Social media gatekeeping: An analysis of the gatekeeping influence of newspapers' public Facebook pages

Kasper Welbers
VU University Amsterdam, The Netherlands; KU Leuven, Belgium

Michaël Opgenhaffen
KU Leuven, Belgium

Abstract
Due to the rising importance of social media platforms for news diffusion, newspapers are relying on social media editors to promote the distribution of their news items on these platforms. In this study, we investigate how much of an impact these social media editors really have, focusing on the impact of newspapers' public pages on Facebook. Since the actions of individual users are not visible on many platforms due to privacy consideration, we propose a method that leverages time series of aggregated scores for total user engagement, which are available for various platforms. We use this method to study and compare the influence of Facebook pages for six newspapers from the United Kingdom, the Netherlands, and Flanders, for all news items published over 2 weeks in 2017.

Social media platforms have become important sources of information for many people (Gottfried and Shearer, 2016; Newman et al., 2016). In response, newspapers have

Corresponding author:
Kasper Welbers, VU University Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands.
Email: k.welbers@vu.nl

become active on social media to reach out to these people and to attract them to their websites by distributing links to their own news items (Bastos, 2015; Hille and Bakker, 2013). The social media editors who manage the social media accounts of these newspapers, who can be traditional journalists and also more specialized social media experts, thereby fulfill an important role in the competition between news outlets. Given the difficult transition of newspapers into the digital age, that has seen revenues plummet and giants tumble (Hamilton, 2004; Ryfe, 2012), the question has been posed whether these social media endeavors could save newspapers (Ju et al., 2014).

Previous studies have investigated how newspapers use social media (Hille and Bakker, 2013) and whether having many subscribers to a newspaper's social media account has certain benefits (Ju et al., 2014). However, the audience of newspaper content on social media is not limited to direct subscribers. A defining characteristics of social media platforms is that the diffusion of content largely occurs through technologically enhanced social network processes (Klinger and Svensson, 2015). Each active user is a participant in the diffusion process, because their engagement with news content (i.e. liking, commenting, and sharing) makes the content visible to their social alters and thereby promotes its propagation throughout the network (Hermida et al., 2012). Furthermore, anyone can share links to news items, regardless of whether the newspaper itself shares these links, which is also made easy through the widespread availability of social media buttons—social media plug-ins on websites that allow people to directly like, comment, and share content (Gerlitz and Helmond, 2013). This involvement of individual users in the selection and distribution of news leads to "increasingly complex relationships between news production and consumption" (Goode, 2009: 1304) that warrant new research and debate. In this article, we focus on how the importance of user engagement for news distribution on social media makes it difficult to determine how much of the total audience of newspaper items on social media can actually be attributed to the work of social media editors.

To investigate this, we propose a method to estimate the proportion of audience engagement with news content on social media that can be attributed to specific users and apply this method to study the influence of social media editors on Facebook. Building on gatekeeping theory, we define the *social media editor channel* as the flow of newspaper items on Facebook that trace back to the publication of these items by the newspaper's social media editors (Barzilai-Nahon, 2008; Shoemaker and Vos, 2009). We refer to all other ways through which newspaper items enter circulation on social media (e.g. individual users, other news pages, interest groups) as the *alternative channel*. By measuring how influential these competing channels are, we address what Lewin (1947) stated to be the first diagnostic task of gatekeeping research: finding out who the gatekeepers are. If social media editors are indeed influential gatekeepers, this warrants investigation into their norms and routines regarding the selection and presentation of news items. If a substantial portion of the distribution stems from alternative channels, this calls for research into who the gatekeepers of these channels are (for which the method proposed in this study can be used). Based on our analysis, we also address how the role of social media editors in the gatekeeping process intersects with the role of individual users and what this means for the definition of gatekeeping on social network sites.

Our analysis covers six newspapers, two from the United Kingdom, two from the Netherlands, and two from Flanders, over the course of 2 weeks in 2017. For each newspaper, we monitored all news item publications on the website in real-time via the RSS feed. For each news item, we then monitored the Facebook engagement scores by querying the Facebook Graph API with their canonical URL, for every 30 minutes since the publication time until up to 3 days. The resulting time series per news item are then analyzed with multilevel pooled time series analysis. In addition to providing new empirical insights into the role of social media editors for newspapers on Facebook, we argue that this method has promising applications for other questions regarding news diffusion on social media.

Gatekeeping on social media

Gatekeeping theory addresses how the news messages that circulate throughout society are selected and shaped (Shoemaker and Reese, 1996; Shoemaker and Vos, 2009). Given the huge amount of events that occur each day, and the virtually countless number of ways to describe them, why do certain news messages spread like wildfire while other are left untold? To understand this, one of the most important factors is to understand the people, organizations, and institutions that control the most far and wide-reaching communication channels, such as television, newspapers, and online platforms such as Facebook and Twitter. The actors that control these channels can be conceptualized as *gatekeepers*, referring to their power to decide which messages may and may not pass through their channels. Perhaps the most vivid example stems from the seminal gatekeeping study of David Manning White (1950), who investigated how the wire editor of a local newspaper, referred to as *mr. Gates*, selected which messages were published.

Kurt Lewin (1947), who coined gatekeeping, argued that gatekeepers operate in a complex field, in which the gatekeeper and its environment "have to be considered as one constellation of interdependent factors" (p. 338). There are many studies that show that journalists are often influenced by the work of their colleagues (Cook, 2005; Crouse, 1972; Vliegenthart and Walgrave, 2008). A particularly strong form of interdependence occurs when gatekeepers operate within the same channels. For example, mr. Gates guarded the final gate before news reaches the audience, but before a news item arrived at mr. Gates' desk, a reporter had already decided that an event was worth writing about. This direct interdependence of gatekeepers needs to be taken into account when we analyze who the most important gatekeepers of society are. Bass (1969) argued, for instance, that when we study the gatekeeping process by investigating individual news outlets, we tend to underestimate the gatekeeping role of news agencies on which these news outlets often rely for much information. This claim might be even more accurate now, as many online channels lack their own news-gathering apparatus and mainly seek to curate and reinterpret news that is already circulating (Baum and Groeling, 2008; Welbers et al., 2018).

With the rising popularity of social media as a popular platform for news distribution, complex networks of interdependent gatekeepers are emerging (Goode, 2009). Someone, possibly a news organization itself, can *post* a news item or a link to a news item on a platform such as Facebook. A person who has a direct social network tie to the original

poster can see this post and can interact with the post (e.g. liking, commenting, sharing).[1] Other people connected to this person can see these interactions, due to which the content can diffuse further throughout the network. Essentially, this makes every actor that is exposed to the content a potential gatekeeper, but with different levels of influence (Shoemaker and Vos, 2009). Actors with a central position in the network, such as news organizations with many *followers*, can reach many people at once, akin to traditional mass communication. Yet, as we will address shortly, due to the speed of communication and high level of interconnectedness on social media, news with a high level of "share-worthiness" (Trilling et al., 2017) can diffuse rapidly even without mass communication, similar to how a contagious virus can spread rapidly throughout a dense population.

