International Journal of Politics, Culture, and Society
https://doi.org/10.1007/s10767-026-09589-x

Populism in Algorithmic Publics: Reassessing the Spiral of Silence Under AI-Driven Visibility

Lourenço Silva Ferreira[1]

Received: 19 February 2026 / Accepted: 29 June 2026
© The Author(s) 2026

Abstract

This article offers a critical theoretical reassessment of how political expression, dissent, and silence are structured within algorithmically mediated publics.Moving beyond the assumption of digital platforms as neutral arenas of participation, the article examines how AI-driven visibility infrastructures and platform-specific interactional cultures reshape the conditions under which opinions become visible or legitimate to express, as well as the risks associated with their articulation. Drawing on an illustrative, case-based analysis of highly visible user interactions on Facebook, TikTok, and X surrounding posts by the Portuguese populist leader André Ventura, the article mobilises empirical material to illuminate how distinct sociotechnical environments generate differentiated opinion climates. Engaging critically with Spiral of Silence theory, the analysis suggests that contemporary silencing dynamics may be driven less by interpersonal fear of isolation than by the interplay between algorithmic filtering and emotionally charged interactional environments. By reconceptualising the Spiral of Silence as an affective–algorithmic process, the article contributes to debates in communication theory and digital political communication, while also reflecting on the normative implications of algorithmic visibility for both publicness and democratic participation, particularly in relation to platform governance.

Keywords Populism · Algorithmic visibility · Spiral of silence · Digital public sphere · Political communication · Democratic participation

✉ Lourenço Silva Ferreira
glourencosilvaferreira@gmail.com

[1] Department of Communication and Media Studies, Escola Superior de Educação de Coimbra, Polytechnic Institute of Coimbra, Coimbra, Portugal

Published online: 11 July 2026

Introduction

Algorithmically mediated visibility has become a central condition shaping contemporary public expression. On digital platforms, what appears as dominant opinion is increasingly structured less through interpersonal observation and more through platform-mediated signals of visibility. Under these conditions, silence and self-censorship do not simply emerge from social pressure, but from the perception of visibility itself—what is rendered visible and sustained within platform-mediated environments. This dynamic introduces a fundamental tension: while digital platforms expand opportunities for participation, they simultaneously produce asymmetries of exposure that render some voices highly visible and others structurally marginal.

Within this context, understanding how silence and self-censorship emerge in digital environments has become an increasingly relevant problem for communication research. Spiral of Silence theory (Noelle-Neumann, 1974, 1984), originally formulated to explain how fear of social isolation leads individuals to suppress minority views, has provided the foundational framework for investigating this problem. Although social media platforms are often associated with openness and participation, a growing body of research confirms that silencing dynamics persist in online environments. Hampton et al. (2014) demonstrated that users are significantly less willing to express dissenting political views on social media than in face-to-face settings, while Kim (2016) found that fear of social sanctions on Facebook remains a strong predictor of self-censorship. Neubaum and Krämer (2018) further showed that algorithmic amplification of consonant viewpoints generates distorted perceptions of consensus—what they term "spectral majorities"—that constrain expression even when actual opinion distributions are more plural. Taken together, this body of work establishes that silencing processes are operative in digital environments and are shaped by platform-specific dynamics that classical formulations of the theory did not anticipate.

However, two important limitations persist in existing scholarship. First, platform architectures are frequently treated as contextual backdrops rather than as active forces shaping opinion visibility and perceived dominance—the infrastructural mechanisms through which perceived consensus is produced remain under-specified. Second, the affective dimensions of digital communication—particularly the role of emotional intensity and polarisation in conditioning expressive risk—have not been systematically integrated into the analysis of silencing processes. These limitations are analytically consequential: without attending to the ways algorithmic systems actively produce opinion climates, and without accounting for the role of affect in amplifying expressive costs, accounts of digital silencing remain partial.

It is from these limitations that the present investigation departs. By advancing a critical reassessment of how political expression and silence are produced in algorithmically mediated publics, this article argues that silencing dynamics cannot be adequately understood without attending to the infrastructural and affective conditions through which opinion climates are actively produced within platform-mediated environments. Rather than testing causal relationships, it develops a theoretical argument supported by illustrative empirical material, examining how

AI-driven visibility mechanisms and affective interactional dynamics interact to shape the conditions under which opinions become legitimate or risky to express. To this end, the present investigation draws on a case-based, interpretive analysis of highly visible comments responding to posts by the Portuguese populist leader André Ventura across Facebook, TikTok, and X—mobilised in an illustrative capacity as an analytical site through which broader conceptual dynamics can be examined, rather than as a basis for empirical generalisation.

Building on this approach, the study makes three main contributions. First, it proposes a reconceptualisation of the Spiral of Silence as an affective–algorithmic process, in which expressive risk emerges from the interaction between algorithmic visibility cues and emotionally charged communication environments. Second, it advances a theoretical re-specification of opinion climate formation, arguing that perceptions of dominance are increasingly shaped by algorithmically structured signals of visibility rather than by interpersonal observation. Third, it contributes to ongoing debates on digital political communication by highlighting how platform architectures and affective dynamics jointly shape the conditions under which public expression becomes possible in contemporary mediated environments.

Conceptual Framework

The section outlines the conceptual model advanced in this article—an affective–algorithmic reconceptualisation of the Spiral of Silence—and specifies its three constituent pillars. References to prior scholarship are mobilised to situate the model's contribution, rather than to provide an extensive review of the field. The proposed framework conceptualises the Spiral of Silence as an affective–algorithmic process, analytically structured around three interrelated dimensions: algorithmic visibility cues, affective dynamics, and interactional hostility. As summarised in Table 1, these dimensions form a self-reinforcing dynamic in which platform architectures, emotional economies, and interactional sanctions jointly produce conditions of expressive risk and silence within platform-mediated populist publics. The discussion proceeds by developing the theoretical grounding of each pillar before examining their articulation as an integrated model and clarifying how this formulation extends existing accounts.

