Original research article

Beyond the share button: How partisan alignment, journalistic quality, and algorithmic governance shape what millions see on Facebook

Platforms & Society
Volume 3: 1–15

DOI: 10.1177/29768624261452529
journals.sagepub.com/home/pns

Fabio Giglietto[1] **and Giada Marino**[1]

Abstract

Social media platforms serve as critical gatekeepers in democratic information ecosystems, yet the mechanisms governing political content amplification remain poorly understood. This study examines how Facebook amplifies news by analysing 130,448 highly circulated URLs shared between 2017 and 2022, using Meta's Privacy-Protected Full URLs Dataset. We investigate how user-driven sharing translates into viewership and how audience partisan alignment, source credibility and algorithmic evolution moderate this relationship. Our analysis yields three key findings. First, although sharing reliably predicts viewing, this relationship weakens significantly for content with intensely partisan audiences, even after controlling for engagement levels. Second, adherence to professional journalistic standards independently increases reach – sources with higher credibility scores receive substantially more views than those with lower scores, holding shares constant. Third, and most critically, temporal analysis reveals that these effects are algorithmically modulated rather than structurally fixed. During Facebook's 2020 election 'break the glass' interventions, the partisan penalty doubled while rewards for quality journalism surged simultaneously. These volatile coefficients provide compelling evidence that reduced reach for partisan content stems from active platform intervention, not organic network limitations. Our findings demonstrate that Facebook operates as an active curator, not a neutral conduit. Its algorithmic choices create measurable disparities in content visibility that profoundly shape democratic discourse. As algorithmic curation increasingly determines citizens' information exposure, the temporal instability of these effects underscores the urgent need for platform transparency and accountability in democratic governance.

Keywords

Partisan alignment, algorithmic governance, Facebook, amplification, professional journalistic standards, NewsGuard

Introduction

Social media platforms are not neutral infrastructures for information exchange but governing actors that actively shape the circulation and visibility of political information. Through evolving moderation policies, ranking systems and design choices, platforms intervene in the conditions under which content is circulated, amplified or suppressed. These interventions have profound implications for democratic discourse, as they structure not only what information spreads but also how citizens encounter and engage with it. Despite a growing body of scholarship on platform governance, we still lack systematic empirical evidence on how governance mechanisms interact with user behaviour to produce large-scale patterns of political information visibility.

This study examines Facebook's role in mediating political information flows through sharing during a critical phase in the platform's governance history. Between 2017 and 2022, Facebook occupied a dominant position in the digital media ecosystem, functioning as one of the primary gateways for news distribution and political communication (Smith, 2024). In the aftermath of the 2016 US presidential election, the platform's influence over democratic discourse became the object of unprecedented public, regulatory and scholarly scrutiny, prompting direct governance interventions. Most notably, in February 2021 Facebook announced

[1]University of Urbino Carlo Bo, Urbino, Italy

Corresponding author:
Giada Marino, University of Urbino Carlo Bo, via Aurelio Saffi, 15 61029 Urbino, Italy.
Email: giada.marino@uniurb.it

that it was actively testing measures to reduce the distribution of political content in the News Feed (Frenkel and Isaac, 2024; Stepanov and Gupta, 2021). This period therefore offers a unique opportunity to study amplification dynamics under conditions of intensified platform intervention, where governance decisions are not a background condition but a visible and consequential component of information circulation.

Such interventions are part of broader patterns of platform governance that encompass algorithmic curation, content policies and enforcement practices, through which platforms actively shape the visibility and circulation of content (Gillespie, 2021; Gorwa, 2019). Unlike traditional media gatekeeping, platform governance operates through a combination of explicit rule enforcement and implicit algorithmic ranking, producing layered systems of visibility that remain largely opaque to users and researchers alike (Gorwa et al., 2020). Understanding these governance mechanisms is essential for interpreting observed relationships between user behaviour and news circulation on Meta platforms – particularly given that Meta remains the only major social media company to have provided researchers with a dedicated dataset covering this period. While constrained by privacy protections and limited in scope, this dataset offers valuable aggregated individual-level information, including age, gender and – uniquely for the US case – page-level political alignment (Messing et al., 2020).

Within this high-stakes information ecosystem, the share button represents a central mechanism of amplification. Originally introduced to formalize users' habitual copy-and-paste practices, sharing has evolved into a key affordance that not only structures individual news consumption but also drives broader patterns of visibility and engagement (Adamic et al., 2016). Yet, despite its importance, the relationship between exposure and amplification through sharing remains only partially understood. In particular, it is unclear how feedback loops between user behaviour and algorithmic ranking shape which content ultimately gains prominence. Prior research has shown that partisan content can achieve high engagement despite limited reach (Kalsnes and Larsson, 2018) and that credible sources generally outperform untrustworthy ones in both sharing and viewing (Guess et al., 2021). However, how these dynamics interact within Facebook's algorithmic ecosystem – especially under conditions of active governance intervention – remains underexplored.

Recent large-scale studies conducted through Meta's research partnerships, focusing on Facebook's role during the 2020 US presidential election, reported limited effects on political polarization and the circulation of untrustworthy content (González-Bailón et al., 2023, 2024; Guess et al., 2023a, 2023b). While these studies benefited from unprecedented access to internal platform data, their findings have generated sustained debate regarding the assumptions embedded in their research designs. Critics have noted that such analyses often treat platform governance interventions as stable background conditions rather than as dynamic variables shaping content visibility (Bagchi et al., 2024).

These concerns are compounded by the inherent challenges of studying platform governance empirically. Algorithmic adjustments and policy changes introduce discontinuities in information flows that complicate causal inference (Gillespie, 2022). Particularly consequential are emergency 'break the glass' interventions – temporary measures Meta deploys during periods of heightened political risk. Following the 6 January 2021 Capitol invasion and amid the COVID-19 infodemic, Facebook activated emergency protocols that reportedly doubled the penalty for content from politically extreme sources and reduced the algorithmic distribution of civic and political content more broadly (Isaac and Frenkel, 2020; Jackson, 2024). These 63 emergency measures, implemented around the 2020 election, reduced views of misinformation by at least 24% but were reversed by March 2021 (Bagchi et al., 2024). Such interventions can substantially reshape the amplification landscape during the periods when platform effects on public discourse may be most consequential. When platforms simultaneously suppress specific content types while researchers attempt to measure organic circulation patterns, disentangling user preferences from governance effects becomes analytically difficult.

These methodological challenges point toward longitudinal research designs that explicitly account for governance dynamics. This study contributes to such an approach by examining the relationship between sharing and viewing for widely circulated URLs in the United States across a nearly 6-year period (January 2017–October 2022) that spans multiple governance regimes – including pre-election periods, the implementation of emergency interventions and their subsequent relaxation. Building on Trilling et al. (2022), who analysed amplification patterns across four European countries, we apply a comparable analytical framework while extending attention to how partisan alignment and adherence to journalistic standards moderate the sharing–viewing relationship. The US context is particularly valuable because the Facebook Privacy-Protected Full URLs Dataset includes Political Page Affinity (PPA) scores – measures of ideological alignment unavailable for other countries – enabling analysis of how partisan intensity shapes amplification dynamics.

We conceptualize news amplification on social media as a hybrid phenomenon – a now well-established view in media scholarship (Chadwick, 2013) – shaped by user decisions but mediated by algorithmic logics that remain largely opaque. To capture this dynamic, we examine four factors structuring the amplification loop between viewing and sharing: (a) the partisan intensity of audience alignment with shared links; (b) the journalistic trustworthiness of

those links; (c) the interaction between partisanship and credibility; and (d) temporal variation, which captures changes in amplification dynamics under shifting governance conditions.

The paper proceeds as follows. We first develop a theoretical framework on the relationship between news exposure and amplification, outlining our expectations regarding these drivers. We then describe our data and methodology before presenting the findings and discussing their implications for research on platform governance and democratic discourse.

Theoretical background

Social media and Facebook in particular have become central infrastructures for the circulation of high-stakes political information, especially content that achieves rapid and large-scale visibility and shapes public discourse (Newman et al., 2024). While much existing research focuses on routine patterns of social media use, this study centres on the dynamics governing widely circulated, high-salience content, namely, URLs that break through everyday informational noise and reach mass audiences. Examining how such content circulates is critical for understanding large-scale opinion formation during politically salient moments, when platform affordances and governance decisions are most consequential.