The notion of gatekeeping on social media requires a different conceptualization than the one used by White (1950) to describe the work of mr. Gates. In the traditional gatekeeping literature, the gatekeeper is someone who guards discrete gates that determine which news does and does not reach the audience. In a strict following of this definition, it can be argued that there will not be any gatekeepers in the digital age because the redundancy of channels "undermines the idea that there are discrete gates through which political information passes: if there are no gates, there can be no gatekeepers" (Williams and Carpini, 2000: 61). But this does not mean that gatekeeping theory is no longer useful. In debating the continued relevance of gatekeeping, we must cautiously distinguish between gatekeeping as a theoretical tradition and metaphor and remember that the metaphor is not set in stone, but serves as an interpretative tool (Heinderyckx, 2015). In contemporary gatekeeping literature, a broader interpretation of gatekeeping is often used, which is more in line with the original use of the term by Lewin (1947). Building on Lewin's field theory, Shoemaker and Vos (2009) demonstrated how gatekeeping is still a valuable lens for understanding how news is created and circulated today, both as a theory and as a metaphor.

Still, there is a need to "reimagine gatekeeping as a concept in the digital era" (Vos, 2015: 7), and several scholars have proposed alternative frameworks and metaphors to supplement or replace gatekeeping. Bruns (2005) argued that the gatekeeping concept does not adequately describe the work of many new participants in the news circulation process. Influential blogs and individuals on social media often do not *keep* gates of their own, but keep watch of existing gates to create a curated hub for their audience. Bruns (2005) conceptualizes this practice as "gatewatching" and argues that it is slowly replacing the traditional role for journalists. Alternatively, Singer (2014) argues that audiences have become important participants in the gatekeeping process and defines their role in relation to news media as a form of "secondary gatekeeping." Taking a broader focus, Thorson and Wells (2015) developed the "curation of flows" framework in which journalistic curation is positioned next to four sets of curating actors: "individual media consumers themselves; social others embedded in online and offline networks; strategic communicators; and algorithms designed to shape the discovery and presentation of content in many digital contexts" (p. 31). The focus on *curation* is a deliberate step away from *gatekeeping*, which in this framework more specifically refers to the "curation practices of journalistic organizations" (Thorson and Wells, 2015: 31).

To conceptualize and understand gatekeeping in the context of social networks, we can also build on social network theory, where gatekeeping in a network context has long

since been a field of study (Freeman, 1980). It is interesting to note that this gatekeeping tradition can also be traced back to Kurt Lewin (1947), in particular through the work on network centrality by Bavelas (1948), who was also a student of Lewin (Scott, 2011). Despite having these shared roots, the gatekeeping traditions in journalism research and network theory have evolved mostly as two separate branches. Where gatekeeping in journalism research has for a long time revolved mainly around unidirectional mass communication, gatekeeping in network theory focused on multi-directional interactions between many people. With the rise of news diffusion through social media, where mass communication merges with niche media and interpersonal communication, the social network literature has become more relevant for journalism.

For this study, the concept of network diffusion is of particular interest (Valente, 1995). For a news organization to reach many people on social media, it does not only matter how many people it can reach directly, but also whether these people themselves pass on the item, and the continuation of this diffusion process. This is related to secondary gatekeeping (Singer, 2014) and gatewatching (Bruns, 2005), but network theory contributes a more general framework for how diffusion through the interaction of many individuals works. A common metaphor for this process is that of a *virus*, which is typically passed on via dyadic ties between individuals, but can nevertheless diffuse rapidly throughout a population. Within this metaphor, the concept of *contagion* is used to describe the transfer of information or ideas between individuals (Lerman and Ghosh, 2010), which for news diffusion shares common ground with the concept of shareworthiness (Trilling et al., 2017). On social network sites, this virus-like propagation of content through connected individuals explains how news can quickly reach many people—popularly known as "going viral"—even without the use of mass communication (Klinger and Svensson, 2015). As the mechanism for contagion relies on user engagement with content, each individual in the network participates in the curation of flows (Thorson and Wells, 2015).

Thus, it follows that to determine the gatekeeping influence of social media editors, we cannot simply look at the size of their direct contacts (e.g. friends or followers on Facebook, followers on Twitter). Rather, we need to look at how their use of social media to publish their own news items (or links to these items) affects the overall diffusion of these items. For this, we need to have some way to measure the success of the diffusion (i.e. often being shared or otherwise interacted with) of these items and also need to be able to control for other factors—in particular other gatekeepers—that could affect the diffusion. In this study, we propose a method to investigate this, which we use to investigate to what extent newspapers can influence the diffusion of their news items on Facebook through their public Facebook pages.

Previous studies have used the popularity of news items on social media, for instance, to investigate what content characteristics are correlated with successful diffusion (Bastos, 2015; Trilling et al., 2017). Other studies investigated how news media use Facebook (Hille and Bakker, 2013), and at least one study also addressed a question similar to ours: whether newspapers' distribution of news items via Facebook (and Twitter) is an effective way to reach the audience (Ju et al., 2014). However, to our best knowledge, there are no studies that investigate how the diffusion of specific newspaper items on Facebook is influenced by publication of these items on the newspaper's
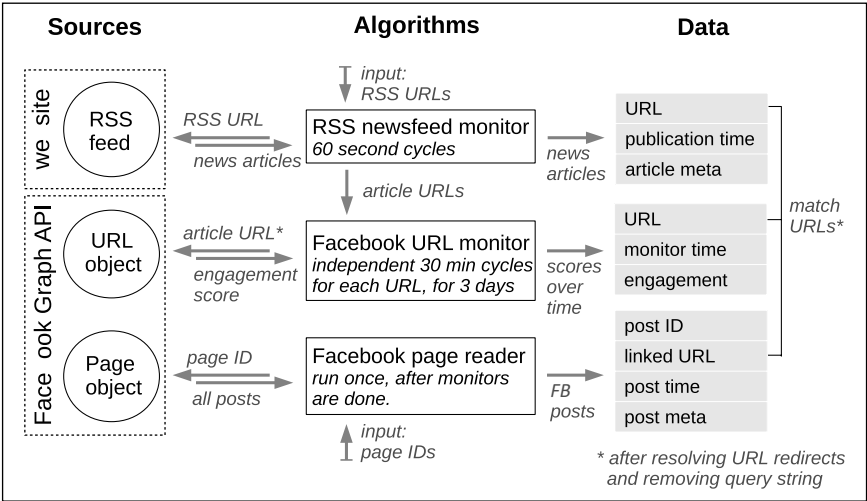
Figure 1. Conceptual illustration of data collection process.