Algorithmically mediated visibility emerges as a central mechanism through which contemporary opinion climates are produced. AI-driven curation systems actively structure which positions attain prominence within platform environments; in doing so, they generate visibility-based proxies for social legitimacy that increasingly substitute for the interpersonal observation of majority opinion central to classical Spiral of Silence theory (Napoli, 2019). Positions that accumulate engagement cues—such as likes, replies, and algorithmic ranking—come to appear normative within a given comment space, irrespective of their actual numerical prevalence, while dissent is rendered structurally marginal. By prioritising engagement and retention over deliberative relevance (Tufekci, 2015, 2017; Gillespie, 2018), these systems shape interactional prominence and peripheralisation alike. The model's contribution here lies in specifying this infrastructural mechanism as an active co-producer of opinion

Table 1 An affective–algorithmic model of the spiral of silence: theoretical pillars, description, and operationalisation in comment spaces

Pillar	Description	Operationalisation in Comment Spaces
(a) Algorithmic visibility cues	AI-driven ranking and filtering systems (e.g., engagement metrics, recommender algorithms) produce cues of popularity, dominance, and legitimacy. These cues substitute for interpersonal observation of majority opinion.	Highly visible comments (top-ranked by likes, replies, or algorithmic prominence) signal "what is acceptable". Dissenting views that receive low engagement become effectively invisible.
(b) Affective dynamics	Emotional intensity—especially moral indignation, outrage, and in-group solidarity—is privileged by both populist communication styles and engagement-based algorithms. Emotionally charged content circulates more widely and normalises polarised stances.	Comments expressing strong alignment or hostility toward out-groups accumulate visibility, while neutral or ambivalent positions are marginalised. Affective reinforcement consolidates in-group cohesion.
(c) Interactional hostility	Discursive hostility (ridicule, ad hominem attacks, intimidation, coordinated dismissal) raises the expressive costs of dissent. Anticipating these sanctions discourages users from voicing minority views, even in formally open spaces.	Dissenting comments are met with rapid dismissal, mockery, or moral disqualification. On some platforms (e.g., X), hostility is horizontal (user-to-user); on others (e.g., Facebook, TikTok), it is vertical (directed at out-groups). Both forms suppress sustained deliberation.

climates—rather than a contextual backdrop—thereby extending prior accounts focused on individual perception (Neubaum & Krämer, 2018) or informational filtering (Pariser, 2011; Sunstein, 2017) toward the platform-level architectures that generate such distortions as a structural condition.

This infrastructural dynamic is closely intertwined with the affective logics that govern visibility in contemporary digital communication. Emotional intensity is not incidental to algorithmic prominence but structurally privileged by it, particularly within the context of populist communication. Populism, understood as a discursive logic organised around the antagonism between "the people" and "the elite" (Mudde, 2004), mobilises communication styles marked by moral indignation, authenticity performance, and emotional saturation (Moffitt, 2016). These styles align closely with platform logics that reward emotionally charged and polarising content (Brady et al., 2017; Tufekci, 2017), producing a recursive dynamic in which affective expression drives visibility, visibility reinforces perceived consensus, and perceived consensus discourages dissent. The analytical contribution of this dimension lies in specifying this recursive structure as a systemic feature of the platform–populism interface rather than a contingent outcome of individual communicative behaviour.

Translation of perceived dominance into expressive constraint is further mediated through interactional dynamics of hostility. Discursive hostility, ranging from ridicule and *ad hominem* attacks to coordinated dismissal and moral disqualification, operates not only as a reactive sanction against dissent—as documented in research on hostile communicative environments (Benesch, 2014)—but, crucially, as an anticipatory deterrent that precedes expression. This anticipatory function is the model's central contribution at this pillar. What is at stake is not simply the presence of hostility, but its visibility within algorithmically organised comment spaces. The display of hostile responses renders the social cost of disagreement legible to other users before they themselves choose whether to speak, thereby shaping expressive decisions in

advance. This anticipatory dimension marks a significant extension of classical Spiral of Silence theory: whereas Noelle-Neumann's formulation emphasised gradual withdrawal under perceived minority status, the present model highlights more immediate, localised, and temporally compressed dynamics in which silencing is embedded within the interactional environment itself.

Taken together, these dimensions operate not as isolated mechanisms but as an integrated and mutually reinforcing system. Algorithmic visibility structures perceived opinion climates; affective dynamics amplify emotionally consonant positions; and interactional hostility translates perceived dominance into concrete expressive costs. The result is a self-reinforcing loop in which visibility, affect, and sanction co-produce conditions of expressive risk and silence. What distinguishes this model from prior accounts is not the identification of these elements individually, but the specification of their joint operation as a structurally embedded dynamic within platform-mediated populist communication.

Data and Methods

Analytical Framework and Interpretive Guiding Questions

This article adopts a conceptual and interpretive analytical framework in place of a conventional empirical methodology. Its purpose lies in advancing theoretical understanding rather than in testing formal hypotheses or establishing causal claims, and it does not aim to produce probabilistic generalisations. Instead, the analysis is guided by a set of theoretically informed analytical orientations derived from Spiral of Silence theory, in conjunction with contemporary scholarship on algorithmic governance and the affective dimensions of polarisation and digitally mediated hostility. These frameworks operate as sensitising concepts that structure interpretation and comparison, without predetermining outcomes or imposing deductive explanatory claims on the material under analysis.

Spiral of Silence theory (Noelle-Neumann, 1974, 1984) constitutes the central conceptual reference for examining how perceptions of opinion dominance and marginality shape political expression, particularly in relation to perceived social risk. In this article, the theory is approached as a heuristic and diagnostic framework for re-examining contemporary silencing dynamics under conditions of algorithmic curation and AI-driven visibility, shaped by processes of emotional amplification and accelerated interactional rhythms, rather than as a predictive or falsifiable model subjected to empirical testing. Consistent with this interpretive orientation, quantitative indicators—such as engagement metrics and visibility signals—are treated solely as contextual cues that help delineate the interactional environments in which communicative exchanges unfold. Their role is limited to situating these environments, without grounding claims about consensus, homogenisation, or silencing at the population level. The analysis instead centres on recurring discursive patterns and interactional sequences, attending to contrastive cases and to the qualitative texture of exchanges through which perceptions of dominance, marginality, legitimacy, and expressive risk are produced and negotiated.

The empirical component of this study should be understood as illustrative—it does not aim to generate theory inductively from emergent patterns in the data, nor to establish probabilistic claims about platform behaviour, but to exemplify, through systematic engagement with a bounded corpus, a theoretical model developed independently of the empirical material. The relatively elaborate empirical apparatus—including a substantial corpus, a structured coding grid, and intercoder reliability assessment—serves this illustrative function precisely: it ensures that the examples mobilised throughout the analysis emerge from a systematic and intersubjectively consistent engagement with the material, thereby reinforcing the credibility of the illustrations without repositioning them as a basis for inference.

