# NEWS ORGANIZATIONS' NEWS LINK SHARING STRATEGIES ON TWITTER: ECONOMIC THEORY AND COMPUTATIONAL TEXT ANALYSIS

By

Chankyung Pak

## A DISSERTATION

Submitted to Michigan State University in partial fulfillment of the requirements for the degree of

Information and Media – Doctor of Philosophy

2018

### ABSTRACT

## NEWS ORGANIZATIONS' NEWS LINK SHARING STRATEGIES ON TWITTER: ECONOMIC THEORY AND COMPUTATIONAL TEXT ANALYSIS

### By

### Chankyung Pak

This dissertation explores news organizations' social media strategies to disseminate their news stories as emergent *quasi-editorial decisions*. As in traditional editorial decisions, how news organizations share their news links on social media determines the visibility of certain stories and makes certain aspects of news stand out or be unobtrusive. By illustrating how these practices on Twitter resemble and deviate from traditional ones, I argue that social media open up a new path by which news organizations mediate news information, relatively free from journalistic norms and routines embedded in the older media and traditional editorial processes.

To answer a fundamental question, "Is there a reason for a news organization to be strategic on social media?" I presented an economic model based on competition for limited attention of social media users. In this model, the scarcity of users' attention capacity relative to the volume of information propagated via social media creates competition between news organizations. The model illustrates that one news organization's attempt to capture users' attention undermines the chance for other organizations to do so. Thus, a news organization should *strategically* decide how many news links it is going to share considering how many *others* would share. A simple empirical test confirms the model's prediction that news organizations will reduce the proportion of news links they share on social media as more news is published by all organizations.

Computational text analyses shed light on the more qualitative aspects of news dissemination strategies on social media. First, using a recently developed machine learning technique, Structural Topic Model (STM), I investigate news organizations' selective news link sharing as a new layer of gatekeeping. The result indicates that the common concern that commercialized media drives news toward human interest rather than newsworthiness is crystallized more visibly on Twitter than on news websites. Further, a comparison of the 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. Even though a news organization may consider a certain topic to be important in its editorial decision, so that the organization has become *specialized* in that topic throughout its history, it would not share much about the topic on Twitter because popularity in the short term dominates link sharing strategies.

I found that regional media convey less negative sentiment through news stories than other types of news organizations. This seems to be associated with the less controversial news topics they frequently cover compared to national and online media. However, news paraphrasing for Twitter homogenizes emotional framing across different types of organizations. In particular, regional media catch up to other types by adding even more negativity on news paraphrases for Twitter. This finding provides another significant indication that social media strategies are governed by different logic than that which governs traditional editorial practices.

Major empirical findings provide evidence that the social media strategies of news organizations are already functioning as a separate information-mediating process. I argue that the distinctiveness of social media strategies as *quasi-editorial decisions* raises a practical need to publicly monitor news organizations' behaviors on social media to learn whether they will provide news that is informative and diverse enough for news readers' informed decisions. The automated data collection schemes and computational text analysis techniques I adopted in this dissertation will inform the design of infrastructure for such public monitoring.

Copyright by CHANKYUNG PAK 2018 To J

#### ACKNOWLEDGEMENTS

One important lesson I learned from years for this dissertation is that a single person's hard work is not even close to enough for a doctoral degree. I would like to thank all who supported me in both academic and nonacademic communities during my PhD work. Here I would like to acknowledge a few of those who made specific contributions to this dissertation.

First of all, I would like to thank my family I have in Korea and here in US. I am grateful for having an amazing life-parter, Jina Yu. Not only did she support me sharing everyday, but also gave many insights that helped me improve research designs and ways to present my work to broader audience. Chatting with her at a coffee shop and during commuting time was more valuable than anything I learned from academic sources.

I was fortunate to have a supportive dissertation committee. How Rick Wash, my advisor and co-author, nurtured me throughout my PhD work is beyond a simple word, support. He taught me how to think, read and write as an interdisciplinary scholar, by serving as a role model. He also provided generous financial support for a long time. Steve Wildman has been a great mentor giving me invaluable wisdom about research and life. His lectures and countless conversations with him inspired my model work in Chapter 3, which was the inception of this dissertation. He carefully read multiple versions of drafts and provided me with very detailed comments. As my committee member and a chairperson of Department of Media and Information, Johannes Bauer has been a great defender of interdisciplinary research in media field, and always enthusiastic in suggesting ways to expand my work. Esther Thorson provided me with many insights about how I can situate my work in the realm of journalism studies, which greatly helped me forge my academic identity.

In addition to my committee, other faculty members contributed to my dissertation in various ways. Emilee Rader, as a co-PI of Behavior, Information, and Technology Lab (BITLab) at MSU and my co-author, gave me generous support and a great deal of advice about research and being a good scholar. Kjerstin Thorson, one of my co-authors, helped me expand my perspective to political communication. Winson Peng carefully read my proposal draft and gave me valuable comments.

I owe thanks to BITLab members, Yumi Jung, Wenjuan Ma, Chris Fennell, Janghee Cho, and Hope Chidziwisano to name a few, who shared brilliant insights in weekly meetings and everyday conversations at the lab space for years. I was fortunate to have two old friends, Youngwoo Jeung and Jungyoung Kim, who work on their own PhD works here in US. They have been generous enough to catch my random phone calls any time and share their thoughts about my unorganized ideas.

I want to say a special thank you to THE friends from Information and Media doctoral program at MSU; Robin Blom, Kelley Cotter, Julia DeCook, Shaheen Kanthawala, Mel Medeiros and Janine Slaker. They have been my great conversation partners, co-authors, and above all, enthusiastic cheerleaders. They are the best luck that happened to my PhD life.

Finally, College of Communication Arts and Sciences financially supported this dissertation in the form of a Dissertation Completion Fellowship.

