

# News Organizations' Selective Link Sharing as Gatekeeping

*A Structural Topic Model Approach*

Chankyung Pak

CCR 1 (1): 45–78

DOI: 10.5117/CCR2019.1.003.PAK

## Abstract

To disseminate their stories efficiently via social media, news organizations make decisions that resemble traditional editorial decisions. However, the decisions for social media may deviate from traditional ones because they are often made outside the newsroom and guided by audience metrics. This study focuses on *selective link sharing* as quasi-gatekeeping on Twitter – conditioning a link sharing decision about news content. It illustrates how selective link sharing resembles and deviates from gatekeeping for the publication of news stories. Using a computational data collection method and a machine learning technique called Structural Topic Model (STM), this study shows that selective link sharing generates a different topic distribution between news websites and Twitter and thus significantly revokes the specialty of news organizations. This finding implies that emergent logic, which governs news organizations' decisions for social media, can undermine the provision of diverse news.

**Keywords:** selective link sharing, gatekeeping, Structural Topic Model

## 1. Introduction

Many actors who control the news flow in the digital era (Barzilai-Nahon, 2008) have made the privileged position of news organizations as a gatekeeper (Shoemaker & Vos, 2009) and even the relevance of the metaphor of gatekeeper questionable (Thorson & Wells, 2015). This theoretical

withdrawal is not merely an academic word game. When the *New York Times* eliminated the Public Editor position, Arthur Sulzberger Jr., a former publisher of the *New York Times* acknowledged that news organizations have no role in filtering information that emerges from social media users (Jackson, 2018). However, recently reported evidence suggests that news organizations still occupy a differential position in the propagation of news via social media. A research team at the University of Washington discovered that, despite the distributed nature of the propagation of misinformation on social media (Lazer et al., 2018), the official Twitter accounts of news organizations help stop the spread of misinformation (Andrews, Fichet, Ding, Spiro, & Starbird, 2016). Explicitly examining the relevance of the gatekeeper concept for news organizations, Welbers and Opgenhaffen (2018b) showed that the Facebook posts of news organizations are a substantial determinant of the news engagement of online readers.

The acknowledgment of the gatekeeper role of news organizations on social media does not necessarily mean that the traditional mechanism by which gatekeeping decisions were made simply extends to the new news dissemination platform. Although the modality and functionality of news organizations' decisions for social media closely resemble their traditional editorial decisions, the newsroom or the desk are often involved in a limited way with such decisions (Elizabeth, 2017). Further, those decisions are driven by audience metrics rather than the normative evaluation of newsworthiness (Tandoc, 2016). In sum, news organizations are now making extra decisions that potentially have a great impact on news reading, but they are likely to be governed by a logic that is different from traditional norms of journalism.

In this paper, I label the news organization's decisions for social media that resemble the traditional journalistic practices as *quasi-editorial decisions*. Among them, I specifically investigate *selective link sharing* – conditioning a link sharing decision on news content, as a new form of gatekeeping (Welbers & Opgenhaffen, 2018b). This is important because if the new gatekeeping deviates from the traditional pattern, it would be reasonable to question whether the new logic behind the decisions is reliable enough to expect socially desirable provision of news. However, because news organizations produce a prohibitively large number of news stories and social media posts everyday, the random sampling and the manual coding for analysis are nearly infeasible. This study presents computational data collection and machine learning techniques that can overcome such a challenge. The suggested tools can further provide infrastructure for constant monitoring of the emergent social media editing of news organizations, which is still in flux.

The results from a Structural Topic Model (STM) (Roberts, Stewart, & Airoldi, 2016) applied to 280 thousand news stories and 130 thousand tweets indicate that the deviation between gatekeeping for publication and link sharing is indeed visible for several news topics. Further, a comparison of selective link sharing across different media types shows that topic selection differs, depending on a given topic's popularity on Twitter and a news organization's specialty in the topic. The dominant momentum from selective link sharing tends to suppress different specialties of news organizations, which either homogenize news topic distribution on Twitter or make organizations invisible from their non-specialized topics, rather than focusing on their specialized topics. This finding warns that news dissemination via social media may un-diversify the provision of news information, which calls for further public monitoring.

## 2. Social Media as a News Distribution Platform

Although there is a broad consensus that social media are fundamentally transforming journalism, what they afford news organizations is still debatable. In previous years, media scholars expected that the Web would afford news organizations a channel for the mutual interaction between media and audiences (Chan-Olmsted & Park, 2000), and be an effective promotion tool to attract younger audiences who do not regularly access traditional media (Chan-Olmsted, Rim, & Zerba, 2013; Palser, 2009).

However, evidence that news organizations are using these opportunities is rather weak. For example, Greer and Ferguson (2011) analyzed tweets from 488 local TV stations in the U.S., and found that only 23.3% of 455 commercial TV stations tweet for interaction with news readers. Similarly, Meyer and Tang (2015) analyzed tweets from 60 local news organizations, and only 7.4% of the tweets from local television stations and 11.6% from local newspapers were intended for interaction. Cleary, al Nashmi, Bloom, and North (2015) reported a similar finding from tweets from the CNN International Channel. These studies also generally conclude that traditional news companies are not engaged in promotion for either their websites, or the organizations' brands (Greer & Ferguson, 2011; Meyer & Tang, 2015). Armstrong and Gao (2010) concluded that traditional news organizations generally extend their conventions to Twitter rather than adopting strategies that are customized to social media.

Traditional news organizations use social media as an additional news distribution platform. Greer and Ferguson (2011) found that 94.9% of the

commercial TV stations tweet to disseminate their news articles whereas only 17.6% tweet to promote their programs. Similarly, Meyer and Tang (2015) reported that 94.4% of the tweets from TV stations and 96.3% from newspapers are for news link sharing. Further, news dissemination through social media is often a profitable tactic although it also depends on how the platform companies adjust and redesign their content curation algorithms (Rashidian et al., 2018). Hong (2012) found that social media use and the number of tweets by news organizations induced more traffic toward their news websites. According to Newman (2011), in 2011, the BBC switched from an automatic feed to manual choice and editing optimized for Twitter, and the number of followers has doubled since then. Recognizing this potential, Facebook published a guideline for strategic news posts based on its own user news engagement studies in the same year.<sup>1</sup> Therefore, news link sharing has become the main concern when news companies contemplate their social media use.

Indeed, it is well understood that news organizations strive to increase Web traffic from their social media posts, and boost virality on the platforms. Given the decline in the traditional ways of disseminating news, digital advertising revenue in the U.S. market reached 31% in 2017. This was an increase from 17% in 2011 (Barthel, 2018). Furthermore, the majority of digital revenue comes from only a few platform companies, such as Facebook, Google, and Twitter (Barthel, 2018). Although better-resourced news organizations, such as the *New York Times* and the *Wall Street Journal*, try to decouple from the giants and diversify social media platforms through which they disseminate their news links, smaller outlets, which take up a large proportion of news organizations, still focus on a few representative platforms (Rashidian et al., 2018). Thus, news organizations adopt different kinds of social media strategies to stand out on social media. As Tandoc (2016) illustrated, social media editors and journalists widely adopt audience metrics to quickly determine which news stories to expose to social media users and how (Roston, 2015; Welbers, Van Atteveldt, Kleinnijenhuis, Ruigrok, & Schaper, 2016). This practice results in changes in what news readers will see on social media. For example, García-Perdomo, Salaverría, Kilgo, and Harlow (2018) showed that odd topics are prevalent on news organizations' Twitter accounts, and Welbers & Opagenhaffen (2018a) also found that news paraphrases to which Twitter users are exposed have writing styles that differ from traditional news writing.

Social media editing is often disconnected from traditional news editing in many news organizations although the degree of separation differs

across organizations. Recent interviews and ethnographic works illustrate that two jobs are organized as separate units in many organizations (Elizabeth, 2017; Roston, 2015; Tandoc, 2016). Further, using social media as a news distribution platform and audience metrics for gatekeeping call for renegotiating journalists' job as 'marketers' (Tandoc & Vos, 2016). Although analyzing big data often entails open source culture, journalists trained for traditional newsroom culture tend to struggle to adopt such a new culture. (Lewis, 2012; Lewis & Westlund, 2015). This heterogeneity within a news organization that has been brought in by information technologies also creates tension between different types of employees; social media editors feel that editors in the newsroom are locked in their own routines or 'comfort zone' (Elizabeth, 2017). This discrepancy explains the gap between journalists' skepticism about news dissemination via social media and its widespread practice (Rashidian et al., 2018; Welbers et al., 2016).

### 3. News Link Sharing on Social Media as Gatekeeping

The decisions about what to share also imply its flip side decisions about what *not* to share (García-Perdomo et al., 2018). Indeed, as the recent *TOW Center's* analysis about how news organizations upload posts on different social media platforms (Rashidian et al., 2018) shows that they share far fewer than all the news stories they publish. This is naturally explained by an economic theory proposed by Anderson and De Palma (2012). In the face of competition for the limited attention (of social media users), relative to the tremendous amount of information on social media, simply pouring out more information is likely to cannibalize their other news links, which might lead to less overall attention drawn to their website. (Anderson & De Palma, 2013). Thus, under competition for limited attention, certain types of news stories are likely to be 'prioritized' over others (Kalsnes & Larsson, 2018). As a result of the prioritization, social media editors practically control the accessibility to news information, which has traditionally been an exclusive function of the newsroom. This naturally raises the question of whether to consider social media editors as a new type of gatekeeper (Welbers & Opgenhaffen, 2018b).

Gatekeeping – the selection of messages to reach readers (Shoemaker & Vos, 2009)-has long been recognized as a fundamental function of journalism. Although it has often been considered a practice to choose news stories from *raw* information in the context of mass communication, recent conceptualization has attempted to encompass multiple gates, which exist

between an information source and final readers (Barzilai-Nahon, 2008). News organizations decide whether to cover news stories discovered in an inter-media setting (Meraz, 2011; Vargo, Guo, & Amazeen, 2017), and the algorithmic decisions of search engines to post links to news stories can act as a gatekeeper (Bozdag & van den Hoven, 2015). In fact, in the past, the concept of gatekeeping originated from an inter-media setting, in which a wire-editor selected stories from a mass of wire copies from wire services (White, 1950). In the modern context, hyperlinks are at the center of the development of the many steps of gatekeeping in online news consumption. Dimitrova, Connolly-Ahern, Williams, Kaid, and Reid (2003) earlier found that major American newspaper companies use hyperlinks as a gatekeeping tool to control the number of external links to information sources. Further, bloggers and social media users can participate in the gatekeeping process using hyperlinks as a facile path to certain information (Meraz & Papacharissi, 2016).

