

Publishing, sharing, and spreading online news

A case study of gatekeeping logics in the platform era

Michael Karlsson,^I Elizabeth Van Couvering,^I & Johan Lindell^{II}

^IDepartment of Geography, Media and Communication, Karlstad University, Sweden

^{II}Department of Informatics and Media, Uppsala University, Sweden

Abstract

News dissemination online is driven by three gatekeeping logics: the gatekeeping logic of the news media (publishing), the gatekeeping logic of social actors (sharing), and the gatekeeping logic of platform algorithms (spreading), each guided by different values and with a different relationship to content. Using a reverse engineering approach, this study applies a 2015 dataset to empirically explore how a Facebook algorithm changed the overall composition of the news users saw, highlighting the ongoing issue of how the different gates and associated gatekeeping logics – especially that of platforms – influence news distribution. In contrast to previous studies, we find the relationship between news properties and the distribution of news online to be non-linear. Results point to Facebook’s role in the overall composition of online news, both directly and in interaction with other gatekeepers. As news stories become more widely spread online, algorithmic logics take precedence over user behaviour and preferences.

Keywords: content analysis, gatekeeping, online news dissemination, algorithm, reverse engineering

Introduction: The role of news media, audience, and platforms in the online distribution of news

Social media in general, and Facebook in particular, have become major venues for news consumption (Newman et al., 2020). The ascendance of platforms also means that there is a complex and changing dynamic between news media, platforms, and audiences, where news distribution is increasingly dependent on and filtered by a platform’s algorithms (Meese & Hurcombe, 2021; Plantin et al., 2018; Wang, 2020).

Karlsson, M., Van Couvering, E., & Lindell, J. (2022). Publishing, sharing, and spreading online news: A case study of gatekeeping logics in the platform era. *Nordicom Review*, 43(2), 190–213. <https://doi.org/10.2478/nor-2022-0012>

The power relations between the different actors are asymmetrical: News organisations must adjust to the platforms rather than vice versa (Ihlebeak & Sundet, 2021; Meese & Hurcombe, 2021; Kleis Neilsen & Ganter, 2018). Platform involvement in news distribution is an issue that extends far beyond the immediate interest of media business (Helberger, 2019), since it impacts the news content that citizens need in order to make informed decisions. For instance, the news media could be publishing a complete and well-composed depiction of social and political life, while individual news stories that are distributed by platforms could paint a very different picture, or vice versa. This may have consequences for citizens, hindering their ability to be informed, and also for journalism, reducing its capability to fulfil its public role and tarnishing its image and legitimacy (Williams & Delli Carpini, 2011).

Our understanding of news in this article is a pragmatic one: It is the non-advertising information that news organisations select, process, and make available for consumers (see Bennett, 2009). This includes fact-based material as well as opinion pieces. Previous studies on how news spreads on social media have focused on content, specifically on which characteristics of news articles cause them to go viral – that is, be widely spread (e.g., Berger & Milkman, 2012; Khuntia et al., 2016; Trilling et al., 2017). Implicit or explicit in this research on virality is the argument that a viral story reveals user judgements about what deserves to be shared. Our argument is that viral spread cannot be explained solely by content features or user judgment. There is a growing recognition among scholars that platform algorithms also guide the production and circulation of cultural commodities, including news (e.g., Gerlitz & Helmond, 2013; Nieborg & Poell, 2018; Vos & Thomas, 2019; Wallace, 2018). Both the Facebook newsfeed algorithm and the dynamics of social interaction are implicated in sharing and non-sharing behaviours and subsequent message exposure (see Thorson & Wells, 2016). This implies that individual preferences and judgements are but a partial explanation of online news distribution. Empirically, in order to create a reliable study, we must also investigate the content properties of unshared news (see Kümpel et al., 2015). In this study, we bridge the existing empirical studies, which focus on content characteristics as the explanatory variable, and theoretical contributions, which have developed the tradition of gatekeeping studies to shed light on the role of “journalists, individuals and algorithms in a shared news dissemination process” (Wallace, 2018: 288; see also Thorson & Wells, 2016).

The study employs a multi-step quantitative research design based on a set of news articles in Sweden with Facebook sharing metrics. The dataset is from 2015, a time when Facebook was the dominating platform (Plantin et al., 2018). While the data used for this study was collected in 2015 – before Facebook (in 2016) changed its algorithms to downrate news items – the general issue of the platform dependency of news distribution remains, and so does Facebook’s position as a dominant platform. We show empirically that, in Sweden in 2015, news sharing on Facebook followed irregular patterns with multiple gates and different gatekeeping logics, which we term the gatekeeping logics of publishing, sharing, and spreading.

Gatekeeping in transition

The idea of gatekeeping originated with Lewin (1947; see also Katz, 1957) and has historically been used within journalism studies to understand the factors that shape what information gets published by news media. By extension, in modern media settings, gatekeeping generally refers to the processes that affect the type of information that flows to, through, and from different stakeholders and the conditions under which those processes operate (Barzilai-Nahon, 2008; Coddington & Holton, 2014; Hemsley, 2019; Thorson & Wells, 2016; Vos & Thomas, 2019). Traditionally, the gatekeepers of news were editors and journalists who controlled what stories were deemed newsworthy. However, according to the theory of networked gatekeeping (Barzilai-Nahon, 2008; see also Coddington & Holton, 2014), the roles of the gatekeepers and the “gated” have shifted: Those who were previously “gated” (e.g., people without influence over information selection and distribution) have effectively become gatekeepers through the development of digital and social media, while the influence over the flow of information of previous gatekeepers (e.g., news media) has decreased. Relatedly, Thorson and Wells (2016; see also Braun & Gillespie, 2011; Hemsley, 2019; van Dijk & Poell, 2013; Wallace, 2018) propose that the gates in this complex network can best be understood as curated flows in which several acts of curation simultaneously shape the flow of information. Thorson and Wells (2016) further identify four curating actors in addition to journalists: individual media consumers, social others embedded in networks, strategic communicators, and – most interesting for our study – algorithms that affect the display of content. Each of these curators are driven by different logics and preferences. We adopt the idea of different gatekeeping logics, but instead of five actors, we collapse strategic communicators, social others, and individual consumers into a single category of platform users, which we call social actors. We are thus left with three categories of curators or gatekeepers: news media (journalists), social actors (in the literature, these are mainly understood as media consumers), and platform algorithms.

The logics of publishing, sharing, and spreading

In this section, we develop the gatekeeping logics that correspond to the three categories of gatekeepers – news media, social actors, and algorithms. Following Altheide and Snow (1979; see also Strömbäck, 2008; van Dijk & Poell, 2013), we use the term logic to refer to the processes, principles, and practices by which a gatekeeper transforms and transmits information – including the substance, format, and style of the information. All three gates follow their own logic and, hence, affect the flow of information. However, their guiding principles vary depending on their relationship to the information and the purpose it serves. We call the three logics publishing (the logic of the news media gate), sharing (the logic of the public gate), and spreading (the logic of the algorithmic gate).

The publishing logic operates through journalists’ and editors’ assessment of news values (cognitive and normative concepts of what constitutes news) and news judgment (the application of those concepts on bits of information) as they select events to write about and publish (Strömbäck et al., 2012). Classic studies (Galtung & Holmboe Ruge, 1965; Harcup & O’Neill, 2017) have discovered that not all events have the same chance of passing through the news media gate and making the transition from information to an actual news story. Events that are, among other things, negative,

unexpected, and involving people (especially elites) have an increased chance of becoming news stories. This process makes a certain content composition available for media consumers to share.