# TABLE OF CONTENTS

LIST O	F TABLES	Х
LIST O	F FIGURES	xi
KEY T	O ABBREVIATIONS	xv
CHAPT	TER 1 INTRODUCTION	1
1.1	News Organizations' Adaptation to Social Media	1
1.2	Social Media Strategies on Twitter as Quasi-Editorial Decisions	4
CHAPT	TER 2 BACKGROUND AND RELATED WORK	8
2.1	Social Media and New Journalism	8
	2.1.1 Individual and Organizational Use of Twitter for Journalism $\ldots$	8
	2.1.2 Twitter as a News Distribution Platform	11
	2.1.3 Emerging News Outlets and Social Media	12
2.2	News Dissemination Strategies on Social Media	13
	2.2.1 What Makes Social Media Different as a News Distribution Platform?	13
	2.2.2 Competition for Limited Attention on Social Media	15
	2.2.3 Selective News Link Sharing as Gatekeeping	17
	2.2.4 Sentiment and News Framing	19
2.3	Contribution of Dissertation	22
~~~ ~ ~ ~ ~		
CHAP'I	TER 3 A CONCEPTUAL MODEL: AN ECONOMIC MODEL OF NEWS LINK SHARING	25
CHAP'I 3.1	TER 3       A CONCEPTUAL MODEL: AN ECONOMIC MODEL OF NEWS         LINK SHARING         Introduction	$\frac{25}{25}$
3.1 3.2	TER 3       A CONCEPTUAL MODEL: AN ECONOMIC MODEL OF NEWS         LINK SHARING         Introduction	25 25 26
CHAP'I 3.1 3.2 3.3	TER 3       A CONCEPTUAL MODEL: AN ECONOMIC MODEL OF NEWS LINK SHARING         Introduction	$25 \\ 25 \\ 26 \\ 28$
3.1 3.2 3.3	TER 3 A CONCEPTUAL MODEL: AN ECONOMIC MODEL OF NEWS LINK SHARING         Introduction	25 25 26 28 28
3.1 3.2 3.3	TER 3 A CONCEPTUAL MODEL: AN ECONOMIC MODEL OF NEWS LINK SHARING         Introduction	25 25 26 28 28 28
3.1 3.2 3.3 3.4	<ul> <li>TER 3 A CONCEPTUAL MODEL: AN ECONOMIC MODEL OF NEWS LINK SHARING</li> <li>Introduction</li></ul>	25 25 26 28 28 29 30
3.1 3.2 3.3 3.4	TER 3 A CONCEPTUAL MODEL: AN ECONOMIC MODEL OF NEWS LINK SHARING         Introduction	25 25 26 28 28 29 30 32
CHAPT 3.1 3.2 3.3 3.4 3.5	<ul> <li>TER 3 A CONCEPTUAL MODEL: AN ECONOMIC MODEL OF NEWS LINK SHARING</li> <li>Introduction</li></ul>	25 25 26 28 29 30 32 33
3.1 3.2 3.3 3.4 3.5 CHAPT	<ul> <li>TER 3 A CONCEPTUAL MODEL: AN ECONOMIC MODEL OF NEWS LINK SHARING</li> <li>Introduction</li></ul>	25 25 26 28 29 30 32 33 33
CHAPT 3.1 3.2 3.3 3.4 3.5 CHAPT 4.1	<ul> <li>TER 3 A CONCEPTUAL MODEL: AN ECONOMIC MODEL OF NEWS LINK SHARING</li> <li>Introduction</li></ul>	25 25 26 28 29 30 32 33 33 35 36
CHAPT 3.1 3.2 3.3 3.4 3.5 CHAPT 4.1	TER 3 A CONCEPTUAL MODEL: AN ECONOMIC MODEL OF NEWS LINK SHARING         Introduction	25 25 26 28 29 30 32 33 33 35 36 36
CHAPT 3.1 3.2 3.3 3.4 3.5 CHAPT 4.1	TER 3       A CONCEPTUAL MODEL: AN ECONOMIC MODEL OF NEWS LINK SHARING         Introduction	25 25 26 28 29 30 32 33 35 36 36 37
CHAPT 3.1 3.2 3.3 3.4 3.5 CHAPT 4.1 4.2	TER 3       A CONCEPTUAL MODEL: AN ECONOMIC MODEL OF NEWS LINK SHARING         Introduction	25 25 26 28 29 30 32 33 33 35 36 36 37 40
CHAPT 3.1 3.2 3.3 3.4 3.5 CHAPT 4.1 4.2	TER 3 A CONCEPTUAL MODEL: AN ECONOMIC MODEL OF NEWS LINK SHARING         Introduction         Competition for Limited Attention Model         News Link Sharing under Users' Limited Attention         3.3.1 A Model Setting         3.3.2 Symmetric Nash Equilibrium         Results: Over-sharing Hypothesis         3.4.1 Comparative Statics         Discussion of the Conceptual Model and An Empirical Hypothesis         TER 4 AUTOMATED DATA COLLECTION AND COMPUTATIONAL ANALYSIS OF NEWS DISTRIBUTION ON TWITTER         Automated Data Collection         4.1.1 Data Collecting Software         4.1.2 Description of Data         Computational Text Analysis         4.2.1 Text Preprocessing	25 25 26 28 29 30 32 33 35 36 36 36 37 40 41
CHAPT 3.1 3.2 3.3 3.4 3.5 CHAPT 4.1 4.2	<ul> <li>FER 3 A CONCEPTUAL MODEL: AN ECONOMIC MODEL OF NEWS LINK SHARING</li> <li>Introduction</li></ul>	25 25 26 28 29 30 32 33 33 35 36 36 36 37 40 41 42

СНАРТ	TER $5$	A SIMPLE TEST OF THE STRATEGIC LINK SHARING MODEL	49
5.1	Introd	uction	49
5.2	Analys	Sis	50
	5.2.1	Empirical Models	50
	5.2.2	Results	51
5.3	Discus	sion	55
СНАРТ	TER 6	SELECTIVE NEW LINK SHARING ON TWITTER: A STRUC-	
		TURAL TOPIC MODEL	57
6.1	Introd	uction	57
6.2	Motiva	ation and Related Works	58
	6.2.1	Social Media as A News Distribution Platform	59
	6.2.2	News Organizations' Online News Choice	60
6.3	Compi	utational Identification of Selective News Link Sharing	64
	6.3.1	News Topic Categorization	64
	6.3.2	Overall Patterns of Selective Link Sharing	67
	6.3.3	Difference Across Organizations	72
64	Discus	sion	82
0.1	Diseus		02
СНАРТ	TER 7	NEWS PARAPHRASING AND SENTIMENT	87
7.1	Introd	uction: News Paraphrasing as a Strategy	87
7.2	Relate	d Works	89
	7.2.1	Paraphrasing News on Social Media	89
	7.2.2	Negative Framing in Online News	91
7.3	Result	§	94
1.0	731	Negativity of News by Organizational Types	94
	732	Strength of Extra-Negative Framing	96
7.4	Discus	sion	98
1.1	Diseus		00
СНАРТ	TER 8	CONCLUSION: SOCIAL MEDIA STRATEGIES AS QUASI-EDITORI.	AL
		DECISIONS	103
8.1	News 1	Link Sharing as a Strategic Choice	104
8.2	Selecti	ve News Link Sharing as Gatekeeping	107
8.3	Extra-	negative Framing for Paraphrases	110
8.4	Movin	σ Forward	112
0.1	1110 111	5 loi wara li	
APPEN	DICES		115
APP	ENDIX	A PROOFS OF LEMMAS AND PROPOSITIONS IN CHAP-	
		TER 3	116
APF	PENDIX	X B DETAILS OF DATA COLLECTION	118
APF	PENDIX	C ADDITIONAL STM GRAPHS AND FULL OUTPUT	124
APP	ENDIX	X D SENTIMENT ANALYSIS OUTPUTS BASED ON LSD AND	
4 <b>11</b> 1		AFINN	134
			104
BIBLIO	GRAPI	HY	135

# LIST OF TABLES

Table 4.1:	The list of news organizations included in the dataset	38
Table 5.1:	Regression of the proportion of news stories shared by news organizations on Twitter	52
Table 6.1:	Words likely to appear given each topic	68
Table 6.2:	Summary of the selective news link sharing strategies. Numbers in paren- thesis are number of topics categorized as a given strategy.	82
Table 7.1:	Linear regression of sentiment gap on news sentiment and news organizations type.	98
Table B.1:	List of Twitter accounts included in the dataset.	119

# LIST OF FIGURES

Figure 3.1:	Conditions for SNE.	32
Figure 4.1:	A computational method to identify shared and unshared news on Twitter.	37
Figure 4.2:	The average proportion of news sharing of each news organization against the average number of news stories daily published with a smoothed trend. Each dot represents a news organization	39
Figure 4.3:	Daily Number of Published Online News Stories. The data weren't collected during the period between the blue dotted lines	40
Figure 4.4:	Text pre-processing procedure	41
Figure 4.5:	A graphical representation of sharing-dependent topic model	45
Figure 6.1:	A bar graph for estimated overall proportions of 30 topics. The unit is fraction	69
Figure 6.2:	Relative propensity to be shared on social media of 30 topics. The propensity is measured as a fractional deviation of the proportion of a news topic among shared stories from that among published stories. The topics are ordered by the frequency on news websites. The propensity to be shared and frequency on websites are not notably associated	70
Figure 6.3:	News publication and link sharing patterns by news organizations: Pol- itics/President topic	74
Figure 6.4:	News publication and link sharing patterns by news organizations: Sports/Basketball topic	75
Figure 6.5:	News publication and link sharing patterns by news organizations: Im- migration topic	75
Figure 6.6:	News publication and link sharing patterns by news organizations: In- ternational/Middle East topic	76
Figure 6.7:	News publication and link sharing patterns by news organizations: In- ternational/Far East topic	76
Figure 6.8:	News publication and link sharing patterns by news organizations: Sports/Football topic.	77

Figure 6.9:	News publication and link sharing patterns by news organizations: Crime.	78
Figure 6.10:	News publication and link sharing patterns by news organizations: Transportation.	79
Figure 6.11:	News publication and link sharing patterns by news organizations: Pol- itics/Congress topic.	80
Figure 6.12:	News publication and link sharing patterns by news organizations: So- cial/Gender topic	81
Figure 6.13:	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.	85
Figure 7.1:	Average sentiment scores of news stories and tweets: LIWC dictionary.	95
Figure 7.2:	Sentiment level of news stories by news organization type	96
Figure 7.3:	Sentiment level of tweets by news organization type	99
Figure 7.4:	Additional negative sentiment to tweets by news organization type	99
Figure B.1:	An example of shortened news URL in a tweet.	123
Figure C.1:	News publication and link sharing patterns by news organizations: Life/In- terview topic	124
Figure C.2:	News publication and link sharing patterns by news organizations: Econ- omy/Technology East topic	124
Figure C.3:	News publication and link sharing patterns by news organizations: So- cial/Identity Conflict topic.	125
Figure C.4:	News publication and link sharing patterns by news organizations: Legal/Regulation topic.	- 125
Figure C.5:	News publication and link sharing patterns by news organizations: Econ- omy/Finance topic	126
Figure C.6:	News publication and link sharing patterns by news organizations: UK topic.	126

Figure C.7:	News publication and link sharing patterns by news organizations: Up- date topic	127
Figure C.8:	News publication and link sharing patterns by news organizations: Sports/Others topic	127
Figure C.9:	News publication and link sharing patterns by news organizations: En- tertainment topic	128
Figure C.10	News publication and link sharing patterns by news organizations: Re- gional Politics topic.	128
Figure C.11	News publication and link sharing patterns by news organizations: Weather topic.	129
Figure C.12	News publication and link sharing patterns by news organizations: Health topic.	129
Figure C.13	News publication and link sharing patterns by news organizations: Sub- scription topic.	130
Figure C.14	News publication and link sharing patterns by news organizations: ITax topic.	130
Figure C.15	News publication and link sharing patterns by news organizations: So- cial Media topic.	131
Figure C.16	News publication and link sharing patterns by news organizations: Education topic.	131
Figure C.17	News publication and link sharing patterns by news organizations: Sport- $s/College$ topic.	132
Figure C.18	News publication and link sharing patterns by news organizations: Me Too topic	132
Figure C.19	News publication and link sharing patterns by news organizations: French topic.	133
Figure C.20	News publication and link sharing patterns by news organizations: Advertising topic.	133
Figure D.1:	Average sentiment scores of news stories and tweets: AFINN dictionary. Zero score indicates neutral sentiment.	134

Figure D.2:	Average sentiment scores of news stories and tweets: LSD dictionary.	
	Zero score indicates neutral sentiment.	134

## **KEY TO ABBREVIATIONS**

- **LDA** Latent Dirichlet Allocation
- ${\bf NAS}\,$  Network Agenda Setting
- ${\bf SEO}~{\rm Search}~{\rm Engine}~{\rm Optimization}$
- ${\bf SMO}$ Social Media Optimization
- ${\bf STM}$ Structural Topic Model