Facebook page. Since it is very unlikely that there is no effect of the publication (i.e. zero gatekeeping influence) or that it explains all diffusion (i.e. full gatekeeping influence), these are not relevant hypotheses. We therefore dedicate this study to exploring and statistically estimating the influence of social media editors, posing the following research question:

> *RQ.* What proportion of the total user engagement with a news item on Facebook is the result of the publication of this item on the newspaper's own public Facebook page?

Data collection algorithms

To perform the analysis for this study, we needed to measure the engagement on Facebook for each news article that is published on a newspaper's website, with steady intervals starting from the time of publication. The news articles can be collected directly from the website, and the engagement score for a URL can be obtained through the Facebook Graph API. However, the engagement score for a URL can only be obtained for the current moment in time, meaning that to measure engagement over time the data have to be collected in real-time. Also, we needed to know all the posts the newspaper made on its Facebook page that link to a news article on their own website. To get this information, we wrote algorithms that monitored the newspaper website updates and Facebook engagement statistics in real-time.

The data collection process is summarized in the conceptual illustration presented in Figure 1. We used RSS to monitor the publication of new website articles. This is a format for web feeds that many news sites use, which enables people or applications to

receive a list with the most recent publications. We first manually made a list of the available RSS feeds for the newspapers in our study. Given this list as input, the *RSS newsfeed monitor* collected the RSS feeds with 60-second cycles. Whenever a unique new URL was monitored, it was written to a data file, together with the publication time and available article metadata, and put on a queue for the *Facebook URL monitor*. In this second monitor, the Facebook Graph API is used to get the engagement scores for a URL with 30-minute intervals, with the first time being 30 minutes after publication (the engagement score at the time of publication can only be zero). The cycles are independent, and parallelization was used to ensure that URLs with coinciding cycles are still monitored within only a few seconds of the intended monitoring time. Each cycle, the URL, the monitoring time, and the engagement score are written to a data file. This is repeated for each URL for 3 days. In total, the *RSS newsfeed monitor* ran for 2 weeks, and the *Facebook URL monitor* was allowed to run for 3 days more to finish the last URLs. Note that while it is possible to use shorter cycles for more accurate observations, these settings were used taking the strain on our server into account.

The last piece of information is whether the newspaper posted a link to the news article on its own Facebook page, and if so at what time. These data can relatively easily be obtained through the Facebook Graph API. We manually created a list of Facebook page IDs for the same newspapers as covered in the list of RSS feed URLs. Given this list, the *Facebook page reader* queries the Facebook Graph API to get all the posts from each page. The URLs that are linked to in each post are written to a data file, together with the time and ID of the post and additional post metadata.

Measuring diffusion on Facebook

To measure the diffusion of news articles on Facebook, we used the Facebook Graph API to get the *engagement* of article URLs. Facebook indexes all URLs that are posted and keeps a single count for URLs that point toward the same content. The engagement is based on the counts for reactions (i.e. like and the alternatives for like), comments, and shares. The counts for these indicators include those made by private users, and as such provide a rare insight into the total diffusion of news articles. These scores are also directly relevant for news diffusion on Facebook because they affect how likely people are to be exposed to these articles. Facebook does not reveal its formula for how the different indicators are balanced. When websites report engagement for their content, the sum of these scores is often used, which is how older versions of the Facebook Graph API report it. As this is measured consistently and includes all of Facebook, it is to our knowledge the best available indicator for news diffusion on Facebook. We therefore also use this score for this study.

When using these data, it is important to take two considerations into account regarding how Facebook deals with different URLs that point toward the same article. The first is whether and how engagement scores for these URLs are added together as a single score. The second consideration is how the Facebook Graph API matches URLs when we look up the engagement. This actually works in a different way, with more strict criteria, and if these are not taken into account, the retrieved engagement scores will not be

accurate. Based on the Facebook documentation for webmasters and some manual tests, we took the following measures.

We first looked into the behavior of URL redirection or URL forwarding. URL redirection means that a URL does not directly point to a web page, but instead redirects to a next URL. When a URL is posted on Facebook, the Facebook crawler (that also gets the title, publication date, picture, etc.) will resolve to a URL that is designated as the canonical URL, following a chain of multiple redirects if needed (Facebook, 2016). It is also possible to use a redirecting URL in a Graph API query, which can return the engagement score of the original. However, this only works if that specific redirecting URL has been posted at least once on Facebook—otherwise Facebook will not recognize it and return zero. For this study, we thus resolved all URLs.

Websites are often not sensitive to certain variations in how a URL is written (e.g. casing, duplicate forward slashes). This has similar implications as redirection: engagement for URLs with these variations is added up and attributed to the original URL, but querying the Graph API only works with the original URL or if the specific variation has been posted at least once. It is thus safest to always query the Graph API with the original or canonical version of the URL, and not to normalize these elements of URLs, for instance, by lowercasing them.

Third, URLs can have parameters in the form of a query string, which is typically included at the end of the URL, separated by a question mark. While traditionally used to pass arguments, query strings can be added to URLs as a way to monitor traffic using Google Analytics, and this is often used on social media. Regarding the counting of URL scores, the query string is simply ignored, meaning that all scores are attributed to the original URL. However, if a query string is included when querying the Graph API, it will never find a match, even if it has been posted before on Facebook. Accordingly, we always removed the query string before querying the Graph API and for matching the URLs from the RSS feeds and the Facebook posts.

Finally, there is the complicated scenario in which a website changes the actual URL and redirects the old URL to the new one. Whether Facebook recognizes the new URL as the same content, and thus carries over the URL scores, depends on whether the canonical URL is designated as specified in the guidelines for webmasters (Facebook, 2016). Since newspapers would want to carry over engagement scores, they tend to follow these guidelines, and the newspapers in our study do provide the required open graph meta information tags. Still, as addressed below, we encountered rare cases with suspicious changes or starting values in engagement, which are likely related to incorrect designation of canonical URLs, which we had to delete from the analysis.

Data

We used the RSS newsfeed monitor to collect all news items for 12 newspapers over a period of 2 weeks, from 28 July 2017 till 10 August. The engagement for each news item URL was monitored every 30 minutes for 3 days after publication.

The initial selection of newspapers consisted of four UK, four Flemish, and four Dutch national newspapers. The focus on Flanders (the Dutch-speaking part of Belgium) and Dutch newspapers is part of a larger research project about the role of social network

sites in professional news organizations in these countries. As described by Aelst et al. (2008), the countries have a similar media system and political history, but with notable differences, such as the more competitive market in the Netherlands (Van Aelst, 2007) and the stronger perceived influence of media organizations in Belgium (Aelst et al., 2008). The newspapers in both countries publish in Dutch and focus on a relatively small national market, with the Netherlands being a larger market than Flanders. We included the United Kingdom in this study to compare this to newspapers that publish in English, as a widely spoken international language, and that have many readers from both inside and outside of the United Kingdom. A thorough comparison of these countries is out of the scope of this study, but we expect that these differences can have a strong impact on the dynamics of engagement on Facebook, and therefore included them for a preliminary investigation.