A similar process of selection or gatekeeping takes place among members of the public, leading some scholars to refer to the public as secondary gatekeepers (Braun & Gillespie, 2011; Shoemaker & Vos 2009; van Dijck & Poell, 2013). Previous research indicates that different types of content may be appealing to share (Berger & Milkman, 2012; Khuntia et al., 2016; Larsson, 2018; Trilling et al., 2017), and as we describe in the methods section, these content elements are considered in the design of our study; however, the logic of sharing operates differently. Factors involved in the logic of sharing relate to both content and context. For example, a person thinking of sharing an article may be contemplating “Do I like this article?” but also “What will other people think of me if I share it?” Much previous research in journalism studies has emphasised content rather than context. However, sharing news on Facebook is often related to the management of social relationships (Picone et al., 2016). For example, praising leaders or supporting a particular viewpoint influence sharing, as well as the position of the sharer and their own personal social goals in terms of self-presentation (e.g., to entertain or to perform identity work) (see Bessi et al., 2015; Costera Meijer & Groot Kormelink, 2015; Katz, 1957; Robinson, 2014). By analogy to news values and news judgement (operating in the publishing logic), we might say that selection in the sharing logic is mainly driven by social values and social judgement.

News items become available to Facebook’s platform gate and the spreading logic operating through its newsfeed algorithm once a news item is (first) published by news media and (second) shared by someone on Facebook. As soon as a news item is shared, it becomes intertwined with the algorithms of the platform. Thus, it is impossible for a news item to spread either narrowly or widely without the influence of the platform algorithm. Exactly how the algorithm influences the onward distribution is an empirical question linked to the configurational change within each platform; thus, it is highly susceptible to change. Once shared, the algorithm mediates the display of news content to Facebook users. Thus, when a user sees a given item in their newsfeed, it may happen as a result of another user actively sharing. It may also happen due to behaviour associated with the user or the content, such as liking, commenting, reading, or any other behaviour that leaves data traces. Algorithmic decision-making then further depends upon who the sharer is and their previous clicking, liking, ranking, commenting, and sharing behaviour (Bakshy et al., 2015; DeVito, 2017; O’Brien et al., 2014), as well as a range of other factors in what is a highly complex and personalised machine-learning system (DeVito, 2017). A Facebook algorithm instantiates the spreading logic, where the collection of data to enable advertising sales is the paramount goal (Van Couvering, 2017). Content which generates engagement (and therefore, user data) is more valuable to the platform and will be spread more widely (see Vaidhyanathan, 2018), acquiring a platformised character in which information about how people behave when viewing and interacting with it become increasingly important in its distribution (see also Helmond, 2015; Nieborg & Poell, 2018), regardless of its editorial subject or tone. Again, by analogy, the algorithmic gate where spreading occurs is marked by commercial values and commercial judgement. This means, in the overall output, that certain stories are

amplified or muted compared with the selection that news organisations would have made when relying on professional news values and news judgement.

The publishing, sharing, and spreading logics described above are useful analytical categories. In practice, they are overlapping and interdependent and change continuously depending on platform priorities, audience habits, and media business strategies (Meese & Hurcombe, 2021). As Altheide and Snow (1979) argued, the logic of one actor can be imposed on the other actors. Historically, this was understood as the news media imposing its logic on political actors and wider society – today, platforms seem to have the upper hand. For example, we can see the logic of social media platforms being imposed on news media: In response to the platforms, publishers have made adjustments by making content more “engaging” and have developed new norms that gravitate away from journalistic objectivity to taking a position (Costera-Meijer, 2020; Ferrer-Conill & Tandoc Jr., 2018; Hurcombe et al., 2021). News organisations have also implemented data-driven personalisation technologies originating from platforms like Google, Amazon, and Facebook, while trying to adapt them to the logic permeating news work (Bodó, 2019). Meanwhile, actors on social media have also followed platform cues prompting engagement with content (Docherty, 2020), in the process generating more data for the platforms. Evidence in the reverse direction is sparse, but it is likely that some features from media logic (e.g., a high value of fresh news) might transfer to platforms. One thing we do know, however, is that amongst the general factors which Facebook cites as important in its newsfeed, content ranks among the lowest in prioritising what should be shown to users (DeVito, 2017).

In summary, the value of actual news content in each of the different gatekeeping logics is rather different. News content features are extremely important in the publishing logic, while content features function as a proxy for social network maintenance in the sharing logic, where personal context gains importance. Finally, in the spreading logic, content features per se are one factor among many, as the platform algorithm makes content *about* the item as important as content *within* the news item. Nevertheless, the democratic importance of news content remains, and the news that is spread has many social consequences. In this exploratory study, we empirically investigate the relationship between content properties and the distribution of news at the three gates.

Our overall methodological orientation is one of reverse engineering, using the content properties of shared, unshared, and widely shared stories to tell us something about the operation of the Facebook algorithm when it comes to news selection. Obviously, a direct examination of the algorithm’s logic would tell us more; however, researchers are unable to observe proprietary company algorithms. This is the “black box challenge”, to which there is no universally applicable method (Ashby, 1956/2015; Kitchin, 2017). Reverse engineering is one strategy often used to tackle the challenge, entailing experimentation with the inputs and outputs of the black box in order to, to quote Bucher (2018: 64), “find ways to make the algorithm talk”. The point of reverse engineering is not to “unveil the exact formula” of the algorithms but to “develop a critical understanding of the mechanisms and operational logic of the software” (Bucher, 2018: 61), while at the same time being aware that what we observe “is only a portion, or an aspect, of the whole truth” (Ashby, 1956/2015: 107). The research design, method, and result sections below provide details on our attempts to see the Facebook algorithm at work.

Research design

In this article, we empirically identify and compare the overall composition of the information available to the public as it is passed through the three different gates and was subjected to the different gatekeeping logics.

We began by replicating previous studies and investigating the explanatory strength of different news properties in the sharing of news items in our corpus. Our first research question therefore sought data that can both be compared with previous research and form a baseline for further inquiry:

RQ1: What characteristics of news items (in our sample) predicted a greater degree of sharing?

This first step in our research design examined how different news properties may explain sharing, using a linear regression analysis.

In previous work, much of the debate on the influence sharing may have on journalism is derived from a discussion of increased audience orientation: When readers act as gatekeepers, the expectation has been that news characteristics usually associated with journalism and that cater to audience preferences (e.g., tabloid journalism) will be more prevalent. This expectation has been confirmed by research showing that increased sharing is correlated with tabloid features such as expressing emotions, using a subjective writing style, and visual content (Berger & Milkman, 2012; Harcup & O'Neill, 2017; Khuntia et al., 2016; Trilling et al., 2017). Building on these earlier studies, the dataset we used was coded for tabloid features.

The second step in our design was to identify and compare the chosen news stories at the different gates by analytically separating the news stories most associated with the respective gates. We approached this by assessing the characteristics of items that were never shared by the public, using the same characteristics as in the first research question. Using a metaphor of explosion, we called these unshared items "duds" – potential hits that just don't detonate. Duds, we reasoned, passed the first gate (journalists) but not the second gate (readers) in terms of being shared. Thus, we developed a second, three-part research question, the first aspect of which asks the following:

RQ2a: What are the characteristics of unshared news items?

Since we measured unshared and shared news items in a dichotomous way, our choice of method was binary logistic regression analysis, which estimates how a number of independent variables may impact the odds of moving from one category to another (e.g., the factors that increase the likelihood of a news item being a dud). Next, we compared duds with shared items to see potential differences in news judgement. If content guides sharing, we reasoned, there should be large differences between what gets shared and what does not. Hence, the second aspect of research question 2 asks the following:

RQ2b: How do the characteristics of shared and unshared news items differ, if at all?