Three interrelated guiding questions structure the interpretive analysis:

> RQ1: How do algorithmically mediated visibility cues and platform-specific interactional norms shape the formation of localised opinion climates within highly visible comment spaces?
> RQ2: How do users orient their expressive behaviour—including conformity, dissent, and anticipatory self-censorship—in relation to positions perceived as dominant or marginal within these opinion climates?
> RQ3: How do affective dynamics and discursive hostility contribute to the production of expressive costs and the emergence of silencing processes across different platform environments?

Case-Based Qualitative Design and Data Collection

The analysis proposed in this study focuses on comments posted in response to publications by André Ventura, leader of the populist right in Portugal, on the platforms X, Facebook, and TikTok. The aim is to understand whether, how, and under what conditions users express dissenting opinions in digitally mediated environments shaped by emotionally charged and strongly polarising discourse. The study seeks to produce sociotechnical interpretive knowledge on how platform architectures and emotional dynamics condition dissent.

Choice of André Ventura is justified by his centrality within the Portuguese political–media ecosystem, comparable in European and global terms due to the recurrent frames and thematic orientations he promotes—most notably moral indignation, security, and national identity (Di Fátima, 2023; Mudde, 2024; Papacharissi, 2021). Additional relevance derives from his role as a contemporary catalyst of affective polarisation and online conflictuality, in line with broader supranational dynamics. His publications thus constitute a critical case for observing mechanisms of silencing and opinion dynamics in populist contexts (Flyvbjerg, 2006).

Empirical delimitation focuses on a strategically defined period—from 1 January to 30 April 2025—characterised by dual political–communicational relevance. Beginning of the civil year typically coincides with agenda setting and discursive repositioning, alongside the strategic formulation of electoral priorities. Proximity to the legislative elections of 18 May 2025 further captures a phase of rhetorical intensification and affective mobilisation, marked by the instrumentalisation of polarisation typical of pre-electoral contexts. Such temporal framing does not assume a rigid

annual reset in Portuguese politics; instead, early months of the year are approached as a phase of discursive signalling and strategic repositioning, particularly in electoral cycles scheduled for the spring, as occurred in 2025. Sample construction comprises nine emblematic posts, selected according to a principle of contextual representativeness rather than statistical generalisation. Distribution across platforms ensures thematic diversity and communicational transversality, while also reflecting patterns of algorithmic representativeness.

Identification of posts followed a dual criterion: (i) algorithmic popularity, operationalised through systematic assessment of engagement metrics recorded at the time of data collection—specifically, likes, comments, shares or reposts, views, and saves, depending on platform affordances (see Table 2); and (ii) interactional density, understood as the capacity of a post to generate sufficiently rich comment threads to constitute observable opinion microclimates. Posts were identified through direct manual consultation of André Ventura's public profiles on each platform, conducted systematically across the collection period. Candidate posts were assessed through visual comparison with other publications by the same actor during the same period, with posts included when their engagement metrics—particularly comment volume and reaction counts—visibly and substantially exceeded the typical performance of non-selected content, without applying a fixed quantitative threshold.

No inferences are drawn about platform characteristics; the focus lies instead on examining how comparable forms of populist performance are negotiated within distinct interactional ecologies. Post selection therefore prioritised analytically relevant instances aligned with the objectives of the investigation, without requiring thematic equivalence across platforms. High visibility within each platform provides a basis for comparing how similar styles of populist communication are received and negotiated across differing interactional environments, in line with previous research (Gearhart & Zhang, 2015).

Selection of X, Facebook, and TikTok is justified by their distinct communicational ecologies: Facebook hosts extensive political interactions and broad publics; X operates as a real-time arena of political conflict, marked by contestation and

Table 2 Engagement metrics of selected posts at the time of data collection. Note. Engagement figures recorded at the time of data collection. sv. = saves; vi. = views. Facebook does not display share counts publicly. TikTok saves replace view counts as the primary retention metric. X saves reflect bookmarking behaviour. Dashes indicate metrics not publicly available on the respective platform

Platform	Date	Topic	Likes	Comments	Shares	Views/ Saves
TikTok	25/02/2025	Housing crisis /immigration	71,853	2,003	10,609	3,457 sv.
TikTok	19/04/2025	Authoritarian governance scenario	8,491	102	236	137 sv.
TikTok	29/04/2025	Muslim vs. Portuguese students	13,152	747	2,196	803 sv.
Facebook	23/01/2025	Housing policy/Roma communities	56,049	11,477	—	2,652 vi.
Facebook	23/03/2025	Muslim cultural practices/TV	33,634	3,804	—	1,276 vi.
Facebook	19/04/2025	Hypothetical governance agenda	17,732	1,166	—	752 vi.
X	21/04/2025	Pope Francis condolences	1,502	494	166	24 sv.
X	28/04/2025	National blackout/governance	2,582	154	285	42 sv.
X	30/04/2025	TV censorship/interview	3,322	1,091	307	39 sv.

counter-discourse; TikTok privileges emotional virality through recommendation algorithms. Combining the three platforms enables the observation of differentiated patterns of visibility, hostility, polarisation, and silencing. For each post, fifty comments with the highest interaction were collected, yielding a total of 450 comments. Comment visibility was operationalised through each platform's native default ranking at the time of data collection: on Facebook, comments were retrieved under the "Most relevant" sort order, which privileges reactions and replies; on TikTok, under the default sort order privileging total interactions; and on X, under "Top" ranking, which weights engagement and recency. Data collection was conducted manually through each platform's native interface, without the use of automated scraping or third-party tools, in order to faithfully capture the algorithmically ranked comment environment as experienced by ordinary users at the moment of collection. Data collection took place between 7 and 10 February 2026, retrospectively accessing publicly available comment threads from posts published during the delimitated period (January–April 2025). This procedure—acknowledged as limited yet appropriate to the study's aims—seeks to balance discursive representativeness and cross-platform comparability, while maintaining analytical feasibility, following Gearhart and Zhang's (2015) emphasis on highly visible comments and Bardin's (2011) guidelines for qualitative organisation.