*Selective link sharing* for social media by news organizations can be understood as inter-media gatekeeping *between* news websites and social media *within* an organization. Given stories published on a website, a news organization prioritizes news stories that should be posted on its accounts. Although it is possible to criticize that the multiplicity of information 'gates' makes privileging news organizations as gatekeepers less accurate (Bruns, 2005; Thorson & Wells, 2015), Welbers and Opgenhaffen (2018b) showed that the decisions of social media editors for Facebook have a significant impact on the news engagement of online readers, thus validating selective link sharing as the gatekeeping practice of news organizations. In other words, despite the secondary and networked multiple gates that prevail on social media (Diakopoulos & Zubiaga, 2014), the decisions of news organizations still matter as a substantial determinant in the diffusion of news.

Although the evaluation of news topics across different media platforms has been a classical subject in the field of media study (Maier, 2010; Stempel, 1985), we have little knowledge about how this influential new form of gatekeeping changes what we see on social media as news readers. As discussed in the previous section, the organizational separation (Elizabeth, 2017; Roston, 2015) of the tasks and audience-centric motivation (Tandoc, 2014, 2016; Welbers et al., 2016) indeed imply that news shared on news organizations' social media accounts is likely to deviate from what journalists publish on a news website. Building on the previous finding that selective link sharing indeed has a power as gatekeeping (Welbers & Opgenhaffen, 2018b), this study focuses on how the outcomes of the new gatekeeping

differ from those of the traditional outcomes. Thus, I pose the following research questions:

*RQ 1. To what extent does selective link sharing resemble or deviate from traditional gatekeeping?*

*RQ 2. In which news topics is the deviation between selective link sharing and traditional gatekeeping more significant?*

The previous literature, which focused on how news stories on social media spread among users hints at the deviation pattern; Berger and Milkman (2012) and Stieglitz and Dang-Xuan (2013) found evidence that emotional arousal (intensity) is associated with virality on social media. In a similar vein, Armstrong and Gao (2010) speculated that the abundance of links to news stories with 'sensational' topics, such as crime, and many life-style news links by local media compared to national and regional media are a result of consideration for the targeted demand. García-Perdomo et al. (2018) also found that conflict/controversy, human interest and odd news categories are more likely to be shared by Twitter users. Similarly, Kalsnes and Larsson (2018) showed that what is often categorized as "soft news" tends to go viral although "hard news" was more popular on Twitter than it was on Facebook. Thus, assuming that news organizations are adapting to such news demand on Twitter, I suspect that news organizations will share news topics that are likely to contain great emotional arousal, controversy, human interest, or oddity.

How news shared on Twitter deviates from the publication pattern may differ depending on the organization's audience base, internal structure, and financial status. For example, whereas *The New York Times'* social media strategy leans toward a marketing approach as of 2015 (Roston, 2015), Elizabeth (2017) found that the *Washington Post* takes a different path, which encourages social media specialists and news rooms to work together. Further, because the emergent online news organizations have bigger stakes in social media to disseminate their stories, they are likely to be more engaged in strategies to draw the attention of social media users, which is likely to make them lean toward sharing sensational news topics on their posts. It is harder to foresee the general tendency in the case of regional newspapers because how social media editing is done within those organizations is irregular. Sometimes a very limited number of staff members, even a single person, decides the social media posting. But in other cases, social media editing is entirely dissipated among individual reporters without the aid of social media specialists, due mainly to tight financial

constraints. Kalsnes and Larsson (2018) reported “the high degree of commonality for the most popular news items across the majority of the studied news providers.” In sum, although it is very likely that selective link sharing deviates from traditional gatekeeping decisions in general, it is difficult to predict how the pattern differs across different organizations. Thus, I pose the following research question.

*RQ 3. Does selective link sharing differ depending on the types of news organizations?*

## 4. Method

The analysis of the potential discrepancy between traditional gatekeeping and selective link sharing poses methodological challenges, that mainly involve the scale of the data. To test selective news link sharing, a researcher must have a sample of published news stories that do not misrepresent the distribution of news contents because such a sampling bias will change the estimated proportion of shared news content. However, it is impossible to distinguish news contents before reading the news. That is, a researcher must first analyze the data to get a good sample.

In addition, even though automated data collection allows for bypassing such a sampling challenge by collecting all the data available online, the sheer amount of data also prevents the use of a traditional hand-coding approach. For example, the data set used for this study includes more than 280K news stories and more than 130K tweets that embed news links. Further, the analysis requires the comparison of the pairs of news stories and tweets, as well as separate stories and tweets on their own. Thus, answering the questions in this study requires computational text analysis methods as well as the automated data collection.

### 4.1. Data Collection Using *Media Cloud*

The essential information needed to uncover selective link sharing requires knowing whether a news story published on a news website was shared on the social media accounts of news organizations. To acquire this information, a researcher needs to know whether a hyperlink attached to each news story published on news websites is also found on the news organization’s social media account. Although many proprietary news databases, such as *Factiva* or *LexisNexis*, are available, these databases often lack information that is needed to link the two platforms. Thus, I instead used *Media*

*Cloud*, an open sources database developed and maintained by the *Harvard Berkman Klein Center*<sup>2</sup>. *Media Cloud* monitors the RSS feeds of approximately 60,000 news organizations around the world every thirty minutes, and provides meta-information of the news stories including news URLs.

Because *Media Cloud* does not provide the content of news stories, I developed a news scraper using *Python* language, which opens all the news URLs from *Media Cloud* API and downloads the news text. When news websites have a paywall, I used *Selenium* library to log in through a *JavaScript* login interface. To collect data about which stories a news organization shared, I also developed a Twitter scraper that collected tweets from official news websites of organizations<sup>3</sup> via Twitter's Search API.

Another piece of software, URL matcher, matches URLs of published news (from *Media Cloud*) with URLs embedded in news tweets. Figure 1 visualizes the process. Because URLs are 'messy' in the sense that they are shortened for the sake of saving space, and include query terms to specify traffic sources, etc., the URLs had to be 'normalized' (Welbers & Opgenhaffen, 2018a). The normalization process took three steps: first, the URL matcher sent HTTP HEAD requests using URLs from *Media Cloud* and Twitter API recursively to get the final form of the URL. Although this is a relatively slow process because it has to wait to receive responses over the network (potentially multiple times per URL), this method avoids possibly different forms of a URL toward the same content. This is because the process uses the same method that readers would use to read the news stories through their browsers. Second, the query terms that mark redirecting histories were removed. Finally, only the unique identifiers of each URL were retrieved. Because of the common Search Engine Optimization (SEO) practice, the unique identifiers have forms of abridged titles, e.g. "evans-calls-report-wayne-co-treasurer-extremely-troubling." Because how news URLs are structured differs across news organizations, the regular expressions

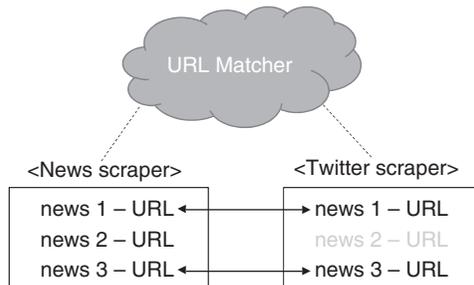


Figure 1. A computational method to identify shared and unshared news on Twitter.

to detect the identifiers had to be customized for each news organization. After the identifiers had been retrieved, 100 random samples from each news organization were manually inspected to check that the retrieval had been correctly performed. Then, the URL matching was conducted using only the identifiers *within* each news organization.

News organizations in the data set were collected from *Alexa* online traffic ranking. I retrieved the top 200 websites from the *Alexa* news media section. I augmented this list using a list of media that were influential during the 2016 US presidential election according to the recent *Harvard Berkman Klein Center* report (Faris et al., 2017). Then, I chose only U.S. news websites except the *BBC*, the *Guardian*, *Reuters* and *AlJazeera*, which have potentially substantial impact on news consumption in the United States. Among 116 news organizations in the original list, I removed news organizations whose RSS feed is not well monitored by *Media Cloud* (28 organizations), and whose online news stories were not parsed by the News Scraper (15 organizations). This process resulted in 73 news organizations. These news organizations are categorized as in Table 1.

The U.S. media market provides an interesting test case to gauge how social media editing evolves depending on the media competition between diverse media species. Given the weak public media, the market competition is likely to yield a strong drive toward audience-centric social media editing for more online advertising revenue. Second, American news readers use social media as a news source more often than those in other Western countries (U.S. 45%, U.K. 39%, France 36%, Germany 31% according to Newman, Fletcher, Kalogeropoulos, Levy, and Nielsen (2018)), which makes social media editing more relevant for news organizations' online revenue. Lastly, the U.S. media ecosystem is composed of more diverse news outlets than many other countries; the share of regional broadcasters and newspapers is relatively high in the readers' media diet, and news readers in the U.S. rely more on online-born outlets than other countries (Newman et al., 2018). This heterogeneity particularly provides the valuable variation in the data that is potentially associated with different practices of social media editing.