Our theoretical framework addresses the challenge of disentangling platform-mediated visibility from the intrinsic characteristics of content and the users who engage with it. Although amplification dynamics and shifts in moderation policies are empirically intertwined, they are discussed separately in the literature review to align theory development with the structure of the findings. The relationship between viewing and sharing is therefore introduced first as the core amplification mechanism (Trilling et al., 2022) and forms the basis of the first hypothesis. The framework then turns to content-level factors, specifically source ideological alignment and journalistic credibility, that condition how amplification unfolds. Temporal variation associated with changes in platform moderation and ranking policies is addressed separately through the third research question rather than a directional hypothesis. Conceptually, the paper examines how amplification dynamics, source partisanship, source credibility and changes in platform moderation over time jointly shape the popularity of news content on Facebook.

The viewing and sharing feedback loop for high-reach content on Facebook

Although viewing and sharing are central to social media activity, we still lack a clear understanding of how they interact in a reinforcing loop, in which shares both result from prior visibility and function as signals that platforms use to further amplify news content, especially in cases of viral circulation (Trilling et al., 2022).

In this essay, we thus intend *viewing* as exposure to content, whether accidental or intentional, while *sharing* as amplification, i.e. a user-driven action that directly contributes to the circulation of content, sometimes with civic or ideological motivations. Facebook's distribution algorithm weighs shares – along with other measurable engagement signals – when determining which posts should be considered popular and thus circulated more widely (Adamic et al., 2016; Smith et al., 2020). Each new share drives more views, a process of 'content amplification' (Giglietto et al., 2020) driven by users' intentions but crucially reinforced by algorithmic and content moderation logics that are not only opaque but also in constant flux.

Sharing plays a central role in news exposure, with its impact heavily dependent on the source – particularly strong-tie connections (Kaiser et al., 2021; Karnowski et al., 2017; Kümpel, 2019). However, viewing and sharing reflect different levels of involvement: viewing signals passive exposure, whereas sharing indicates a higher degree of engagement, often with the explicit intention of amplifying a message (Schäfer, 2023; Trilling et al., 2022). Crucially, their relationship is reciprocal. Content must first be viewed before it can be shared, and once shared, it reaches new audiences. More views increase the likelihood of re-shares, fuelling a cycle of amplification.

The decision to share, however, depends on both platform-level and content-level factors. Interface features such as popularity cues can enhance algorithmic amplification (Messing and Westwood, 2014), while characteristics of the content – emotional charge, ideological resonance or partisan framing – affect shareability (Brown et al., 2020; Karnowski et al., 2021; Trilling et al., 2017; Wischnewski et al., 2021).

At the same time, prior research highlights that sharing itself is a powerful driver of visibility. For example, Trilling et al. (2022) found a strong positive correlation between the number of shares and the number of views on Facebook. Political content, in particular, tends to be shared more frequently than expected – especially by older users – even when it receives fewer views overall. These findings inform the study's first hypothesis:

> **H1.** For highly circulated links on Facebook, the frequency of link sharing is positively correlated with the frequency of link viewing.

Factors driving algorithmic reach

This section examines how content characteristics, specifically source partisanship and credibility and platform governance factors, including the evolution of moderation policies and algorithmic ranking systems, shape the feedback loop between viewing and sharing on social media.

Ideological alignment in high-stakes information flows. While ideological alignment influences content sharing across all levels of social media activity, its effects are magnified for high-reach content that can shape political narratives. A growing literature highlights the distinctive amplification patterns of hyperpartisan news, especially on Facebook. While such outlets generally have a smaller overall reach than legacy media, they often outperform them in terms of shares – particularly during election cycles. This pattern suggests that symbolic circulation, through sharing, matters more than broad-based approval expressed via likes or comments (Kalsnes and Larsson, 2018).

This dynamic is largely driven by a small but highly active group of users who frequently share ideologically aligned content – predominantly but not exclusively from extreme right-wing sources. Topics that are emotionally charged, such as immigration, tend to generate particularly high levels of engagement (Kalsnes and Larsson, 2018; Kümpel et al., 2015) and emotional intensity – regardless of whether it is positive or negative – has been shown to enhance the likelihood of sharing (Larsson, 2017). These behaviours are especially common among politically engaged users, who often act as opinion leaders in their networks (Karlsen, 2015).

Hyperpartisan outlets capitalize on this by reframing mainstream news with ideological cues that reinforce in-group narratives (Haller and Holt, 2019; Karlsson and Holt, 2016). This strategy boosts shareability not by adding informational value but by aligning content with users' political and emotional predispositions (Trilling et al., 2017). Motivated reasoning plays a key role here: users are more likely to share content that confirms their beliefs and identities (An et al., 2013; Chaiken et al., 1996; Kunda, 1990). Additionally, emotionally and morally charged content – especially that marked by outrage, disgust or conflict – can spread more rapidly through social contagion mechanisms (Baumeister et al., 2001; Brady et al., 2017). Notably, ideological alignment can even outweigh concerns about source credibility, as users sometimes share congruent content from untrustworthy outlets (Clemm von Hohenberg, 2018).

The tension between viral shareability and actual reach becomes particularly pronounced for politically charged content. While partisan content may excel in shareability, its reach is more limited. Selective exposure plays a role here: users predominantly interact with ideologically aligned content (Bakshy et al., 2015). Yet legacy media still dominate in terms of views (Kalsnes and Larsson, 2018). This is partly due to platform design and partly because reliable data on content views has only recently become available (Meta Platforms, Inc., 2025), opening new avenues for studying how ideological alignment influences both symbolic and actual exposure at societally relevant scales.

It is crucial to distinguish between two potential mechanisms that could limit the reach of partisan content even at a viral scale. First, structural network effects arising from homophilic clustering: highly partisan content may face inherent ceiling effects when circulating within ideologically sealed communities, as these 'partisan walled gardens' naturally limit organic reach beyond the core audience. Second, active platform intervention through algorithmic dampening: platforms may deliberately suppress the reach of partisan content through content moderation policies. While both mechanisms produce similar cross-sectional patterns – reduced reach for partisan content – their temporal dynamics differ markedly. Structural network effects should remain relatively stable over time, whereas platform interventions should produce volatile patterns coinciding with policy changes or external pressures.

> **RQ1.** For widely circulated URLs shared on Facebook, how does the relationship between viewing and sharing vary based on the intensity of the audience's partisan affiliation?

Trustworthy versus untrustworthy news. Another key research strand concerns how source credibility affects news circulation and reach. While the rise of social media has intensified concerns about misinformation, large-scale studies suggest that credible news sources are still prevalent in the information diets of most users.

For instance, Guess et al. (2021) found that articles from reputable outlets were shared more than five times – and viewed more than seven times – as often as those from low-credibility domains. Similarly, earlier work by Guess et al. (2020) showed that untrustworthy websites accounted for only a minor share of online news consumption during the 2016 US election. Even ideologically motivated users regularly engage with mainstream news (Guess et al., 2021).

Indeed, laypeople generally care about accuracy when sharing (Pennycook et al., 2021), yet some researchers caution against imposing researcher definitions of 'credible news' onto audiences, given that credibility is itself an increasingly polarized concept (Kramer, 2021). Journalism is notoriously self-defined by outlets (Carlson, 2017), making adherence to professional standards – rather than audience trust – a more defensible analytical frame.

However, sharing and viewing do not always align. In a European study, Rossi et al. (2023) analysed Facebook URL sharing in Germany, France and Italy between 2019 and 2022. They found that while total sharing declined, the proportion of links from untrustworthy sources rose – particularly during national elections. Still, this uptick in symbolic circulation did not translate into greater visibility in terms of views. These findings underscore a disconnect between the circulation of content and its actual consumption.

This disconnect is often exacerbated by the actions of small but highly active partisan communities

(Green et al., 2025). These groups can successfully manipulate platform algorithms – sometimes through coordinated behaviour or other manipulative tactics – to disproportionally inflate the 'share-of-voice' of their preferred content, making it appear more prevalent than its actual viewership would suggest.