The final step in our design was to compare the news stories most influenced by members of the public (the second gate) with those news stories spread more by the platforms (the third gate), to detect the effect of platform gatekeeping logic. According to our theoretical examination of platform logic, very widely spread content must be more embedded in the platform. Thus, the third aspect of research question 2 asks the following:

RQ2c: How do the characteristics of news shared a few times differ from those shared many times, if at all?

We call the news items that are shared only a few times “squibs”, as they, in contrast to duds, detonate but do not “blow up” in terms of social media distribution. We use the conventional name “hits” for widely spread stories. We addressed this empirically in a three-part process: First, we examined the dataset to see where changes in sharing patterns occurred (using a logistic regression). Next, we divided the dataset into three cohorts: zero shares (duds); shared 1–10 times (squibs); and shared 100+ times (hits). Finally, we used correlation tests between cohorts for each variable. This process allowed us to identify and collate news characteristics as they appeared at the three gates. This process is detailed more fully in the results section.

Data collection

Studying Facebook presents several challenges in obtaining data. For this study, we used one publicly available dataset made available by the independent foundation Institutet för Mediestudier in Sweden (2016) [The Swedish Institute for Media Studies] (IMS), which also coded the characteristics of the articles, providing the raw data used in our analysis. The raw data consists of published articles appearing on the websites of traditional media outlets that included counts of Facebook sharing and contained both shared and unshared articles. Eighteen media organisations – a mix of tabloid, broadsheet, local papers, free dailies or weeklies, public service broadcast media, and commercial broadcast media – are included in the data. An overview of the media organisations, their origin, and web traffic (where available) can be found in Appendix A.

The IMS project is designed to investigate the quality of news in relation to an informed citizenry. The assumption is that a citizenry that has access to thorough, unbiased original reporting of contemporary societal issues has a better chance of making informed decisions compared with one that does not (for an elaborate discussion, see Aalberg & Curran, 2012). The codesheet was largely derived and adapted from the review and suggestions made by Reinemann and colleagues (2012) on how to study thematic, focus, and style dimensions in news content. But the IMS project also draws from other research related to the quality of news: news topics (Quandt, 2008), tonality (Messner & South, 2011; Michaelson & Griffin, 2005), and news origin, or original reporting (Williams & Delli Carpini, 2011).

The dependent variable is the number of shares a news item received on Facebook, a number that was visible at the bottom or side of the news items. Three trained coders collected the data, and an account of all the variables used in this article and their reliability scores – reported as both average Holsti and Fleiss Kappa – can be found in Appendix B.

The drawbacks of this secondary data include a limited choice of variables and a limited visibility into the reliability of the coding. In terms of reliability, the study has good scores when evaluated against the Holsti standard, as the average score is 0.9 and all variables meet the 0.8 standard suggested by authoritative sources (e.g., Lombard et al., 2002), but it is problematic for some variables if the less generous Fleiss Kappa is used. However, we are not primarily interested in the performance of individual

content characteristics, but rather in the overall pattern of how content characteristics fare at the different gates. Moreover, only 3.5–4.9 per cent of the sample was checked for inter-reliability in the last revision of the codesheet. Nevertheless, the dependent variable received a good score (e.g., Fleiss Kappa = 0.83), as did the average results.

Article data in the IMS study was collected using The Wayback Machine, hosted by the online repository *Archive.org*. The Wayback Machine essentially downloads and stores websites, in effect preserving the way they looked at a specific point in time. While online content analysis is in principle difficult to conduct, and archives cannot be fully trusted, The Wayback Machine has been validated and used in previous research (John, 2013; Karlsson, 2016; Murphy et al., 2007). The sample consists of seven randomised days (of the days available in the repository) from each news site in 2014. Every third news article on the landing page of the website was coded. The data were assessed and coded in 2015 and 2016.

To answer the first research question addressing potentially important characteristics, a subset consisting of only the news stories that were shared at least once ($n = 1,319$) was analysed. To answer the other research questions, the full dataset ($N = 3,144$) was used.

Results

In the following section, we present the results: first, the content characteristics of shared items; second, the content characteristics of unshared items; third, a comparison of the two groups; fourth, differences in content characteristics between widely shared items and narrowly shared items; and finally, differences in content characteristics between all three groups (unshared, narrowly shared, and widely shared items).

Content characteristics of shared items

We first conducted a linear regression of shared items to examine how much of the variance in sharing could be attributed to various content characteristics (RQ1), using the number of shares on Facebook as the dependent variable (see Table C1 in Appendix C).

The coefficients predict how much a given variable changes the number of shares. For example, a news item written in an interpretative style ($\beta = 1335.793^{***}$) is more likely to be shared compared with news written in descriptive style. The content variables included in the linear regression together explained 9.4 per cent of the variance in sharing (in line with most previous studies) (e.g., Berger & Milkman, 2012; García-Perdomo et al., 2017; Khuntia et al., 2016). In our sample, news written in an interpretative style, topics related to culture and science or technology, thematic framing, and citizen contribution increased sharing. On the other hand, news identified as advertorial, written from a neutral position, and items written by a man or an author of unidentifiable gender decreased sharing. In general, we see that few variables seem to affect sharing when a linear regression (which is common in previous research) is used on shared news items.

Content characteristics of unshared items (news duds)

A different pattern arose when we contrasted shared news items with unshared news items (see Table C2 in Appendix C). The first part of the second research question

(RQ2a) concerned the characteristics of news items that are not shared on Facebook. Previous studies have not included duds, but they provide important clues as to the potential importance of content in sharing. In order to answer RQ2a (establishing the characteristics of news duds) and RQ2b (contrasting the duds with news that were shared at least once), we ran a binary logistic regression analysis. This model, seen in Table C2, predicts the likelihood for a news item not to be shared depending upon a range of characteristics. The dependent variable was coded as a binary: unshared news was coded as “1” (56.2% of the sample; $n = 1,689$) and shared news as “0” (43.8% of the sample; $n = 1,319$). An odds ratio above 1.0 indicates that a particular characteristic lessens the likelihood for a news item to be shared on Facebook. So, an item with an odds ratio of 1.5 is 50 per cent less likely to be shared. Conversely, an odds ratio below 1.0 suggests that a characteristic makes the news item less likely to be a dud – that is, more likely to be shared on Facebook at least once.

Of all the characteristics tested, our analysis found only three items that contributed to a news item’s likelihood of being a dud: First, if the geographical location was not identifiable, the odds for the given news item to be shared on Facebook were twice as low, compared with news items with a geographical anchoring (odds ratio = 2.083**). Second, stories told with a positive tone were 1.3 times less likely to be shared compared with stories lacking a tone. Finally, news items that cover lifestyle issues were also less likely to be shared on Facebook. These three characteristics were the only ones that explained the odds of news *not* being shared on Facebook at a statistically significant level ($p < .05$). These characteristics are thus unique to news duds. On the other hand, the logistic regression also demonstrates that many content characteristics are significant when it comes to *avoiding* becoming a dud (which is a complicated – but statistically correct – way of saying that these content characteristics contribute to sharing).

Comparing duds with shared news

Looking at the variables that increase sharing in the two forms of regression (e.g., Table C1 & Table C2), some noteworthy patterns are visible (see Table C3 in Appendix C for an overview).