The data were collected between 2017/11/20 and 2018/01/01, and between 2018/01/09 and 2018/01/28, for 63 days. This data collection process resulted in 281,508 news stories. Among them, 131,939 new stories were shared by news organizations on their Twitter accounts, resulting in a 46.87% sharing proportion. The number of published news stories by news organizations varied from 60 (*factcheck.org*) to 18,239 (*Reuters*). This variation stems from a difference in their media types, from websites that grew from opinion blogs, such as *factcheck.org* or *townhall.com*, to magazines, such as the

**Table 1:** *The list of news organizations included in the dataset.*

National	Regional	Magazine	Online
ABC News	Newsday	Economist	AlterNet
Al Jazeera	NJ.com	Forbes	Bipartisan Report
BBC	The Baltimore Sun	New York Magazine	Breitbart
Bloomberg	The Boston Herald	The Atlantic	BuzzFeed News
CBS News	The Chicago Sun Times	The Christian Science Monitor	Common Dreams
CNN	The Chicago Tribune	The New Republic	Conservative Tribune
Fox News	The Dallas Morning News		FactCheck
MSNBC	The Denver Post		FiveThirtyEight
NBC News	The Des Moines Register		Free Beacon
New York Times	The Detroit Free Press		Gateway Pundit
NPR	The Detroit News		Huffington Post
PBS Newshour	The Indianapolis Star		IBTimes
Reuters	The Mercury News		Inquisitr
The Guardian	The New York Post		Media Matters
The Wall Street Journal	The Oregonian		Politico
USA Today	The Orlando Sentinel		PoliticusUSA
Washington Post	The Philadelphia Inquirer		Slate
	The Press Democrat		Talking Points Memo
	The Stranger		The Daily Beast
	The Sun-Sentinel		The Daily Caller
	The Tennessean		The Hill
			The Intercept
			The Root
			Vanity Fair
			Vice
			Vox
			Washington Examiner
			Western Journalism
			Westword

*Economist* or *Time*. On the other hand, daily news outlets with blogs as a part of their website such as the *Washington Post*, or newswire services such as *Reuters* publish far more news stories on their websites. The proportion of shared news stories also varied from 10.96% (*MSNBC*) to 95.43% (*The Economist*). This variation depends partially on how many news stories they publish on their own. Magazines tend to share a majority of what they publish whereas outlets with partnered content and blog posts by journalists, such as *Washington Post* or *The Press Democrat*, tend to share less. This relationship is visualized in Figure 2.

Figure 3 shows the daily pattern of the number of published online news stories by the 73 news organizations. It shows a clear fluctuation, in which

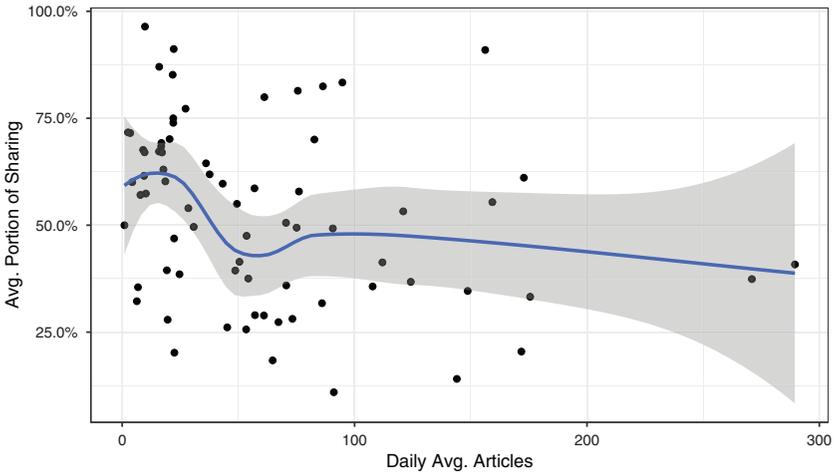


Figure 2 The average proportion of news sharing of each new organization against the average number of news stories daily published with a smoothed trend. Each dot represents a news organization.

the number of published news stories is lower on weekends and holidays. That is, in addition to the lower number of published news on Saturday and Sunday, the figure also shows that news organizations published much less news on Thanksgiving (November 23, 2017), Christmas, and New Year’s Day. Moreover, heading toward the end of 2017, the number of news stories took a downward trend, which seems natural for the holiday season, and a sharp

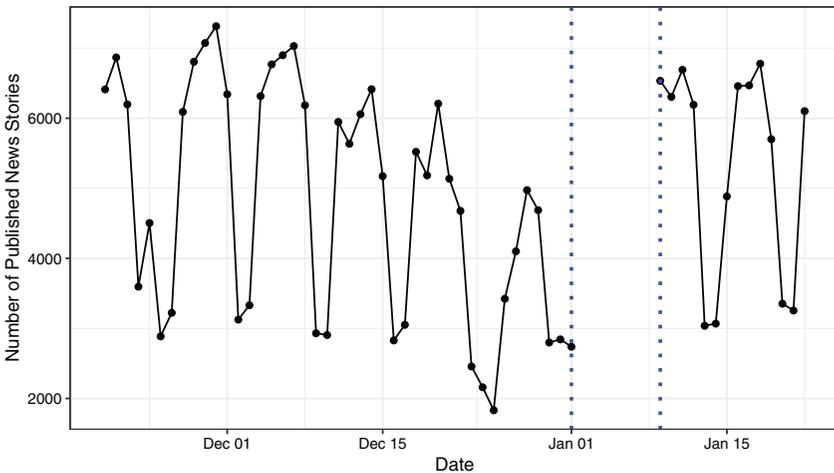


Figure 3 Daily number of published online news stories. The data were not collected during the period between the two dotted vertical lines.

recovery after the New Year. The graphical observation illustrates that the daily pattern is consistent with weekly and seasonal working patterns, and provides a useful face validity of the data.

#### 4.2. Structural Topic Model for Selective Link Sharing

To see if news organizations tend to share more popular topics on Twitter, I applied a machine learning algorithm, Structural Topic Model (STM) to the collected data (Roberts, Stewart, & Airolidi, 2016). STM is a recent extension of topic models (Blei & Lafferty, 2006; Blei, 2012; Blei, Ng, & Jordan, 2003) that are designed to identify latent topics from the co-occurrence of words. STM extends topic models by embedding a regression model within a Correlated Topic Model (CTM). That is, it simultaneously finds topics and regresses the topics to other observable variables. This allows for testing the association between latent topics and the link sharing probability in this study. In the context of this study, by applying STM, I replaced the research question with a computational task that identifies topics correlated with the link sharing decision to circumvent the implausible manual approach.

Since STM was proposed by political scientists in 2013 (Roberts, Stewart, Tingley, & Airolidi, 2013), it has been widely applied to news and social media text where observable meta data are available. For example, Kim (2018) relates the information about whether an automotive recall case is from a foreign automaker or a domestic one to the categorization of news topics. Using the *Youtube* channel category as a label, Schwemmer and Ziewiecki (2018) also found that particular categories of influencers play a significant role in product promotion on social media.<sup>4</sup> Compared to previous topic modeling methods, STM generally produces better categorization results because (a) it is built on CTM that allows correlation between different topics unlike standard LDA, and (b) it utilizes observable variables to produce topics (Roberts, Stewart, & Airolidi, 2016). Although this model relaxation costs non-conjugacy and non-convexity of the optimization problem, more computational power and systemized initialization help overcome these issues (Roberts, Stewart, & Tingley, 2016). In the case of this particular study, I found no noticeable difference after multiple runs with different initializations.

Categorizing news based on its content is a longstanding methodological issue of media studies. Depending on news samples and the scope of research questions, some traditional news categories may be irrelevant or collapsible, but other unusual categories may become important (Sjøvaag & Stavelin, 2012). A class of topic models, including STM, overcomes this challenge by using a bottom-up approach whereby topics are *discovered*

from the co-occurrence of words across related documents rather than pre-supposing news categories.

However, that the topic models require researchers to pre-specify the number of topics (i.e. news categories) poses a challenge to this study as well. The challenge is slightly different, but conceptually similar to the traditional question for manual content analysis: are the discovered topics based on the number of topic assumptions appropriate for the research question? This is a harder question to answer when topic models are applied to a highly heterogeneous corpus of text such as a body of news stories published by multiple news organizations. If a researcher sets the number of topics too large, then the model is likely to identify only types of news organizations, such as online, national, entertainment oriented, Texas based, etc., using organization-specific textual traits rather than common topics. In this case, a researcher will be unable to compare different behaviors for the same news topics because topic overlap across different organizations would be limited. On the other hand, if the number of topics is too low, a researcher is likely to miss interesting variations across different news contents.

Researchers have suggested multiple indices to compare the performance of different topic number assumptions from the perspective of the statistical model comparison, such as the held-out log-likelihood approach (Blei et al., 2003), coherence criteria (Mimno, Wallach, Talley, Leenders, & McCallum, 2011; Newman, Lau, Grieser, & Baldwin, 2010) and exclusivity criteria (Bischof & Airoldi, 2012). However, these automatic procedures to choose the optimal number of topics are not often adopted for applied work because they tend to deviate from choices of human coders (Chang, Gerrish, Wang, Boyd-Graber, & Blei, 2009). Moreover, the different proposed indices are often inconsistent with each other, as in this study, and theory does not dictate which index should be prioritized. As Schwemmer & Ziewiecki (2018) illustrated, the different choices of topic number also resulted in monotonically decreasing coherence and monotonically increasing exclusivity. After all, the number of topics tends to be determined by the researcher's careful choices based on his/her research questions. In this case, results from topic models are interpreted as answers to a question, "What is the best categorization of news topics *given the number of topics*?"

For this study, I chose 25 topics. The primary reason for this choice was to not break down topics to the level at which news topics belong solely to specific regional markets. This is not to ignore regional topics; in fact, the motivation is the opposite. How different types of news organizations treat regional topics differently is of a great interest to this study. However, if the

model categorizes news topics to finely, so that New York news is distinguished from Texas news, one ends up with a trivial finding that, for example, New York-based news organizations are likely to share New York news topics. Indeed, with 27 topics from this study's data set, I found that STM begins to identify news about specific regions. Thus, I decided to limit the number of topics to 25 so that the model can identify regional topics that are more broadly labeled.

The regression within the STM model associates portions of the latent topics with observed covariates – (a) whether a news story was shared on Twitter, (b) indicators of news organizations (binary), and (c) interactions between the two. Thus, the estimated coefficients for the covariates mean (a) how much more frequently a given topic is *shared overall* on Twitter compared to the proportion of the topic among the total news stories published on websites, (b) how much more frequently a given topic is *published on a news website* by a news organization relative to other organizations, and (c) how much more frequently a given topic is *shared* on their Twitter account by *each* news organization relative to other organizations.