Overall, while trustworthy sources may still dominate in total sharing and reach, the artificially amplified algorithmic reach of untrustworthy content – and its persistence and cumulated widespread exposure – remains a crucial concern for political communication and misinformation research (Allen et al., 2024; Goel et al., 2025).

RQ2. For widely circulated news URLs shared on Facebook, how does the relationship between viewing and sharing vary based on the news source's adherence to professional journalistic standards?

Shifting moderation and algorithmic reach policies over time. Content circulation on digital platforms is shaped by layered forms of governance, understood as the evolving interplay of rules, algorithms and external pressures that structure interactions among platforms, users, content producers and regulatory, civic and political actors (Gorwa, 2019). Governance operates internally through codified policies and algorithmic curation and externally through regulatory, civic and commercial attempts to steer platform behaviour (Gillespie, 2017; Gorwa, 2019). Crucially, these arrangements are not static. Their effects unfold over time, making longitudinal analysis necessary to distinguish structural properties of information networks from shifts induced by governance interventions.

Platform governance operates through two interrelated mechanisms. Explicit moderation intervenes directly by removing, suspending or labelling content according to formal standards. Algorithmic curation, by contrast, governs visibility without removal, shaping exposure through ranking, weighting and demotion while leaving content formally accessible (Gillespie, 2022). An example is the visibility downgrade of political content by Meta between 2021 and 2025 (Kaplan, 2025). Thus, visibility is not a neutral reflection of user interests: content meeting production standards may be algorithmically constrained, generating misalignments in the amplification mechanisms between reach and sharing that mix governance intervention and audience preference (Bozdag, 2013; Schneiders and Stark, 2025).

Algorithmic reach is itself unstable. Platforms continuously recalibrate selection and ranking criteria – often opaquely – in response to shifting strategic priorities and external pressures (Gorwa, 2019; Schneiders and Stark, 2025). Relevance is inferred from signals such as prior interactions, social proximity and content format rather than from journalistic quality standards alone (DeVito, 2017). This opacity intensifies uncertainty about how visibility is produced (Petre et al., 2019). Proprietary thresholds for weighting and demotion obscure the determinants of reach, such that observed circulation patterns may reflect an indeterminate combination of governance intervention, distribution logic and audience engagement (Gillespie, 2022; Guess et al., 2021). Temporal discontinuities in circulation thus encode the imprint of governance, offering observable traces of platform-mediated control over distribution.

The temporal dimension is therefore central to identifying these effects. Structural constraints such as homophily, audience segmentation and echo-chambers tend to generate relatively stable patterns of circulation, whereas changes in algorithms and moderation policies introduce longitudinal volatility. By leveraging temporal variation in large-scale circulation data, this study examines how evolving platform governance reshapes the relationship between exposure and sharing, addressing the following research question:

RQ3. How has Facebook's evolution in content distribution algorithms and moderation policies affected the relationship between viewing and sharing patterns for widely circulated URLs over time?

Data, variables and measures

Dataset and unit of analysis. This study draws on Meta's 'Facebook Privacy-Protected Full URLs Dataset' (v10), released through the Social Science One initiative (Messing et al., 2020). The dataset covers approximately 68 million URLs shared on Facebook between January 2017 and October 2022, providing aggregate counts of user interactions – views, shares, clicks and reactions – while protecting individual privacy through differential privacy, a technique that adds calibrated statistical noise to each variable before release (D'Orazio et al., 2015; Evans and King, 2023). Different levels of noise are applied to different interaction types depending on their sensitivity: for instance, the standard deviation of the added noise is $\sigma = 2228$ for views but only $\sigma = 14$ for shares, reflecting that users tend to view far more URLs than they actively share (Messing et al., 2020).

The unit of analysis is the individual URL. Each observation represents one web address and its lifetime aggregate engagement metrics across the study period. Because the dataset is restricted to URLs shared publicly at least approximately 100 times, our analysis captures widely circulated content rather than the full universe of links posted on Facebook. We note that this threshold means the dataset captures content that achieved substantial circulation – what we term 'widely circulated' URLs – rather than content that spreads rapidly and simultaneously across networks in the manner typically associated with virality (Nahon and Hemsley, 2014). While heavily shared content and viral content frequently overlap, the dataset's structure does not allow us to distinguish the two, as it aggregates engagement

over each URL's full lifetime rather than tracking the temporal acceleration of its spread.

We restrict our analysis to US-based engagement during the 2017–2022 period. From an initial pool of approximately 14.8 million candidate URLs, we applied a series of quality filters designed to retain only URLs where the true engagement signal substantially exceeds the privacy noise. The filtering strategy, detailed in Annex I, applies signal-to-noise ratio thresholds informed by Buntain et al. (2021) and Evans and King (2023) to ensure that our key measures – particularly the audience partisanship score – are statistically reliable. This process yielded a Phase 1 dataset of 236,341 URLs with reliable partisanship measures.

To address RQ2, which requires a measure of journalistic standards, we merged the Phase 1 dataset with NewsGuard scores, retaining only URLs whose parent domain had a valid rating. This produced the final analysis dataset of 130,448 URLs from identifiable news sources. The two datasets show very similar distributions on key variables (see Annex II), confirming that the NewsGuard filter does not introduce systematic selection bias.

Variables and operationalization. Our model examines how the number of times a URL was shared on Facebook relates to the number of times it was viewed and how that relationship varies depending on the audience's partisan intensity and the news source's professional quality. Table 1 reports the privacy-corrected descriptive statistics for all variables; Table 2 reports the corrected pairwise correlations.

The dependent variable URL views (url_views) captures the estimated number of unique users who saw a post containing the URL in their Facebook News Feed. Importantly, 'viewing' in this dataset reflects algorithmic exposure – it counts users who encountered the URL in their feed, regardless of whether they clicked through to the linked content. This measure thus captures the joint product of user sharing behaviour and Facebook's content distribution algorithm: a URL can accumulate views only if the algorithm surfaces it to users' feeds. Views therefore serve as our indicator of a URL's total reach on the platform.

The primary predictor URL shares (url_shares) counts the number of unique users who shared or re-shared the URL. Unlike passive exposure (views) or private consumption (clicks), sharing is a public act of amplification in which a user re-broadcasts content to their own network. Each share creates new potential pathways for algorithmic distribution, making shares the primary mechanism through which users actively extend a URL's reach beyond its original audience. The share-to-view relationship is our core measure of amplification – how many additional viewers each act of sharing generates – and is the focus of H1.

Audience partisan intensity (std_audience_alignment_strength) measures how ideologically concentrated each URL's sharing audience is (RQ1). This variable is derived from the Political Page Affinity (PPA) score, a measure developed using Barberá et al.'s (2015) methodology that estimates each user's ideological position based on which political Facebook Pages they follow. Each URL's raw PPA score is the share-weighted average of the PPA scores of the users who shared it, ranging from −2 (very liberal) to +2 (very conservative). We then take the absolute value of this score to capture the intensity of partisan alignment regardless of ideological direction and standardize it to a mean of zero and a standard deviation of one. Higher values indicate that a URL's audience was more ideologically concentrated – whether on the left or the right – while lower values indicate more ideologically diverse sharing. The calculation of URL-level PPA scores from the disaggregated data, including the reliability filters that ensure each score is based on a representative share of the URL's audience, is detailed in Annex I.

Adherence to professional journalistic standards (ng_score) measures the professional quality of the source at the domain level (RQ2). We operationalize these concepts using the NewsGuard score, a 0–100 rating that evaluates news websites on nine criteria covering credibility and transparency practices (Lühring et al., 2025). This is an assessment of the domain's adherence to professional journalistic norms – such as clearly labelling advertising, correcting errors and avoiding deceptive headlines – rather than a measure of how credible audiences perceive the source to be (Carlson, 2017). Each URL inherits the NewsGuard score of its parent domain.

The control variable URL clicks (url_clicks) counts the number of unique users who clicked through to the linked content. Like sharing, clicking is an intentional act of engagement, but unlike sharing, it is primarily a private act of content consumption that does not expose the URL to new audiences. Including clicks allows us to separate the factors that drive baseline audience interest from those that motivate active dissemination. If a variable's effect on views persists after controlling for clicks, its influence operates through the amplification pathway (shares generating new exposure) rather than through simple content popularity.

