbaldovi	Compromìs	lluisrecoder	Convergencia
jordi_canyas	Cs	junqueras	ERC	sergiosayas	PP
lozanoirene	PSOE	luzseijo	PSOE	mariadolorosa	CUP
martapascal	JxCAT	mestremanuel	VOX	micaela_navarro	PSOE
monasterioR	VOX	pablocasado_	PP	rosadiezglez	UPyD
salvadorilla	PSOE	sanchezcastejon	PSOE		
susana_ros	PSOE				

Table B4: Party ideology scores as classified by ChatGPT.

Party	**Spectrum**	**Party**	**Spectrum**	**Party**	**Spectrum**
Alternatiba	-1	CCanaria	0	Compromis	-0.5
Cs	0	CUP	-1	EnComúPodem	-1
EHBildu	-1	ERC	-1	ForoAsturias	0.5
IzqUnida	-1	JxCat	0.5	PCPV	-1
PDCat	0.5	PFeminista	-1	PNV	0.5
Podemos	-1	PPCantabria	0.5	PP	0.5
PRCantabria	0.5	PSOE	-0.5	UMallorquina	0.5
UPYD	0	VOX	1		

B.8 Tweet text tables

Table B5: Top tweets in Castellano originating from *Castellano* users, their English translations, and the number of retweets by *Catalan* users 30th of September categorized by topic.

Tweet in Castellano	**Translation into English**	**Retweets by *Catalan* users**	**Topic**
Les quitan censo, urnas, papeletas, mesas, les cierran Telecomuniciones...¡Y quieren votar en paz! Toda Europa debe homenajear a Catalunya.	They take away census, ballot boxes, ballots, tables, they close Telecommunications... And they want to vote in peace! All of Europe should pay tribute to Catalonia.	2668	Referendum
ATENCIÓN! Empiezan a salir del Puerto, cada vez en grupos más grandes. Se ha oído por el muelle que la hora clave son las 00.00	ATTENTION! They start leaving the Port, each time in larger groups. It has been heard on the dock that the key time is 00.00	2574	Police
- Mi Teniente, hemos interceptado el horario de apertura de los colegios.- Buen trabajo. ¿Cuál es?- "Tres quarts i cinc de set"- Mierda.	- Lieutenant, we have intercepted the opening hours of the schools. - Good job. What is it? - "Tres quarts i cinc de set" - Damn.	1782	Referendum
Parece más legal cantar el cara al sol que poner urnas. Tenemos un gran problema y no es con Cataluña precisamente	It seems more legal to sing the "cara al sol" than to put up ballot boxes. We have a big problem and it's not with Catalonia precisely	1738	Realignment
Hola @InesArrimadas. Te paso una imagen por si tienes que volver a la Fiscalía de Menores para denunciar lo de la manipulación de los niños. https://t.co/GRTh1Z5zzf	Hello @InesArrimadas. I send you an image in case you need to go back to the Juvenile Prosecutor's Office to report the manipulation of children. https://t.co/GRTh1Z5zzf	1496	Realignment
Denunciamos la militarización del @portdebarcelona en estos momentos.Acceso y movilidad restringido incluso xra los Trabajadores Portuarios	We denounce the militarization of @portdebarcelona at this time. Access and mobility restricted even for Dock Workers	1444	Police
QUÉ HORROR. Madrid, 2017.óvenes y mayores hacen saludo fascista y cantan alegremente 'Cara al Sol' en Cibeles.://t.co/H9To0vs0Rd https://t.co/5dWhTDX66y	WHAT A HORROR. Madrid, 2017. Young and old make fascist salute and cheerfully sing 'Cara al Sol' in Cibeles. https://t.co/H9To0vs0Rd https://t.co/5dWhTDX66y	1325	Realignment
Hallan muerto en el jardín de su casa a otro testigo de la corrupción del PP, que tenía que declarar estos días. Ya van 11 en cuatro años...	Another witness of PP corruption, who was supposed to testify these days, is found dead in his garden. There are already 11 in four years...	1169	Realignment
Los adolescentes que cantan en Madrid el cara al sol fueron adoctrinados en colegios catalanes.	The teenagers who sing the "cara al sol" in Madrid were indoctrinated in Catalan schools.	1142	Realignment
Un Cara al Sol en Cibeles? Qué será lo siguiente? Qué gane el PP las elecciones?	A Cara al Sol in Cibeles? What will be next? The PP winning the elections?	1125	Realignment
TVE no ha dado ni lo de el Cara al Sol en Cibeles ni a Esperanza Aguirre a su lado.Ya somos tratados como a una dictadura en Europa. Bravo https://t.co/lzxEPR0f8f	TVE hasn't shown either the Cara al Sol in Cibeles or Esperanza Aguirre next to it. We are already treated like a dictatorship in Europe. Bravo https://t.co/lzxEPR0f8f	1005	Realignment
Sectorial de x la Independencia d ANC pide ayuda a bomberos d EH para asegurar referendumá vamos ElsBombersSeremSempreVostres https://t.co/AJz9DDgLkc	Sectorial of for the Independence of ANC asks for help from EH firefighters to ensure referendumwe go ElsBombersSeremSempreVostres https://t.co/AJz9DDgLkc	888	Referendum
Qué manía la gente llamando fascistas a gente que canta himnos fascistas y hace el saludo fascista con banderas fascistas	What an obsession people have calling fascists people who sing fascist hymns and do the fascist salute with fascist flags	884	Realignment
En el Puerto hay 5 salidas:(de S a N)-X el Prat-X cocheras de TMB-X pg Zona Franca-Frente al cementerio-X Grimaldieso m han dicho...	There are 5 exits at the Port: (from S to N)-X el Prat-X TMB garages-X pg Zona Franca-In front of the cemetery-X Grimaldiso I was told...	860	Police