#### CHAPTER 1

### INTRODUCTION

## 1.1 News Organizations' Adaptation to Social Media

Technological changes on a platform create platform users' new economic incentives. As a recent report about Facebook's algorithm change in 2018 pointed out, Facebook's longstanding, thoughtful design effort to make its platform a new public sphere unexpectedly discovered a niche for rampant misinformation about political issues (Ingram, 2018, January 22). As in many cases involved with technological uncertainty, this unexpected discovery even appears to have caused the platform owners to give up on their vision for their business, establishing their platform as an online public square (Warzel, 2018, January 12). However, despite the indispensable role of media organizations for journalism, the academic community and industry still lack an understanding of the economy behind the organizations' adaptation and survival to social media, compared to human users' reactions. This dissertation aims to illustrate how new economic incentives spurred by social media as the new news distribution platform for news organizations led to an additional layer of information mediating processes, in addition to the traditional editorial decisions, potentially with different logic.

Journalism is a socio-technical network in that human actors – typically journalists and their audience – and a variety of technologies related to information generation and dissemination collaborate to define and re-define information mediating practices (Turner, 2005). The multiplicity of the human actors and technologies and their interdependence makes it hard to evaluate the impact of a technological advance on journalism. When a new technology emerges, both information generators and audiences adapt their cultures, routines and norms, which in turn gives feedback to the technological design.

To understand this complexity, scholars have focused on specific aspects of journalism as an ever-changing, socio-technical network. Some focus on how audiences' cultures coevolve with technologies, mutating the social role of journalism (Singer, 2005; Deuze, 2008; Hermida, 2010). Some scholars zoom in on journalists' relationships within an organization and focus on changes in their explicit and implicit rules and news-making processes (Cottle, 2007; Domingo & Paterson, 2011). Other scholars pay attention to determining the power of new, information-curating technologies, such as recommendation algorithms (Pariser, 2011; Bakshy et al., 2015) and social networking features, rather than human actors (Hermida et al., 2012).

In this discussion, which tends to focus on the novel aspects of transforming the media environment, news organizations as a business are often portrayed merely as victims losing their jurisdiction. However, news organizations are still powerful actors retaining scarce resources, compared to other emergent actors, such as participatory users and platform companies. Although their privilege is being undermined, news organizations have reporting infrastructure, longstanding institutional and personal relationships with news sources with power, and editorial processes, which give them authority for factual information. Indeed, news readers still put more weight on "official" reporting from news organizations than on individual journalists or participatory audiences, especially when information the readers seek is more personally relevant (Schmierbach & Oeldorf-Hirsch, 2012; Thomson et al., 2012). Thus, how news organizations react to new technologies, such as social media, still has significant implications for users' news consumption in the transforming media environment.

A focus on news organizations' adaptation to new online technologies presnet a need to consider their economic goals. The traditionally conceived goal was to provide objective information that was newsworthy enough to inform citizens' decisions (Hermida, 2010). Hierarchical editorial decisions and a fact checking process by the professional desk have been an outcome of the social and organizational evolution of objectivism based on cheaper printing and delivery cost that reach back to the late 19th century (Gentzkow et al., 2006). However, there has always been a tension between journalistic norms and the pursuit of profitability in journalism (Gentzkow & Shapiro, 2010). By offering a new news dissemination platform that is relatively free from the traditional editorial decisions, social media are likely to add weight to the profitability motivation.

As more readers find news from social media rather than traditional sources (Newman et al., 2016; Shearer & Gottfried, 2017), news organizations have recognized the importance of social media strategies (Newman, 2011). Because the distribution of news is under the influence of the unique socio-technical traits of social media, news organizations strive to understand what happens to their news as it is disseminated via social media (Diakopoulos, 2017). These traits include information overload by pushing technology (Weltevrede et al., 2014), information cascade (Lerman & Ghosh, 2010), selective exposure based on social networks (Bakshy et al., 2015), and filter bubble by recommendation algorithms (Pariser, 2011; Bozdag & van den Hoven, 2015). However, optimizing online news dissemination for their audience as they understood using Web analytics techniques increasingly detaches journalism practices from the traiditional social goal of journalism (Tandoc Jr, 2016).

The shifting goal of news organizations is likely to result in new decision rules. This point has a powerful implication for the discussion about the improvement of social media algorithms. Researchers often suggest design improvements based on users' current behaviors, which they have observed by surveys or observational data analysis. Yet, in an influential paper in economics, Robert Lucas Jr. argued that a policy evaluation based on the current behavior of economic actors is fallible because the policy also changes actors' decision rules (Lucas Jr, 1976). Similarly, to understand how actors will react to algorithmic advances, one needs to base the discussion on new incentives and motivations, which will govern the new decision rule. In this vein, I relate empirical findings from the behavior of news organizations on Twitter, one of the most popular social media platforms used as a news distribution platform, to newly found economic incentives, relying on the economic modeling and predictions from media studies in this dissertation.

## 1.2 Social Media Strategies on Twitter as Quasi-Editorial Decisions

Although different actors – not only news organizations, but also readers and news aggregators – contribute to news dissemination in a variety of forms in the new media environment, outcomes of such contributions often boil down to a few common functions that select, reduce, and filter information. When an actor chooses to share a news story on an information platform, the news story becomes more salient to others. Further, as an actor rephrases what she reads from other sources, a certain aspect of the news becomes more salient. These outcomes are analogous to news organizations' traditional functions, such as gatekeeping, agenda-setting, and framing, except that those functions are now distributed across different types of actors rather than exclusively occupied by news organizations.

Recognizing the similarity in outcomes between traditional journalism practices and distributed information mediating processes, scholars have tried to extend concepts built on traditional journalism to a variety of ways for actors to intervene in news dissemination. Gatekeeping theory is extended as multiple steps by journalists, individuals, strategic communicators, and algorithms (Wallace, 2018), or a network of multi-directional processes of those steps (Barzilai-Nahon, 2008). Whether to provide a hyperlink to certain information is a new primary form of online gatekeeping according to these studies (Dimitrova et al., 2003). As a gatekeeping role is distributed across different actors, the power of agenda setting is also distributed as an outcome (Russell Neuman et al., 2014; Vargo et al., 2014, 2017). Scholars also found that non-journalists have the power of news framing as they interpret and reinterpret news as it is propagated through networked paths (Meraz & Papacharissi, 2013). I define *quasi-editorial decisions*, as an overarching term that encompasses these various forms of news information mediating processes, which are happening nontradiationally outside the customary boundary of journalism.

In this dissertation, I consider the news dissemination strategies on social media of news organizations as a kind of quasi-editorial decisions. News organizations decide what information to show readers by choosing news links to share on social media, as in gatekeeping. They focus on specific aspects of a news story as they paraphrase the story for a social media post, e.g. as in mkaing headlines. Thus, to the extent that news readers migrate to social media, and that the news organizations' choices on social media deviate from their traditional editorial decisions, the social media strategies are likely to plae an additional layer of information mediating process on the audience. This is a significant phenomenon, which may impact everyday news reading, if one considers that social media have become an major news source for readers. In 2017, two-thirds of U.S. adults identified social media as their news source, and twenty percent of them reported that they *often* get news from social media (Shearer & Gottfried, 2017). Social media come ahead of TV as a news source for the 18 to 24 year-old age group (Newman et al., 2016).

I relate this additional layer of the information mediating process to the economic incentives of news organizations, which social media newly established as a news distribution platform. Although social media provide a new opportunity to distribute news stories, they also create a new competitive environment for news organizations. In additiona to news stories disseminated by the traditional media organizations, readers also produce news-related content in varying degrees on social media; they create their own news stories, write blog posts, share news links, and comment on the news. Consequently, news organizations must compete with this variety of information to reach readers. From the perspective of readers, allocating their *limited attention* to desirable news information is becoming an important task. Thus, capturing readers' limited attention poses a challenge to news organizations that are forming media strategies (Anderson & De Palma, 2013). News organizations strive to capture users' attention with a variety of methods. Recent reports show that they increasingly hire data scientists for viral marketing and social media optimization (SMO) and adopt social media monitoring platforms (Rowan, 2014, January 2; Diakopoulos, 2017). In many organizations, SMO specialists work in the context of audience management rather than in the editorial process, although the degree to which this is done varies (Roston, 2015, January 22; Elizabeth, 2017, November 14). This evidence strongly suggests that the behavior of news organizations on social media is likely to be governed by economic incentives stemming from competition for social media users' limited attention.