Analytical strategy

We estimate privacy-aware linear regression models using the lmdp function from the PrivacyUnbiased R package (Evans et al., 2023), which corrects for the bias introduced by differential privacy noise – a form of measurement error that can attenuate coefficients, exaggerate them or even reverse their sign if left uncorrected. Following Evans and King's (2023) reasoning, linear regression provides the best linear approximation to the underlying relationships while remaining computationally feasible for large differentially private datasets, where the additional uncertainty from privacy noise may make the advantages of more complex functional forms undetectable. Where the noise creates

Table 1. Privacy–corrected descriptive statistics – Full URLs dataset filtered by the NewsGuard score ($N = 130,448$).

Variable	Mean	SD	Skew	Kurtosis
url_views	15,333,617.42	23,255,292.51	10.30	307.55
url_shares	137,567.82	303,455.06	10.92	206.73
std_audience_alignment_strength	0.00	1.00	−0.49	1.94
url_clicks	867,175.54	1,902,932.40	12.76	244.55
ng_score	79.14	23.08	−1.27	4.10

numerical instability in the variance-covariance matrix, we employ bootstrap variance estimation with 1000 replications – a documented fallback for high-noise datasets within the PrivacyUnbiased framework.

To address H1, RQ1 and RQ2, we fit four nested models (Table 3) on the full analysis dataset ($N = 130,448$), progressively adding variables to assess each factor's independent contribution:

Each coefficient in these models has a concrete interpretation. The shares coefficient (in all models) estimates how many additional views each additional share generates, on average, holding other variables constant – this is the amplification rate at the heart of H1. The audience alignment strength coefficient captures whether URLs shared by more ideologically concentrated audiences tend to reach more or fewer total viewers, controlling for their number of shares – this is the partisan exposure gradient central to RQ1. The journalistic standards coefficient estimates how much additional reach a URL gains for each point increase on the NewsGuard scale, independent of its sharing volume – the quality premium that addresses RQ2. Model comparisons follow Evans and King's framework, using estimated error variance ($\sigma^{\wedge 2}$) as the primary fit metric, supplemented by adjusted R^2 for interpretive context.

It is important to clarify what these models do and do not estimate. The relationship between sharing and viewing is inherently bidirectional: shares generate views (through amplification), and views may prompt further shares (through exposure). As Trilling et al. (2022) note, disentangling this reciprocal influence would require time-series data with fine temporal resolution. The monthly granularity of our dataset does not support such analysis. Our models therefore estimate the cross-sectional association between sharing volume and total reach, which captures the net amplification effect but does not isolate the causal direction. Similarly, the audience alignment and journalistic standards variables enter the model as additive predictors. They capture how these URL-level characteristics co-vary with reach after accounting for sharing volume – not a mediation or moderation pathway in the formal causal sense.

To address RQ3, we estimate Model 4 separately for each quarter from 2017-Q1 to 2021-Q3, allowing all coefficients to vary freely across time periods. We adopt this approach rather than incorporating time as a single factor (e.g. through period dummies or a time trend) for two substantive reasons. First, our theoretical interest is in how the entire relationship between sharing and viewing shifts in response to Facebook's evolving governance – including changes in the amplification rate, the partisan penalty and the journalistic quality premium – rather than in a single main effect of time. Estimating the full model within each quarter allows every coefficient to change independently, capturing the possibility that platform interventions affected different aspects of content distribution in different ways and at different times. Second, key platform interventions – such as the emergency measures during the 2020 US election – were discrete, targeted events rather than smooth temporal trends. A pooled model with period dummies would impose parametric constraints on how these effects unfold, while quarter-by-quarter estimation allows the data to reveal abrupt shifts, gradual drifts and post-intervention rebounds in each coefficient independently. The full quarterly coefficient table is reported in Annex III.

Findings

Our analysis employs privacy-aware regression models using the lmdp function from the PrivacyUnbiased R package. This approach generates statistically valid inferences from differentially private data by correcting for measurement error introduced by privacy noise. As noted by

Table 2. Privacy–corrected correlations – Full URLs dataset filtered by the NewsGuard score.

Variable pair	Corrected *r*
Views and shares	0.731
Views and clicks	0.757
Shares and clicks	0.448

Table 3. Model specifications.

M	Formula
1	url_views∼ url_shares + std_audience_alignment_strength
2	url_views∼ url_shares + std_audience_alignment_strength + url_clicks
3	url_views∼ url_shares + std_audience_alignment_strength + ng_score
4	url_views∼ url_shares + std_audience_alignment_strength + ng_score + url_clicks

Table 4. Hierarchical regression models predicting URL views.

Predictor	Model 1	Model 2	Model 3	Model 4
URL shares	83.55 (1.39)	56.07 (1.29)	56.04 (1.25)	56.04 (1.23)
Partisan alignment strength	–	—	−2.441e + 06 (60,464.28)	−2.295e + 06 (61,754.60)
Journalistic standards score	—	—	—	28,756.74 (1,158.55)
URL clicks	—	6.57 (0.24)	6.36 (0.24)	6.37 (0.25)
Intercept	3.839e + 06 (178,503.78)	1.927e + 06 (167,417.17)	2.109e + 06 (171,322.09)	−177,573.8 (206,017.8)
Adjusted R^2	0.53	0.76	0.78	0.78
Sigma-squared ($\sigma^{\wedge 2}$)	2.52e + 14	1.27e + 14	1.21e + 14	1.21e + 14
N	130,448	130,448	130,448	130,448

Evans and King (2023), this noise can sometimes create numerical instability. Our models address this by implementing bootstrap variance estimation – a documented fallback for high-noise datasets – to ensure reliable standard errors. Results are reported in Table 4 and discussed below.

H1: The shares-to-views relationship. To test H1, we first estimated a bivariate linear model predicting URL views from URL shares using the Evans and King bias-correction method implemented through the PrivacyUnbiased package to estimate linear models on differentially private data containing 130,448 observations. The baseline specification (Model 1) regressing URL views on URL shares yielded a bias-corrected coefficient of $\beta = 83.55$ (SE = 1.39, $p < 0.001$), indicating that each additional share is associated with approximately 84 additional views after accounting for differential privacy noise. Given the substantial noise in the data ($\sigma = 2228$ for url_views as specified in the dataset codebook), we employed bootstrap variance estimation with 1000 simulations to ensure numerical stability when the variance-covariance matrix approached singularity.

To assess robustness and account for potential confounding, we extended the specification to include URL clicks (Model 2), a variable mechanically related to both shares and views within the Facebook platform. The augmented model maintained a highly significant share–view relationship, with each additional share associated with 56.07 additional views ($\beta = 56.07$, SE = 1.29, $p < 0.001$), while clicks contributed an additional 6.57 views per click ($\beta = 6.57$, SE = 0.24, $p < 0.001$). Using the Evans and King primary fit metric (Sigma_sq_hat), Model 2 demonstrated superior statistical fit with substantially reduced unexplained variance. The adjusted R^2 of 0.765 provides supplementary evidence of model adequacy, though we prioritize the bias-corrected error variance estimates following the Evans and King framework. These results provide robust support for H1, demonstrating a strong positive correlation between link sharing and viewing that persists even when controlling for user engagement through clicks.

RQ1: The moderating role of partisan alignment. To address RQ1, we extended the model to include standardized audience alignment strength, assessing how the viewing–sharing relationship varies with audience partisanship. The analysis was restricted to URLs with partisan audience (PPA) coverage of at least 0.5, ensuring that partisanship was estimated from a representative subset of each URL's sharing audience.

The effect of shares remains stable and highly significant ($\beta = 56.04$, $p < 0.001$), confirming the robustness of H1. Crucially, the coefficient for audience alignment strength is large, negative and significant ($\beta = -2{,}441{,}132$, $p < 0.001$). Holding shares and clicks constant, URLs with more intensely partisan audiences receive significantly fewer views.