## 5. Results

### 5.1. News Topic Categorization

Table 2 summarizes the identified topics and words with the highest FREX score in a given topic. FREX score is a weighted average between word frequency in a given topic and the degree to which the word appears *only* in the given topic (i.e. exclusivity) (Bischof & Airoidi, 2012). I labeled each topic by inspecting the list of words with high FREX scores and randomly chosen news stories among those estimated most likely to be about each topic. For most cases, the high FREX words provide a straightforward interpretation of each topic category. For example, the output is precise enough to distinguish different sports: Sports/Basketball topic is associated with words such as 'rebound', 'basketbal' and 'nba' whereas 'nfl', 'quarterback', 'touchdown' represent the Sports/Football topic. The sports-related topics will be particularly important for the regional media's strategy in what follows. Furthermore, the output distinguishes topics related to national politics – e.g. Politics/President ('trump', 'fbi', 'flynn') and Politics/Election ('moor', 'elect', 'vote').

However, the manual inspection also revealed that the high FREX words did not necessarily guarantee proper labeling. For example, although the high FREX scores of 'attorney', 'complaint', 'statement', 'lawsuit' seemed to

**Table 2: Words highly associated with each topic (High FREX scores).**

Label	Words with High Probability
Life/Fun	realli, thing, think, know, im, dont, that, someth, lot, go, feel, want, kind, hes, ive, didnt, just, there, get, cant
Crime	polic, arrest, kill, gun, murder, suspect, prison, fire, sheriff, crash, counti, crime, car, shoot, jail, dead, death, man, offic, driver
Economy/Business	compani, text, firm, million, busi, inc, ceo, industri, custom, oper, execut, energi, contract, employe, financi, manag, consum, electr, own, servic
Technology	internet, exampl, technolog, user, data, creat, neutral, idea, googl, platform, self, understand, comput, digit, connect, experi, system, tool, communic, content
Sports/Basketball	score, point, rebound, basketbal, shot, half, game, nba, guard, second, ball, pointer, assist, minut, lead, warrior, rewritten, straight, redistribut, play
Entertainment	music, movi, film, episod, charact, star, song, artist, album, actor, singer, fan, award, seri, comedi, band, netflix, studio, disney, perform
Sports/Football	nfl, quarterback, yard, bowl, touchdown, playoff, footbal, coach, eagl, patriot, lion, offens, season, vike, defens, game, giant, tackl, intercept, bradi
Royal Family	mother, father, famili, daughter, son, princ, children, husband, babi, child, parent, wife, harri, mom, friend, wed, sister, kid, life, coupl
Sports/Others	goal, leagu, sport, match, player, club, champion, team, cup, round, titl, england, won, win, athlet, competit, championship, winner, tournament, basebal
Local/Outdoor	de, park, citi, hotel, beach, island, town, resid, area, squar, museum, mayor, san, build, mountain, river, land, locat, mile, lake
Economy/Finance	stock, market, price, percent, growth, trade, investor, bank, rate, gain, index, economi, rise, analyst, currenc, global, billion, averag, oil, higher
Food/Cooking	amazon, store, shop, appl, app, food, gift, holiday, card, cook, restaur, wine, item, pinterest, retail, eat, coffe, christma, devic, toy
International/Middle East	jerusalem, israel, protest, palestinian, muslim, iran, islam, peac, saudi, isra, capit, refuge, syria, embassi, countri, al, war, terror, turkey, terrorist
Media	news, media, fox, post, daili, facebook, nbc, video, cbs, fake, abc, tweet, network, host, page, twitter, tv, mail, journal, journalist
Politics/President	trump, fbi, flynn, obama, donald, mueller, presid, clinton, white, counsel, advis, administr, russia, russian, investig, intellig, campaign, presidenti, lie, hillari
Sexual Harassment Investigation	alleg, accus, resign, franken, attorney, complaint, statement, deni, lawsuit, claim, file, abus, investig, inappropri, lawyer, ethic, former, apolog, misconduct, conduct
International/Europe	rule, court, uk, ban, legal, eu, law, britain, brexit, commiss, justic, union, appeal, copyright, suprem, caption, imag, judg, british, constitut
Medical/Health	patient, medic, drug, studi, health, doctor, npr, diseas, cancer, research, treatment, climat, hospit, dr, transcript, mental, brain, care, anim, risk
Local/Weather	com, photo, detroit, iowa, http, michigan, dec, https, pic, usa, snow, nashvill, jan, decemb, twitter, today, nj, weather, nov, www
Politics/Congress	immigr, shutdown, lawmak, daca, congress, senat, sen, legisl, democrat, bipartisan, schumer, republican, illeg, rep, hous, mconnel, border, fund, leader, bill
Economy/Tax	tax, incom, cut, corpor, deduct, insur, bill, benefit, pay, wage, budget, reform, repeal, debt, worker, rate, money, taxpay, individu, trillion

Politics/Election	moor, elect, voter, alabama, roy, candid, poll, vote, parti, democrat, jone, polit, seat, republican, governor, race, ballot, campaign, doug, senat
International/East Asia	korea, korean, nuclear, missil, north, china, flight, kim, chines, airlin, airport, south, plane, navi, weapon, japan, russia, sanction, putin, russian
Education	student, school, newsllett, educ, mr, univers, colleg, teacher, graduat, subscrib, pleas, ms, campus, sign, class, recruit, inbox, click, board, program
Me Too	women, weinstein, sex, men, femal, sexual, harass, rape, black, woman, male, gender, harvey, assault, gay, hollywood, metoo, movement, racial, rose

suggest topics related to legislation, the news articles with a high chance of containing the corresponding topic category were mainly about the sexual misconduct of high profile figures (Sexual Harassment Investigation topic). Similarly, whereas the high FREX words, such as ‘amazon’, ‘store’, ‘shop’, ‘appl’, ‘app’ seemed to imply a topic related to shopping, the manual inspection suggested that representative news articles in this category were mainly about cooking. This discrepancy suggests that although FREX provides useful information about words that *distinguish* one topic from another, it does not necessarily mean that those words *represent* the topic. This emphasizes the importance of manual inspection. To show the importance of such a manual validation process, Appendix B lists the titles of news articles with the highest probability to contain each topic.

Figure 4 shows the frequency of each topic among all the news in the dataset. The distribution shows a sensible mix of news topics: topics often classified as *soft news* such as Life/Interview, Crime, and the sports related topics occupy top ranks of the list. However, there is a good amount of *hard news* with the topics related to national politics, the economy and international relations as well. Additionally, there are a few topics that are likely to contain regional interests, such as Sports/Basketball, Sports/Football, Education, Local/Outdoor.

## 5.2. Overall Patterns of Selective Link Sharing

Figure 5 presents the sharing propensity of each topic aggregated over all news organizations. The zero in the middle is the normalized topic-wise proportion among all the news stories *published on websites*, which in fact varies across different topics before the normalization. The points represent the deviation of the 25 topics on Twitter from the proportions on websites. For example, a point at approximately 0.05 for Crime topic means that, on Twitter, there is 5% point *more* Crime news on Twitter than on websites, which, in turn, implies that *given* Crime topic was published on websites, it is likely to be shared on Twitter.

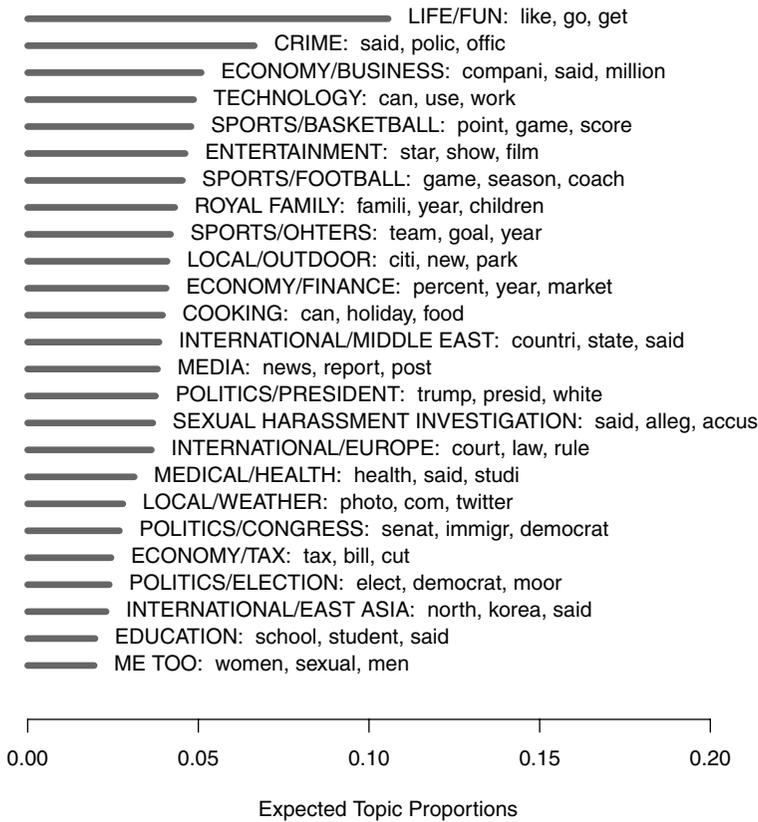
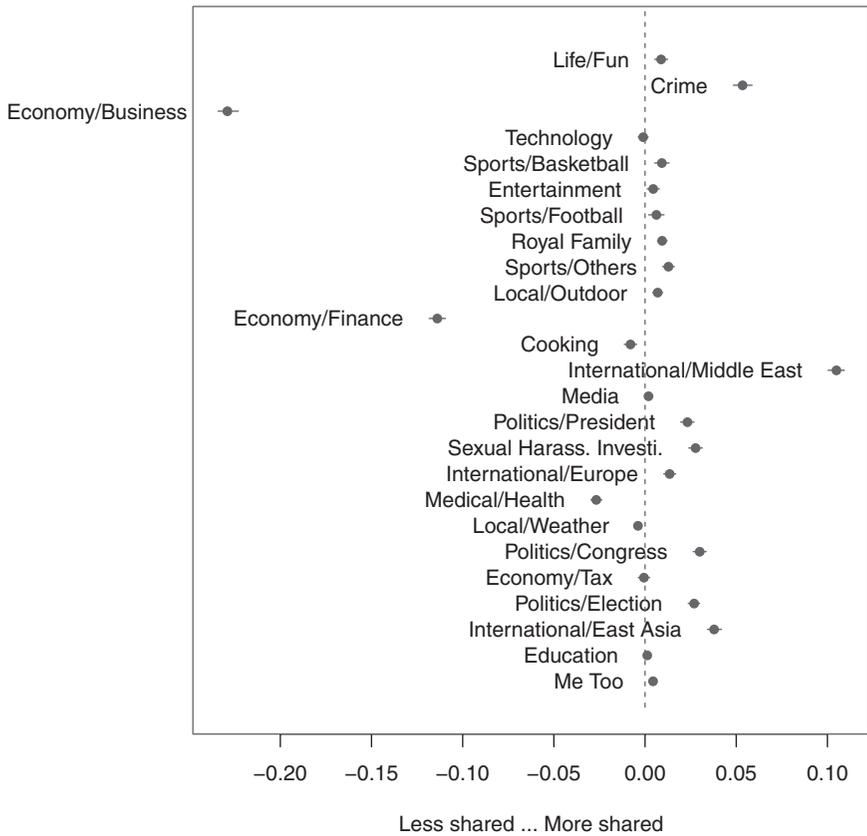


Figure 4 A bar graph for estimated overall proportions of 25 topics.