Analytical procedures followed a deductive categorical grid grounded in the theoretical frameworks mobilised: strategies of polarisation (Mudde, 2004), the emotional economy of visibility (Tufekci, 2017), silencing dynamics in hostile climates (Neubaum & Krämer, 2018), and populist performances (Moffitt, 2016). From this foundation, five analytical dimensions were defined: (i) agreement or disagreement with the content of the post; (ii) degree of polarisation (unconditional support, aggressive rejection, moderation, irony); (iii) presence of negative emotional language (insults, sarcasm, aggressiveness, ad hominem attacks); (iv) expressions of silencing (delegitimation, intimidation, explicit attempts at censorship, comment deletion); and (v) forms of interaction among commenters (identity reinforcement, confrontation, exclusion, or discursive isolation). Within this framework, "irony" was coded as rejection when it functioned interactionally to delegitimise, ridicule, or undermine the target of the post or another commenter, whereas "moderation" designates comments expressing neither unconditional support nor aggressive rejection. The five dimensions operated as a coding grid later consolidated into four higher-order interpretive axes, including qualified agreement, ambivalent positioning, and non-hostile disagreement. No instances of affiliative or supportive irony were identified in the corpus.

Given that the five analytical dimensions were conceived as independent and non-mutually exclusive coding axes—each applied separately to the same comment—a single comment could legitimately receive codes on multiple dimensions simultaneously (e.g., coded as expressing both unconditional support and negative emotional language). Each dimension was therefore coded independently, without requiring categorical priority or conflict resolution between dimensions. The percentage values reported in Table 3 reflect the proportion of comments within each platform sample ($n = 150$) that received a positive code on each dimension independently; given the

Table 3 Distribution of comment types across analytical dimensions by platform (Facebook, TikTok, and X)

Analytical dimension	Facebook ($n=150$)	TikTok ($n=150$)	X ($n=150$)
Agreement with post	74%	82%	38%
Unconditional support	52%	61%	21%
Aggressive rejection	14%	9%	44%
Moderation/ambivalence	12%	8%	18%
Negative emotional language	46%	51%	58%
Expressions of silencing	24%	21%	31%

non-mutually exclusive nature of the coding axes, these figures should be read as dimension-specific rates rather than as components of a whole.

To enhance analytical consistency, an intercoder reliability assessment was conducted on a subset of the corpus independently coded by a second trained coder. Krippendorff's alpha coefficient indicated substantial agreement across the five analytical dimensions ($\alpha=0.78$), supporting a satisfactory level of reliability for interpretive analysis, consistent with the illustrative status of the empirical component outlined in Sect. 3.1. Discrepancies were discussed and resolved, leading to minor refinements of category definitions prior to full corpus coding. Conceptual consistency was further pursued through an iterative piloting process, in which the grid was repeatedly tested and refined against the corpus. Such an approach aligns with the interpretive and heuristic orientation of the study, oriented toward theoretically grounded qualitative interpretation rather than replicable measurement (Neuendorf, 2002; Krippendorff, 2004).

Illustrative Corpus and Platform-Specific Contexts

The three videos corresponding to the selected TikTok posts stage moralised crisis narratives structured around antagonistic dichotomies—immigration, housing, public security, and cultural belonging. The first video (19 April) presents a hypothetical scenario of authoritarian leadership, outlining radical policy measures in the domains of justice and immigration, alongside issues of security. The second video (29 April) contrasts "privileged Muslim students" with "disadvantaged Portuguese students" in the context of school meals. The third video (25 February) mobilises the housing crisis to oppose "legitimate Portuguese citizens" to immigrants portrayed as "illegal squatters."

Comments on these posts are characterised by high levels of alignment with the author and a recurrent use of evaluative and moral language, often accompanied by direct address to occasional opponents. Interactional forms include expressions of agreement, aggressive language, sarcasm, intimidation, and replies directed at dissenting comments. Comment threads tend to develop through short exchanges, with reinforcement among aligned users and episodic confrontations with critics. Manifestations of dissent appear in isolated and intermittent forms, without the emergence of sustained dialogical sequences capable of transcending rejection from commenters who share the dominant opinion.

The first Facebook post selected (23 January) constructs a moral opposition between "law-abiding Portuguese citizens" and "privileged Roma communities" in

the context of municipal housing policies (in Portugal, this community is typically designated as an internal enemy by the populist radical right—the "Them" in the antagonistic "Us versus Them" framework). The second post (23 March) consists of a video excerpt from a television programme appearance, in which cultural practices associated with Muslim-majority countries are criticised and a narrative of Western normative superiority is asserted. The third post (19 April) presents a hypothetical governance agenda centred on justice, immigration, corruption, and security—issues central to the Portuguese populist agenda, as well as to broader European and global populist discourses, and increasingly salient in the Portuguese context through frames of threat and danger. Comments reveal dense interactional activity, with extensive use of evaluative language, direct replies between users, and frequent addressing of groups perceived as external. Interactions include agreement, verbal aggression, sarcasm, confrontations between commenters, and attempts to marginalise dissenting positions. Some comment threads show clustered exchanges among aligned users, while others involve short-lived confrontations that do not develop into sustained dialogue.

In the case of X, the first post (21 April) consists of a condolence message following the death of Pope Francis, in which Ventura praises the Pope's "closeness and simplicity." The second post (30 April) accuses a television channel of censorship following the interruption of an interview. The third post (28 April) criticises the government's handling of a national blackout and questions public sector governance. The material examined on this platform exhibits a dynamic distinct from those observed on the other platforms analysed: high levels of direct address, extensive use of irony and verbal aggression. The dominant opinion oscillates between rejection and support—the post of 21 April, in fact, displays total rejection, with no supportive or even neutral comments. The other two posts (30 April and 28 April) show a divided set of opinions between supportive and rejecting positions. Interactions frequently take the form of short replies directed either at the post author or at other commenters, often marked by tones of intimidation and hostility, with limited accumulation of shared arguments across the debate. In some cases, aligned commenters reinforce one another; in others, interaction is dominated by confrontational or dismissive responses that prevent the development of sustained engagement.

Discursive Microclimates and Platform-Specific Regimes of Expression

A comparative reading of comment spaces is presented in an explicitly illustrative register. Patterns discussed below consist of theoretically informed interpretations drawn from a bounded heuristic corpus and are intended to elucidate underlying dynamics, without supporting generalisations about platform behaviour or user populations. Analytical attention centres on how specific affordances and affective norms generate differentiated interactional regimes within highly visible comment environments, with visibility appearing to be shaping these dynamics, examined here as heuristic sites through which the theoretical reconceptualisation advanced in this article can be rendered analytically tractable.