This 'dampening' effect could reflect either structural network limitations – partisan content hitting organic ceiling effects within homophilic communities – or active algorithmic suppression. The static analysis alone cannot distinguish between these mechanisms. However, the temporal analysis in RQ3 provides crucial evidence: if the penalty were purely structural, we would expect stable coefficients over time. Instead, as we demonstrate below, the partisan penalty shows marked volatility, particularly intensifying during periods of known platform intervention.

RQ2: The role of journalistic standards. To examine RQ2 and assess how professional journalistic quality moderates viewing and sharing, we incorporated the NewsGuard score as an indicator of source credibility into the full additive model (Model 4). The model yields clear, significant results for all predictors and the best overall fit (Adjusted $R^2 = 0.78$). The effect of shares remains robust ($\beta = 56.04$ $p < 0.001$), and the negative influence of audience alignment persists ($\beta = -2{,}295{,}054.72$, $p < 0.001$). Most importantly, the NewsGuard score is positive and highly significant ($\beta = 28{,}756.74$, $p < 0.001$). A one-point increase on NewsGuard's 100-point scale corresponds to over 28,700 additional views, even after controlling for shares, clicks and audience polarization. This finding shows that adherence to professional journalistic standards – as measured by NewsGuard scores – is a strong, independent driver of content reach, while audience polarization continues to constrain exposure across all content types.

RQ3: Algorithmic evolution over time. To investigate how Facebook's evolving content distribution algorithms and moderation policies shaped these relationships over time (RQ3), we estimated Model 4 separately for each quarter from 2017 to 2022. The full quarterly coefficients are reported in the Supplementary Appendix (Annex II), with key trends visualized in Figure 1. The temporal analysis reveals marked shifts in the effects of engagement and content quality.

While the click coefficient remained stable (consistently yielding between 6 and 7.5 views per click), the shares effect showed significant volatility. It peaked at 70.8 views per share in 2017-Q4 and 70.1 in 2019-Q2, before declining sharply to its lowest point (45.6) in 2020-Q3, during the run-up to the US presidential election. Rather than continuing to drop after the election and the January 2021 Capitol Hill invasion, the coefficient rebounded modestly to 53.9 in 2021-Q1. This volatility strongly suggests algorithmic responses to politically sensitive events.

The fluctuations in the audience alignment penalty provide compelling evidence of active platform intervention rather than mere structural network effects. A temporal pattern inconsistent with structural homophily – which would presumably produce stable coefficients – emerged. Instead, the sharp fluctuations, including a dramatic intensification of the penalty in 2020-Q3 (−2.90 M), coincide with the onset of the COVID-19 pandemic and the election run-up. This pattern strongly suggests deliberate algorithmic modulation of partisan content reach.

Patterns in content-quality signals are particularly revealing. A clear temporal synchronization emerged: the audience alignment penalty was most intense in 2020-Q3 (−2.90 M), precisely as the reward for high-quality journalism (NewsGuard score) began to surge. This positive effect of journalistic quality climbed from 2020-Q2 (31,545), strengthened significantly through 2020-Q4 (41,148) and 2021-Q1 (58,203) and ultimately peaked in the second quarter of 2021 (76,903).

This synchronization suggests that, in response to these major political and public health crises, Facebook's algorithm amplified the reach of content from sources rated highly by NewsGuard for their adherence to professional journalistic standards, while simultaneously maintaining a significant (if volatile) penalty for partisan content. Overall, while the core logic of Facebook's content distribution may have remained consistent, its strategic weighting of partisanship and credibility shifted in response to perceived threats to the information ecosystem.

Discussion and conclusion

Despite their widely recognized relevance in contemporary society, basic questions about the role social media platforms play as information gateways remain largely unanswered. The existing literature offers conflicting insights regarding social media's role in amplifying partisan and problematic information (Fraxanet et al., 2025). In this paper, we addressed these crucial and complex questions by analysing over 130,000 highly circulated links shared on Facebook in the USA between 2017 and 2022, using the largest privacy-protected dataset of platform engagement data made available to independent researchers. While the USA focus limits the generalizability of our results, the intrinsic complexity of social media platforms often necessitates such a focused approach to yield clear findings. Taken together, our results reveal that Facebook operates not as a neutral conduit for user-driven content circulation, but as an active curator whose amplification rules are complex, non-neutral and temporally dynamic.

Our first and most fundamental finding is that sharing reliably predicts increased viewership: each additional share of a widely circulated URL is associated with approximately 56 additional views, even after controlling for clicks. This relationship, while perhaps intuitive, has rarely been demonstrated empirically at this scale, in large part because viewing data is commercially sensitive and has historically been unavailable to researchers outside platform companies (Bakshy et al., 2015b). Prior research on content circulation has relied primarily on sharing metrics as a proxy for reach, yet the relationship between sharing and actual exposure cannot be assumed to be linear or uniform (Trilling et al., 2022). Notably, the latest generation of platform research APIs – including the Meta Content Library and the TikTok Research API – now routinely provides view counts alongside other engagement metrics, opening a productive avenue for replicating and extending this type of analysis across platforms and national contexts. Our analysis demonstrates that for the most widely circulated content on Facebook, sharing functions as a powerful amplification lever – but one whose effects are substantially shaped by what Petre et al. (2019) have termed 'algorithmic visibility': the platform's capacity to govern what users see without removing content outright.

This finding speaks to a core tension in platform studies between user agency and algorithmic mediation. The two-step flow model, as updated for digital environments (Bakshy et al., 2012), suggests that opinion leaders drive information diffusion through their sharing decisions. Our results partially support this view: sharing does predict reach. However, the substantial variation in amplification effects across content types and time periods, which we document below, suggests that the relationship between a user's decision to share and the resulting audience exposure is far from mechanical. Rather, it is filtered through algorithmic systems that assess, weight and selectively distribute content according to criteria that remain largely opaque to both users and researchers (DeVito, 2017; Gillespie, 2022). In this sense, sharing on Facebook is better understood as a necessary but not sufficient condition for