The inclusion of 12 newspapers was deliberately broad because we anticipated that for some newspapers there would be either structural or incidental problems with the data collection. First, a portion of Facebook URLs could not be matched to the website URLs, which indicates that the RSS monitor missed some articles or that some articles were not published in the RSS feeds.[2] We deleted four newspapers from the analysis (*The Mirror, The Telegraph, Het Laatste Nieuws*, and *Nieuwsblad*) for which the RSS feed coverage was poorest (<80% coverage). For the remaining newspapers, we could match 94.8% of the Facebook URLs to the RSS feed, with the lowest coverage being for *De Standaard* (83.7%) and *De Volkskrant* (89.4%). We excluded articles for which no match could be found. This is unlikely to bias our findings because problems in the RSS monitor or RSS feed are not likely to be correlated to sharing rates on Facebook. Second, on 2 days during our monitoring period, certain cycles in the *Facebook URL monitor* failed. This happened randomly across all media, which suggests that the problem lies with server issues or Facebook Graph API rate limiting (i.e. too many calls to the API per time period). We deleted 208 news items (0.95%) for which less than 80% of the monitor cycles was observed. Third, for two newspapers, *De Volkskrant* and *NRC Handelsblad*, the publication time on the RSS feed often did not include the exact publication time, due to which they could not be used for the analyses. Finally, there are rare cases (0.44%) for which the monitored engagement over time suddenly reset to zero or started immediately with a score that was improbably high. We could not find a clear pattern in the cases where this happens, but suspect that it is related to incorrect specification of the canonical URL. We excluded these cases from the analysis.

For the six newspapers that remained, we collected the Facebook posts from their public page, using the Facebook Graph API as described above. Three of the newspapers in our study also have *sub-organizational* pages that focus on more specific types of content. For *De Telegraaf*, this is only one page for financial news with very few followers, *The Guardian* has multiple smaller pages, and *The Sun* has one page for football news that is very popular. We did not include these pages in our analysis because we cannot group them together due to the vast differences in the number of followers, but this does impose limitations for some of our findings for these newspapers.

The number of articles after cleaning our data is presented in Table 1 in the N articles columns. Each of these articles has a time series with engagement scores for 30-minute intervals over a period of 3 days. For reference, the table includes the number of

Table 1. Newspaper publication and Facebook statistics during period of investigation.

| Medium	N articles		FB statistics	
	Website	FB	Followers	FB posts
UK newspapers				
The Guardian	2915	547	7,378,457	1039
The Sun	7202	1114	2,709,146	2148
Flemish newspapers				
De Morgen	3386	201	203,349	391
De Standaard	1267	253	251,558	586
Dutch newspapers				
Trouw	568	101	43,292	194
De Telegraaf	3931	228	376,107	516

FB: Facebook.

followers each newspaper has on Facebook and the total number of Facebook posts. Note the stark difference in the number of followers, with the UK newspapers clearly in the lead.

A notable observation in the data was that there are strong differences between the countries regarding the time between the original publication on the website and the publication on the Facebook page. In Figure 2, we see for every hour the proportion of Facebook posts with news items for the time between the website and Facebook publication. In particular, we see that in the Netherlands the vast majority of times that a news item was published on Facebook, this occurred shortly after the website publication— more than 50% within 1 hour. In contrast, in the United Kingdom, we see a much longer tail, with many posts about news items many hours after the website publication, sometimes more than a day. Often these are types of news items with less temporal urgency, such as opinion pieces.

Results

Table 2 presents the amount of engagement per newspaper and shows the proportions for three conditions. The first condition is *website only*, which refers to the news articles that the newspaper did not publish on their Facebook page. Accordingly, this engagement is not the result of the newspaper's main Facebook page, but stems from individuals who shared the articles directly from the website or from other Facebook pages such as interest groups, nongovernmental organizations (NGOs), or sub-organizational pages of the newspaper (only for *The Sun, The Guardian*, and *De Telegraaf*). The engagement scores for articles that the newspaper did publish on Facebook are split into two columns: the the engagement score before (*before FB*) and after (*after FB*) the Facebook publication.

The *after FB* score is the upper limit of the engagement that resulted from the newspaper's Facebook page, for which the percentages vary around 53.2%. Two newspapers for which the *after FB* percentage was clearly lower than the rest were *The Sun* (38.6%)
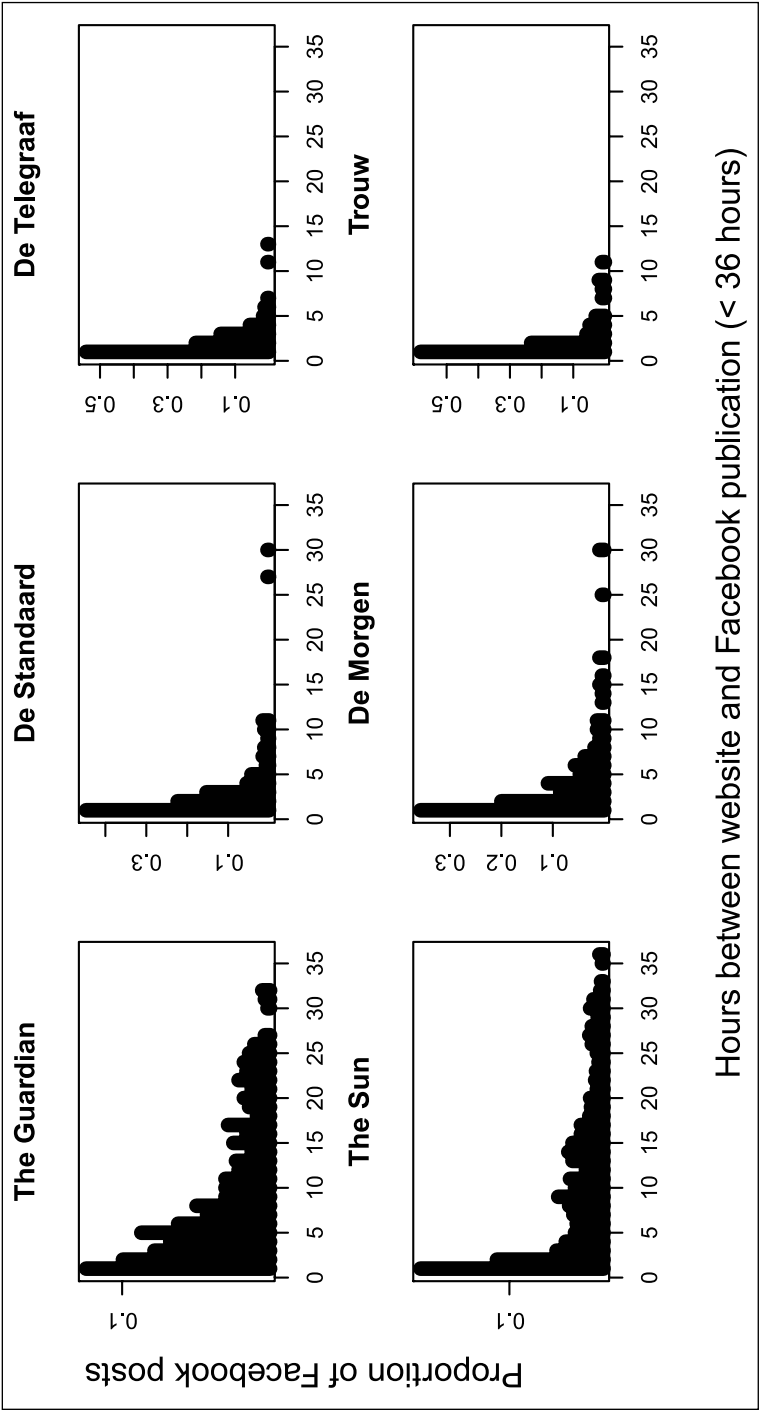
Figure 2. Time in hours (x-axis) between website and Facebook publication.