Table 3 summarises the distribution of coded comments across the five analytical dimensions by platform. These figures are presented as descriptive approximations intended to contextualise the qualitative patterns discussed below, rather than as a basis for inferential claims. Values represent percentages of comments ($n = 150$ per platform) coded according to the analytical dimensions defined in the study: agreement with post, degree of polarisation (unconditional support, aggressive rejection, moderation/ambivalence), presence of negative emotional language, and expressions of silencing. Categories are not mutually exclusive; therefore, percentages do not sum to 100 within each column.

In response to RQ1—focused on how algorithmically mediated visibility cues shape the emergence of localised opinion climates—the material examined across Facebook and TikTok suggests that comments aligned with the post's framing are disproportionately represented among the most visible positions: among the ten most visible comments per post on these two platforms (60 comments in total), 78% expressed alignment with the post's framing, while dissenting comments accounted for fewer than 12% of equivalently visible positions. Qualitative inspection of this subset indicates that these highly visible, aligned comments are also predominantly characterised by emotionally charged and morally polarising language, consistent with the substantial proportion of comments coded as containing negative emotional language across these platforms (Facebook: 46%; TikTok: 51%; see Table 3), functioning as reference points for subsequent interaction and reinforcing perceptions of consensus among participants. Comment threads frequently open with statements of agreement that reiterate the post's moral framing and quickly establish consonant positions as normative, followed by brief replies that sustain the same evaluative orientation. This opening dynamic is observable across several threads: in the Facebook post of 23 January, where municipal housing policy is framed through a moral opposition between 'true Portuguese' citizens and Roma communities receiving state support, the ensuing responses display immediate and dense alignment—comments such as "This is already long overdue", "This is the future of Portugal… congratulations to mum and dad", and "Cleaning Portugal is what we want" reproduce and intensify the post's moral vocabulary, consolidating consonance as the dominant interactional register, while contributions expressing dissent remain isolated and fail to generate sustained dialogue; in the post of 23 March, comments such as "This is exactly what's happening and nobody talks about it" and "Finally someone says what everyone thinks"—both accumulating over 200 reactions—establish the same dynamic, with responses questioning the post's framing appearing only further down the thread with minimal engagement. Expressions of dissent appear sporadically and tend to surface at later stages, rarely developing into sustained exchanges—across the Facebook posts analysed, dissenting comments represented approximately 26% of the coded corpus, and fewer than one in five generated any reply. In this context, when disagreement does emerge, it is commonly met with moral disqualification or direct confrontation, after which interaction dissipates. Contributions expressing dissent remain isolated and fail to generate sustained dialogue, while visibility concentrates around this prevailing evaluative frame. Such patterns are consistent with the formation of localised opinion microclimates within the material analysed here.

Interactional dynamics further shape these patterns, particularly within the TikTok material examined, where speed and performativity appear to amplify their effects within fragmented interactional environments. In the post of 25 February, the dramatization of deportation—performed through the throwing of a plastic aeroplane and accompanied by the statement "Are you seeing this? This is what will happen to all illegal immigrants. Not 15, not 16, not 4,000. All of them"—elicits responses that replicate and escalate the video's affective register. Comments such as "I convinced all my family to vote for you. Please don't let me down. Portugal needs to change" convey emotional investment and personal commitment, while others intensify antagonistic framing—"They can also be sent back to their origins by boat, raft, or rocket"—by directing hostility explicitly towards the out-group invoked in the post. Interaction is dominated by performative responses that mirror the tone and rhythm of the video, while extended dialogical sequences are rare. A representative exchange from the TikTok post of 25 February illustrates this: a comment expressing strong alignment—"There has to be someone with eyes in their head"—attracted 47 replies, none of which engaged with its substance; instead, replies functioned as successive endorsements ("exactly", "couldn't agree more", "share this") or as hostile deflections directed outward at the post's designated out-group. When dissenting remarks do appear, they tend to receive little or no engagement, as in isolated comments such as "Not everyone thinks like that" (TikTok, 19 April), which accumulated no replies and minimal engagement, remaining effectively invisible within a thread dominated by affectively consonant responses. The absence of interactional uptake is analytically significant here: it is not that dissent is actively suppressed, but that the microclimate effectively leaves it structurally inert within the interactional environment observed.

A comparable analysis of opening dynamics on X reveals a markedly different pattern from that observed on TikTok. Where TikTok threads tend to consolidate rapidly around a single affectively consonant register—replies functioning as successive endorsements of the post's tone rather than independent contributions—X threads more frequently open with the immediate juxtaposition of opposed positions. This is observable in the post of 30 April, where alignment ("This kind of militant journa-trash is numbered") and explicit rejection ("Ignorance at its finest…") appear in close succession from the outset, with neither position establishing itself as dominant before the other intervenes. A contrasting case is evident in the post of 21 April, where the opening responses display total rejection, with no aligned or moderating comments appearing at any point in the thread. This variability in how X threads open—sometimes immediately contested, sometimes immediately uniform, but rarely consolidating into a stable evaluative register either way—is consistent with the absence of sustained community reinforcement noted elsewhere in this analysis, and offers a preliminary account of why silencing on X takes the form of interactional fatigue rather than the rapid normalisation of a dominant frame observed on Facebook and TikTok. The post of 30 April, examined above in terms of its opening dynamics, also offers a useful entry point into how users orient their expressive behaviour once these climates are established (RQ2).