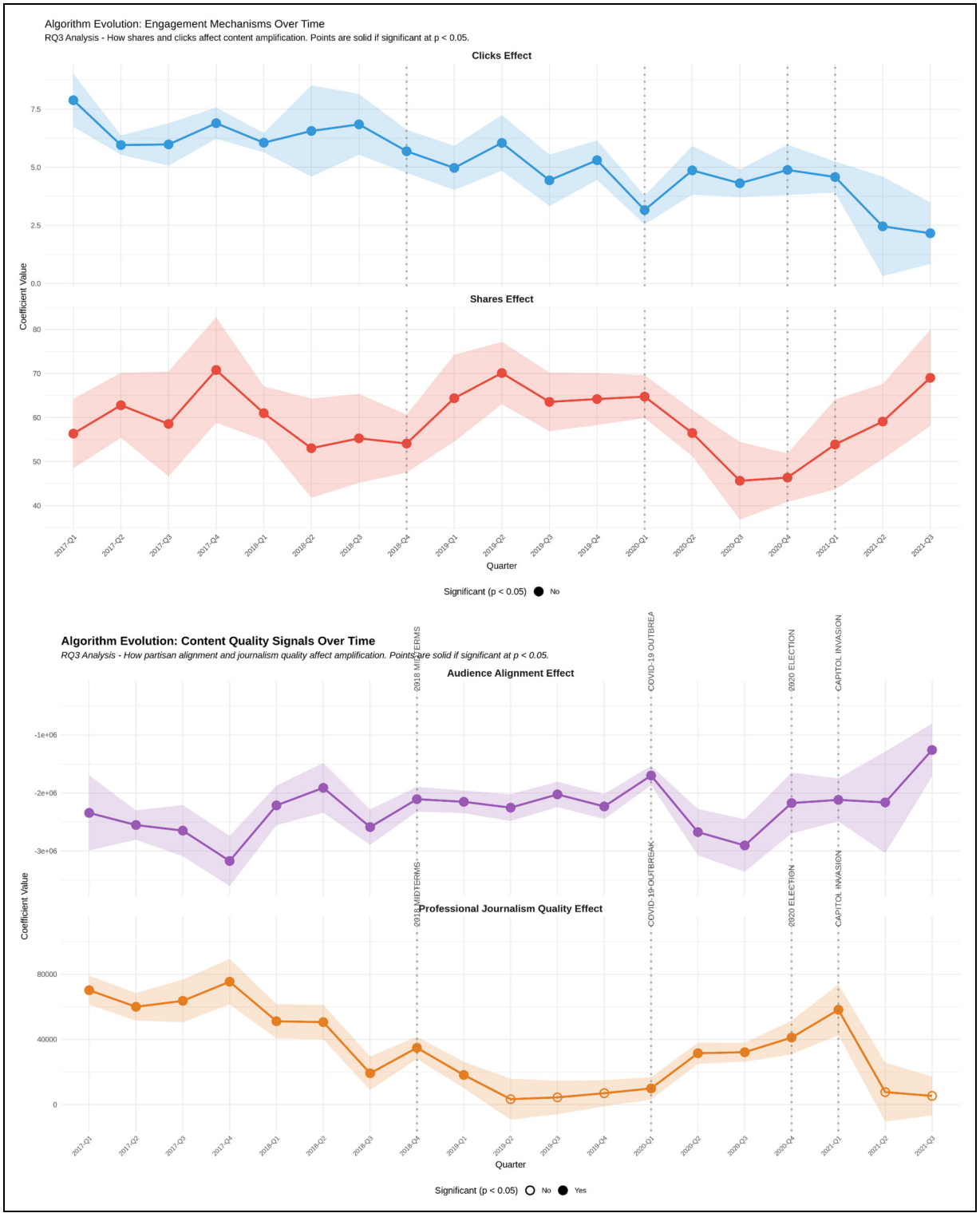

Figure 1. Algorithm evolution: engagement and content signals over time.

wide reach – a signal that enters the platform's distribution machinery rather than a direct channel to audiences.

Our analysis of how audience partisanship shapes the sharing–viewing relationship yields what is perhaps the study's most consequential finding: URLs shared by more intensely partisan audiences receive significantly fewer views, even when the volume of shares and clicks is held constant. This reach penalty associated with partisan

content is both statistically robust and substantively large as a one-standard-deviation increase in audience partisan alignment is associated with over 2.3 million fewer views.

This finding contributes to ongoing debates about the role of platforms in political polarization. A substantial body of research has examined whether social media creates 'echo-chambers' or 'filter bubbles' that reinforce pre-existing political views (Bakshy et al., 2015; Barberá et al., 2015; Guess et al., 2023b). Our results add a distinct dimension to this debate by demonstrating that the platform's distribution system itself appears to constrain the reach of content with highly partisan audiences, independently of any organic ceiling effects produced by homophilic network structures. This matters because it positions the platform as an active participant in shaping which political content gains wide visibility, rather than merely reflecting the preferences and network structures of its users.

The mechanism behind this penalty, however, cannot be fully disentangled through our cross-sectional analysis alone. The reduced reach could reflect structural network limitations, where partisan content circulates within ideologically homogeneous communities that impose natural ceilings on viewership. Alternatively, it could reflect active algorithmic suppression, with the platform deliberately reducing the distribution of content that scores high on partisan engagement signals. As we discuss in relation to our temporal findings below, the evidence strongly favours the latter interpretation, but with important caveats.

What is at stake here extends beyond the technical question of algorithmic design. If users who engage with polarized, conspiracist or anti-scientific content can influence others through their sharing practices, the degree to which platform architectures amplify or attenuate that influence becomes a matter of democratic concern (Pennycook et al., 2021). Research on misinformation sharing suggests that people do not always share content out of belief – they may do so out of partisan identity signalling, concern for peers or simple inattention to accuracy (Pennycook et al., 2021). Our findings suggest that regardless of the motivations behind sharing, the platform's distribution system imposes a differential penalty on content associated with strongly partisan audiences. Whether this penalty is calibrated to serve democratic values or primarily to manage platform reputation is a question our data cannot answer directly, but the existence of the penalty itself is significant and worthy of further scrutiny.

Our findings show that adherence to professional journalistic standards is a strong, independent predictor of content reach and adds an important layer to understanding Facebook's amplification logic. A one-point increase on NewsGuard's 100-point scale corresponds to approximately 28,700 additional views, even after controlling for shares, clicks and audience partisanship. This suggests that the platform's distribution system does not treat all content equally: sources that meet established standards of journalistic practice appear to receive a measurable visibility advantage.

This finding has implications for scholarly and policy debates about platform responsibility toward news quality. Research on algorithmic curation has documented how platform logics can privilege engagement-maximizing content at the expense of journalistic quality (Bozdag, 2013; Schneiders and Stark, 2025), potentially creating incentive structures that reward sensationalism over accuracy. Our results offer a partial counterpoint: at least during the period under study, Facebook's algorithm appears to have rewarded content from sources that adhere to professional standards, suggesting an active quality signal in the platform's ranking criteria.

However, a crucial distinction must be drawn between a source's adherence to journalistic standards – especially when these are assessed by just one commercial rating service – and its credibility as perceived by audiences. As Carlson (2017) has argued, journalism is notoriously self-defined, and the boundaries of what constitutes 'real' or 'reliable' journalism are themselves sites of political contestation. NewsGuard's assessments reflect expert evaluation of whether outlets follow established production practices – such as distinguishing news from opinion, disclosing ownership and correcting errors – but they do not capture whether audiences perceive those outlets as trustworthy. A source rated highly by NewsGuard may be dismissed as biased by segments of the public, while a source with low professional standards may enjoy high trust among its partisan audience. Our study can demonstrate that the platform's algorithm appears to favour content from professionally vetted sources, but it cannot speak to whether this algorithmic preference aligns with or contradicts the trust judgments of the platform's users. This gap between institutional quality assessment and audience perception represents an important limitation of our analysis and a productive avenue for future research that could combine platform trace data with survey-based measures of media trust.

The quarterly analysis of our model across the 2017–2022 period provides what we regard as the study's most distinctive contribution: empirical evidence of platform governance in action. Structural properties of information networks – such as homophily and audience segmentation – tend to produce relatively stable patterns of circulation over time. By contrast, governance interventions – algorithmic adjustments, policy changes and emergency measures – introduce longitudinal volatility that can serve as an observable trace of platform-mediated control over distribution (Gillespie, 2022; Gorwa, 2019). Our temporal analysis reveals precisely this kind of volatility: While the click coefficient remained remarkably stable across the entire period, the amplification effect of shares showed dramatic fluctuations. They in fact peaked at approximately 71 views per share in late 2017 and mid-2019 before declining sharply to their lowest point of roughly 46 views per share in the third quarter of 2020,

during the convergence of the COVID-19 pandemic and the US presidential election campaign. The partisan reach penalty showed even more pronounced instability, intensifying dramatically in 2020-Q3 before partially recovering. These patterns are inconsistent with purely structural explanations, which would predict temporal stability, and instead strongly suggest active algorithmic recalibration in response to perceived threats to the information ecosystem.

Particularly revealing is the temporal synchronization between the partisan penalty and the journalistic quality reward. During the period when the algorithm most aggressively penalized partisan content (2020-Q3), the positive effect of NewsGuard scores began a sustained surge, climbing from approximately 31,500 additional views per quality point in early 2020 to over 76,900 by mid-2021. This matching pattern suggests a deliberate dual strategy: simultaneously suppressing the reach of highly partisan content while amplifying content from professionally vetted news sources. The timing aligns with Facebook's reported deployment of emergency 'break the glass' measures, i.e. protocols that reportedly doubled the penalty for politically extreme content and reduced the algorithmic distribution of civic and political content more broadly (Bagchi et al., 2024; Isaac and Frenkel, 2020; Jackson, 2024). Our findings provide independent quantitative evidence consistent with these reported interventions, offering a rare empirical window into governance measures that are otherwise known only through journalistic accounts and leaked internal documents.

This evidence of active algorithmic intervention has important implications for the broader debate on platform governance. The distinction between structural and algorithmic constraints on content reach is consequential: if reduced visibility of partisan content were solely a by-product of network topology, platforms could plausibly claim neutrality in the distribution process. However, the temporal volatility we observe – closely synchronized with political crises – complicates this interpretation. Instead, it suggests that Facebook functions as an active gatekeeper, with interventions that appear to intensify during periods of heightened political sensitivity.