Throughout this dissertation, I use Twitter as a typical case of social media as a news dissemination platform. There are several reasons for this choice: First, the interest in the organization level news dissemination naturally limits the choice to Facebook and Twitter, where most news organizations regularly share hyperlinks to their published stories via their official accounts. In 2017, these two services are among the three social media websites used most frequently as a pathway to news stories (Shearer & Gottfried, 2017). Although Youtube and Instagram are also significant news sources, the specific purpose of those sites - video and image sharing – prevents them from being used as a general news distribution platform. Second, Twitter retains the primary features I try to capture as the main drivers of the dissemination strategies on social media of news organizations as distinguished from traditional editorial decisions. Those features are information overload intensified by the data stream and information suppliers' sensitivity to user reactions mainly due primarily to the viral nature of the information dissemination process on social media. It is easier to tie observations from Twitter with the main features of interest compared to Facebook because Twitter relies less on algorithmic curation and has been relatively stable in sophisticating their algorithms. In addition, Twitter API provides a more accessible way to collect news organizations' behaviors.<sup>1</sup> Apparently, observed patterns from Twitter cannot be directly generalized to Facebook. However, my general conjecture that information overload and virality on social media as a news dissemination platform give rise to an additional information mediating process is still valid for Facebook, if the conjecture is empirically verified.

The purpose of this dissertation is to illuminate the significance of the emergent information mediating processes governed by profitability from social media rather than from traditional journalistic norms. To do so, I first formalize an economic theory; social media users' limited attention creates a strategic situation wherein news organizations should

<sup>&</sup>lt;sup>1</sup>Although Twitter has curating functionalities such as a separate section for recommended tweets and promoted tweets, a major part of its interface remain real-time.

consider other organizations' choices in Chapter 3. Using a statistical analysis of a Webscraped dataset described in Chapter 4, I show that news organizations' behaviors approve the model's prediction in Chapter 5. Subsequently in Chapter 6, I analyze how news organizations selectively share news links conditioned on the news content. In Chapter 7, using computational text analysis techniques, I show how they add an emotional frame as they paraphrase news stories for social media posts. I further discuss how these behaviors differ depending on types of news organizations, in gauging the potential impact of the emergent news outlets, which actively use social media as their major news distribution platform.

Major empirical findings reveal that news organizations selectively share popular news topics on Twitter, deviating from their editorial decisions for their website. They add extranegative framing on news as paraphrasing for tweets, which provides evidence that news organizations' social media strategies are already functioning as a separate information mediating processes. These results imply a practical need to publicly monitor the social media strategies of news organizations as quasi-editorial decisions to learn whether the profitability motivation from social media will lead to the provision of news that is informative and diverse enough for citizens' informed decisions, as we have expected from traditional journalism. The automated data collection schemes and computational text analysis techniques I proposed to apply in Chapter 4 will inform the design of the infrastructure for such public monitoring.

### CHAPTER 2

### BACKGROUND AND RELATED WORK

This dissertation focuses on *organizational level* social media use of news organizations to disseminate news stories that potentially impacts news reading. To motivate the research, I begin by reviewing journalism literature about the potential for new journalism that is emerging from social media that focuses primarily on *individual journalists*' Twitter use that bypasses the traditional editorial decisions. Although these works raise a fundamental question – whether social media weaken the gatekeeper role of journalism by offering a means to unveil the news-making process – they tend to presume that information from individual journalists has a significant impact on readers. Built on an assessment that information on social media suffers from a credibility issue, I suggest redirecting the focus to the behaviors of news organizations, and their news distribution strategies via social media. I mainly use Twitter as a case of *social media as a news distribution platform* where news organizations regularly share hyperlinks to their published news stories. I subsequently review new media theories and behavioral economic theories that hint at socio-technical traits, which may influence news organizations' news dissemination strategies on social media. Finally, I motivate my research by arguing that the use of social media by news organizations can emerge as an important information mediating process that influences what news readers think about, and how they think about news.

# 2.1 Social Media and New Journalism

### 2.1.1 Individual and Organizational Use of Twitter for Journalism

Journalism studies about the emergence of new journalism on social media, particularly on Twitter, tend to confine their attention to individual journalists' social media use rather than that of news organizations (Lasorsa et al., 2012). By observing journalists' personal opinions, the reporting processes of covering ongoing events, and how they gather information from other social media users, news readers get to observe how news is produced. On the one hand, this transparency makes news more accountable. On the other hand, news readers may reduce their belief that journalists and news they produce are objective (Lawrence et al., 2014; Mourão, 2015).

These works reveal the fundamental tension between more accountability for the news making process and adherence to the journalistic norms, based on manual content analyses of a relatively small number of sample tweets from journalists. For example, Lawrence et al. (2014) hand-coded 1,946 sample tweets by 430 political reporters during the 2012 US presidential election season, and argue that journalism became somewhat transparent insofar as reporters often expressed their opinions and shared anecdotes from the reporting process with tweets. However, the authors conclude that the news making process is still one-way in the sense that reporters do not often seek information from readers or share substantive information about news making, such as fact-checking processes. Analyzing 5,700 tweets about the first presidential debate in 2012 from 430 reporters, Mourão (2015) draws a similar conclusion; although journalists are willing to depart from the objectivity norm by sharing their opinions, humor and sarcasm, they tend to reinforce their authority by building a community around themselves rather than communicating with an audience. Lasorsa et al. (2012) summarizes the main tension between the traditional norm of objectivity and a newer tendency of opinion sharing and humorous comments as the normalization process that was Singer (2005) earlier characterized political blogs by journalists. As Lawrence et al. (2014) put it, journalists use new affordances of Twitter to break from the norm while the normalizing power of conventions and routines is recapturing the tweets as in the one-way news reporting of the traditional news outlets. Indeed, Parmelee (2013) revealed journalists feel this tension because they are using Twitter as a news reporting tool.

How news organizations use Twitter on the organizational level may be different, however, from how individual journalists intentionally utilize the affordances of Twitter. Empirical research shows that the organizational decision of American news media tends to conform to economic incentives (Gentzkow & Shapiro, 2010) whereas it has been widely studied that individual journalists comply with conventions and routines established in the long run (Lowrey, 2009). Further, in general, journalists' perception of the role of the media deviates from the structure of news organizations with which they are affiliated (Zhu et al., 1997). This result resonates with more general findings from organization studies, in which norms within an organization facilitate the deviation of employees' behavior from employer's expectations, which stems from changes in incentives in a new business environment (Kaplan & Henderson, 2005). Although individual employees in news organizations ultimately compose social media posts, there are indeed difference between journalists' posts and what they write "on behalf of' news organizations. For example, Cleary et al. (2015) reported that news organizations' official tweets are mostly news link sharing whereas journalists' individual tweets were more frequently about promotion and interaction with users. The difference is likely to become clearer as news organizations hire social media specialist and data scientists to customize their social media posts rather than relying on reporters (Roston, 2015, January 22; Elizabeth, 2017, November 14).

The potential disagreement between individual journalists' Twitter use and that of news organizations raises a question about whether the use of social media use on an individual level, however new it is, is the most fundamental change that potentially transforms the everyday reading of the news. Individual journalists' tweets are likely to be susceptible to the credibility issues from which information on social media generally suffers (Schmierbach & Oeldorf-Hirsch, 2012). Thomson et al. (2012) report that people believe official news outlets more than individual journalists. An et al. (2014) found that news readers believe news information significantly less on Twitter when the news is tweeted by individual friends in disaster situations. This assessment brings us to a need to explore organizational level strategies for disseminating news stories via social media to thoroughly evaluate how social media as a new information platform can transform journalism and everyday news reading.