Table 2. Facebook engagement for articles without, and before and after publication on Facebook page.

Medium	Total	Website only[a]		Before FB[b]		After FB[b]	
	N	N	%	N	%	N	%
UK newspapers							
The Guardian	4,555,934	1,898,364	41.7	318,001	7.0	2,339,569	51.4
The Sun	3,183,960	1,914,418	60.1	40,238	1.3	1,229,304	38.6
Flemish newspapers							
De Morgen	151,293	68,325	45.2	1242	0.8	81,726	54.0
De Standaard	183,918	63,567	34.6	2350	1.3	118,001	64.2
Dutch newspapers							
Trouw	78,337	23,848	30.4	886	1.1	53,603	68.4
De Telegraaf	452,368	253,218	56.0	5348	1.2	193,802	42.8

[a]Articles that the newspaper itself did not publish on Facebook.
[b]Engagement before and after the newspaper published an article on Facebook.

and *De Telegraaf* (42.8%). The highest *after FB* percentages were found for *De Standaard* (64.2%) and *Trouw* (68.4%). These results signify that newspapers are far from being the only gatekeepers on Facebook regarding the diffusion of their own news articles.

Still, given that only a small portion of the news items is published on Facebook, these numbers indicate that these items have substantially higher engagement scores. To find out whether this is related to the publication on the Facebook page, two factors need to be taken into account. First, not all engagement that follows after the newspaper's Facebook publication can be attributed to the effect of this publication. Second, it is likely that social media editors take the shareworthiness (Trilling et al., 2017) of news articles into account, so these articles might already have had a greater likelihood to be successful based on their content.

To control for these factors insofar as possible, we used time series of the engagement scores for news items over time. Thus, instead of comparing the engagement scores *between* articles—which is likely to be strongly affected by the content of articles—we analyze changes in engagement scores over time for each article. This gives us much more variance to model and allows us to measure the effect of the publication of a news item on the newspaper's Facebook page on engagement *within* articles. A visualization of the time series for six articles is presented in Figure 3. The graphs show the increase in engagement during every 30-minute monitor cycle. The big tick marks on the *x*-axis indicate the time at which the newspaper published the article on Facebook. In addition to the observed values, we present the fitted values for our model, as addressed shortly.

The presented cases are from *The Guardian* and illustrate a range of scenarios. The first scenario is the *immediate Facebook publication*, in which the article was published on Facebook shortly after the website publication. Here, we see a strong initial peak and a smooth decay. Although in this scenario we have little information about how popular the article would have been without the Facebook publication, the sudden burst of engagement would have been unlikely without the involvement of a powerful
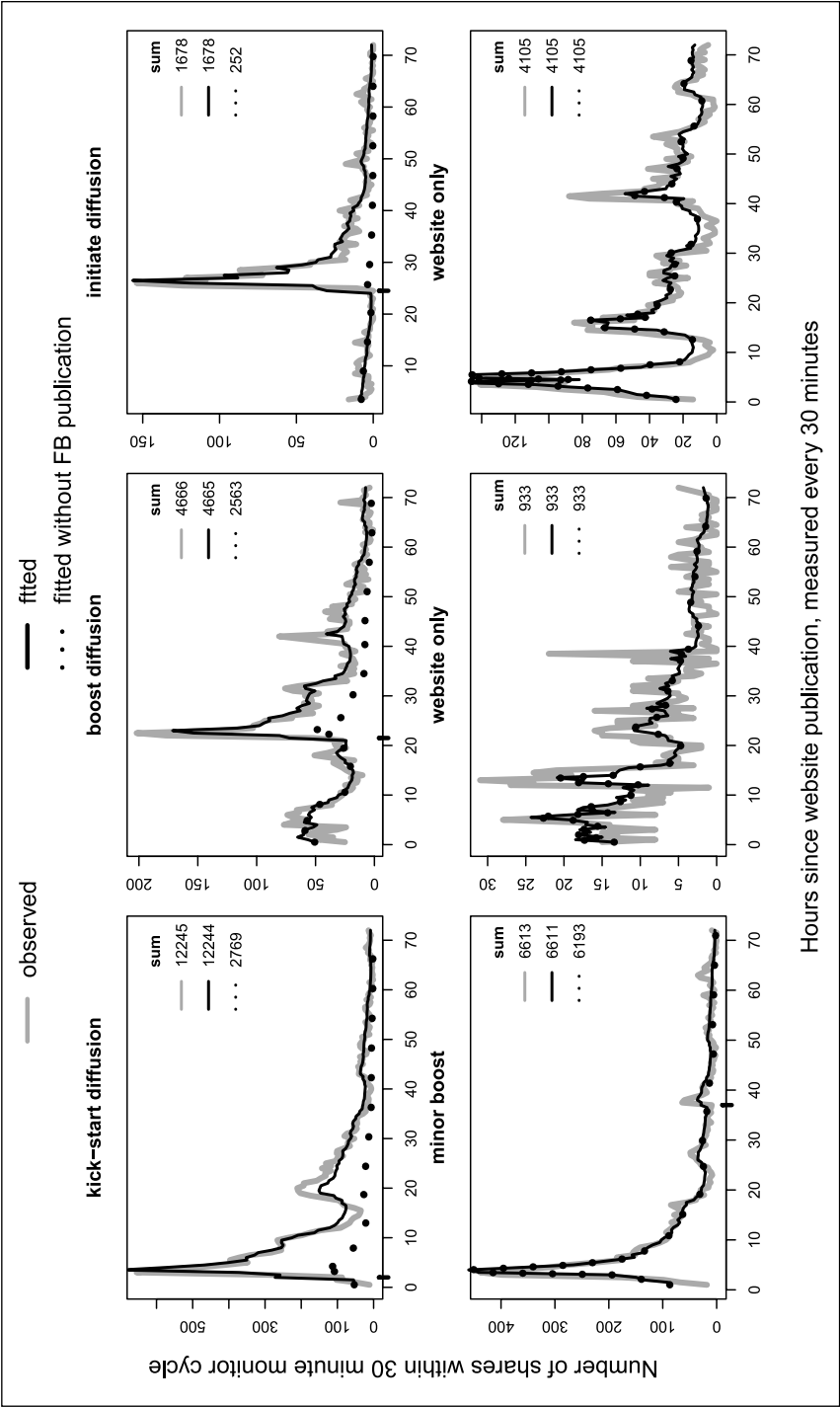
Figure 3. Facebook engagement over time for six articles from *The Guardian*.