The motivations underlying such interventions remain difficult to assess empirically. Rather than attributing them to a single driving logic, it is more plausible to view these actions as the outcome of multiple, overlapping considerations. These may include efforts to safeguard the integrity of public discourse and democratic processes, responses to heightened public and institutional scrutiny and concerns related to brand reputation, user trust and commercial sustainability. As a publicly traded company, Meta operates within a complex set of incentives that can align – or coexist – with broader societal expectations. The intensification of content moderation observed during events such as the 2020 election and the COVID-19 pandemic occurred under conditions of exceptional visibility and pressure from policymakers, civil society and advertisers. In this context, 'break the glass' interventions can be interpreted as part of a broader repertoire of platform responses to systemic risk, without allowing for definitive conclusions about the relative weight of normative versus strategic motivations.

Our temporal analysis reinforces this interpretation by showing that shifts in governance are reflected in measurable changes in amplification dynamics. The responsiveness of the sharing–viewing relationship to these shifts indicates that platform interventions – whether more or less intensive – are closely linked to how information circulates. Importantly, this insight extends beyond moments of explicit moderation: even in their absence, algorithmic ranking systems and design choices continue to structure visibility. From this perspective, non-intervention can also be understood as a form of governance, insofar as it leaves in place mechanisms that differentially amplify content.

The trajectory of Meta's approach to news and political content since our observation period further contextualizes these findings. Between 2021 and 2025, the company progressively reduced the visibility of political content in user feeds (Giglietto, 2025; Stepanov and Gupta, 2021), before announcing in January 2025 a shift toward a more non-interventionist approach to content moderation (Kaplan, 2025). This sequence illustrates that platform governance is not stable or uniform but adaptive and context-dependent. Content governance emerges from the ongoing – and shifting – alignment of normative considerations (such as information quality and electoral integrity), political pressures (including regulatory scrutiny and public accountability demands) and economic imperatives (such as user engagement, advertiser trust and platform sustainability). As these considerations are reweighted in response to changing circumstances, observable shifts in moderation practices and content visibility follow. Our temporal analysis suggests that such shifts are likely to produce measurable changes in amplification dynamics.

Taken together, these results reveal platform governance as a system of calibrated adjustments shaped by shifting political and economic conditions. Our data cannot determine whether any given intervention is driven by democratic concern, reputational risk management or commercial calculation – but they demonstrate that the effects on content distribution are real and consequential. In this sense, the platform's role in shaping which information citizens encounter during periods of democratic contestation is not a background condition but an active and evolving exercise of power (Gillespie, 2022; Gorwa, 2019).

Acknowledgements

This work has received funding from the European Union under the Horizon Europe vera.ai project, Grant Agreement number 101070093.

ORCID iDs

Fabio Giglietto https://orcid.org/0000-0001-8019-1035
Giada Marino https://orcid.org/0000-0002-9087-2608

Author contributions

Both authors contributed to the conceptualization of the paper. FG led the design and implementation of the empirical analysis and conducted all data analyses. He was responsible for the full drafting of the Data, variables and measures and Findings sections. GM developed the theoretical background and drafted all subsections of the Theoretical background section. The Introduction, Discussion and Conclusion sections were jointly written by both authors.

Funding

The authors disclosed receipt of the following financial support for the research, authorship and/or publication of this article: This work was supported by horizon Research and Innovation Actions under the call a human-centred and ethical development of digital and industrial technologies 2021 (Project name Verification Assisted by Artificial Intelligence - vera.ai; Grant Agreement No. 101070093).

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Supplemental material

Supplemental material for this article is available online.

References

Adamic LA, Lento TM, Adar E, et al. (2016, February 8) Information evolution in social networks. *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*. WSDM 2016: Ninth ACM International Conference on Web Search and Data Mining, San Francisco California USA. https://doi.org/10.1145/2835776.2835827.

Allen J, Watts DJ and Rand DG (2024) Quantifying the impact of misinformation and vaccine-skeptical content on Facebook. *Science (New York, N.Y.)* 384(6699): 1–9.

An J, Quercia D and Crowcroft J (2013) Fragmented social media: a look into selective exposure to political news. *Proceedings of the 22nd International Conference on World Wide Web*: 51–52. https://doi.org/10.1145/2487788.2487807.

Bagchi C, Menczer F, Lundquist J, et al. (2024) Social media algorithms can curb misinformation, but do they? In *arXiv*. https://doi.org/10.5281/zenodo.13787981.

Bakshy E, Messing S and Adamic LA (2015) Exposure to ideologically diverse news and opinion on Facebook. *Science* 348(6239): 1130–1132.

Bakshy E, Rosenn I, Marlow C, et al. (2012) The role of social networks in information diffusion. *Proceedings of the 21st International Conference on World Wide Web*: 519–528. https://doi.org/10.1145/2187836.2187907.

Barberá P, Jost JT, Nagler J, et al. (2015) Tweeting from left to right: Is online political communication more than an Echo chamber? *Psychol Sci* 26(10): 1531–1542.

Baumeister RF, Bratslavsky E, Finkenauer C, et al. (2001) Bad is stronger than good. *Review of General Psychology: Journal of Division 1, of the American Psychological Association* 5(4): 323–370.

Bozdag E (2013) Bias in algorithmic filtering and personalization. *Ethics Inf Technol* 15(3): 209–227.

Brady WJ, Wills JA, Jost JT, et al. (2017) Emotion shapes the diffusion of moralized content in social networks. *Proc Natl Acad Sci U S A* 114(28): 7313–7318.

Brown DK, Lough K and Riedl MJ (2020) Emotional appeals and news values as factors of shareworthiness in ice bucket challenge coverage. *Digital Journalism* 8(2): 267–286.

Buntain C, Bonneau R, Nagler J, et al. (2021) *A summary statistic for political ideology of web domains using differentially private engagement data with applications to Social Science One*. *SSRN*. https://doi.org/10.2139/ssrn.3765240.

Carlson M (2017) *Journalistic Authority: Legitimating News in the Digital Era*. New York City, United States: Columbia University Press.

Chadwick A (2013) *The Hybrid Media System: Politics and Power*. Oxford, United Kingdom: Oxford University Press.

Chaiken S, Giner-Sorolla R and Chen S (1996) Beyond accuracy: Defense and impression motives in heuristic and systematic information processing. In: Gollwitzer Peter M and Bargh John A (eds) *The Psychology of Action: Linking Cognition and Motivation to Behavior*. New York City, United States: Guilford Press, 553–578.

Clemm von Hohenberg B (2018) An ocean of possible truth: biased processing of news on social media. *SSRN*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3281038.

DeVito MA (2017) From editors to algorithms: A values-based approach to understanding story selection in the Facebook news feed. *Digital Journalism* 5(6): 753–773.

D'Orazio V, Honaker J and King G (2015) Differential privacy for social science inference. *SSRN*. https://ssrn.com/abstract=2676160 or http://dx.doi.org/10.2139/ssrn.2676160 https://doi.org/10.2139/ssrn.2676160.

Evans G and King G (2023) Statistically valid inferences from differentially private data releases, with application to the Facebook URLs dataset. *Political Analysis: An Annual Publication of the Methodology Section of the American Political Science Association* 31(1): 1–21.

Evans G, King G, Schwenzfeier M, et al. (2023) Statistically valid inferences from privacy-protected data. *Am Polit Sci Rev* 117(4): 1275–1290.

Fraxanet E, Kaltenbrunner A, Germano F, et al. (2025) Analyzing news engagement on Facebook: Tracking ideological segregation and news quality in the Facebook URL dataset. *EPJ Data Sci* 14(1): 1–24.

Frenkel S and Isaac M (2024, September 24) How Meta distanced itself from politics. *The New York Times*. https://www.nytimes.com/2024/09/24/technology/meta-election-politics.html.

Giglietto F (2025) "A pretty blunt approach": Meta's political content reduction policy and Italian parliamentarians' Facebook visibility. In *SocArXiv*. https://doi.org/10.31235/osf.io/8dqag_v2.