#### 2.1.2 Twitter as a News Distribution Platform

Despite the potential of social media to facilitate new types of journalism, traditional news reporting still plays a major role in news distribution and consumption on social media. Pew Research's 2009-2010 study reported that, among online news readers, only 6% responded that special interest news sites are their favorite news site, and 5% answered bloggers' sites are their favorite. A substantial portion still prefers websites run by traditional news organizations (Purcell et al., 2010). Collecting trending topics from the *Tweetersphere*, Kwak et al. (2010) found that the majority of tweets were reactions to headline news.

Moreover, traditional news outlets do not appear to take advantage of the new opportunities that many observers expect Twitter to provide. Media scholars expected that the Web would afford news organizations a channel for the mutual interaction between the media and its audience (Chan-Olmsted & Park, 2000), and an effective promotion tool to attract younger audiences who do not regularly access traditional media (Palser, 2009; Chan-Olmsted et al., 2013). However, evidence that news organizations are using these opportunities is rather weak. For example, Greer & Ferguson (2011) analyzed tweets from 488 local TV stations, and found out that only 23.3% of 455 commercial TV stations tweet for interaction with news readers. Similarly, Meyer & Tang (2015) recently hand-coded 4,507 tweets from 60 local news organizations (TV and newspaper), and only 7.4% of the tweets from local television stations and 11.6% from local newspapers were intended for interaction. Cleary et al. (2015) drew a similar finding in tweets from CNN International channel. These studies also generally conclude that traditional news companies are not engaged in promotion for either their website, or the organization's brands (Greer & Ferguson, 2011; Meyer & Tang, 2015).

News organizations seem rather to use Twitter as another news distribution platform. Greer & Ferguson (2011) found 94.9% of the commercial TV stations tweet to disseminate their news articles whereas only 17.6% tweet to promote their programs. Similarly, Meyer & Tang (2015) reported that 94.4% of the tweets from TV stations and 96.3% from newspapers are for news link sharing. The news distribution through Twitter is a profitable tactic. Hong (2012) found that the Twitter use and the number of tweets by the news organizations induce more traffic toward their news websites. Therefore, news link sharing would be a major concern when news companies contemplate their social media use.

### 2.1.3 Emerging News Outlets and Social Media

As a new news distribution platform, social media bring fresh players into the media ecosystem by sustaining alternative news outlets (Nicholls et al., 2016; Rauch, 2015). Research has described various kinds of emerging media enabled by the Internet's capability of matching the new services with niche demand in the long tail (Anderson, 2007). Earlier works focused on 'niche news' that targets narrowed interests, such as technological gadgets and subcultures. However, many of the niche news ventures became a part of the traditional news ecosystem by providing narrowly focused news stories to traditional media organizations (Grueskin et al., 2011; Cook & Sirkkunen, 2013).

Other types of emerging news outlets that followed the niche news are news curation services, which aggregate news information from other news organizations. Many news curation services followed the normalization process by which they became a regular news organization. In 2011, for example, Weber & Monge (2011) found that Huffington Post was playing a role of hub, similar to Google News or Yahoo! News by applying a modified version of the hub-authority model (Hyperlink-Induced Topic Search; HITS) (Kleinberg, 1999) to a content sharing network between online news outlets. Since then, however, Huffington Post has grown as one of the most popular news sources in the United States that produces its own original content even ahead of CNN (Newman et al., 2016). Fact-checking websites emerged as an alternative form of information sources, but according to Lowrey (2017)'s assessment, now their organizational behaviors partially conform to traditional journalistic conventions and routines.

The success of the emerging news websites is due primarily to the use of social media.

These emerging news organizations are vigorously using social media as their major distribution platform (Newman et al., 2016), and unlike many legacy media, they explicitly aim at the virality of stories to gain popularity among social media users. According to Faris et al. (2017)'s analysis, tweets from online only news websites such as Huffington Post, Breitbart and The Hill were shared more frequently by users than those from most of the legacy media during the 2016 US presidential election season. The only legacy media whose tweets were shared more by small margins were the New York Times and CNN. This is disproportional to the significantly smaller number of followers the emerging news organizations retained; Huffington Post, The Hill and Breitbart had 11.3M, 2.93M and 896K followers respectively as of January, 2018 whereas the New York Times, CNN and Fox had 40.8M, 39.1M and 16.9M, respectively. This observation indicates that the emerging news organizations' news distribution through social media has been successful.

Further, recent works also show that their news links disseminated via social media play an agenda-setting role in political discourses. Benkler et al. (2017) found that the vast majority of political agenda during the 2016 US presidential election emerged from the right-wing online media, such as Breibart much ahead of legacy media such as Fox News. Additionally, Starbird (2017) found that the emerging partisan news websites are playing a leading role on Twitter in forming conspiratorial alternative political narratives about sensitive political issues, such as mass-shootings. Because some of the emerging media are explicitly denying the traditional journalistic conventions and norms, it is more important to know how the emerging media are utilizing social media to gauge the impact of social media on journalism.

## 2.2 News Dissemination Strategies on Social Media

### 2.2.1 What Makes Social Media Different as a News Distribution Platform?

If social media are a news distribution platform for news organizations, does consuming news on social media affect the experience of news readers? Because social media have many sociotechnical components, their impact on news consumption can be delivered through multiple paths. However, the role of news organizations' social media strategies has not drawn much attention as one of the paths. Instead, a social network (An et al., 2011; Bakshy et al., 2015; Hermida et al., 2012) and recommendation algorithms (Bakshy et al., 2015; Pariser, 2011) on social media have been the main foci of literature that investigates the impact of social media on news consumption. The findings differ, or even contradict each other, depending on the context in which each study was conducted.

Results about the impact of a social network on news consumption differ depending on whether the *weak tie* or the *strong tie* (Granovetter, 1973) dominates. An et al. (2011) and Hermida et al. (2012) commonly report that social networks formed on an online social media platform increase news diversity. Their results conform to the "weak tie" argument; users are exposed to information that they would not know about without social media from their friends' recommendation who do not share preferences. However, a Facebook data science team recently reported contradictory evidence analyzing users' news link selection on Facebook (Bakshy et al., 2015). They found that the proportion of hard news stories with the opposite political view available to users decreases dramatically as they friend with other users. For example, 45% and 40% of hard news available to conservative and liberal users are with the opposite political view if the users are randomly exposed to news on Facebook. However, only 24% and 35% of hard news shared by users' friends are "cross-cutting" – i.e. with the opposite political view – for liberal users and conservative users respectively. This is because Facebook users tend to friend users who share similar political views. In other words, the strong ties dominate in this case.

A few recent human-computer interaction (HCI) studies have focused on algorithms that automatically curate information for social media users. Previous research has discussed the power of the algorithms to govern access to information (Kitchin, 2016). Biased access to information online due to the algorithmic curation may cause users to encounter only viewpoints that reinforce their existing attitudes, which prevents minority opinions from being expressed and deliberated (Bozdag & van den Hoven, 2015). Bakshy et al. (2015) also measured how much the algorithm reduces consumption of the opposite view on Facebook. They found that conservative users see 5% less cross-cutting content in their News Feeds compared with what friends share, whereas liberals see 8% less.

### 2.2.2 Competition for Limited Attention on Social Media

This dissertation introduces additional structural factor in news consumption through social media that may affect news readers' experience to the discussion. That factor is news organizations' news dissemination strategies to attract social users' limited attention. As an economic entity, a news organization is likely to adapt its behavior to the socio-technical characteristics of social media as a news distribution platform to attract more news readers. Its social media strategies, as a result, may affect content that news readers end up consuming through social media in turn.