While the linear regression that only takes shares into account indicated that very few variables were relevant in explaining sharing, the binary logistic regression comparing duds and shared items revealed something rather different. In the latter case, 13 variables were found to increase the odds of being shared. It is also worth noting that some of the variables that affected sharing in the linear regression – interpretative style, thematic framing, science and technology topics (positively); advertorial or no detectable positioning (negatively) – no longer affected sharing in the binary logistic regression. Male authors went from affecting the sharing negatively in the linear regression to affecting the odds of sharing positively in the logistic regression. Only news dealing with culture predicts sharing positively in both regressions. Finally, 24 of the variables have no significant correlation to sharing regardless of regression method. Thus, depending on the measurement used – and adding emphasis on unshared news items – there are divergent answers to the apparently simple question of which content properties affect sharing and how.

Separating hits from squibs

In the analyses thus far, we examined sharing as a one-dimensional concept, in effect assuming, as did previous research, that content characteristics affect sharing in a uniform manner. But as discussed previously, our theoretical proposition is that news shared only a few times is less likely to be influenced by network effects and platform algorithms. Thus, RQ2c asks if there is any difference in news properties between news items that are shared only a few times (squibs) compared with news items that are shared many times (hits).

In order to answer this, we first set out to find out an operational solution to the question of how to distinguish between squibs and hits – in other words, to identify a cutoff point (Hemsley, 2019). Our procedure was to group news items in different intervals (e.g., using 1–5 shares, 1–10 shares, or news shared only once as proxies for squibs) to find feasible ways of sorting the data. Irrespective of how we grouped them (with the reservation that we have not exhausted all possibilities), the same patterns occurred at roughly the same place. Somewhere at about 50–100 shares (since this occurred more often over 100 than 50, we use “100+” onwards for clarity), we observed that the data began to behave differently. That is to say, the content composition changed rather markedly compared with patterns among narrowly shared and unshared news stories. This was in line with our theoretical proposition that Facebook’s algorithms would become more visible with an increasing number of shares – assuming that they would be discernible from user behaviour. We therefore categorised the data as follows: unshared news (duds), news shared 1–10 times (squibs), and news shared 100+ times (hits). Using the interval between 1–10 for squibs and 100+ for hits allows us to highlight the contrast between squibs and hits while simultaneously decreasing the risk of overlaps, as the change of pattern occurred somewhere over 50 shares. Having determined a reasonable group interval, the next step was to examine and contrast the variables in the different groups.

Comparing content characteristics between duds, squibs, and hits

As described above, we divided sharing into duds, squibs, and hits, and examined the 49 different content characteristics contained in the dataset. We reasoned as follows: If there were a linear or power-law relationship between content characteristics and news spread, and sharing were the function of the aggregate decisions of individual users (e.g., the second gate), then preferred content characteristics would either continuously increase or decrease as news was shared more widely. Similarly, we might expect characteristics which seem to have no significance on whether items are narrowly shared (squibs) to continue to have no effect on whether an item is widely shared (hits). Thus, if we represented these hypothetical results graphically, we would see lines of a continuous upward, downward, or flat slope – that is, linear patterns.

Tables C4 and C5 (see Appendix C) offer an overview of the content characteristics and the significant differences between duds, squibs, and hits. It shows that there were only three content variables (photography, emotionality, and news with own journalists) that displayed significant increases from duds to squibs to hits (labelled continuous rise). There were no instances of a continuous decline of content properties, and there were nine cases when the presence of the content property remained unchanged at the

different levels of sharing (labelled no change). The most striking observation is that linear patterns (see Table C4) seem to be the exception, rather than the rule, occurring only 12 out of 49 possible times. Instead, we often observed abrupt transitions in the relationship of news characteristics to Facebook shares, primarily between squibs and hits. This, in turn, suggests that different forces (e.g., Facebook's algorithms) are at work when news are spread widely.

Relatedly, there were often significant differences between the hits and the other columns (see Table C5). There were fewer differences between the duds and squibs than between the squibs and hits. Accordingly, many of the content characteristics seem to play a larger role when news stories are *widely* spread. To be specific, 38 of the content characteristics showed no significant difference between duds and squibs, while between all three levels of sharing, there were only 9 content characteristics that showed no significant difference at all. Hits, it seems, are subject to different selection pressures than duds or squibs.

An interesting comparison with the results from the regression analyses in Table C3 was also evident. In the regression analyses, very *few* of the content variables had any significant relation to sharing, while Table C4 shows that *most* of the content variables seem relevant when news is widely spread.

Conclusion: Platform gatekeeping and its consequences

Our study is limited in some regards: It concerns Facebook only, and it builds on data collected in 2015 (before an algorithm change at Facebook) and from one country, with a particular media and political system. Another issue is that digital media – both news media and platforms – change rapidly and, subsequently, this might impact the composition of the news stories available to, and distributed through, the platforms. The numerous correlation tests in the results may result in false positives (Type 1 errors), but that should not distract from the rather robust findings in the overall patterns and descriptive data. Yet another issue for future research to validate is to what extent the reverse engineering approach used here can be applied to other cases. Nevertheless, while the data has limitations and platforms change, the larger issue of different gates and associated gatekeeping logics remains, especially the gatekeeping by platforms. Moreover, this study also provides a relevant point of comparison for future research. For instance, did the change in Facebook's algorithm in 2016 and subsequent downgrading of news stories produce different patterns compared with those observed in this study?

There is a paradox in the results. According to DeVito (2017), content characteristics are one of the least important things for Facebook's algorithms. A feasible outcome – if content features were indeed unimportant – would be that Facebook would simply forward the news items with the same composition as the news media and the audience. Yet, as the results show and the theory predicts, rather than simply passing along news, Facebook's algorithms actively affect the way news is spread. Our proposed interpretation of this ostensible paradox is that the algorithm is reacting to content not detected by content analysis – in other words, metadata *about* the news items. Widely shared news appears to be promoted using these invisible features in accordance with commercial values and judgements. To put it another way, if it were only up to the news media and the audience, the circulation of news would look different. Relatedly,

if we only infer audience judgement based on the news items that get widely spread, we would be mistaken.

Our results point to the role of Facebook as a *de facto* moderator of the public sphere at the time this data was collected. This supports the proposition of scholars such as Helberger (2019: 1000), who points to a potential “creation of new concentrations of market or opinion power”, and Wallace (2018), who warns about non-journalistic actors transforming the public sphere. This concern is echoed by scholars outside journalism studies as well; for example, Vaidhyathan (2018: 3) claims that “anti-social” media foster “the deterioration of democratic and intellectual culture around the world”. As a consequence, efforts made by news media to follow a professional credo and adhere to the democratic function of news through their gatekeeping become less important when people find the news through social media platforms and platform curation is involved.

According to our data, Facebook’s commercial gatekeeping logic differs substantially from the professional gatekeeping logic of news media and the social gatekeeping logic of citizens (the latter two seem to be more aligned). This might also have implications for the legitimacy and authority of news media as providers of information in democracies, since citizens might use their values – that are rather like the values of news media, judging by our results – to assess the professional logic of the news organisations on the basis of the distribution of news caused by the commercial logic of the platforms. Equally, news organisations might mistake the (wide) spread of news stories for user preference and provide more of these to fulfill a perceived need of citizens, while really accommodating the commercial logic of Facebook.

The journalistic gate applies a publishing gatekeeping logic, its output can be duly noted, and it can be held accountable to acknowledged professional standards. This is more challenging with platforms, both because their impact on news distribution is difficult to assess and because of the opaque values that guide the filtering. Given the proposed influence that platforms have in the distribution of news stories, a key policy issue should be transparency in the decision-making process and the regulation of the role of platforms in public speech. This debate is currently developing around the world, with governments proposing specific laws related to content shared on platforms, as happened in France (Cox, 2020); platforms attempting to self-regulate by publicly removing content, as Facebook and Twitter did with misinformation related to Joe Biden’s 2020 American presidential campaign (Paul, 2020); and platforms resisting government action by threatening a withdrawal of services, as happened when Facebook threatened to remove Australian users’ ability to share content when the Australian Competition & Consumer Commission proposed regulation and monitoring the use of algorithms and advertising charges (Duke, 2019; Meade, 2020).