Giglietto F, Righetti N, Rossi L, et al. (2020) It takes a village to manipulate the media: Coordinated link sharing behavior during 2018 and 2019 Italian elections. *Information, Communication and Society* 23(6): 867–891.

Gillespie T (2017) Governance of and by platforms. In: Burgess J, Poell T and Marwick A (eds) *SAGE Handbook of Social Media*. Thousand Oaks, California, United States: Sage Publishing, 254–278.

Gillespie T (2021) *Custodians of the Internet*. New Haven, Connecticut, United States: Yale University Press.

Gillespie T (2022) Do not recommend? Reduction as a form of content moderation. *Social Media + Society* 8(3): 1–13.

Goel P, Green J, Lazer D, et al. (2025) Using co-sharing to identify use of mainstream news for promoting potentially misleading narratives. *Nat Hum Behav* 9(2025): 1–18.

González-Bailón S, Lazer D, Barberá P, et al. (2023) Asymmetric ideological segregation in exposure to political news on Facebook. *Science (New York, N.Y.)* 381(6656): 392–398.

González-Bailón S, Lazer D, Barberá P, et al. (2024) The diffusion and reach of (mis)information on Facebook during the U.S. 2020 election. *Sociol Sci* 11(41): 1124–1146.

Gorwa R (2019) What is platform governance? *Information, Communication and Society* 22(6): 854–871.

Gorwa R, Binns R and Katzenbach C (2020) Algorithmic content moderation: Technical and political challenges in the automation of platform governance. In *SocArXiv*. https://doi.org/10.31235/osf.io/fj6pg.

Green J, Mccabe S, Shugars S, et al. (2025) Curation bubbles. *Am Polit Sci Rev* 119(4): 1–19.

Guess A, Aslett K, Tucker J, et al. (2021) Cracking open the news feed: Exploring what U.S. Facebook users see and share with large-scale platform data. *Journal of Quantitative Description: Digital Media* 1: 1–49.

Guess AM, Malhotra N, Pan J, et al. (2023a) How do social media feed algorithms affect attitudes and behavior in an election campaign?. *Science (New York, N.Y.)* 381(6656): 398–404.

Guess AM, Malhotra N, Pan J, et al. (2023b) Reshares on social media amplify political news but do not detectably affect beliefs or opinions. *Science (New York, N.Y.)* 381(6656): 404–408.

Guess AM, Nyhan B and Reifler J (2020) Exposure to untrustworthy websites in the 2016 US election. *Nat Hum Behav* 4(5): 472–480.

Haller A and Holt K (2019) Paradoxical populism: How PEGIDA relates to mainstream and alternative media. *Information, Communication and Society* 22(12): 1665–1680.

Isaac M and Frenkel S (2020, August 21) Facebook braces itself for Trump to cast doubt on election results. *The New York Times*. https://www.nytimes.com/2020/08/21/technology/facebook-trump-election.html.

Jackson D (2024, October 1) *We Know a Little About Meta's "Break Glass" Measures. We Should Know More*. Tech Policy Press. https://techpolicy.press/we-know-a-little-about-metas-break-glass-measures-we-should-know-more.

Kaiser J, Keller TR and Kleinen-von Königslöw K (2021) Incidental news exposure on Facebook as a social experience: The influence of recommender and media cues on news selection. *Communic Res* 48(1): 77–99.

Kalsnes B and Larsson AO (2018) Understanding news sharing across social media: Detailing distribution on Facebook and Twitter. *Journal Stud* 19(11): 1669–1688.

Kaplan J (2025, January 7) *More Speech and Fewer Mistakes*. Meta. https://about.fb.com/news/2025/01/meta-more-speech-fewer-mistakes/.

Karlsen R (2015) Followers are opinion leaders: The role of people in the flow of political communication on and beyond social networking sites. *European Journal of Communication* 30(3): 301–318.

Karlsson M and Holt K (2016) Journalism on the Web. In: Matthew P (ed) *Oxford Research Encyclopedia of Communication*. Oxford Academic.

Karnowski V, Kümpel AS, Leonhard L, et al. (2017) From incidental news exposure to news engagement. How perceptions of the news post and news usage patterns influence engagement with news articles encountered on Facebook. *Comput Human Behav* 76: 42–50.

Karnowski V, Leiner DJ, Sophie Kümpel A, et al. (2021) Worth to share? How content characteristics and article competitiveness influence news sharing on social network sites. *Journal Mass Commun Q* 98(1): 59–82.

Kramer MH (2021) *Freedom of Expression as Self-Restraint*. Oxford, United Kingdom: Oxford University Press.

Kümpel AS (2019) The issue takes it all?: Incidental news exposure and news engagement on Facebook. *Digital Journalism* 7(2): 165–186.

Kümpel AS, Karnowski V and Keyling T (2015) News sharing in social media: A review of current research on news sharing users, content, and networks. *Social Media + Society* 1(2): 1–14.

Kunda Z (1990) The case for motivated reasoning. *Psychol Bull* 108(3): 480–498.

Larsson AO (2017) Going viral? Comparing parties on social media during the 2014 Swedish election. *Convergence* 23(2): 117–131.

Lühring J, Metzler H, Lazzaroni R, et al. (2025) Best practices for source-based research on misinformation and news trustworthiness using NewsGuard. *Journal of Quantitative Description: Digital Media* 5: 1–53.

Messing S, DeGregorio C, Hillenbrand B, et al. (2020) *Facebook Privacy-Protected Full URLs Data Set* [Dataset]. Harvard Dataverse. https://doi.org/10.7910/DVN/TDOAPG.

Messing S and Westwood SJ (2014) Selective exposure in the age of social media: Endorsements trump partisan source affiliation when selecting news online. *Communic Res* 41(8): 1042–1063.

Meta Platforms, Inc (2025) *Meta Content Library API version v5.0*. https://developers.facebook.com/docs/content-library-and-api/citations.

Nahon K and Hemsley J (2014) *Going viral*. Cambridge, United Kingdom: Polity Press.

Newman N, Fletcher R, Robertson CT, et al. (2024) Reuters Institute digital news report 2024. *Reuters Institute for the Study of Journalism*. https://doi.org/10.60625/RISJ-VY6N-4V57.

Pennycook G, Epstein Z, Mosleh M, et al. (2021) Shifting attention to accuracy can reduce misinformation online. *Nature* 592(7855): 590–595.

Petre C, Duffy BE and Hund E (2019) "Gaming the system": Platform paternalism and the politics of algorithmic visibility. *Social Media + Society* 5(4): 1–12.

Rossi L, Giglietto F and Marino G (2023) Cracking open the European newsfeed. *Journal of Quantitative Description: Digital Media* 3(2023): 1–45.

Schäfer S (2023) Incidental news exposure in a digital media environment: A scoping review of recent research. *Ann Int Commun Assoc* 47(2): 242–260.

Schneiders P and Stark B (2025) Ensuring news quality in platformized news ecosystems: Shortcomings and recommendations for an epistemic governance. *Media Commun* 13(10042). https://doi.org/10.17645/mac.10042.

Smith B (2024) *Traffic: Genius, Rivalry, and Delusion in the Billion-Dollar Race to Go Viral*. London, United Kingdom: Penguin.

Smith M, McAweeney E and Ronzaud L (2020) *The COVID-19 "infodemic". A preliminary analysis of the online conversation surrounding the Coronavirus pandemic* (No. 04.2020). Graphika. https://graphika.com/reports/the-covid-19-infodemic/

Stepanov A and Gupta A (2021, February 10) *Political content in feeds*. Meta. https://web.archive.org/web/20251011124003/https://about.fb.com/news/2021/02/reducing-political-content-in-news-feed/.

Trilling D, Kulshrestha J and De Vreese C (2022) Is sharing just a function of viewing? The sharing of political and non-political news on Facebook. *J Quant. Descr Digit Media* 2(2022): 1–103.

Trilling D, Tolochko P and Burscher B (2017) From newsworthiness to shareworthiness: How to predict news sharing based on article characteristics. *J Quant. Descr Digit Media* 94(1): 38–60.

Wischnewski M, Bruns A and Keller T (2021) Shareworthiness and motivated reasoning in hyper-partisan news sharing behavior on Twitter. *Digital Journalism* 9(5): 549–570.