Table B6: Top tweets in Castellano originating from *Castellano* users, their English translations, and the number of retweets by *Catalan* users 1st of October categorized by topic

Tweet in Castellano	**Translation into English**	**Retweets by *Catalan* users**	**Topic**
Como periodista de TVE siento vergüenza por el tratamiento que ha dado RTVE al 1-O. De nuevo, la tele pública al servicio del gobierno #SOS	As a TVE journalist, I feel ashamed of the treatment RTVE gave to 1-O. Again, public television at the service of the government #SOS	10196	Realignment
Los que idearon este plan para evitar el referéndum, igual no saben que lo que han provocado es que hoy Cataluña se vaya definitivamente.	Those who devised this plan to prevent the referendum might not know that what they have caused is for Catalonia to leave definitively today.	9359	Referendum, Police violence
¿Alguien puede defender esto? https://t.co/WOVIYCnmAS	Can anyone defend this? https://t.co/WOVIYCnmAS	7142	Police violence
Brutal comienzo del especial TV3:habla Soraya,salen imágenes de la represión en Cataluña desmontando lo que dice. #CatalanReferendum https://t.co/Qqz0tX90Yl	Brutal start to the TV3 special: While Soraya speaks, images of the repression in Catalonia appear, dismantling what she says. #CatalanReferendum https://t.co/Qqz0tX90Yl	4560	Police violence
No ha habido suficientes porras para tanta dignidad. #CatalanReferendum	There haven't been enough batons for so much dignity. #CatalanReferendum	4204	Referendum, Police violence
Esta es vuestra proporcionalidad @zoidoJI? No tenéis vergüenza. https://t.co/pP3A0g31iR	Is this your proportionality @zoidoJI? You have no shame. https://t.co/pP3A0g31iR	4098	Police violence
Jamás olvides el silencio del PSOE.	Never forget the silence of the PSOE.	3895	Realignment
Difundid porfavor, Aiguaviva, Girona https://t.co/vbZFZXEufg	Please spread this, Aiguaviva, Girona https://t.co/vbZFZXEufg	3688	Police violence
Si esto no te da vergüenza es que no la tienes. #CatalanReferendum https://t.co/zTZFnBqchU	If this doesn't make you ashamed, you have no shame. #CatalanReferendum https://t.co/zTZFnBqchU	3620	Referendum, Police violence
The Guardian: El Estado español ha perdidoMonde: La derrota de Rajoy: La vergüenza de Europa: Somos el ejemplo del mundo	The Guardian: The Spanish State has lostMonde: Rajoy's defeat: The shame of Europe: We are the example of the world	3406	Realignment
Catalunya será la tumba del fascismo. #CatalanReferendum https://t.co/qrkyMyEyn1	Catalonia will be the tomb of fascism. #CatalanReferendum https://t.co/qrkyMyEyn1	3026	Realignment, Police violence
900 imputados65 casos de corrupción5 tesoreros investigados465 heridosde Rajoy.#CatalanReferendum	900 indicted65 cases of corruption5 treasurers investigated465 injured's record.#CatalanReferendum	2954	Realignment, Police violence
Lo más triste de hoy no es solo ver al cuerpo policial gopear ancianos. Es ver al pueblo español aplaudirles por ello. #CatalanReferendum	The saddest thing today is not just seeing the police force beating up elderly people. It is seeing the Spanish people applaud them for it. #CatalanReferendum	2812	Police violence
Tal vez el mejor vídeo de la labor de los bomberos de todo el día.A primera línea y a empoderar al resto #ElsBombersSeremSempreVostres https://t.co/aq6alRID5G	Perhaps the best video of the firefighters' work all day. On the front line and empowering the rest #ElsBombersSeremSempreVostres https://t.co/aq6alRID5G	2805	Police violence
La policía nacional está usando munición prohibida en Cataluña. Para los que hablan de ley y legalidad.	The national police are using prohibited ammunition in Catalonia. For those who talk about law and legality.	2787	Police violence