These deviating patterns answer RQ 1 and RQ 2 about the extent to which selective link sharing is different from traditional gatekeeping, and in what topics. Many topics seem to have a proportion on Twitter similar to their proportion on websites, which is visualized as points aligned on the vertical line at zero. This means that, for these topics, such as Technology, Entertainment and Media, news organizations tend to share news stories *proportional* to how many they published on their websites. However, there are a few topics that significantly deviate from the publication pattern: stories about Crime and International/Middle East related are very likely to be shared on Twitter given that they are published. On the other hand, Economy/Business and Economy/Finance topics are far less likely to be shared than other topics.

The observation that the chance of a topic being shared on Twitter does not necessarily match with its frequency on news websites implies that the publication decision and the link sharing decision do not necessarily follow



*Figure 5 Overall sharing propensity of 25 topics on Twitter. The propensity is measured as a fractional deviation of the proportion of a news topic among shared stories from the proportion among published stories. The topics are ordered by the frequency on news websites. There are a few topics that largely deviate from their proportions on websites.*

the same logic (Tandoc, 2014; Welbers et al., 2016). There is inertia by which traditional news organizations simply extend their convention into social media as a new news distribution platform (Armstrong & Gao, 2010) and attempt to integrate social media editing with traditional reporting and editing (Elizabeth, 2017). However, the observed topic deviation between the websites and Twitter shows that the emergent logic from social media editing is indeed different enough to result in different news topic proportions.

However, types of news topics more likely to be shared do not simply fall back into the traditional distinction between soft and hard news, and the relative popularity of the former. Although Crime, a topic typically categorized

as 'soft' is indeed very likely to be shared by news organizations, the overall sharing propensity is not large for other soft topics such as Sports/Basketball and Entertainment. On the other hand, there are a few 'hard' topics with a high propensity for sharing, such as International/Middle East and Politics/President. There are several possibilities that can explain this pattern. First, personnel who take charge of social media postings have their own agendas or standards for 'newsworthiness.' Yet, the previous literature generally indicates that the criterion they use to decide newsworthiness is primarily audience reaction. Thus, a more convincing possibility is that the pattern reflects the particular news preference of Twitter users. This interpretation accords with the *Pew Research Center's* report that Twitter has more highly politically engaged users than does Facebook (Gottfried, 2014).

### 5.3. Differences across Organizations

The previous analysis of overall news link sharing reveals a few topics whose link sharing pattern on Twitter deviates from their publication pattern on websites. This deviation may differ depending on who is responsible for social media editing and how the task is organized and distributed. This may in turn depend on the size of organization, its financial constraints, target audience, ownership structure, etc. (Elizabeth, 2017). This possibility points to RQ 3 about the difference in the deviating patterns between selective link sharing and traditional gatekeeping across different news organizations. To distinguish each organization's social media editing from its publication patterns, one needs to be able to separately measure (a) the likelihood for each organization to publish each news topic, and (b) the likelihood for each organization to share each topic *given* that the topic was published. As previously shown, the organization indicator dummies and the interaction terms between the organization indicators and the sharing decisions embedded in the STM model allow the measurement of (a) and (b) separately.

To illustrate the result, I first estimated the expected proportions of each topic on each news organization's website and on its Twitter accounts, using the regression coefficients for the two sets of dummy variables (e.g. the news organization indicators and the interaction terms). Then, I visualized (a) how much each organization's expected topic proportions on its website *deviated* from the *overall* topic proportions aggregated over all news organizations (Figure 4), and (b) how much each organization's topic proportions on its Twitter accounts *deviated* from those on *its own website*. The former deviations (firm-specific publication propensity) represent each organization's publication strategy, and the latter deviations (firm-specific

link sharing propensity) represent each organization's selective link sharing strategy, which is distinct from its own publication pattern.

Figure 6 shows the results. The firm-specific publication propensity (x-axis) and the firm-specific link sharing propensity (y-axis) do not necessarily align. In addition, there are a few cases in which the same type of news organizations behaves in a similar way. For example, the purple cross dots

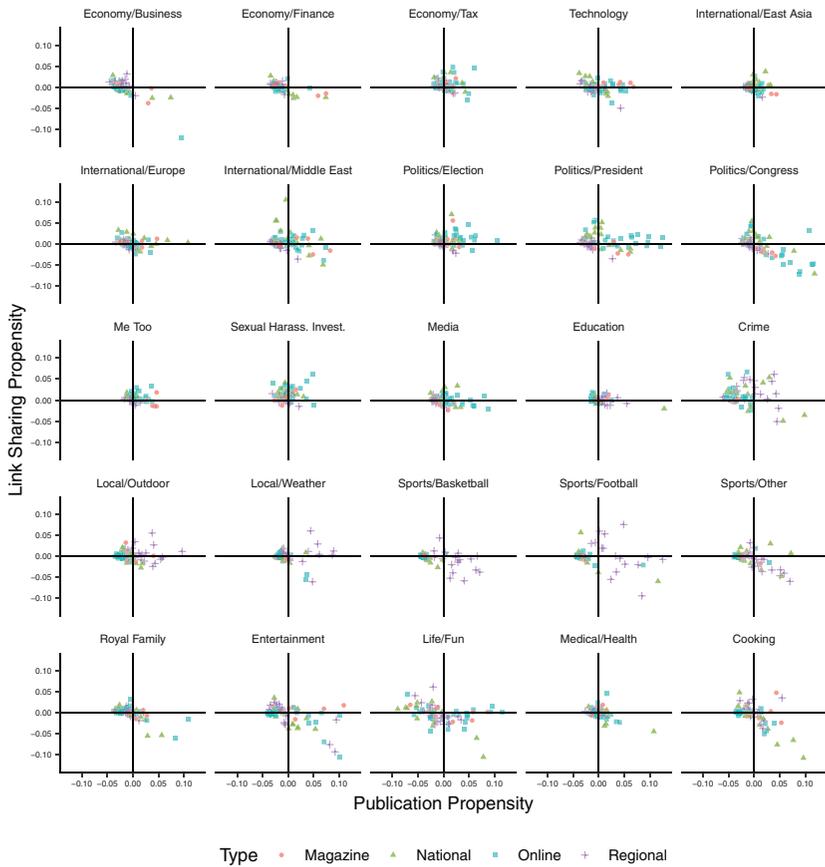


Figure 6 A visualization of firm-specific publication propensity and firm-specific sharing propensity broken down by types of news organizations. The firm-specific publication propensities are measured by how much proportion of each topic on each news organization's website deviates from the overall topic proportions. And the firm-specific sharing propensities are measured by how much proportion of each topic on each news organization's Twitter accounts deviates from its own website.

concentrated on the fourth quadrants show that regional newspapers publish more about Sport/Basketball topics than others, but they refrain from sharing the links toward those news stories on Twitter. Using this visualization, I categorized news organizations topic-wise publication and link sharing patterns.

**5.3.1. News publishing patterns accord with organization type.** The publication patterns on websites (x-axis in Figure 6) accord with commonsense patterns; national media (green triangle dots) tend to publish more about national politics, e.g. Politics/President and international affairs, e.g. International/East Asia, than do regional media. And regional media (purple cross dots) are located on the right hand side of graphs for sports topics, e.g. Sports/Basketball and Sports/Football, and regional topics, e.g. Local/Outdoor and Local/Weather, which means that they publish more about such topics compared to overall. Online media tend to publish more about national politics, such as Politics/President and Politics/Election. Unlike national media types, they do notably publish more about international affairs. This observation is sensible given that offline born national media tend to have a longer history, and are thus likely to have a more robust infrastructure for international reporting.

**5.3.2. A reversionary strategy is the most common.** When it comes to individual organizations' selective news link sharing patterns, the most common strategy across different media types is a *reversionary strategy*, whereby news organizations share less about a particular news topic when they publish a lot about the topic or vice versa. This strategy is visualized as dots in the second quadrant (low publication/high link sharing) or the fourth quadrant (high publication/low link sharing). For example, Figure 6 shows that magazines (red circle dots) and online media (blue square dots) tend to publish more about the Politics/Congress topic than overall, but they refrain from sharing news about the Immigration topic on Twitter. On the other hand, regional media publish a lot about Sports/Basketball. This is probably because data were collected during the winter basketball season, but they tended to share fewer of their links on Twitter. A natural outcome of this strategy is to make different types of news organizations look alike on Twitter, "reversing back to the average."

This strategy is reasonable when news organizations compete on social media with different types of news organizations rather than targeting a specific audience base via their traditional media platform. For example, although the *Wall Street Journal* publishes far more about finance than other news organizations do, an overwhelming number of general social media users with the professional analysis of finance may not be a profitable

approach because, given the users' limited attention, the *Wall Street Journal* is likely to miss a chance to attract their attention with more broadly popular news topics. Indeed, the STM result shows that the *Wall Street Journal* publishes 26% points more about Economy/Finance topic than overall, but it shares 16% point fewer news links about the same topic on Twitter compared with what it published. A similar explanation can be applied to regional media's reversionary strategy for the Sports/Basketball topic. Although their specialty is in covering local issues, and they are likely to expect regional basketball fans to visit their websites, they may have different incentives on Twitter, i.e. to reach out to broader social media users. If this is the case, regional media are unlikely to share too much about regional basketball games. The opposite motivation of the reversionary strategy is also plausible. When a news organization is not specialized in a given topic, it may want to share more to attract a broad range of social media users if the organization believes that the given topic is popular. National and online media tend to use such a reversionary strategy for the Crime topic whereas regional media tend to do so with Economy/Business topic.