News organizations have adopted strategies to cut the clutter on other online news platforms, such as a search engine. Search engines are a previously dominant online news platform for readers overloaded by the mixture of individual news articles from diversified sources and all other kinds of information available online (Hermida, 2010). As Manovich (2012) characterizes them, search engine users query relevant information with keywords they expect to be associated with the content they look for to extract it from databases. Knowing this, news organizations have widely used news distribution tactics to maximize a chance that their content is visible in a search result, which is often called search engine optimization (SEO) (Dick, 2011; Newman, 2011; Giomelakis & Veglis, 2015). Leading news companies such as the New York Times and BBC hired SEO specialists and trained their journalists to publish news content fitted to the Web (Giomelakis & Veglis, 2015). These SEO techniques include customizing news titles for search, HTML tagging, tailoring meta data for images, URL optimizing, etc. They are partly technical choices to conform to a search algorithm, but also strategic choices that reflect an organization's online performance and competitors. Using an ethnographic approach and interviews, Dick (2011) describes that SEO specialists in news organizations engage in a click stream analysis and a competitor analysis to find the best practices of SEO. He also found that SEO involves significant cost in training journalists and facilitating communication between SEO specialists and the desk.

As more audiences are moving toward social media to read news, news organizations are likely to adapt to different socio-technical traits of social media, in addition to search engines. Manovich (2012) contrasts it with search engines, asserting that information on social media is organized as a *data stream* rather than a database. On an information platform that a data stream rules, a user experiences a continuous flow of information. Weltevrede et al. (2014) relates the stream on social media with the traditional distinction between 'push' technology and 'pull' technology on the Web. As information is continuously 'pushed' to a user, new information immediately replaces old information in a data stream. Thus, a user is likely to miss relevant information swept by irrelevant information pushed by a stream. This can create severe information overload compared to a search-based platform where a user pulls relevant information.

News organizations need to consider the information overload that their own and other news organizations' news sharing can generate on social media. Anderson & De Palma (2013) mathematically model a competition among multiple companies to sequentially propagate information to consumers. Having an advertising market in their mind, the authors assumed a scenario where a consumer becomes aware of the company's product only when she chooses to see information from the company. Anderson and De Palma call this situation *competition for limited attention* because other companies also try to reach the consumers who cannot process all the information. One conclusion of their model is that companies are better off by refraining from excessive information propagation because it will mitigate consumers' information overload.

However, whether to propagate information may not be the only strategic concern in sharing news on social media. Anderson and De Palma's model only reflects a limited aspect of information quality in that all consumers in the model are assumed to agree on how good or bad the information quality is, and this simplification allows firms to decide only how many information goods they are going to propagate to consumers depending on the information quality they produce and competition with other firms. However, the quality of news information, in reality, may not be something everyone can agree upon; a news story typically contains different topics, has a specific viewpoint on a certain issue of interest, and may be factual or opinion loaded while there must still be a dimension of being better or worse. Therefore, news organizations' strategic concerns are likely to be richer than only their link sharing decisions. For example, news companies may want to summarize and paraphrase a news story (e.g. within 280 characters on Twitter) shared on social media so that the composed social media post efficiently represents the original content, but is succinct enough not to overwhelm social media users' attention.

Although news organizations' news dissemination strategies on social media have not been widely analyzed, recent marketing literature revealed that, in general, companies are strategic about using social media as their branding platform. Interviewing 14 marketing managers responsible for the social media activity of their company, Tsimonis & Dimitriadis (2014) found that the presence of competitors on social media is one of the main motivations of running their own social media account. With a similar approach, Parveen et al. (2015) revealed that social media teams run by many companies learn information about competitors via social media, and strive to develop an innovative social media use different from their competitors'.

### 2.2.3 Selective News Link Sharing as Gatekeeping

Built on the strategic concerns news organizations consider for social media, an immediate question is, "Which news do they choose to share on social media?" If news organizations select a certain subset of news stories to share their links on social media, and if the selective news link sharing is a strategic choice, it will manifest itself as certain patterns in shared news content and unshared news content. Then, the pattern will effectively act as an additional layer of gatekeeping imposed on social media users. Link sharing as gatekeeping is not a new phenomenon in journalism. Dimitrova et al. (2003) earlier found that American major newspaper companies use hyperlinks as a gatekeeping tool to control the number of external links to an information source.

The selective news link sharing on social media can be understood as an inter-media agenda-setting via gatekeeping between news websites and social media. The selective choice of information by news organizations has been explained under the framework of agenda-setting by gatekeeping (Shaw & McCombs, 1977; Shoemaker & Vos, 2009). In other words, news organizations influence public opinion not only by carrying through their opinion, but also by choosing issues for the public to think about (McCombs, 2014). This theory expanded toward inter-media agenda-setting theory (Meraz, 2011; Vargo et al., 2017) and network agenda-setting theory (NAS) (Vargo et al., 2014), in which sequential choices of information from media by other actors such as other news organizations or social media users, also exert the agenda-setting power by selecting information. Empirical studies based on these frameworks often track changes of issues across different media outlets or media platforms to know which media were influencers, or those who set the news agenda of others (Meraz, 2011; Vargo et al., 2017).

Whereas these frameworks focused on the actors' intentional choices, such as news organizations' editorial decisions, another development of the agenda-setting theory, acknowledges that technological factors can impact the salient agenda as well. Extending the agendasetting theory that focuses on a linear propagation of information from news organizations to a public, agenda-building theory considers reciprocal influences between multiple entities (Kiousis & Ragas, 2015). Because different entities, such as news readers, political parties, and industry come into play to interact in this framework, how technological innovations change how each kind of entity communicate is likely to also influence how a public agenda is built in an unpredicted way. This assessment resonates with the recent discussion to view a technical choice of personalized information curation algorithms as a public matter that requires monitoring (Kroll et al., 2016; Sandvig et al., 2016). In a similar manner, this dissertation explores the possibility that using social media as a news distribution platform functions as a socio-technological factor that imposes another layer of information restriction on the formation of public opinion.

### 2.2.4 Sentiment and News Framing

McCombs et al. (1997) expanded the agenda-setting theory to the second level – the selection of *attributes* of a news object. They argue that news reporting influences *how* to think about a reported issue by focusing on or neglecting specific aspects of a news object, such as a politician or a policy. The second-level agenda-setting has a close relationship with framing (McCombs et al., 1997) although the former tends to focus more on the choice of information whereas the latter focuses more on the attribution of responsibility for an social issue (Scheufele, 2000).

In their seminal study, McCombs et al. (1997) acknowledged the importance of sentiment in shaping how to think about news. By connecting social issues with positive / negative / neutral sentiment, news reporting informs readers' judgment about real-world issues (McComas & Shanahan, 1999). Research has particularly focused on negativity in news framing. Trussler & Soroka (2014) hypothesized that news readers psychologically have a preference for negative news framing because negativity is further from humans' innately positive expectation and is, thus, considered a signal for more useful information (Kahneman, 1979). Indeed, Trussler & Soroka (2014) found that politically motivated news readers prefer negative news framing under an experiment condition. They suggest this result as evidence for the demand side explanation of the prevalence of negativity in political reporting. Built on this assessment, researchers have applied automated measurements of sentiment in news stories. For example, using human coders and multiple lexicons, Young & Soroka (2012) found that news framing is consistently biased toward the negative, particularly for crime and foreign policy topics.

For traditional newspaper reporting, journalists must compose headlines that summarize news stories they have written. However, the headline is not only a mere summary of news content, but also the main 'hook' of a news story to readers (Molek-Kozakowska, 2013). In the same vein, Bell (1991) listed the main purposes of headlines as a) summarizing, b) framing and c) attracting. In other words, how the content is summarized in headlines influences news readers' choice to actually read the story. The similarity between headlines and news paraphrasing for social media implies that journalists' choices made for headlines that have been observed in previous studies can inform us about potential choices made for social media posts.

Online news consumption is likely to make the reader attracting role of news paraphrasing more important, due to the competition between individual news stories online. In traditional news consumption where news stories are consumed as a bundle (e.g. a newspaper or TV news program), headlines signal news content to compete with other articles within the same bundle. Thus, the headlines have to be to written to show a story's relevance to a reader (Dor, 2003). However, in the online environment, news readers typically navigate across different news sources. In this circumstance, news organizations have a strong incentive to signal their stories' appeal to readers. Moreover on social media, this signaling incentive is likely to be even larger because users' attention is easily distracted. Indeed, researchers have detected a proliferation of sensational expressions in online news headlines for *click-bait*. For example, Chakraborty et al. (2016) and Potthast et al. (2016) showed sentiment polarity (valence) can be used as a predictor of an automated click-bait detection algorithm.