The discrepancies found in the priorities between news organisations and citizens, on the one hand, and platforms, on the other, also pose challenges for researchers. Especially problematic in light of this is the tendency in previous empirical research to attribute news distribution to user preferences and content characteristics. While the theoretical interest can be redirected, great challenges remain for research in developing appropriate methodological approaches to study how news is filtered through the different gates. The reversed engineering employed in our study revealed some interesting findings but is far from standardised and needs to be tested in other contexts. Furthermore, as illustrated in the results section, there are remarkable differences depending on whether unshared

news items are included and on which type of statistical analysis is performed. Further methodological innovation, exploration, and validation is needed.

An understanding of the interaction between the people, content, and platforms is critical, precisely as Lewin (1947) and Katz (1957) both observed over 60 years ago. But this interaction is changing: People are increasingly getting their news diet through platforms, and as scholars, we must seek to better understand the flow of information in contemporary society and the forces behind the patterns.

Acknowledgements

We wish to thank the Anne-Marie and Gustaf Ander stiftelse för mediaforskning [Anne-Marie and Gustaf Ander foundation for media research] for the financial support making this research possible.

References

- Aalberg, T., & Curran, J. (2012). *How media inform democracy*. Routledge. <https://doi.org/10.4324/9780203803448>
- Altheide D. L., & Snow, R. P. (1979). *Media logic*. Sage. (Original work published 1956)
- Ashby, R. (2015). *An introduction to cybernetics*. Wiley & Sons.
- Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239), 1130–1132. <https://doi.org/10.1126/science.aaa1160>
- Barzilai-Nahon, K. (2008). Toward a theory of network gatekeeping: A framework for exploring information control. *Journal of the American Society for Information Science and Technology*, 59(9), 1493–1512. <https://doi.org/10.1002/asi.20857>
- Bessi, A., Coletto, M., Davidescu, G. A., Scala, A., Caldarelli, G., & Quattrociocchi, W. (2015). Science vs conspiracy: Collective narratives in the age of misinformation. *PLoS ONE*, 10(2), e0118093. <https://doi.org/10.1371/journal.pone.0118093>
- Bennett, W. L. (2009). *News: The politics of illusion*. University of Chicago Press.
- Berger, J., & Milkman, K. L. (2012). What makes online content viral? *Journal of Marketing Research*, 49(2), 192–205. <https://doi.org/10.1509/jmr.10.0353>
- Bodó, B. (2019). Selling news to audiences – A qualitative inquiry into the emerging logics of algorithmic news personalization in European quality news media. *Digital Journalism*, 7(8), 1054–1075. <https://doi.org/10.1080/21670811.2019.1624185>
- Braun, J., & Gillespie, T. (2011). Hosting the public discourse, hosting the public: When online news and social media converge. *Journalism Practice*, 5(4), 383–398. <https://doi.org/10.1080/17512786.2011.557560>
- Bucher, T. (2018). *If... then: Algorithmic power and politics*. Oxford University Press. <https://doi.org/10.1093/oso/9780190493028.001.0001>
- Coddington, M., & Holton, A. E. (2014). When the gates swing open: Examining network gatekeeping in a social media setting. *Mass Communication and Society*, 17(2), 236–257. <https://doi.org/10.1080/15205436.2013.779717>
- Costera Meijer, I. (2020). Understanding the audience turn in journalism: From quality discourse to innovation discourse as anchoring practices 1995–2020. *Journalism Studies*, 21(16), 2326–2342. <https://doi.org/10.1080/1461670X.2020.1847681>
- Costera Meijer, I., & Groot Kormelink, T. (2015). Checking, sharing, clicking and linking. *Digital Journalism*, 3(5), 664–679. <https://doi.org/10.1080/21670811.2014.937149>
- Cox, K. (2020, May 14). *Failure to delete hate speech could cost Facebook, Google billions in France*. ArsTechnica. <https://arstechnica.com/tech-policy/2020/05/failure-to-delete-hate-speech-could-cost-facebook-google-billions-in-france/>
- DeVito, M. A. (2017). From editors to algorithms: A values-based approach to understanding story selection in the Facebook news feed. *Digital Journalism*, 5(6), 753–773. <https://doi.org/10.1080/21670811.2016.1178592>
- Docherty, N. (2020). Facebook's ideal user: Healthy habits, social capital, and the politics of well-being online. *Social Media + Society*, 6(2), 1–13. <https://doi.org/10.1177/2056305120915606>
- Duke, J. (2019, February 18). 'People not regulators should decide': Facebook slams calls for crack-down. *The Sydney Morning Herald*. <https://www.smh.com.au/business/companies/people-not-regulators-should-decide-facebook-slams-calls-for-crackdown-20190217-p50ybo.html>