In other words, a specialized topic without broad popular appeal is likely to gravitate toward the average proportion by the force of the reversionary strategy on Twitter. This corresponds to García-Perdomo et al. (2018)'s recent observation that topics on social media accounts are amazingly similar across different organizations. This is also in contrast with the common conclusion of economists that competition increases news diversity (Gentzkow, Shapiro, & Stone, 2014). The finding that reversionary strategy is the most common strategy implies that competition on social media acts as a homogenizing momentum, at least in terms of news topics.

**5.3.3. A strategic retreat is a strategy for non-specialized topics.** News organizations tend to share even fewer non-specialized topics, whereas they often counter-balance specialized news topics with less link sharing. This is visualized as dots in the third quadrant in Figure 6. For example, regional media publish less about International/Middle East topics, and share even less than that on Twitter. They also behave in a similar manner with International/Europe topic. This is understandable because if a news organization does not have enough resources to cover a given topic, and broad social media users are not particularly interested in a given topic, news links about the topic are unlikely to attract either the specific target group or the broad news demand on social media.

On the other hand, regional topics in which a limited pool of social media users are likely to be interested, such as Local/Outdoor and Local/Weather, are

topics from which national and online media retreat. This finding is also sensible because covering these regional issues is not in their specialty. Sharing news with those topics would not help appeal to broad social media users compared to sharing national news. In addition, online news media also tend to retreat from international topics, e.g. International/Europe and International/East Asia presumably for a reason similar to that of the regional media.

Unlike the reversionary strategy, strategic retreat implies differentiation between different types of news organizations, particularity in terms of locality/internationality. This result speaks to what Berger (2009) described as a 'paradox' between the globalism of the Internet and the hyperlocalism of news outlets. Online news distribution opens the door for global competition for an international audience. Because of the openness of the Internet, there is no such thing as true local media. However, local news organizations often lack capital to compete against national media. Hence, hyperlocalism is often suggested as a feasible sustaining strategy for regional papers (Lowrey & Kim, 2016). But the hyperlocalism turn will widen the gap between national news organizations and regional news organizations in terms of geographical coverage (Berger, 2009). The finding here suggests that news dissemination via social media accelerates the gap from the distribution side.

**5.3.4. A concentration strategy is used for specialized *and* popular topics.** There are fewer topics about which news organizations publish more *and* share more links. For both national and online media, Me Too is such a hot topic. Sexual Harassment Investigation was another hot topic for online media, but not for national media. The finding that sexual misconduct cases are hot topics might suggest greater attention to gender related topics, which may have been intensified by the Me Too movement during the data collection period. This might be a desirable change, but it can also pose a danger. As McDonald and Charlesworth (2013) pointed out, the sexual misconduct issue is often susceptible to sensational/scandalous and fight framing. If the popularity of these gender issues for news link sharing is merely an outcome of social media optimization to attract more attention, the issues are not likely to be covered in a socially desirable way, at least on social media. However, how the gender issues were framed as a result of the social media editing is beyond the scope of this study, and thus remains a future research question.

In sum, the results of the structural topic model spotted three notable news link sharing strategies on Twitter: (a) reversion to the average, (b) strategic retreat, and (c) concentration. Choices from these strategies differ depending on the link between topics and types of news organizations. The reversion strategy was common when a certain topic was a relative

specialty of the news organization (comparative advantage), but the topic's popularity was modest – e.g. Immigration for national media and Sports/Basketball for regional media. When news organizations do not have a specialty in a certain topic, and the topic is not particularly popular for the expected audience, news organizations seem to retreat from such news topics – e.g. International/Middle East for regional media. However, when news organizations are specialized in a certain topic, and it is popular, they will concentrate on those hot topics.

The categorization of the topics by the different strategies is summarized in Table 3. The numbers in parentheses indicate the number of topics that fall into each category. As mentioned above, the reversionary strategy is the most common category. There is much similarity between online and national media in regard to the reversionary strategy whereas the regional media are disconnected from the two. There are fewer hot topics on which each type of organizations concentrates. Again, regional media are disconnected from national and online media. Regional media retreat from a few topics that are not their traditional focus of coverage. It was harder to detect discernible patterns from magazines because fewer media are included in the dataset compared to other types.

## 6. Conclusion

Although traditional editorial decisions have been considered to be governed by journalistic norms (Hermida, 2010), there has also been a concern that economic incentives drive traditional reporting toward popularity. This is especially true because Internet technology greatly transforms the news industry (Lischka, 2014). However, the relatively high proportion of several hard topics, such as Economy/Business, Economy/Finance and International/Middle East, on websites illustrates that the market drive is not influential enough to banish hard news from online news dissemination in general. Yet, the deviation of news link sharing via Twitter shows that the different logic emerging from social media as a news distribution platform, particularly spurred by the wide adoption of an audience metric (Tandoc, 2017; Welbers et al., 2016) can greatly influence the kinds of news available to readers. Although the results show that Twitter readers will see a fair amount of politics-related hard news, such as International/Middle East and Politics/President topics, they are likely to be exposed to far less news about the economy, such as Economy/Finance or Economy/Business topics. The substantial deviation between websites and Twitter seems to

**Table 3:** *Summary of the selective news link sharing strategies. Numbers in parenthesis are number of topics categorized as a given strategy.*

	Reversion	Retreat	Concentration
Magazine	Immigration Entertainment International/Europe (3)	Local/Outdoor Sports/Basketball Local/Weather Sports/Football (4)	Economy/Tax (1)
National	Life/Interview Politics/Congress International/ Middle East Politics/President International/East Asia Crime Entertainment Media Family Subscription (10)	Local/Outdoor Sports/Basketball  Local/Weather Economy/Business Sports/Football (5)	Me Too (1)
Online	Immigration Sports/Basketball Politics/Congress Crime Entertainment Sports/Football Family Sports/Soccer (8)	Local/Outdoor Local/Weather Economy/Business International/East Asia International/Europe (5)	Sexual Harassment Investigation Me Too (2)
Regional	Sports/Basketball Economy/Business Politics/Congress Politics/President Crime Entertainment Family Economy/Finance Me Too (9)	Life/Interview International/Middle East Technology International/Europe Media (5)	

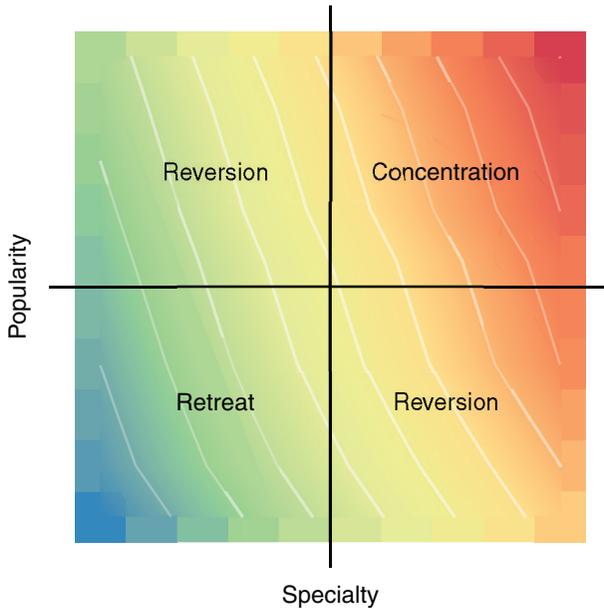
confirm the new logic that governs the quasi-gatekeeping that emerges from the decision-making process that is relatively separate from traditional editorial decisions (Elizabeth, 2017; Roston, 2015), and relies on audience metric technologies (Diakopoulos, 2017; Tandoc, 2014, 2016). Further, the lack of economic reporting raises a concern that popularity driven link sharing can exacerbate the watchdog function of financial journalism, which was triggered by journalists' failure to warn about the financial crisis (Strauß, 2018).

The organization-level analysis reveals more details of selective link sharing. The finding that link selection differs across organizations shows

that their criteria for selective link sharing cannot be reduced to a one-dimensional degree of popularity on Twitter. Further, the observation that the same type of news organization occasionally adopts a coherent selective link sharing strategy implies that the inherent characteristics of an organization, as well as its popularity as an external factor, are also significant determinants of the selective link sharing strategy. I tentatively characterize the other dimension as a *specialty* of a news organization. The publishing patterns of organizations seem to make the characterization appropriate; national media publish significantly more about international topics and national politics, and regional media publish more about sports and regional issues. I use the term “specialty” to roughly mean “what a news organization is good at.” It is related to a news organization’s capabilities, such as the training of reporters at different levels of expertise, the number of reporters, the expertise of editorial staffs, and its connections to certain news sources, etc. Although these capabilities are changeable, they will characterize news organizations at least in the short term because they tend to co-evolve slowly with the long-term expectations of readers.

The intersection of the popularity of a news topic and a news organization’s specialty seem to determine specific selective link sharing strategies. The most conspicuous strategies from this intersection were the *reversionary strategy* for specialized and *non-popular* topics and the *strategic retreat* for non-specialized and non-popular topics. The reversionary strategy acts as a *homogenizing momentum* for unpopular topics in the sense that different news organizations will share similar proportions of a given topic under this strategy. The strategic retreat makes the news organization invisible with respect to a certain topic. News organizations may also concentrate on their specialized *and* popular topics, which can be diversifying because different types of news organizations will share more about their specialized hot topics. Thus, these three strategies imply that *news on social media can be diverse (specialized) only for popular topics*. This prediction is visualized in Figure 7.