gatekeeper. In the second scenario, *boost diffusion*, the article was already popular before the Facebook publication, but the publication clearly sparks a second wave of engagement. The third scenario, *initiate diffusion*, shows how an article that initially received little attention is launched by the Facebook publication. Overall, these first three scenarios show that social media editors of newspapers can indeed be powerful gatekeepers by selecting which articles are given that extra push.

In the fourth scenario, *minor boost*, we see that this is not always the case. Again, we see a stark initial rise, but this was before the Facebook publication. Around 38 hours later, the newspaper did publish the article on Facebook—possibly trying to boost it—but it did not get any traction. The fifth scenario, *website only*, is presented in two cases. In both cases, we see a gradually declining increase in the engagement score, but no involvement of the newspaper was required to trigger the diffusion.

At face value, the graphs in Figure 3 convincingly demonstrate that there are cases where the newspaper's Facebook publication had a causal effect on article engagement. The next step is to try to formalize this with a statistical model and apply it to all cases to get a quantitative estimate of this causal effect. For this causal inference, we need to get an estimate of the *counterfactual* of the Facebook publication. In other words, we need to estimate how the engagement score of an article would have developed if the Facebook publication had not occurred. For time series, we can use information from the time series prior to the condition and comparable time series to estimate how the time series would have continued without the condition (Brodersen et al., 2015). In our case, we estimate the counterfactual of the Facebook publication by using information about the time series from observations without or before the Facebook publication.

To model this, we used a multilevel generalized linear model with a Poisson link function. The dependent variable is the increase in the engagement score during a 30-minute cycle, which is a count variable that resembles a Poisson distribution. The independent variables are the time since publication, the autoregression, and a dummy variable that indicates whether the article has been published on the newspaper's Facebook page. To account for differences between URLs, we added random intercepts and slopes for all three independent variables. Furthermore, there is seasonality in the time series due to the time of the day, especially because there is much less engagement during the night (these are the u shapes in the decay tails in Figure 3). To control for this seasonality, we added random intercepts for the hour of the day. The results are presented in Table 3. The random effects are only controlled for and not presented in the model.

The models show that the engagement per cycle decays over time (negative effect), meaning that in general people engage more with the latest news, as one would expect. Also, there is significant, positive autoregression, which indicates that the amount of engagement is more likely to be high if engagement was high during the previous observation. This makes sense, because high engagement means more likelihood of exposure for more people. Overall, these effects indicate that engagement comes in waves that die out over time, as we also saw in the time series graphs (Figure 3). The positive, significant effect of the independent variable *Facebook publication* indicates that one of the triggers for a wave is when a news item is published on the newspaper's own Facebook page.

Table 3. Multilevel Poisson regression predicting the Facebook engagement score for a news item over time.

	Dependent variable:					
	Engagement (difference)					
	The Guardian	The Sun	De Morgen	De Standaard	Trouw	De Telegraaf
	(1)	(2)	(3)	(4)	(5)	(6)
Time	−3.561***	−6.139***	−4.943***	−5.171***	−2.102***	−7.507***
	(0.064)	(0.148)	(0.326)	(0.245)	(0.352)	(0.201)
Facebook publication	0.633***	1.480***	1.387***	1.677***	1.938**	0.833***
	(0.047)	(0.080)	(0.293)	(0.174)	(0.704)	(0.133)
Autoregression	167.998***	146.717***	16.596***	41.212***	13.132	57.827***
	(4.976)	(7.157)	(1.995)	(4.599)	(10.210)	(4.543)
Constant	2.881***	2.459***	1.347***	1.562***	1.054**	2.640***
	(0.050)	(0.107)	(0.311)	(0.272)	(0.336)	(0.213)
Observations	183,582	193,939	9981	10,560	5524	29,262
Log likelihood	−890,265	−921,418	−31,633	−30,801	−19,084	−84,803
Akaike IC	1,780,560	1,842,867	63,297	61,632	38,198	169,635
Bayesian IC	1,780,712	1,843,020	63,405	61,741	38,297	169,759

IC: information criterion.
Levels: news item (intercepts and slopes) × hour (intercepts).
**$p < .01$; ***$p < 0.001$.

We can now use the model to estimate how engagement would have developed without the Facebook publication. This is illustrated in Figure 3, where the black line shows the prediction of the model with the actual data and the dotted line shows the prediction if we set the dummy variable for Facebook publication to zero, as an estimate of the counterfactual. Sum scores for each line are added to show the total difference. An advantage of the way these predictions are visualized is that we can see whether the way that the counterfactual is estimated matches our own intuition. In our interpretation, this is clearly seen in the boost diffusion, initiate diffusion, and minor boost scenarios. Note that in the website-only scenarios the counterfactual is identical to the observed sum score, which indicates (correctly) that there was no effect of a Facebook publication.

By subtracting the counterfactual from the predicted value, we can now get an estimate of the engagement that was caused by the Facebook publication. The percentages that follow from this calculation are presented in Figure 4. Overall, we see that scores are all below 60%, based on which we can conclude that a substantial portion of the engagement for news items is not the direct result of the newspaper's own social media editing activities. More than 40% is the result of alternative channels, such as individual users who share news items directly from the website or more influential gatekeepers such as Facebook pages of interest groups.
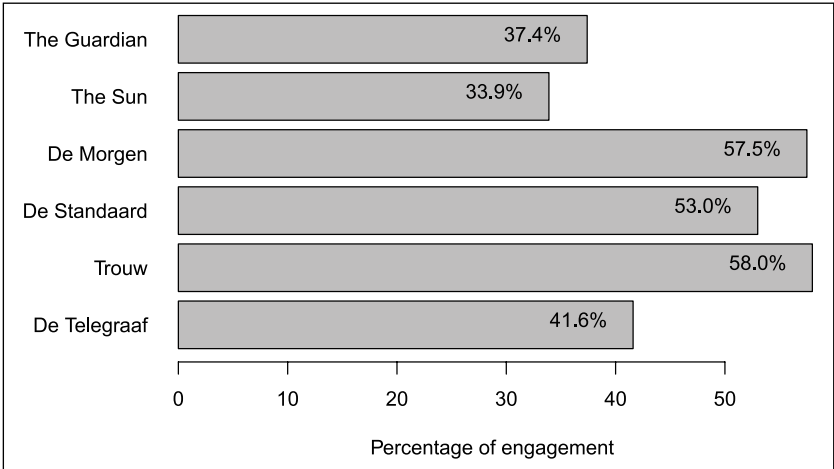
Figure 4. Estimated percentage of engagement attributed to the publication on the newspaper's own Facebook page.