- Ferrer-Conill, R., & Tandoc Jr., E. C. (2018). The audience-oriented editor: Making sense of the audience in the newsroom. *Digital Journalism*, 6(4), 436–453. <https://doi.org/10.1080/21670811.2018.1440972>
- Galtung, J., & Holmboe Ruge, M. (1965). The structure of foreign news: The presentation of the Congo, Cuba and Cyprus crisis in four Norwegian newspapers. *Journal of Peace Research*, 2(1), 64–91. <https://doi.org/10.1177/002234336500200104>
- García-Perdomo, V., Salaverría, R., Kilgo, D., & Harlmow, S. (2017). To share or not to share. *Journalism Studies*, 19(8), 1180–1201. <https://doi.org/10.1080/1461670X.2016.1265896>
- Gerlitz, C., & Helmond, A. (2013). The like economy: Social buttons and the data-intensive web. *New Media & Society*, 15(8), 1348–1365. <https://doi.org/10.1177/1461444812472322>
- Harcup, T., & O'Neill, D. (2017). What is news? News values revisited (again). *Journalism studies*, 18(12), 1470–1488. <https://doi.org/10.1080/1461670X.2016.1150193>
- Helberger, N. (2019). On the democratic role of news recommenders. *Digital Journalism*, 7(8), 993–1012. <https://doi.org/10.1080/21670811.2019.1623700>
- Helmond, A. (2015). The platformization of the web: Making web data platform ready. *Social Media + Society*, 1(2). <https://doi.org/10.1177/2056305115603080>
- Hemsley, J. (2019). Followers retweet! The influence of middle-level gatekeepers on the spread of political information on Twitter. *Policy & Internet*, 11(3), 280–304. <https://doi.org/10.1002/poi3.202>
- Hurcombe, E., Burgess, J., & Harrington, S. (2021). What's newsworthy about 'social news'? Characteristics and potential of an emerging genre. *Journalism*, 22(2), 378–394. <https://doi.org/10.1177/1464884918793933>
- Ihlebak, K. A., & Sundet, V. S. (2021). Global platforms and asymmetrical power: Industry dynamics and opportunities for policy change. *New Media & Society*, Online First. <https://doi.org/10.1177/14614448211029662>
- John, N. A. (2013). "Sharing and web 2.0: The emergence of a keyword." *New Media & Society*, 15(2), 167–182. <https://doi.org/10.1177/1461444812450684>
- Kantar Sifo. (n.d.). *Räckvidd digitalt – total [Digital reach – total]*. Retrieved December 13, 2016, from <https://tns-sifo.se/rapporter-undersokningar/rackviddsmatningar/orvesto-internet>
- Karlsson, M. (2016). Goodbye politics, hello lifestyle: Changing news topics in tabloid, quality and local newspaper websites in the UK and Sweden from 2002 to 2012. *OBS-Observatorio*, 10(4), 150–165. <https://doi.org/10.15847/obsOBS1042016940>
- Katz, E. (1957). The two-step flow of communication : An up-to-date report on an hypothesis. *The Public Opinion Quarterly*, 21(1), 61–78. <https://doi.org/10.1086/266687>
- Khuntia, J., Sun, H., & Yim, D. (2016). Sharing news through social networks. *International Journal on Media Management*, 18(1), 59–74. <https://doi.org/10.1080/14241277.2016.1185429>
- Kitchin, R. (2017). Thinking critically about and researching algorithms. *Information Communication and Society*, 20(1), 14–29. <https://doi.org/10.1080/1369118X.2016.1154087>
- Kleis Nielsen, R., & Ganter, S. A. (2018). Dealing with digital intermediaries: A case study of the relations between publishers and platforms. *New Media & Society*, 20(4), 1600–1617. <https://doi.org/10.1177/1461444817701318>
- Kümpel, A. S., Karnowski, V., & 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. <https://doi.org/10.1177/2056305115610141>
- Larsson, A. O. (2018). "I shared the news today, oh boy." *Journalism Studies*, 19(1), 43–61. <https://doi.org/10.1080/1461670X.2016.1154797>
- Lewin, K. (1947). Frontiers in group dynamics. *Human Relations*, 1(2), 143–153. <https://doi.org/10.1177/001872674700100201>
- Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2002). Content analysis in mass communication: Assessment and reporting of intercoder reliability. *Human Communication Research*, 28(4), 587–604. <https://doi.org/10.1111/j.1468-2958.2002.tb00826.x>
- Meade, A. (2020, September 1). Facebook threatens to block Australians from sharing news in battle over landmark media law. *The Guardian*. <https://www.theguardian.com/media/2020/sep/01/facebook-instagram-threatens-block-australians-sharing-news-landmark-accs-media-law>
- Mediastudier. (2016). *Mediastudiers innehållsanalys [Content analysis of media studies]*. Institutet för Mediastudier. <https://mediastudier.se/mediastudiers-innehallsanalys/>
- Meese, J., & Hurcombe, E. (2021). Facebook, news media and platform dependency: The institutional impacts of news distribution on social platforms. *New Media & Society*, 23(8), 2367–2384. <https://doi.org/10.1177/1461444820926472>
- Messner, M., & South, J. (2011). Legitimizing Wikipedia. *Journalism Practice*, 5(2), 145–160. <https://doi.org/10.1080/17512786.2010.506060>
- Michaelson, D., & Griffin, T. (2005). *A new model for media content analysis*. Institute for Public Relations. <https://instituteforpr.org/wp-content/uploads/MediaContentAnalysis.pdf>

- Murphy, J., Hashim, N. H., & O'Connor, P. (2007). Take me back: Validating the wayback machine. *Journal of Computer-Mediated Communication*, 13(1), 60–75. <https://doi.org/10.1111/j.1083-6101.2007.00386.x>
- Newman, N., Fletcher, R., Schulz, A., Andi, S., & Nielsen, R. K. (2020). *Reuters Institute Digital News Report 2020*. Reuters Institute for the Study of Journalism, University of Oxford. <https://www.digitalnewsreport.org/survey/2020>
- Nieborg, D. B., & Poell, T. (2018) The platformisation of cultural production: Theorizing the contingent cultural commodity. *New Media & Society*, 20(11), 4275–4292. <https://doi.org/10.1177/1461444818769694>
- O'Brien, H., Freund, L., & Westman, S. (2014). What motivates the online news browser? News item selection in a social information seeking scenario. *Information Research – An International Electronic Journal*, 19(3), 95–112.
- Paul, K (2020, October 15). Facebook and Twitter restrict controversial New York Post story on Joe Biden. *The Guardian*. <https://www.theguardian.com/technology/2020/oct/14/facebook-twitter-new-york-post-hunter-biden>
- Picone, I., De Wolf, R., & Robijt, S. (2016). Who shares what with whom and why? News sharing profiles amongst Flemish news users. *Digital Journalism*, 4(7), 921–932. <https://doi.org/10.1080/21670811.2016.1168708>
- Plantin, J. C., Lagoze, C., Edwards, P. N., & Sandvig, C. (2018). Infrastructure studies meet platform studies in the age of Google and Facebook. *New Media & Society*, 20(1), 293–310. <https://doi.org/10.1177/1461444816661553>
- Quandt, T. (2008). (No) news on the world wide web? *Journalism Studies*, 9(5), 717–738. <https://doi.org/10.1080/14616700802207664>
- Reinemann, C., Stanyer, J., Scherr, S., & Legnante, G. (2012). Hard and soft news: A review of concepts, operationalizations and key findings. *Journalism*, 13(2), 221–239. <https://doi.org/10.1177/1464884911427803>
- Robinson, S. (2014). The active citizen's information media repertoire: An exploration of community news habits during the digital age. *Mass Communication and Society*, 17(4), 509–530. <https://doi.org/10.1080/15205436.2013.816745>
- Shoemaker, P., & Vos, T. (2009). *Gatekeeping Theory*. Routledge. <https://doi.org/10.4324/9780203931653>
- Strömbäck, J. (2008). Four phases of mediatization: An analysis of the mediatization of politics. *The International Journal of Press/Politics*, 13(3), 228–246. <https://doi.org/10.1177/1940161208319097>
- Strömbäck, J., Karlsson, M., & Hopmann, D. N. (2012). Determinants of news content. *Journalism Studies*, 13(5-6), 718–728. <https://doi.org/10.1080/1461670X.2012.664321>
- Thorson, K., & Wells, C. (2016). Curated flows: A framework for mapping media exposure in the digital age. *Communication Theory*, 26(3), 309–328. <https://doi.org/10.1111/comt.12087>
- Trilling, D., Tolochko, P., & Burscher, B. (2017). From newsworthiness to shareworthiness: How to predict news sharing based on article characteristics. *Journalism & Mass Communication Quarterly*, 94(1), 38–60. <https://doi.org/10.1177/1077699016654682>
- Vaidhyanathan, S. (2018). *Antisocial media: How Facebook disconnects us and undermines democracy*. Oxford University Press. <https://doi.org/10.1093/oso/9780190056544.001.0001>
- Van Couvering, E. (2017, January 4–5). The political economy of new media revisited: Platformisation, mediatization, and the politics of algorithms. *Proceedings of the 50th Annual Hawaii International Conference on System Sciences (HICSS)*, 1812–1819. <http://hdl.handle.net/10125/41374>
- van Dijck, J., & Poell, T. (2013). Understanding social media logic. *Media and Communication*, 1(1), 2–14. <https://doi.org/10.12924/mac2013.01010002>
- Vos, T. P., & Thomas, R. J. (2019). The discursive (re)construction of journalism's gatekeeping role. *Journalism Practice*, 13(4), 396–412. <https://doi.org/10.1080/17512786.2018.1478746>
- Wallace, J. (2018). Modelling contemporary gatekeeping: The rise of individuals, algorithms and platforms in digital news dissemination. *Digital Journalism*, 6(3), 274–293. <https://doi.org/10.1080/21670811.2017.1343648>
- Wang, Q. (2020). Differentiation and de-differentiation: The evolving power dynamics between news industry and tech industry. *Journalism and Mass Communication Quarterly*, 97(2), 509–527. <https://doi.org/10.1177/1077699020916809>
- Williams, B., & Delli Carpini, M. (2011). *After broadcast news. Media regimes, democracy and the new information environment*. Cambridge University Press.