Overall, however, Twitter as a news dissemination platform seems to serve as a homogenizing momentum rather than diversifying one. As Table 3 shows, there is only a limited number of topics on which news organizations concentrate on Twitter. National and online media concentrated on the Me Too topic, and online media also focused on Sexual Harassment Investigation. Both are of great social importance, but also susceptible to sensational or “horse-race” framing with clearly defined fighting counterparts (McDonald & Charlesworth, 2013). On the other hand, news organizations adopt reversionary strategy or strategic retreat for many topics, which make them look alike, or invisible for certain news topics on Twitter.



*Figure 7 The gatekeeping momentum generated by the selective news link sharing strategies. Topics falling onto the red areas will be more visible, and topics in the blue area will be less visible on social media compared to traditional outlets.*

I suspect that this is because a selective link sharing decision converges to what the broad Twitter users would like rather than what is important for a targeted audience. Yet, less popular, specialized topics may include topics that are complicated enough to require professional treatment to cover but do not stimulate interest from a broad audience. Some argue that covering these topics helps readers easily understand those complex issues and make informed decisions (Yankelovich, 1991). In that sense, the reversionary strategy may indicate a socio-technical momentum that weakens the role of journalism that informs public decisions. In other words, although news shared on social media includes a fair amount of civic issues, the shared news is likely to highlight the sensational aspect of those issues.

Selective news link sharing as a reaction to its environment has an implication for the design of information curating algorithms on social media. The conclusion that selective link sharing of news organizations is an outcome of their adaptation to inherent characteristics and external situations implies that they will also adapt to potential changes in the demand for

news on social media caused by changes in an algorithm. This adaptation process is likely to become quicker and smoother, determined by the data scientists, whom they hire (Rowan, 2014) and the social media monitoring platforms, which they adopt (Diakopoulos, 2017). However, the quick adaptation is likely to make news organizations vulnerable because their demand depends on how the platforms' decision changes the news reading behavior of users. Based on market foreclosure theory, economists point out that the dependence on monopolized platforms may suppress independent content providers, as was notable in the Microsoft case (Evans, 2003). Indeed, there is evidence that Facebook's algorithm change adds a great impact on the online revenue of news organizations (Brown, 2018). Therefore, the discussion of a desirable algorithm design should include consideration about the potential strategic adaptation of news organizations as well.

In this work, I proposed a computational framework to monitor news organizations' selective news link sharing. It leverages automated data collection using an open source database and a recent development of topic modeling. However, the topic modeling approach is also known to have innate limitations. First, the choice of the number of topics always entails a certain degree of arbitrariness. The manual validation results provided high confidence in the classification results for news stories clearly classified by the 25 topics. However, the results may be missing information from news stories that do not fit well into any of the 25 topics. The STM takes into consideration the uncertainty from the less clear classification by giving less weight to those topics in the regression. However, the low applicability of the automated process of choosing a topic number implies that it is innately hard to know the best topic number to result in the optimal level of classification. In addition, the concept of topics as in topic models does not distinguish news issues from tone and framing. Thus, STM does not give us much information about the dependence between how a news issue is stated and the link sharing decision. This challenge was the most visible in our uncertainty about why gender issues are a hot topic. To address these issues, the use of supervised learning methods that involve the development of a manual coding scheme and distinguish tones and frames will be a fruitful extension of this work as a further validation.

Finally, the results found in this study may be subject to a topic popularity cycle although the dataset includes a fair amount of time – sixty-three days. For example, STM identifies the Me Too topic as a separate topic. Thus, verification of the conclusions that I draw requires an analysis of a longer time period beyond time-sensitive issues. The data collection scheme and the

automated text analysis I apply here can be used as the infrastructure for the constant monitoring of news organizations' selective link sharing strategy.

Despite the limitation, STM provides a semi-supervised extension of topic models, which enables the linking of observable link sharing-behaviors with automated topic categorization. Further, STM provides a way to produce news topic categories that are explicitly relevant to research questions by incorporating information from independent variables whereas the supervised learning techniques requires *ex ante* topic categorization. This approach provides an immediately reusable infrastructure to constantly monitor news organizations' emerging quasi-editorial decisions, which have a potentially significant impact on everyday news reading. Given that the quasi-editorial decisions are still in flux, depending on algorithm changes in social media platforms' and the adaptation of a business model, having such a handy monitoring framework will be useful for a future discussion.

## Author Note

This paper is based on chapters 4 and 6 of my Ph.D. dissertation written in 2018 at Michigan State University. Special thanks to my committee members, Rick Wash, Steve Wildman, Johannes Bauer, and Esther Thorson. Financial support from the Graduate School Dissertation Completion Fellowship at Michigan State University is gratefully acknowledged.

An online appendix is available at [http://pakchank.com/wp-content/uploads/2019/05/CCR\\_appendix.pdf](http://pakchank.com/wp-content/uploads/2019/05/CCR_appendix.pdf)

## Notes

- 1 <https://www.facebook.com/notes/facebook-journalists/study-how-people-are-engaging-journalists-on-facebook-best-practices/245775148767840>
- 2 Available at <https://mediacloud.org>
- 3 A news organization typically has more than one official accounts. See Appendix A for a list of the accounts included in the list.
- 4 The website maintained by the STM developers (<http://www.structuraltopicmodel.com/>) updates the full list of academic research that adopted STM.

## References

- Anderson, S. P., & De Palma, A. (2012). Competition for attention in the information (overload) age. *The RAND Journal of Economics*, 43(1), 1–25. doi: 10.1111/j.1756-2171.2011.00155.x

- Anderson, S. P., & De Palma, A. (2013). Shouting to be heard in advertising. *Management Science*, 59(7), 1545–1556. doi: 10.1287/mnsc.1120.1682
- Andrews, C., Fichet, E., Ding, Y., Spiro, E. S., & Starbird, K. (2016). Keeping up with the tweet-dashians: The impact of 'official' accounts on online rumoring. *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, 452–465. doi: 10.1145/2818048.2819986
- Armstrong, C. L., & Gao, F. (2010). Now tweet this how news organizations use twitter. *Electronic News*, 4(4), 218–235. doi: 10.1177/1931243110389457
- Barthel, M. (2018, June). *State of the news media - data and trends about key sectors in the U.S. news media industry*. Retrieved from Pew Research Center: <http://www.journalism.org/fact-sheet/newspapers/>
- Barzilai-Nahon, K. (2008). Toward a theory of network gatekeeping: A framework for exploring information control. *Journal of the Association for Information Science and Technology*, 59(9), 1493–1512. doi: 10.1002/asi.20857
- Berger, G. (2009). How the internet impacts on international news: Exploring paradoxes of the most global medium in a time of hyperlocalism. *International Communication Gazette*, 71(5), 355–371. doi:10.1177/174804850904977
- Berger, J., & Milkman, K. L. (2012). What makes online content viral? *Journal of Marketing Research*, 49(2), 192–205. doi: 10.1509/jmr.10.0353
- Bischof, J., & Airoldi, E. M. (2012). Summarizing topical content with word frequency and exclusivity. *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*, 201–208.
- Blei, D., & Lafferty, J. (2006). Correlated topic models. *Advances in Neural Information Processing Systems*, 18, 147–154. doi: 10.1214/07-AOAS14
- Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77–84. doi:10.1109/MSP.2010.938079
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *The Journal of Machine Learning Research*, 3, 993–1022.
- Bozdog, E., & van den Hoven, J. (2015). Breaking the filter bubble: Democracy and design. *Ethics and Information Technology*, 17(4), 249–265. doi: 10.1007/s10676-015-9380-y
- Brown, P. (2018, April 18, April). Facebook struggles to promote 'meaningful interactions' for local publishers, data shows. *Columbia Journalism Review*. Retrieved from [https://www.cjr.org/tow\\_center/facebook-local-news.php](https://www.cjr.org/tow_center/facebook-local-news.php)
- Bruns, A. (2005). *Gatewatching: Collaborative online news production*. Peter Lang.
- Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J. L., & Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. *Advances in Neural Information Processing Systems*, 288–296.
- Chan-Olmsted, S., Rim, H., & Zerba, A. (2013). Mobile news adoption among young adults examining the roles of perceptions, news consumption, and media usage. *Journalism & Mass Communication Quarterly*, 90(1), 126–147. doi: 10.1177/1077699012468742
- Chan-Olmsted, S. M., & Park, J. S. (2000). From on-air to online world: Examining the content and structures of broadcast TV stations' web sites. *Journalism & Mass Communication Quarterly*, 77(2), 321–339. doi: 10.1177/107769900007700207
- Cleary, J., al Nashmi, E., Bloom, T., & North, M. (2015). Valuing twitter: Organizational and individual representations at CNN international. *Digital Journalism*, 3(6), 908–924. doi: 10.1080/21670811.2014.990255
- Diakopoulos, N. (2017). Computational journalism and the emergence of news platforms. In B. Franklin & S. A. Eldridge II (Eds.), *The Routledge Companion to Digital Journalism Studies* (pp. 176–184). New York, NY: Routledge.
- Diakopoulos, N., & Zubiaga, A. (2014). Newsworthiness and network gatekeeping on twitter: The role of social deviance. *Eighth International AAAI Conference on Weblogs and Social Media*.