Table B7: Top tweets in Castellano and Catalan originating from *Catalan* users, their English translations, and the number of retweets by *Castellano* users 30th of September categorized by topic

Tweet in Castellano	**Translation into English**	**Retweets by *Castellano* users**	**Topic**
La palabra fachada nunca había tenido tanto sentido. https://t.co/2MirxSDt5W	The word facade had never made so much sense. https://t.co/2MirxSDt5W	955	Realignment
¿Quién puede estar en contra del lema "MÁS DEMOCRACIA"?triste imagen que no necesita palabras.ás que nunca: #MésDemocràcia!! https://t.co/WWZCezE9SR	Who can be against the slogan "MORE DEMOCRACY"? A sad image that needs no words. More than ever: #MésDemocràcia!! https://t.co/WWZCezE9SR	411	Realignment, Referendum
Aquesta fotografia és el discurs més eloqüent per explicar el que ens juguem en el referèndum de demà. https://t.co/apmCThjSGw	This photograph is the most eloquent speech to explain what is at stake in tomorrow's referendum. https://t.co/apmCThjSGw	361	Referendum
Independentistas catalanes = Nazis= Fascistascantando el cara al sol = Manifestantesneolengua on fire.	Catalan independentists = Nazis= Fascistssinging the cara al sol = Demonstratorson fire.	322	Realignment
Más libros y menos banderas.	More books and fewer flags.	296	Realignment
Barcelona. Plaça Catalunya. 30 de setiembre. 15:54h. https://t.co/PZN4wyechG	Barcelona. Plaça Catalunya. September 30. 15:54h. https://t.co/PZN4wyechG	262	Realignment
Balconing contra la democracia://t.co/Kl9uK8yQN6	Balconing against democracy://t.co/Kl9uK8yQN6	206	Referendum
La revolución de las sonrisas apalizando a un catalán que paseaba con la bandera de España. Y todo así https://t.co/bFMitxJ4gK	The revolution of smiles beating up a Catalan who was walking with the Spanish flag. And everything like that https://t.co/bFMitxJ4gK	186	Other
Diálogo, negociación y pacto es la única vía para resolver el problema. Respeto a todas las ideas y sentido común. #DialoguenPorFavor https://t.co/TvDGVL9n4Y	Dialogue, negotiation, and agreement are the only way to solve the problem. Respect for all ideas and common sense. #PleaseDialogue https://t.co/TvDGVL9n4Y	179	Other
Aquesta és la realitat. Els molesten pancartes que diuen "Més democràcia" i a Madrid canten el "Cara al sol". Per això demà guanyarem. https://t.co/rxGfM05xki	This is the reality. They are bothered by banners that say "More democracy" and in Madrid, they sing the "Cara al sol". That's why we will win tomorrow. https://t.co/rxGfM05xki	169	Realignment
Catalunya puede estar domingo sin Internet. Para que la app Firechat funcione sin red necesita un 5% de masa crítica de usuarios. RT please	Catalonia could be without the Internet on Sunday. For the Firechat app to work without a network, it needs a 5% critical mass of users. RT please	150	Referendum
Hemos jugado muchos partidos juntos. Y eres uno de los responsables de que los haya disfrutado tanto. Eres el mejor, Leo. @mundodeportivo https://t.co/CdgvdWiYrs	We have played many matches together. And you are one of the reasons I enjoyed them so much. You are the best, Leo. @mundodeportivo https://t.co/CdgvdWiYrs	138	Other
Familias ganando y protegiendo colegios, mesas electorales. No sé de un país como este que se merezca tanto el derecho a votar.	Families winning and protecting schools, polling stations. I don't know of a country like this that deserves the right to vote so much.	120	Referendum
POR PRIMERA VEZ EN LA HISTORIA DEL FESTIVAL DE SAN SEBASTIÁN UN MUJER DIRECTORA GANA EL PREMIO A LA MEJOR DIRECCIÓN: ANAHÍ BERNERI. https://t.co/JMcwqhUQLi	FOR THE FIRST TIME IN THE HISTORY OF THE SAN SEBASTIÁN FESTIVAL, A WOMAN DIRECTOR WINS THE BEST DIRECTOR AWARD: ANAHÍ BERNERI. https://t.co/JMcwqhUQLi	110	Other
Una àvia, fa 5 minuts, a l'escola: "Com que he sentit que a les iaies no les pegarien, he vingut per si em necessiteu". #llagrimeta	An old woman, 5 minutes ago, at the school: "Since I heard that they wouldn't hit the grannies, I came in case you need me". #teardrop	93	Referendum, Police