Having examined how algorithmically mediated visibility structures the emergence of localised opinion climates (RQ1), the analysis now turns to how users

orient their expressive behaviour within those climates (RQ2), concerning the ways in which users orient their expressive behaviour in relation to positions perceived as dominant or marginal, including patterns of conformity and dissent, as well as anticipatory self-censorship. In keeping with the illustrative and non-experimental design of this study, no causal relationships are advanced; the patterns discussed are best understood as theoretically informed associations. Anticipatory self-censorship, central to Spiral of Silence theory and to the model advanced in this article, poses a particular analytical challenge: by definition, it concerns expression that does not occur, and is therefore not directly observable within a corpus of published comments (Neubaum & Krämer, 2018). Its presence can only be inferred indirectly, through proxies within the material that is available—namely, the severity and speed with which visible dissent is sanctioned, the extent to which isolated dissenting comments fail to generate further engagement, and explicit textual markers of the perceived cost of disagreement. Read together, these proxies suggest that anticipatory self-censorship operates differently across the three platforms in ways that track the perceived stability of dominant positions. On Facebook and TikTok, where opinion climates consolidate rapidly and visibly around a dominant evaluative register (see RQ1, above), the cost of dissent is signalled clearly and early: comments such as "These are also Portuguese citizens who need support" (Facebook, 23 January) are met with immediate moral disqualification, while isolated dissenting comments on TikTok, such as "Not everyone thinks like that" (19 April), generate no engagement at all. The comment "I convinced all my family to vote for you. Please don't let me down" (TikTok, 25 February) is suggestive in this respect: the personal stakes it expresses imply an environment in which the anticipated cost of voicing doubt would be substantial. On X, by contrast, the absence of a stable dominant position appears to weaken this anticipatory mechanism: because no single evaluative register consolidates clearly (see RQ1, above), users may have less basis for anticipating which position carries social risk. This does not indicate an absence of expressive cost on X, but rather its displacement onto a different mechanism—interactional fatigue—discussed further below.

Within the X material examined, a higher degree of discursive plurality is observable relative to the other two platforms in this corpus, with posts attracting both supportive and critical responses and dissent expressed with relative visibility. Here, the post of 30 April—in which André Ventura accuses a television channel of censorship following the interruption of an interview, referring to it as "jornalixo" ("journa-trash")—shows alignment ("This kind of militant journa-trash is numbered") alongside explicit rejection ("Ignorance at its finest… Leftist blindness in wanting to attack other ideas leads to this nonsense being said"). Such plurality, however, seldom develops into deliberative exchange. Highly visible comments illustrate an interactional pattern structured around antagonism and reciprocal dismissal, where positions are asserted and opposed without engagement. This tendency is further evident in reply chains in which users address one another through brief, dismissive interventions—such as ""Go study before commenting", observed in the post of 21 April—that function less as contributions to argumentation than as mechanisms of closure.

Considering this environment, dissenting users appear to orient their expression differently from those on Facebook or TikTok. Withdrawal is less evident; instead,

engagement often assumes a confrontational form. Yet such participation remains volatile and socially costly, rarely coalescing into sustained counter-discourses. This pattern is observable in the X post of 30 April, where a dissenting user responded to three separate aligned commenters in succession without generating any substantive reply; each intervention was met with a single dismissive response or left unanswered, illustrating how confrontational engagement on X tends to produce isolated skirmishes rather than accumulated counter-argument. In the X comments analysed, the Spiral of Silence appears to take on a distinct configuration: not the progressive withdrawal associated with low visibility, as observed on Facebook and TikTok, but a form of disengagement shaped by interactional fatigue, arising from instability and the absence of shared interpretative frameworks capable of sustaining disagreement. This dynamic is directly observable in the X corpus: in the post of 21 April, where Ventura's condolence message following the death of Pope Francis was met with total rejection, dissenting-from-rejection voices—that is, users defending Ventura's gesture—appeared in only three comments, each of which was met with dismissive one-line responses and generated no further exchange. The absence of accumulation is analytically significant: it suggests that on X, silencing operates through the inability of any position—including aligned ones—to consolidate into durable interactional presence.

By contrast, on Facebook and TikTok, expressive orientations tend more clearly toward alignment. As noted above in relation to anticipatory self-censorship, this dynamic is directly observable in the Facebook post of 23 January, where a comment expressing qualified disagreement with the post's framing was met with immediate moral disqualification and generated no further engagement of its own, with subsequent interaction bypassing it entirely to consolidate the dominant evaluative register. Such sequences suggest that the cost of dissent becomes visible to other users before they themselves choose whether to express disagreement—a structurally anticipatory form of silencing.

Beyond the structural formation of opinion climates and individual expressive orientations, a third dimension concerns the affective mechanisms through which expressive costs are produced and sustained (RQ3). In relation to this question—addressing how affective dynamics and discursive hostility, as observable in comment threads responding to highly visible posts, produce differentiated patterns of polarisation and discursive hostility, alongside processes of intra-group reinforcement across platforms—affective polarisation appears across all three platforms in the analysed material as a recurrent structuring feature, though with platform-specific configurations. On Facebook and TikTok, within the selected posts, hostility is predominantly vertical, directed at out-groups—such as immigrants or political opponents—and functioning to reinforce moralised "us versus them" boundaries that consolidate internal cohesion among aligned commenters. This consolidation dynamic is visible in the Facebook post of 23 January, where, following a comment directing hostility outward—"They keep coming here and taking advantage of everything without contributing"—a cluster of replies among aligned users develops not around the post's content but around mutual endorsement: "Exactly, finally someone says it as it is", "People are waking up, we can't stay silent anymore". The thread functions

less as a debate than as a ritual of in-group affirmation. Additional comments such as "They should go back to where they came from", on TikTok post of 29 April, further illustrate how hostility is directed outward, reinforcing group boundaries rather than engaging opposing viewpoints. The comment "Cleaning up Portugal is what we want" on Facebook and "They can also be sent back to their origins by boat, raft, or rocket" on TikTok exemplify this vertical hostility: both direct aggression outward toward the designated out-group, while simultaneously functioning as signals of in-group solidarity and alignment with the populist communicator.

Within the X material examined, hostility appears primarily horizontal, distributed among users through direct confrontation and mutual attacks. This horizontal dynamic is directly observable in the post of 30 April, where aligned and dissenting users address one another in rapid succession without engaging the substance of opposing positions: a supporter's "This kind of militant journa-trash is numbered" draws the response "Ignorance at its finest… Leftist blindness in wanting to attack other ideas leads to this nonsense being said", which in turn produces dismissive closure rather than argumentative development. Neither position accumulates sustained interactional support; both remain isolated assertions in an environment structured by mutual delegitimisation rather than deliberative exchange.

A contrasting pattern within the same platform is observable in the post of 28 April, where criticism of the government's handling of the national blackout attracted both aligned and dissenting responses without either camp consolidating into a coherent bloc. Unlike the Facebook post of 23 January—where aligned commenters developed a recognisable ritual of mutual endorsement—the X thread produced no equivalent structure: users expressing agreement with Ventura's framing and users rejecting it remained equally isolated, each position circulating without accumulating interactional support. This absence of community reinforcement in the X material examined further distinguishes the silencing dynamics observed there from those identified on Facebook and TikTok, where in-group cohesion appears to function as a mechanism of expressive normalisation.