Appendix A

Table A1 Overview of media organisations studied

Media	Origin	Reach (weekly unique visitors, thousands) October 2016
Svt.se	Public service broadcaster (TV)	1,611
Sverigesradio.se	Public service broadcaster (Radio)	879
Aftonbladet.se	National tabloid	3,261
Expressen.se	National tabloid	2,685
DN.se	National broadsheet	1,066
Svd.se	National broadsheet	1,077
Tv4	Commercial national broadcaster	n.a.
Metro	Free daily (mostly distributed in Stockholm but also in some other cities)	199
ABC	Regional public service broadcaster (TV)	n.a.*
SR Värmland	Regional public service broadcaster (Radio)	n.a.*
Värmlands Folkblad	Local daily	70
Nya Wermlandstidningen	Local daily	99
Mittnytt	Regional public service broadcaster (TV)	n.a.*
Sundsvalls Nyheter	Free weekly	n.a.
P4 Västmanland	Regional public service broadcaster (Radio)	n.a.*
Vestmanlands Läns Tidning	Local daily	104
Västerås Tidning	Free weekly	n.a.

* The regional public service broadcasters are organised under the website of the parent organisation and there are no available traffic numbers for them.

Source: Kantar Sifo, n.d.

Appendix B

Table B1 Variables studied and intercoder reliability test results

Variable	Average Holsti	Fleiss Kappa
Photography	0.93	0.87
Video/audio	0.96	0.78
Citizen contribution	1.00	1.00
Shares on Facebook	0.96	0.83
News topic	0.85	0.83
Societal actors*	0.97	0.75
Decision-making authorities*	0.89	0.63
Policy plan*	0.90	0.62
Actors concerned*	0.91	0.53
Framing individual-social relevance*	0.82	0.69
Framing episodic-thematic*	0.92	0.75
Personal-impersonal*	0.95	0.72
Emotional-unemotional*	0.83	0.44
Dominant tonality*	0.82	0.60
Journalistic style interpretive/ descriptive*	0.91	0.63
Journalistic positioning vis-à-vis subject*	0.90	0.67
Timeliness*	0.83	0.62
Main geographical area of reporting	0.85	0.71
Origin of news	0.89	0.78
Gender of author	0.98	0.97
Average score	0.90	0.72

Comments: Three trained coders coded the material. There were several revisions of the codesheet and in the final test, the variables with an asterisk (*) were test-coded on 112 articles, while the rest were test-coded on 153 articles.

Appendix C

Table C1 Content characteristics' effect on news sharing

Content characteristic themes	Content characteristics variables	OLS Regression analysis (standard error in parentheses)
Journalistic Style	Emotional style (ref: not emotional)	332.665 (259.990)
	Impersonal style (ref: personal)	508.563 (388.578)
	Positive tonality (ref: no tonality)	89.495 (373.017)
	Negative tonality (ref: no tonality)	87.491 (325.036)
	Interpretative style (ref: descriptive style)	1335.793*** (402.448)
	Critical positioning (ref: neutral)	-338.068 (496.644)
	Advertorial positioning (ref: neutral)	-1557.548* (751.194)
	No detectable positioning (ref: neutral)	-956.185* (429.770)
Multimodality	Photography attached (ref: no photo)	-123.909 (280.923)
	Video (ref: none)	264.434 (398.651)
	Audio (ref: none)	-619.886 (423.797)
Topic (ref: "other" incl. weather)	Politics	-622.008 (652.234)
	Economy	-561.298 (721.542)
	Social issues	-371.616 (613.186)
	Sports	-194.602 (571.136)
	Accidents	-235.132 (643.155)
	Culture	1353.507* (641.192)
	Entertainment	-320.166 (561.225)
	Science/tech.	2684.851*** (781.693)
	Media	-949.041 (1123.215)
	Environment	-884.136 (887.246)
	Crime	628.135 (590.565)
	Lifestyle/health	-711.946 (615.752)
	War/conflict	-182.370 (787.948)

Geographical scope (ref: local)	Unable to identify geographical scope	-566.163 (760.358)
	Regional	252.898 (766.504)
	National	562.836 (337.299)
	International	119.617 (321.476)
Temporality (ref: current)	Unable to identify temporality	805.516 (549.910)
	Past	-638.522 (396.356)
	Future	-444.175 (576.650)
Original news story (ref: not original news)	–	-408.780 (358.913)
Civic engagement-index	–	204.662 (523.231)
Framing	Social (ref: not applicable)	-530.646 (496.896)
	Individual (ref: not applicable)	-71.347 (415.883)
	Thematic (ref: episodic)	1033.484** (345.608)
Citizen contribution (ref: no contribution)	–	3188.259*** (848.629)
Gender of author (ref: woman)	Unable to identify author gender	-1032.360* (415.388)
	Man	-751.422** (261.811)
	Man and woman	-687.464 (577.210)
Constant	-69.494	
Adjusted R^2	0.065	
R^2	9.4%	
n	1,235	

Comments: Ordinary Least Square Regression Analysis (standard error in parentheses).

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table C2 Odds of being a news dud

Content characteristics themes	Content characteristics variables	Odds ratios	Lower CI	Upper CI
Media type (ref: local PBS)	Tabloid	.115***	.079	.168
	National broadcaster	1.085	.691	1.706
	Urban broadsheet	.573**	.397	.829
	Provincial daily	.531**	.369	.765
Emotionality	Emotional style (ref: not emotional)	.705**	.577	.862
	Impersonal style (ref: personal)	.907	.672	1.224
	Negative tonality (ref: not applicable)	.970	.748	1.257
	Positive tonality (ref: not applicable)	1.363***	1.027	1.809
	Interpretative style (ref: descriptive style)	1.026	.746	1.411
	No detectable positioning (ref: neutral)	.844	.610	1.167
	Critical positioning (ref: neutral)	.818	.534	1.253
	Advertorial positioning (ref: neutral)	.791	.449	1.393
Multimodality	Photography attached (ref: no photo)	.608***	.496	.744
	Audio/visual elements (ref: no audio/visuals)	.600***	.459	.785
News topic (ref: "other" incl. weather)	Politics	.901	.560	1.450
	Economy	.987	.595	1.636
	Social issues	.594*	.374	.942
	Sports	1.353	.891	2.057
	Accidents	1.041	.643	1.685
	Culture	.555*	.341	.902
	Entertainment	1.244	.810	1.910
	Science/tech.	.848	.478	1.505
	Media	1.226	.554	2.714
	Environment	.575	.283	1.169
	Crime	.732	.474	1.130
	Lifestyle/health	1.557*	1.000	2.425
	War/conflict	.769	.417	1.416
Geographical scope (ref: local)	Unable to identify geographical scope	2.083**	1.269	3.419
	Regional	1.197	.719	1.995
	National	1.148	.872	1.511
	International	.672	1.062	.805
Temporality (ref: current)	Unable to identify temporality	1.252	.862	1.820
	Past	1.251	.917	1.707
	Future	1.069	.713	1.603
Framing	Individual (ref: not applicable)	.689*	.517	.919
	Social (ref: not applicable)	.487***	.343	.691
	Thematic (ref: episodic)	.969	.745	1.260
Citizen contribution (ref: no contribution)	–	.341	.683	.312

Gender of author (ref: woman)	Unable to identify author gender	.899	.658	1.229
	Man	.809*	.660	.990
	Man and woman	.510*	.297	.877
Civic engagement-index	–	.655*	.434	.989
Original news story (ref: not original news)	–	.680**	.518	.892
Constant	2.671			
Nagelkerke R ²	25.5%			
Hosmer Lemeshow	.109			
n	2,871			

Comments: Binary logistic regression. CI = Confidence intervals. Civic engagement-index = (societal actors + decision-making authorities + policy plan + actors concerned) / 4 (Cronbach’s alpha = .862).