Table B8: Top tweets in Castellano and Catalan originating from *Catalan* users, their English translations, and the number of retweets by *Castellano* users 1st of October categorized by topic

Tweet in Castellano	**Translation into English**	**Retweets by *Castellano* users**	**Topic**
Vergüenza, es lo que siento como periodista de TVE viendo el tratamiento sesgado que está dando el Telediario. Esto es manipulación.	Shame, that's what I feel as a TVE journalist seeing the biased treatment that the news is giving. This is manipulation.	3640	Realignment
Esta papeleta estaba en mi urna, en el colegio Rafael de Casanovas de Llefiá. lo que consiguen vuestras porras #orgull #vergüenza https://t.co/ToAIZc6CKg	This ballot was in my ballot box, at Rafael de Casanovas school in Llefiá. Look what your batons achieve #pride #shame https://t.co/ToAIZc6CKg	2012	Referendum, Police violence
El Rey se esconde. El PSOE agoniza. La prensa miente. El PP ahostia a la clase obrera.finaliza La Transición.España ha fracasado.	The King hides. The PSOE agonizes. The press lies. The PP beats the working class.The Transition ends.has failed.	1995	Realignment
Votar es democracia!	Voting is democracy!	1950	Referendum
Ja he votat. Junts som imparables defensant la democràcia. https://t.co/mGXf7Qj1TM	I have already voted. Together we are unstoppable defending democracy. https://t.co/mGXf7Qj1TM	1856	Referendum
Soraya habla. Los vídeos desmienten su argumentario. Este inicio del informativo de TV3 ha sido brillante.https://t.co/weI5FUgQw9	Soraya speaks. The videos contradict her arguments. This start of the TV3 news was brilliant.https://t.co/weI5FUgQw9	1821	Realignment, Police violence
Aplausos en colegio electoral de Torrefarrera para chaval que vota envuelto en una rojigualda[Video: @xavicantons] https://t.co/oQpolhYyTE	Applause at the polling station in Torrefarrera for a kid who votes wrapped in a Spanish flag[Video: @xavicantons] https://t.co/oQpolhYyTE	1805	Referendum, Realignment
Las actuaciones policiales contra población pacífica deben parar. Hoy todos, en Catalunya y en el Estado, tenemos que exigir #RajoyDimisión	The police actions against peaceful people must stop. Today everyone, in Catalonia and in the State, we must demand #RajoyResignation	1795	Realignment, Police violence
Un presidente de gobierno cobarde ha inundado de policía nuestra ciudad. Barcelona ciutat de pau, no té por #MésDemocracia @marianorajoy	A cowardly president has flooded our city with police. Barcelona city of peace, is not afraid #MoreDemocracy @marianorajoy	1600	Realignment, Police violence
Patada voladora y lanzamiento de una mujer escaleras abajo por parte de la Policía en el instituto Pau Clarís de Bcn #CatalanReferendum https://t.co/IFTCmfPDAU	Flying kick and throwing a woman down the stairs by the Police at the Pau Clarís institute in Bcn #CatalanReferendum https://t.co/IFTCmfPDAU	1590	Police violence
Ya hay más de 460 heridos en Catalunya. Como alcaldesa de BCN exijo el fin inmediato de las cargas policiales contra población indefensa https://t.co/412z6Jacap	There are already more than 460 injured in Catalonia. As mayor of BCN I demand the immediate end of police charges against defenseless people https://t.co/412z6Jacap	1419	Realignment, Police violence
#CatalanRederendum otro votante del NO acosado y violentado por los de Sí https://t.co/PKpL0h7bSR	#CatalanReferendum another NO voter harassed and violated by those of Yes https://t.co/PKpL0h7bSR	1037	Referendum
ULTIMA HORA. Un ciudadano de Badalona consigue ser el mayor demócrata del mundo al lograr votar 27 veces en 38 minutos.	LAST MINUTE. A citizen of Badalona manages to be the greatest democrat in the world by voting 27 times in 38 minutes.	927	Referendum
La Puerta del Sol ahora mismo. Rajoy, algo me dice que tienes más problemas que los catalanes#MadridASolContraLaRepresión #CatalanReferendum https://t.co/B4l9BUxmqU	The Puerta del Sol right now. Rajoy, something tells me you have more problems than the Catalans#MadridASolAgainstRepression #CatalanReferendum https://t.co/B4l9BUxmqU	920	Realignment
Al final no ha tenido que llegar Podemos al poder para que España se convierta en Venezuela.	In the end, Podemos didn't have to come to power for Spain to become Venezuela.	885	Realignment, Policy violence