Across the three platforms, comments coded as containing negative emotional language represented 52% of the total corpus (Facebook: 46%; TikTok: 51%; X: 58%), consistently outpacing neutral or ambivalent expressions and reinforcing cycles of indignation and antagonism while simultaneously constraining deliberative engagement. This dynamic is theoretically consistent with the affinity of populist communication styles with the algorithmic logics of visibility that privilege emotional intensity and moral polarisation (Moffitt, 2016; Tufekci, 2017; Gerbaudo, 2018). This configuration—confrontational engagement without deliberative accumulation—can be understood here as a platform-specific expression of the anticipatory self-censorship dynamic central to the affective–algorithmic reconceptualisation proposed in this article: the social cost of sustained dissent within the X material examined appears to operate not through withdrawal but through exhaustion, producing a functionally equivalent silencing effect.

These dynamics further illustrate how silencing processes operate through visibility asymmetries and anticipatory self-censorship rather than through explicit forms of repression. Analysis across all three platforms consistently points to mechanisms

that are structural and affective rather than overt. As noted above in relation to anticipatory self-censorship, the TikTok comment expressing strong personal investment in Ventura's electoral success is analytically significant not only as an expression of alignment but as an indicator of the affective climate within which it was produced: it suggests an interactional environment in which the expression of doubt or opposition would require navigating the emotional intensity of the dominant affective register. In the comment spaces analysed, silencing appears to operate less through direct censorship than through the progressive marginalisation of dissenting voices: on Facebook and TikTok, through rapid discouragement as dissenting comments are met with moral disqualification or coordinated dismissal that effectively removes them from visible interactional circulation; on X, through the social and reputational costs of sustained engagement in a hostile and fragmented environment. Visibility asymmetries appear to play an important role in this process: within algorithmically organised comment spaces, positions that align with the dominant affective framing accumulate engagement and visibility, while dissent remains peripheral and subject to symbolic sanctions that discourage further expression.

This pattern recurs across the corpus: as noted above in relation to RQ1, visibility among the most prominent comments on Facebook and TikTok concentrates overwhelmingly around aligned positions, with dissent occupying a marginal share of equivalently visible comments. Such marginality is directly observable in the corpus: in the TikTok post of 29 April, the comment "Not all immigrants are like that—generalising is wrong" accumulated two reactions and no replies, while the immediately preceding comment expressing alignment with the post's framing accumulated over 340 likes and generated a thread of 17 replies. The asymmetry in interactional uptake—not the content of the comments themselves—determines which positions become reference points within the thread and which remain peripheral, providing an empirical illustration of the visibility foregrounded in the affective–algorithmic reconceptualisation proposed here. Anticipatory self-censorship—a central concept in Spiral of Silence theory—thus appears to operate not only in response to perceived majority opinion, but also in response to the algorithmically structured cues of dominance and legitimacy that make certain positions appear naturally hegemonic within a given comment space.

Taken together, these observations suggest that the three guiding questions are analytically interconnected: the emergence of localised opinion climates (RQ1) shapes the conditions under which users orient their expressive behaviour (RQ2); affective dynamics and populist communication styles (RQ3) contribute to the reproduction and intensification of those climates through affective and interactional mechanisms. The contribution of this illustrative analysis lies in rendering visible—through a theoretically informed interpretation of a bounded corpus—how the affective–algorithmic reconceptualisation of the Spiral of Silence proposed in this article can be understood as an integrated and mutually reinforcing set of dynamics within platform-mediated populist communication, without establishing causal chains between these processes.

Discussion: The Affective–Algorithmic Spiral and Its Theoretical Implications

Patterns identified across the three platforms, as observed within the bounded corpus analysed here, support a coherent theoretical argument that remains analytically grounded in the empirical material examined, while also suggesting possible relevance beyond the specific cases analysed. The patterns observed in this study suggest that silencing in algorithmically mediated populist publics may operate as a constellation of mutually reinforcing dynamics: visibility-driven normalisation of consonant positions and the affective discouragement of dissent, with the anticipation of symbolic sanctions further inscribing expressive risk within the interactional environment itself. What appears to distinguish this configuration from the classical Spiral of Silence is not the absence of social pressure but its infrastructural mediation—the fact that the cues signalling majority opinion and legitimacy are increasingly produced by algorithmic systems, with expressive risk inferred from these mediated signals rather than from direct interpersonal observation.

A productive dialogue with existing scholarship emerges through this reconceptualisation, which both aligns with and extends prior contributions. Illustrative patterns suggesting that opinion climates are constituted through algorithmically structured visibility cues rather than interpersonal perception resonate with Hampton et al. (2014) identification of self-censorship on social media and with Neubaum and Krämer's (2018) notion of spectral majorities, understood as distorted perceptions of consensus generated through selective amplification. Greater emphasis is placed here on platform architecture as an active co-producer of opinion climates, moving beyond approaches that treat platforms primarily as contextual variables shaping individual willingness to speak. A similar extension applies to Moffitt's (2016) framework of populist performance, which is here situated within platform-specific sociotechnical conditions, highlighting how the affinity between populism and algorithmic visibility contributes to the consolidation of discursive dominance in digital environments.

From this perspective, the material analytically suggests a threefold transformation of the Spiral of Silence under algorithmic conditions. One dimension concerns the source of opinion climate perception, which shifts from interpersonal observation to algorithmic curation, with engagement metrics and visibility rankings functioning as proxies for majority opinion (Splichal, 2022). Another relates to the trigger of self-censorship, no longer grounded in cumulative social pressure but increasingly associated with rapid affective feedback—or its anticipation—within compressed interactional temporalities. A further dimension involves the structural context of silencing, marked by a transition from a relatively unified public sphere to a mosaic of platform-specific microclimates in which silencing operates locally and situationally, in forms shaped by architecture and interactional culture.