- Dimitrova, D. V., Connolly-Ahern, C., Williams, A. P., Kaid, L. L., & Reid, A. (2003). Hyperlinking as gatekeeping: Online newspaper coverage of the execution of an American terrorist. *Journalism Studies*, 4(3), 401–414. doi: 10.1080/14616700306488
- Elizabeth, J. (2017, November 14, November). Social media teams today: A summary of what we learned. Retrieved from *American Press Institute*: <https://www.americanpressinstitute.org/publications/reports/strategy-studies/social-media-teams-today/>
- Evans, D. S. (2003). The antitrust economics of multi-sided platform markets. *Yale Journal on Regulation*, 20, 325–381.
- Faris, R., Roberts, H., Etling, B., Bourassa, N., Zuckerman, E., & Benkler, Y. (2017). *Partisanship, propaganda, and disinformation: Online media and the 2016 us presidential election*. Retrieved from Berkman Klein Center: [https://dash.harvard.edu/bitstream/handle/1/33759251/2017-08\\_electionReport\\_o.pdf](https://dash.harvard.edu/bitstream/handle/1/33759251/2017-08_electionReport_o.pdf)
- García-Perdomo, V., Salaverría, R., Kilgo, D. K., & Harlow, S. (2018). To share or not to share: The influence of news values and topics on popular social media content in the United States, Brazil, and Argentina. *Journalism Studies*, 19(8), 1180–1201. doi: 10.1080/1461670X.2016.1265896
- Gentzkow, M., Shapiro, J. M., & Stone, D. F. (2015). Media bias in the marketplace: Theory. In S. P. Anderson, J. Waldfoegel, & D. Stromberg, *Handbook of Media Economics* (Vol. 1, pp. 623–645). North-Holland. doi: 10.1016/B978-0-444-63685-0.00014-0
- Gottfried, J. (2014, November). *Facebook and twitter as political forums: Two different dynamics*. Retrieved from Pew Research Center: <http://www.pewresearch.org/fact-tank/2014/11/12/facebook-and-twitter-as-political-forums-two-different-dynamics/>
- Greer, C. F., & Ferguson, D. A. (2011). Using twitter for promotion and branding: A content analysis of local television twitter sites. *Journal of Broadcasting & Electronic Media*, 55(2), 198–214. doi: 10.1080/08883815.2011.570824
- Hermida, A. (2010). Twittering the news: The emergence of ambient journalism. *Journalism Practice*, 4(3), 297–308. doi: 10.1080/17512781003640703
- Hong, S. (2012). Online news on twitter: Newspapers' social media adoption and their online readership. *Information Economics and Policy*, 24(1), 69–74. doi: 10.1016/j.infoecopol.2012.01.004
- Jackson, N. (2018, December). More transparency around newsroom decisions. Retrieved from Nieman Lab: <http://www.niemanlab.org/2018/12/more-transparency-around-newsroom-decisions/>
- Kalsnes, B., & Larsson, A. O. (2018). Understanding news sharing across social media: Detailing distribution on Facebook and Twitter. *Journalism Studies*, 19(11), 1669–1688. doi: 10.1080/1461670X.2017.1297686
- Kim, S. E. (2018). Media bias against foreign firms as a veiled trade barrier: Evidence from Chinese newspapers. *American Political Science Review*, 112(4), 954–970. doi: 10.1017/S0003055418000242
- Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., ... Jonathan L. Zittrain (2018). The science of fake news. *Science*, 359(6380), 1094–1096. doi: 10.1126/science.aa02998
- Lewis, S. C. (2012). The tension between professional control and open participation: Journalism and its boundaries. *Information, Communication & Society*, 15(6), 836–866. doi: 10.1080/1369118X.2012.674150
- Lewis, S. C., & Westlund, O. (2015). Big data and journalism: Epistemology, expertise, economics, and ethics. *Digital Journalism*, 3(3), 447–466. doi: 10.1080/21670811.2014.976418
- Lischka, J. A. (2014). Different revenue incentives, different content? comparing economic news before and during the financial crisis in German public and commercial news outlets over time. *European Journal of Communication*, 29(5), 549–566. doi: 10.1177/0267323114538851

- Lowrey, W., & Kim, E. (2016). Hyperlocal news coverage: A population ecology perspective. *Mass Communication and Society*, 19(6), 694–714. doi: 10.1080/15205436.2016.1201687
- Maier, S. (2010). All the news fit to post? comparing news content on the web to newspapers, television, and radio. *Journalism & Mass Communication Quarterly*, 87(3-4), 548–562. doi: 10.1177/107769901008700307
- McDonald, P., & Charlesworth, S. (2013). Framing sexual harassment through media representations. In *Women's studies international forum* (Vol. 37, pp. 95–103). doi: 10.1016/j.wsif.2012.11.003
- Meraz, S. (2011). Using time series analysis to measure intermedia agenda-setting influence in traditional media and political blog networks. *Journalism & Mass Communication Quarterly*, 88(1), 176–194. doi: 10.1177/107769901108800110
- Meraz, S., & Papacharissi, Z. (2016). Networked framing and gatekeeping. In T. Witschge, C. W. Anderson, D. Domingo & A. Hermida II (Eds.), *The Sage Handbook of Digital Journalism* (pp. 950–112). Sage.
- Meyer, K. M., & Tang, T. (2015). #socialjournalism: Local news media on twitter. *International Journal on Media Management*, 17(4), 241–257. doi:10.1080/14241277.2015.1107569
- Mimno, D., Wallach, H. M., Talley, E., Leenders, M., & McCallum, A. (2011). Optimizing semantic coherence in topic models. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 262–272.
- Newman, D., Lau, J. H., Grieser, K., & Baldwin, T. (2010). Automatic evaluation of topic coherence. *Human language technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, 100–108.
- Newman, N. (2011). *Mainstream media and the distribution of news in the age of social discovery*. Retrieved from Reuters Institute for the Study of Journalism: <https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2017-11/Mainstream%20media%20and%20the%20distribution%20of%20news%20in%20the%20age%20of%20social%20discovery.pdf>
- Newman, N., Fletcher, R., Kalogeropoulos, A., Levy, D., & Nielsen, R. K. (2018). Reuters Institute Digital News Report 2018. Retrieved from *Reuters Institute for the Study of Journalism*: <https://reutersinstitute.politics.ox.ac.uk/sites/default/files/digital-news-report-2018.pdf>.
- Palser, B. (2009). Hitting the tweet spot: news outlets should use twitter to reach elusive and valuable audiences. *American Journalism Review*, 31(2), 54–55.
- Rashidian, N., Brown, P., Hansen, E., Bell, E., Albright, J., & Hartstone, A. (2018). *Friend and foe: The platform press at the heart of journalism*. Retrieved from *Tow Center for Digital Journalism*: [https://www.cjr.org/tow\\_center\\_reports/the-platform-press-at-the-heart-of-journalism.php](https://www.cjr.org/tow_center_reports/the-platform-press-at-the-heart-of-journalism.php)
- Roberts, M. E., Stewart, B. M., & Airolidi, E. M. (2016). A model of text for experimentation in the social sciences. *Journal of the American Statistical Association*, 111(515), 988–1003. doi: 10.1080/01621459.2016.1141684
- Roberts, M. E., Stewart, B. M., & Tingley, D. (2016). Navigating the local modes of big data. *Computational Social Science*, 51.
- Roberts, M. E., Stewart, B. M., Tingley, D., Airolidi, E. M. (2013). The structural topic model and applied social science. *Advances in neural information processing systems workshop on topic models: Computation, application, and evaluation*. 1-20.
- Roston, M. (2015, January 22, January). Don't try too hard to please twitter — and other lessons from the new york times' social media desk. Retrieved from *Nieman Lab*: <http://www.niemanlab.org/2015/01/dont-try-too-hard-to-please-twitter-and-other-lessons-from-the-new-york-times-social-media-desk/>
- Rowan, D. (2014, January 2, Jan). How buzzfeed mastered social sharing to become a media giant for a new era. *Wired*. Retrieved from <http://www.wired.co.uk/article/buzzfeed>
- Schwemmer, C., & Ziewiecki, S. (2018). Social media sellout: The increasing role of product promotion on Youtube. *Social Media+ Society*, 4(3), 1-20. doi: 10.1177/2056305118786720

- Shoemaker, P. J., & Vos, T. (2009). *Gatekeeping Theory*. New York: Routledge.
- Sjøvaag, H., & Stavelin, E. (2012). Web media and the quantitative content analysis: Methodological challenges in measuring online news content. *Convergence*, 18(2), 215–229. doi: 10.1177/1354856511429641
- Stempel III, G. H. (1985). Gatekeeping: The mix of topics and the selection of stories. *Journalism Quarterly*, 62(4), 791–815. doi: 10.1177/107769908506200412
- Stieglitz, S., & Dang-Xuan, L. (2013). Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior. *Journal of Management Information Systems*, 29(4), 217–248. doi: 10.2753/MIS0742-1222290408
- Strauß, N. (2018). Financial journalism in today's high-frequency news and information era. *Journalism*, 274–291. doi: 10.1177/1464884917753556
- Tandoc Jr, E. C. (2014). Journalism is twerking? how web analytics is changing the process of gatekeeping. *New Media & Society*, 16(4), 559–575. doi: 10.1177/1461444814530541
- Tandoc Jr, E. C. (2017). Journalistic autonomy and web analytics. In B. Franklin & S. A. Eldridge II (Eds.), *The Routledge Companion to Digital Journalism Studies* (pp. 293–301). New York, NY: Routledge.
- Tandoc Jr, E. C., & Vos, T. P. (2016). The journalist is marketing the news: Social media in the gatekeeping process. *Journalism Practice*, 10(8), 950–966. doi: 10.1080/17512786.2015.1087811
- Thorson, K., & Wells, C. (2015). Curated flows: A framework for mapping media exposure in the digital age. *Communication Theory*, 26(3), 309–328. doi: 10.1111/comt.12087
- Vargo, C. J., Guo, L., & Amazeen, M. A. (2017). The agenda-setting power of fake news: A big data analysis of the online media landscape from 2014 to 2016. *New Media & Society*, 2028–2049. doi: 10.1177/1461444817712086
- Welbers, K., & Opgenhaffen, M. (2018a). Presenting news on social media: Media logic in the communication style of newspapers on Facebook. *Digital Journalism*, 1–18. doi: 10.1080/21670811.2018.1493939
- Welbers, K., & Opgenhaffen, M. (2018b). Social media gatekeeping: An analysis of the gatekeeping influence of newspapers' public Facebook pages. *New Media & Society*, 20(12), 4728–4747. doi: 10.1177/1461444818784302
- Welbers, K., Van Atteveldt, W., Kleijnijenhuis, J., Ruigrok, N., & Schaper, J. (2016). News selection criteria in the digital age: Professional norms versus online audience metrics. *Journalism*, 17(8), 1037–1053. doi: 10.1177/1464884915595474
- White, D. M. (1950). The “gate keeper”: A case study in the selection of news. *Journalism Bulletin*, 27(4), 383–390. doi: 10.1177/107769905002700403
- Yankelovich, D. (1991). *Coming to Public Judgment: Making Democracy Work in a Complex World*. Syracuse Syracuse University Press.

## About the author

**Chankyung Pak:** Amsterdam School of Communications Research,  
Universiteit van Amsterdam  
Correspondence address: c.pak@uva.nl

Creative Commons License CC BY  
(<https://creativecommons.org/licenses/by/4.0/>)