***p < 0.001, **p < 0.01, *p < 0.05

Table C3 Comparison of significant values in different regressions

Content characteristics themes	Content characteristics variables	Binary logistic regression: Odds of being a dud	Linear regression
Journalistic style	Emotional style (ref: not emotional)	-	n.a.
	Impersonal style (ref: personal)	n.a.	n.a.
	Negative tonality (ref: not applicable)	n.a.	n.a.
	Positive tonality (ref: not applicable)	+	n.a.
	Interpretative style (ref: descriptive style)	n.a.	+
	No detectable positioning (ref: neutral)	n.a.	-
	Critical positioning (ref: neutral)	n.a.	n.a.
	Advertorial positioning (ref: neutral)	n.a.	-
Multimodality	Photography attached (ref: no photo)	-	n.a.
	Audio/visual elements (ref: no audio/visuals)	-	n.a.
News topic (ref: "other" incl. weather)	Politics	n.a.	n.a.
	Economy	n.a.	n.a.
	Social issues	-	n.a.
	Sports	n.a.	n.a.
	Accidents	n.a.	n.a.
	Culture	-	+
	Entertainment	n.a.	n.a.
	Science/tech.	n.a.	+
	Media	n.a.	n.a.
	Environment	n.a.	n.a.
	Crime	n.a.	n.a.
	Lifestyle/health	+	n.a.
	War/conflict	n.a.	n.a.

Geographical scope (ref: local)	Unable to identify geographical scope	+	n.a.
	Regional	n.a.	n.a.
	National	n.a.	n.a.
	International	n.a.	n.a.
Temporality (ref: current)	Unable to identify temporality	n.a.	n.a.
	Past	n.a.	n.a.
	Future	n.a.	n.a.
Framing	Individual (ref: not applicable)	-	n.a.
	Social (ref: not applicable)	-	n.a.
	Thematic (ref: episodic)	n.a.	+
Citizen contribution (ref: no contribution)	–	n.a.	+
Gender of author (ref: woman)	Unable to identify author gender	n.a.	-
	Man	-	-
	Man and woman	-	n.a.
Civic engagement-index	–	-	n.a.
Original news story (ref: not original news)	–	-	n.a.

Comments: Table C3 provides an overview of the different regressions. If the news characteristic is marked with a minus sign (-), there is a significant negative correlation, and if the news characteristic is marked with a plus sign (+), there is a significant positive correlation; “n.a.” means that there was no explanatory strength in any of the regressions. Note that a plus sign in the binary regression is indicative of being a dud. Thus, when there is a minus sign in the binary regression, it is shared more. Observe that it is only “culture” that behaves coherently between the different regressions, where a minus sign in the binary regression indicates that it is not a dud, while a plus in the linear regression demonstrates that it is positively correlated with sharing.

Table C4 *Characteristics of news that are not shared, shared a few times, and shared many times (linear patterns)*

Content variable		Duds (non-shared)	Squibs (Shared 1–10)	Hits (Shared 10+)
Continuous rise (Rise)	Photography, yes	66	73**	80 ^{†††*}
	Emotional (c.f. Unemotional)	38	44*	70 ^{***†††}
	Origin: own journo	67	73*	86 ^{***†††}
No change (Flatline) – 9 cases	Jour positioning advertorial	3	3	3
	Origin: Mix of jour+agency	1	2	1
	Origin: partner	3	3	3
	Topic: Accidents/disaster	6	5	5
	Topic: Crime	11	12	9
	Topic: war/conflict	3	5	3
	Topic: Others	7	5	6
	Topic: Science/tech	4	3	3
	Topic: Entertainment	10	12	10
Continuous decline (Fall)	None			

Comments: Table C4 shows the percentages of different content properties that are either unshared, shared few times, or shared many times, linear patterns (per cent rounded). All significant values have been measured with Cramer's V. Significance values are reported with asterisks when the values are compared to the column to the left of the column with asterisks. They are reported with crosses when the two outer columns are compared. The number of news items used varies due to internal dropout in the different variables, but the approximate values are as follows: unshared (n = 1,600); shared 1–10 times (n = 440); shared 100+ times (n = 410).

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, ^{†††} $p < 0.001$, ^{††} $p < 0.01$, [†] $p < 0.05$

Table C5 Characteristics of news that are not shared, shared a few times, and shared many times (non-linear patterns)

Spreading pattern	Content variable	Duds (Non-shared)	Squibs (Shared 1–10)	Hits (Shared 100+)
Rise then no change (Upper plateau)	News with focus on present	77	84**	79
Rise then decline (Hump)	Audio, yes	4	12***	4***
	International news	31	44***	35**
	Civic news index variables:			
	Societal actors present	10	12	24****†††
	Decision making authorities present	16	19	30****†††
	Policy plan present	11	13	21****†††
	Actors concerned present	6	7	16****†††
	Social framing (c.f. individual)	33	36	46**†††
	Thematic framing (c.f. episodic)	18	19	32****†††
	Personal (c.f. impersonal)	30	31	44****†††
	Negative tonality (c.f. pos. or n.a.)	16	19	38****†††
	Interpretive style (c.f. descriptive)	33	36	47****†††
No change then rise (Hook)	Jour positioning critical	5	6	20****†††
	National news	27	24	43****†††
	Origin: other	3	3	6*†††
	Citizen contribution	1	1	3*†††
	Topic: Politics	12	14	20*†††
	Topic: Media	2	1	2*
	Topic: Environment	1	2	4*†††
	Topic: Culture	5	5	8††
	Topic: Social issues	7	7	10†
	News with historical focus	8	9	11†
	Gender of author: Male	36	40	47*†††
	Gender of author: Female	30	31	38*†††
	Gender of author: Both	2	2	7***†††

No change then decline (Dip)	Jour positioning neutral	92	91	77***†††
	n.a. tonality (c.f. neg. or pos.)	68	69	46***†††
	Local news	31	27	18***††
	Regional news	4	3	1 [†]
	News with future focus	5	3	3 [†]
	Origin: Agency	15	12	3***†††
	Origin: n.a.	11	8	1***†††
	Topic: business	6	5	3 [†]
	Topic: Sports	14	15	8***†††
	Topic: Weather	1	1	0*
	Gender of author: n.a.	32	27	8***†††
Decline than rise (Dim-ple)	Video, yes	13	7***	17†***
Decline then no change (Lower plateau)	Positive tonality (c.f. neg. or n.a.)	17	13*	16
	News without geographical ID	7	3***	3†††
	News without time ID	10	4***	7
	Topic: Lifestyle, fitness	12	8*	8 [†]

Comments: Table C5 shows the percentages of different content properties that are either unshared, shared few times, or shared many times, non-linear patterns (per cent rounded). All significant values have been measured with Cramer's V. Significance values are reported with asterisks when the values are compared to the column to the left of the column with asterisks. They are reported with crosses when the two outer columns are compared. The number of news items used varies due to internal dropout in the different variables, but the approximate values are as follows: unshared ($n = 1,600$); shared 1–10 times ($n = 440$); shared 100+ times ($n = 410$).

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, ††† $p < 0.001$, †† $p < 0.01$, † $p < 0.05$