Normative implications for democratic publicness emerge from these dynamics. As a theoretical implication of the affective–algorithmic reconceptualisation advanced in this article—and independently of the specific corpus examined—AI-mediated visibility regimes may function as proxies for legitimacy and consensus—a dynamic that, if confirmed at broader scale, would suggest the relevance of platform governance questions concerning how ranking systems may differentially condition

the visibility of certain forms of political expression. If the dynamics theorised here are structurally embedded in platform architectures that reward affective intensity and privilege antagonistic engagement, this may indicate a tendency to convert disagreement into expressive cost—a theoretical inference that warrants empirical investigation at larger scale. From a normative standpoint, and as a theoretical inference rather than an empirical prescription, these dynamics tentatively suggest that sustaining genuine pluralism in algorithmically mediated publics may require sociotechnical and institutional conditions—including transparency in ranking mechanisms and the cultivation of algorithmic literacy—capable of addressing the tendencies towards conformity and expressive marginalisation theorised here, pending confirmation through broader empirical inquiry.

Conclusion

This study develops a theoretical reconceptualisation of the Spiral of Silence as an affective–algorithmic process, drawing on an illustrative, case-based analysis of 450 highly visible comments across Facebook, TikTok, and X, collected in response to posts by the Portuguese populist leader André Ventura between January and April 2025. Using a qualitative and interpretive approach guided by a deductive categorical grid, the analysis examined how platform-specific visibility mechanisms and interactional dynamics shape the conditions under which opinions become visible, legitimate, or risky to express.

Building on these observations, the study offers three contributions to communication theory. First, it reconceptualises the Spiral of Silence as an affective–algorithmic process and advances a theoretical re-specification of opinion climate formation, arguing that perceptions of dominance are increasingly shaped by algorithmically structured visibility cues rather than by interpersonal observation. Second, it provides an analytically grounded account of how platform-specific interactional environments shape expressive patterns and silencing dynamics. Third, it illustrates how these dynamics vary across platforms—with the Facebook and TikTok material examined here pointing to evaluatively hegemonic microclimates and the X material suggesting a distinct form of silencing through interactional exhaustion—thereby showing that the affective–algorithmic spiral is not uniform but architecturally differentiated.

Several limitations should be acknowledged. The empirical corpus is deliberately limited and illustrative, consisting of nine posts and 450 comments drawn from a single political actor within a specific national context and time frame, which constrains both comparative scope and generalisability. The focus on highly visible comments systematically excludes low-visibility or non-engaged forms of dissent, including expressions that do not accumulate interactional prominence or remain outside algorithmically ranked visibility. Opacity in platform algorithms further constrains direct observation of ranking mechanisms, which must be inferred from interactional patterns instead of being empirically measured. In addition, the interpretive orientation of the analysis positions the findings as theoretically informed insights, not as causal explanations.

Future research could address these limitations by extending the empirical scope and integrating cross-national comparison with computational approaches to visibility and engagement dynamics. Longitudinal approaches could illuminate how opinion microclimates evolve over time, while multi-actor designs would allow for assessment of generalisability beyond the specific case examined here. These avenues would allow for a more systematic assessment of the dynamics theorised here, contributing to a cumulative research programme on algorithmic visibility and political expression in digitally mediated environments.

Funding Open access funding provided by FCT|FCCN (b-on).

Declaration

Competing Interests The author has no competing interests to declare that are relevant to the content of this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Bardin, L. (2011). *Content analysis*. Edições 70.

Benesch, S. (2014). *Countering dangerous speech: New ideas for genocide prevention*. United States Holocaust Memorial Museum.

Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences, 114*(28), 7313–7318. https://doi.org/10.1073/pnas.1618923114

Di Fátima, B. (2023). *Hate Speech on Social Media: A Global Approach*. LabCom Books.

Flyvbjerg, B. (2006). Five misunderstandings about case-study research. *Qualitative Inquiry*, *12*(2), 219–245. https://doi.org/10.1177/1077800405284363

Gearhart, S., & Zhang, W. (2015). Was it something I said? Shifts in political self-expression on social networking sites. *Computers in Human Behavior*, *53*, 36–44.

Gerbaudo, P. (2018). *The digital party: Political organisation and online democracy*. Pluto Press. https://doi.org/10.2307/j.ctv86dg2g

Gillespie, T. (2018). *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press. https://doi.org/10.12987/9780300235029

Hampton, K., et al. (2014). *Social media and the spiral of silence*. Pew Research Center.

Kim, M. (2016). Facebook's spiral of silence and participation: The role of political expression on Facebook and partisan strength in political participation. *Cyberpsychology Behavior and Social Networking*, *19*(12), 696–702. https://doi.org/10.1089/cyber.2016.0137

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology* (2nd ed.). SAGE.

Moffitt, B. (2016). *The global rise of populism: Performance, political style, and representation*. Stanford University Press.

Mudde, C. (2004). The populist zeitgeist. *Government and Opposition*, *39*(4), 541–563.

Mudde, C. (2024). From the margins to the mainstream: A personal reflection on three decades of studying and teaching far-right politics. *UC Berkeley*. https://doi.org/10.5070/RW3.25030

Napoli, P. (2019). *Social media and the public interest: Media regulation in the disinformation age*. Columbia University Press. https://doi.org/10.7312/napo18454

Neubaum, G., & Krämer, N. C. (2018). What do we fear? Expected sanctions for expressing minority opinions. *Communication Research*, *44*(3), 341–364. https://doi.org/10.1177/0093650215623837

Neuendorf, K. A. (2002). *The content analysis guidebook*. SAGE.

Noelle-Neumann, E. (1974). The spiral of silence. *Journal of Communication*, *24*(2), 43–51. https://doi.org/10.1111/j.1460-2466.1974.tb00367.x

Noelle-Neumann, E. (1984). *The spiral of silence: Public opinion—our social skin*. University of Chicago Press.

Papacharissi, Z. (2021). *After democracy: Imagining our political future.* Yale University Press.

Pariser, E. (2011). *The filter bubble*. Penguin.

Splichal, S. (2022). *Datafication of public opinion and the public sphere: How extraction replaced expression of opinion*. Anthem Press. https://doi.org/10.51480/1899-5101.16.2(34).667

Sunstein, C. R. (2017). *#Republic: Divided democracy in the age of social media*. Princeton University Press.

Tufekci, Z. (2015). Algorithmic harms beyond Facebook. *Colorado Technology Law Journal*, *13*(203), 203–218.

Tufekci, Z. (2017). *Twitter and tear gas: The power and fragility of networked protest*. Yale University Press.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.