For three out of six newspapers in our study, *The Guardian, The Sun*, and *De Telegraaf*, the percentage was only 41.6% or lower. One reason for these relatively low percentages is that these are the only newspapers in our sample that also have sub-organizational Facebook pages. As such, certain categories of news might be mostly published on Facebook through these pages. Furthermore, these were also the three newspapers with the largest audiences. Following network diffusion theory, it is possible that the diffusion of news items from these newspapers through the alternative channel is quicker to reach a tipping point. This indicates that popular news outlets could be less dependent on their own social media editing activities for their success on social media.

Conclusion

The main purpose of this study was to form a starting point for measuring and conceptualizing the gatekeeping influence of social media editors for professional news organizations. Given the limitation that data on individual user engagement with news items are generally not available due to privacy considerations, we leveraged the use of aggregated engagement statistics. By monitoring these statistics over time, we can use time series analysis to investigate whether and when news items diffuse rapidly on social media and whether this can be predicted by certain events. We used this to investigate to what extent the diffusion of newspapers items on Facebook is affected by whether and when the newspaper posted this item on their own Facebook page. In other words, to what extent social media editors have influence, as gatekeepers, on which of their news items become successful on Facebook.

Our analysis shows that newspaper's Facebook pages can indeed have a strong influence on the diffusion of news items on Facebook. We found three common scenarios. First, there are cases where newspapers post a news item on Facebook shortly after it was

published on their website and thereby kick-started an immediate spike of engagement. Second, there are cases where the news item was posted on Facebook with some delay, after it already received moderate engagement, and boost a second wave of engagement. Third, there are cases where the news item was posted on Facebook with some delay, after it received almost no engagement, in which case it could still initiate a wave of engagement. Based on our analysis, we estimate that for different newspapers, between 33.9% and 58% of engagement could be attributed to the publications on the newspaper's Facebook pages.

This also implies that a large portion of the engagement for newspaper items has to be attributed to *alternative channels*. We found that, in particular, there are many news items that were not at all published on the newspaper's own Facebook page, but nevertheless received much engagement. There are also cases where the newspaper did post the news item, but only after the item already became popular. At the very least, this shows that a large portion of the news that circulates on Facebook arrived there through channels other than the newspaper itself. It remains an open question what these channels are. Looking at the time series for these cases, we saw that there are often sudden bursts of engagement, which suggests that there are other influential gatekeepers that can trigger a sudden, large exposure.

Overall, these results reveal a duality in the gatekeeping influence of the newspapers' Facebook pages. On the one hand, they are indeed in a key position to influence news diffusion: their endeavors benefit the total diffusion of news items, and their specific selection of news items is more successful. On the other hand, the newspaper's news items are also shared and engaged with independently from them, which shows that they are indeed not gatekeepers in the traditional sense, where they control discrete gates and decide which items are *in* or *out*.

This interplay between mass communication and rapid interpersonal diffusion requires us to rethink how gatekeeping should be defined and to find new ways to measure who the powerful gatekeepers are. In a context where news diffuses through networks of curating actors, gatekeeping has become a relative concept, where a user's probability to encounter certain content is a function of both mass communication by news providers and the interest in this content (i.e. engagement) from people in the user's social circle. While the redundancy of channels makes it virtually impossible to keep content hidden behind gates, we can define powerful gatekeepers by the ability to push the diffusion of content beyond a tipping point, creating waves where otherwise there would have been mere ripples. Our results show that social media editors are indeed powerful gatekeepers according to this definition. This power correlates with the ability to mass communicate, but the leading role of diffusion mechanics adds new, network-level factors, such as the structure of a network and the algorithms that govern its communication flows. A prime example is the recent announcement by Facebook, on 12 January 2018, that it will change its algorithm to favor interactions between individual users by downplaying the visibility of posts from businesses, brands, and media. Such changes can in an instant reconfigure the balance of gatekeeping power. This form of influence exerted by Facebook could be conceptualized as gatekeeping on a meta level, where instead of controlling communication flows by operating gates, the power of gates is reconfigured by controlling the rules of the communication channels.

In this study, we proposed a new method to measure gatekeeping power in the context of a social network site, but our analysis has several limitations, of which some could be addressed in future research. First, while we can model whether publication on the Facebook page causes a wave, it is difficult to determine whether subsequent waves are still the result of this publication—they might or might not be caused by other gatekeepers that act independently. Second, we cannot account for a possible correlation between the newspaper's selection of items and the inherent shareworthiness of these items. This can to some extent be addressed by including content characteristics that are related to shareworthiness (Trilling et al., 2017). Third, three out of the six newspapers in our analysis (*The Guardian, The Sun*, and *De Telegraaf*) also have sub-organizational pages, such as *guardianscience*, that also publish news items. More generally, for all newspapers in our analysis, there can be other gatekeepers on social media that are associated with the newspaper organization, such as individual journalists. Thus, our analysis underestimates the full influence of newspaper organizations on Facebook. A more in-depth investigation that includes these organizational branches is out of the scope of this study, but our method could be used to investigate this. Finally, note that the estimates presented here should not be interpreted as the percentage of engagement that would be lost if the newspaper decides to take down its Facebook page. It is an estimate of the proportion for which the news passed through the channel of the newspaper's Facebook page under the current conditions and as such serves to indicate the impact of the gatekeeping choices of the editors who operate this page.

Although the use of aggregate engagement scores is only a proxy for news diffusion, it is arguably the closest we can get to measuring it without violating the privacy of individual users. Furthermore, by monitoring these scores over time as time series data, we can get a good indication of the life cycle of news items on social media. In this study, we used this method to investigate the gatekeeping influence of newspapers' Facebook pages, but it can be used more broadly. First is to explore whether there are other important gatekeepers on Facebook, which can be analyzed by including publications on other public pages in the model. Second is to investigate whether the effect of publication on the newspaper's own Facebook page is affected by content characteristics of news items. Third, if the page views of news items on the website are also obtained, which requires collaboration with newspapers, the effect of engagement on readership could also be analyzed. The method is applicable for any social media platform for which it is possible to search for a URL or to obtain engagement scores (see, for example, Bastos, 2015). We thus conclude that analyzing time series of aggregate engagement scores can be a fruitful approach to study gatekeeping and news diffusion on social media. Our computer scripts, written in Python, are published open-source on GitHub.[3]

Funding

Notes

1. Note that on some platforms, one can also receive posts from indirect connections. It is more accurate to say that the closer someone is to someone else in the network, the more likely he or she is to be exposed to other's posts.