Negative news summary has been shown to have a strong impact on news perception. Manipulating headlines and leads, Price et al. (1997) found that experiment subjects' opinion and emotional valence about funding to public universities change depending on how the same information is framed (conflict / human interest / consequence). Further, they found that the changed opinion impacts the decision making about the policy. With a similar experiment design, Zillmann et al. (2004) showed that experiment subjects spend more time reading news stories when they are framed as conflict or victims' agony rather than misfortune or economic loss. These studies imply that news paraphrasing, such as a social media post, has a power of *extra-framing* on top of frames in the main body of news stories.

All in all, the literature I have discussed so far implies that news organizations are likely to be engaged in strategic behaviors on social media through which they regularly disseminate their news stories, and that such behaviors can act as an additional layer of *information mediating process* to news readers. These strategies can influence social media users' news reading as though they are editorial decisions. Just as editorial decisions have influenced news readers about "what to think" and "how to think" (Bennett & Iyengar, 2008) in news production, "what to distribute" and "how to distribute it" through social media may also influence news consumption to the extent that they have distinguished patterns. The goal of this research is to explore such strategic behaviors, particularly on Twitter, and evaluate whether news organizations' news dissemination strategies on social media deserve public discussion as an information mediating process that potentially has a significant social impact.

Althoung a few previous works have categorized types of news tweets by news organizations (Newman, 2011), their attention has been limited to a social media post as such rather than its relationship with original news text, due to the large volume of news texts and the cost of the hand-coding approach. However, the focus on tweets limits the scope of an inquiry only to *shared* news rather than *unshared* news. Further, data exclusively from tweets contains only information about *outcomes* of news organizations' social media strategies rather than *input*. Thus, the previous approach prevents researchers from observing the decision-making of news organizations with regard to their distribution of published news stories on Twitter. I, instead, adopt a large-scale news scraping and a computational text analysis to enable the comparisons between shared *and* unshared news, and between text that is tailored for Twitter and original news. Further, the computational methods I adopt
provide an infrastructure for potential public monitoring of the social media strategies for future discussion.

## 2.3 Contribution of Dissertation

This dissertation explores how news organizations' social media use as a news distribution platform imposes an additional layer of the information mediating process, focusing on Twitter. News organizations still play a central role in audience's access to information on public issues (Purcell et al., 2010), and they use major social media, such as Twitter, mainly as a news distribution platform (Greer & Ferguson, 2011; Meyer & Tang, 2015). News organizations' behaviors are under the effect of the socio-technical traits of social media as a news distribution platform, and they can influence what news readers think and how they think about public issues (Bennett & Iyengar, 2008). Specifically, news organizations are likely to adopt strategies to cope with social media users' limited attention, which may bias their news reading. Thus, as more news readers use social media as their source of information, what news organizations share and how they share news on social media may also influence the information to which news readers are exposed.

News organizations potentially both help and harm society by selectively choosing newsworthy information. They help news readers understand the world by reducing the amount of information to process, but they also may bias readers' perception of public issues due to selectively chosen information in favor of their viewpoints. Thus, how news organizations choose information has been central to the evaluation of the social impact of news. The information restriction by news organizations generally takes place in two ways. First, news organizations decide the selection and salience of public issues (gatekeeping / agendasetting). Second, they decide the selection and salience of particular aspects of an issue (framing) (Scheufele, 1999).

My research question about the quantity of news distributed through social media (Chapter 3 and 5) and a hypothesis about selective news link sharing conditioned on news content (Chapter 6) are to learn how social media as a news distribution platform influences the selection and salience of public issues. A second hypothesis about paraphrasing news for a tweet (Chapter 7) aims at how organizations' news distribution via social media influences the selection and salience of particular aspects of an issue. Further, social media provides opportunities for a diversified news supply. They sustain many online only news organizations as their major news distribution platform (Newman et al., 2016). Thus, this dissertation further aims to illustrate the impact of social media on how blends of news organizations inform readers by inclusively analyzing different types of news organizations, e.g. magazines, national media, regional media, and online media, across the different chapters.

Empirical findings from Twitter cannot be generalized directly to other platforms. However, the general conclusion that news organizations' strategies reflecting socio-technical traits of social media as a news distribution platform creates an additional information mediating process, is expected to be valid for Facebook, the single largest social media platform also used as a prominent news distribution platform. As in Twitter, news organizations have similar needs to address users' limited attention to reach them, and sensitively react to users' engagement in and sharing of news information. Given that Facebook refers the largest user traffic to news websites, organizations are even more likely to customize their news dissemination for Facebook as well. However, the specific patterns of news organizations' choices may differ between Twitter and Facebook because, compared to Twitter, Facebook provides functions for a more closed social network actively curated by algorithms.

Methodologically, I use computational approaches for data achievement and content analysis, which enables answering research questions that are difficult to answer with traditional approaches. Because it is hard to achieve an unbiased sample of connections between news stories and social media posts, the hand coding approach that needs a limited number of sample data could not allow for questions about such connections. The computational data scraping approach and large-scale text analysis techniques I suggest in this dissertation provide tools to analyze what news organizations did to their news stories for social media. These tools can be further applied to additional research questions about relationships between the news and social media. Because my data collection and analysis process are fully automated, they provide tools to constantly monitor news organizations' social media strategies and share the information with the public.

#### CHAPTER 3

# A CONCEPTUAL MODEL: AN ECONOMIC MODEL OF NEWS LINK SHARING

# 3.1 Introduction

In this chapter, I present a simple model in which news organizations choose the number of news links that they share on social media. Although highly stylized, the presented model captures a fundamental aspect of the competition for the limited attention on social media: to *competitively* supply an information good the amount of which exceeds a user's cognitive capacity.

Several observations indicate that this quantitative aspect of the competition is important. Above all, there is simply too much information on social media for a user to process. An average social media user is exposed daily to 54,000 words, the equivalent of an average novel (Bennett, 2013, July 13). As a result, news organizations are failing to attract social media users' attention to their brands. According to a recent Reuters Institute report (Newman et al., 2016), only 52% of US news readers reported that they noticed the news brand when they access to a news article via social media. Even though a user chooses to read a news article, she wouldn't stay within the news organization's website. News readers who visit news organization websites via social media tend to stay for one minute and forty seconds only whereas news readers who visit news websites directly stay on average for four minutes and thirty six seconds (Mitchell et al., 2014). In other words, a typical news reader on social media clicks on a news link, reads the news article, and leaves the news organization's website. In this case, the goal of the news organizations' social media strategy is likely to maximize the number of clicks on news links they share.

A fundamental intuition of the quantitative competition for limited attention is that the amount of other organizations' shared news links on social media can affect the probability that my news links attract social media users' attention. Thus, my decision will depend on other organizations' behavior. For example, if other news organizations share so many news links that users will not pay attention to additional information, I will not share news links if the sharing incurs cost. This situation in which my decision depends on other agents' decisions is often called a *strategic* situation, or simply a *game*. Thus, the simple model I suggest is to show social media users' limited attention can make news link sharing on social media a strategic concern.