2. RSS feeds only show recent publications, so with many publications at once, or if our server or the newspaper website is delayed, some articles can be missed.
3. https://github.com/kasperwelbers/url_engagement_monitor

References

Aelst PV, Brants K, Praag PV, et al. (2008) The fourth estate as superpower? *Journalism Studies* 9(4): 494–511.

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.

Bass AZ (1969) Refining the "gatekeeper" concept: a UN radio case study. *Journalism & Mass Communication Quarterly* 46(1): 69–72.

Bastos MT (2015) Shares, pins, and tweets: news readership from daily papers to social media. *Journalism Studies* 16(3): 305–325.

Baum MA and Groeling T (2008) New media and the polarization of American political discourse. *Political Communication* 25(4): 345–365.

Bavelas A (1948) A mathematical model for group structures. *Human Organization* 7(3): 16–30.

Brodersen KH, Gallusser F, Koehler J, et al. (2015) Inferring causal impact using Bayesian structural time-series models. *The Annals of Applied Statistics* 9(1): 247–274.

Bruns A (2005) Gatewatching: Collaborative Online News Production (Digital formations). New York: *Peter Lang*.

Cook T (2005) *Governing with the News: The News Media as a Political Institution.* 2nd ed. Chicago, IL: The University of Chicago Press.

Crouse T (1972) *The Boys on the Bus: Riding with the Campaign Press Corps*. New York: Random House.

Facebook (2016) *A guide to sharing for webmasters*. Available at: https://developers.facebook.com/docs/sharing/webmasters (accessed 21 June 2017).

Freeman LC (1980) The gatekeeper, pair-dependency and structural centrality. *Quality & Quantity* 14(4): 585–592.

Gerlitz C and Helmond A (2013) The like economy: social buttons and the data-intensive web. *New Media & Society* 15(8): 1348–1365.

Goode L (2009) Social news, citizen journalism and democracy. *New Media & Society* 11(8): 1287–1305.

Gottfried J and Shearer E (2016) News use across social media platforms 2016. *Pew Research Center*. Available at: http://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/

Hamilton J (2004) *All the News That's Fit to Sell: How the Market Transforms Information into News.* Princeton, NJ: Princeton University Press.

Heinderyckx F (2015) Gatekeeping theory redux. In: Vos TP and Heinderyckx F (eds) *Gatekeeping in Transition*. London: Routledge, pp. 253–267.

Hermida A, Fletcher F, Korell D, et al. (2012) Share, like, recommend: decoding the social media news consumer. *Journalism Studies* 13(5–6): 815–824.

Hille S and Bakker P (2013) I like news: searching for the "holy grail" of social media: the use of Facebook by Dutch news media and their audiences. *European Journal of Communication* 28(6): 663–680.

Ju A, Jeong SH and Chyi HI (2014) Will social media save newspapers? Examining the effectiveness of Facebook and Twitter as news platforms. *Journalism Practice* 8(1): 1–17.

Klinger U and Svensson J (2015) The emergence of network media logic in political communication: a theoretical approach. *New Media & Society* 17(8): 1241–1257.

Lerman K and Ghosh R (2010) Information contagion: an empirical study of the spread of news on Digg and Twitter social networks. *ICWSM* 10: 90–97.

Lewin K (1947) Frontiers in group dynamics II: channels of group life; social planning and action research. *Human Relations* 1(2): 143–153.

Newman N, Fletcher R, Levy D, et al. (2016) The Reuters institute digital news report 2016. Available at: http://reutersinstitute.politics.ox.ac.uk/sites/default/files/research/files/Digital%2520News%2520Report%25202016.pdf

Ryfe DM (2012) *Can Journalism Survive: An inside Look at American Newsrooms*. Cambridge: Polity.

Scott J (2011) Social network analysis: developments, advances, and prospects. *Social Network Analysis and Mining* 1(1): 21–26.

Shoemaker PJ and Reese SD (1996) *Mediating the Message: Theories of Influence on Mass Media Content.* 2nd ed. Harlow: Longman.

Shoemaker PJ and Vos T (2009) *Gatekeeping Theory*. London: Routledge.

Singer JB (2014) User-generated visibility: secondary gatekeeping in a shared media space. *New Media & Society* 16(1): 55–73.

Thorson K and Wells C (2015) Curated flows: a framework for mapping media exposure in the digital age. *Communication Theory* 26(3): 309–328.

Trilling D, Tolochko P and Burscher B (2017) From newsworthiness to shareworthiness: how to predict news sharing based on article characteristics. *Journalism & Mass Communication Quarterly* 94(1): 38–60.

Valente TW (1995) *Network Models of the Diffusion of Innovations*. Cresskill, NJ: Hampton Press.

Van Aelst P (2007) *Toeschouwer, speler of scheidsrechter? Een studie naar de rol van de media in de verkiezingscampagne van 2003*. Brugge: Vanden Broele.

Vliegenthart R and Walgrave S (2008) The contingency of intermedia agenda-setting: a longitudinal study in Belgium. *Journalism & Mass Communication Quarterly* 85(4): 860–877.

Vos TP (2015) Revisiting gatekeeping theory during a time of transition. In: Vos TP and Heinderyckx F (eds) *Gatekeeping in Transition*. London: Routledge, pp. 3–24.

Welbers K, Van Atteveldt W, Kleinnijenhuis J, et al. (2018) A gatekeeper among gatekeepers: news agency influence in print and online newspapers in the netherlands. *Journalism Studies* 19(3): 315–333.

White DM (1950) The gatekeeper: a case study in the selection of news. *Journalism Quarterly* 27(4): 383–390.

Williams B and Carpini M (2000) Unchained reaction the collapse of media gatekeeping and the Clinton–Lewinsky scandal. *Journalism* 1(1): 61–85.

Author biographies

Kasper Welbers (PhD) is a postdoctoral researcher at the VU University Amsterdam, with an additional position at KU Leuven. His research focuses on changes in the news diffusion process due to the proliferation of new media technologies. In particular, he uses computational text analysis to trace how news messages diffuse trough networks of old and new media channels.

Michaël Opgenhaffen (PhD) is a professor of New Media & Journalism at the KU Leuven and Lecturer at the Leiden University. His research interests include new/social/digital media, online journalism, information processing, and journalistic competencies. He is director of studies of the Master Journalism KU Leuven at Campus Antwerp.