The model shows that the quantitative competition for the limited attention results in many news link excessively sharing compared to news organizations' collectively optimal level, as an equilbrium. Since an individual company cannot fall out of this equilibrium, it is likely to carefully devise *qualitative* means to attract users' attention to news link it published as in SEO practice. In the following chapters, I will introduce other qualitative social media strategies in which news organization may be engaged. This conjecture about the qualitative strategy will lead to empirical research questions I will explore with a largescale text analysis of news texts and social media posts by news organizations.

## **3.2** Competition for Limited Attention Model

Inspired by Anderson & De Palma (2013)'s advertising market model, this simple model focuses on the short-term news link sharing decision after the news production is finished. The main feature of this model is that there are multiple firms that repeatedly propagate information to users who have a cognitive limit to the amount of information they can process. Anderson & De Palma (2012) call this situation *competition for limited attention*. Users' limited attention or cognitive capacity creates a trade-off for the firms because, on the one hand, they want to propagate more information to increase the probability of reaching users' attention. But on the other hand, sharing too much information will overwhelm users' attention, which in turn makes users unable to process anymore information.

In Anderson and De Palma's model, firms repeatedly send out one advertising message,

only one exposure of which has an impact on consumer. In other words, their model considers a situation where consumers become aware of the advertised product upon the one exposure, which generates a fixed amount of expected profit; any extra exposure does not generate profit for firms. Thus, it resembles the classical "informative advertising" literature in economics wherein the role of advertising is having consumers find out about a good (Stigler, 1961). Unlike "persuasive advertising" models, Anderson and de Palma assume that repetitive exposure to an advertising message does not increase the likelihood of purchasing the advertised product as a model simplification. However, a fixed amount of marginal cost of advertising occurs every time a firm sends out an advertising message. In this setting, firms have an incentive to repeatedly send out the same message because only greater frequency of the message increases the probability that the message is exposed to a consumer, at the expense of greater cost.

Unlike Anderson and De Palma's advertising model, I postulate a situation where a firm sends out hyperlinks to *multiple* online news stories containing advertisements, and news organizations gain fixed profits from the online advertisement when a user clicks the hyperlink. However, an organization sends out each link only once, which prevents a link from incurring the profit more than once. This setting contrasts with Anderson and DePalma's model where one message is repeatedly sent out. The different settings between the two models come from the different realities that they try to capture, e.g. product advertising vs. news dissemination. Despite the difference, the main intuition that drives the outcomes in the two models is similar; one firm's propagation of information generates a negative externality to competitors in that it eats up users' limited attention. However, a selfish firm would not care about this negative externality, which leads to an overuse of users' limited attention. In other words, as an outcome of the competition, news organizations excessively share many news links on social media compared to the collectively optimal level of news organizations.

# 3.3 News Link Sharing under Users' Limited Attention

#### 3.3.1 A Model Setting

The model makes a few simplifying assumptions about identical news organizations and news stories as their products. To be more specific, (a) news organizations publish the same number of news stories, and (b) all the published news stories have the same appeal to homogeneous news readers. (c) Also, the marginal cost of sharing an additional news link is the same for all news stories regardless of which news organization created the story. Thus, an arbitrary news organization, i, does not have other strategic choices other than the number of shared links,  $x_i$ . In other words, I assume that the revenue from social media users' clicks on news links is independent of news organizations' other sources of revenue such as offline subscription or visiting a webpage via a search engine. This assumption means that the present model focuses on the short term decision of sharing news stories that are already published. This assumption may not be too strong because, for most traditional media firms, online revenues are still dwarfed by offline revenues. Not only does this simplification allow us to concentrate on the core of the strategic interaction between news organizations by eliminating many complications in reality, but also it helps develop a research question about the degree of news link sharing on social media. I assume that a fixed cost, c, incurs per a news link created. This cost specification can be interpreted as the effort to optimize each news link to distribution on social media, as in SEO practices (Dick, 2011) such as news title optimization or summarizing the news content for a social media post.

Each of K news organizations publishes M news stories, whose links may be shared on social media. Because the value of news stories is assumed to be the same and independent of other news stories' value, the total  $K \times M$  news stories are identical up to the news reader's choice. There are N news readers who subscribe to the only social medium that exists. Each reader can read only T news stories, which represents their *limited attention*; if there are more news available on social media than a news reader can read, i.e. KM > T, she will read only a subset of the available news. News readers are exposed to news stories shared on social media in a random order and read the news stories in the order they encounter them with probability 1 until they reach the cognitive limit T. Thus, each reader will randomly read T stories among the KM available in the case of KM > T.

In this setting, a risk-neutral news company i's expected profit maximization problem is as follows:

$$\max_{0 \le x_i \le M} E[\pi_i(x)] = \left[ Pr\{a \text{ story is read}\} Np - c \right] x_i$$
$$= \left[ \min\left\{ \frac{T}{\sum_j x_j}, 1 \right\} Np - c \right] x_i, \tag{3.1}$$

where  $x_i$  denotes the number of news articles shared by news organization *i*, and *p* denotes an advertising revenue from a single click on a news article, which I assume to be the same for all clicks. If there are *more* shared news articles than a social media user can read (KM > T), the probability that a news article is chosen is the number of news article a reader can process (T) over the number of news articles shared by all firms in the market  $(\sum_j x_j)$  due to the equal probability assumption. If there are *fewer* news links shared on social media than readers can process  $(KM \leq T)$ , all news articles will be read with probability one. Thus,  $\min\{T/\sum_j x_j, 1\}$  is the probability for a user to read an article in both cases. Given this probability,  $\min\{T/\sum_j x_j, 1\}Np$  is the expected revenue from a single news story through social media. Thus the whole expression,  $E[\pi_i(x)]$ , is news organization *i*'s total expected profit from sharing  $x_i$  news articles on social media given other firms choices.

#### 3.3.2 Symmetric Nash Equilibrium

I focus on the symmetric Nash Equilibrium (SNE) as the most plausible equilibrium because parameters except the choice variable, i.e. the link sharing decision,  $x_i$  are the same across all firms. Let us assume there exists such a choice in the symmetric Nash equilibrium,  $x^*$ . Then, there can be three possible cases.

- 1. No link sharing:  $x^* = 0$
- 2. Full link sharing:  $x^* = M$
- 3. Partial link sharing:  $0 < x^* < M$

# 3.4 Results: Over-sharing Hypothesis

First note that no sharing is the only sustainable SNE if  $Np \leq c$ . By equation 3.1, the marginal profit from sharing a link cannot exceed  $Np - c \leq 0$ . Thus, in the case of  $Np \leq c$ , sharing any link cannot be profitable for all news organizations. Thus in the following, I will focus on the case that Np is sufficiently high relative to c that at least one firm can offer at least one link and not lose money doing so, i.e. NP > c.

In addition, I will only consider a case where  $\frac{Np}{c} \ge \frac{K}{K-1}$  for simplicity. This assumption is not very restrictive; rearranging the right hand side of the inequality, we get  $Np \ge c(1 + 1/(K-1))$ . For K = 2, the right side of this inequality is 2c. Thus, the assumption requires only that there be enough revenue in the market to support two messages, which could be one firm with two messages or two firms with one message each.

Let us define the news supply - attention ratio,  $\phi \equiv \frac{MK}{T}$ . Then the following proposition shows that partial sharing is sustained as a SNE when competition for limited attention is relatively strong.

**Proposition 1.**  $\frac{Np}{c}\frac{K-1}{K}$  is a cut point such that, if  $\phi$  is no greater than that, full-sharing is the symmetric equilibrium, and if  $\phi$  is greater than that, partial sharing characterizes the equilibrium. Under the partial sharing equilibrium, the number of shared news stories per firm is determined as  $x^* = \frac{Np}{c}\frac{T(K-1)}{K^2}$ .

#### *Proof.* See appendix

(Sketch of Proof) If there are fewer news stories than a news reader can process in the market, marginal profit is fixed at Np - c. In this case, as long as this term is positive,

full sharing is SNE. If there are more news stories than readers' cognitive capacity, marginal profit is decreasing in the number of links. Then, partial sharing equilibrium potentially exists at the level where marginal profit is zero. However in some range, marginal profit may stay positive at the maximum number of news stories news organizations can share, M. In that case, full sharing still characterizes SNE. Otherwise, partial sharing is sustained as a SNE.

In sum, both the partial sharing and the full sharing equilibrium are possible depending on the news supply - attention ratio relative to profitability. The relative strength between how many total news stories are published compared to users' limited attention and the profitability of sharing determines the proportion of news stories shared on social media  $\left(\frac{x^*}{M} = \frac{Np}{c} \frac{(K-1)}{K} \frac{T}{MK} = \frac{Np}{c} \frac{(K-1)}{K} \frac{1}{\phi}\right)$  among M published news stories. In Chapter 5, this outcome will be used to test whether a news organization indeed cares about others' quantitative choices.

Note that  $\frac{T}{MK} \frac{(K-1)}{K} Np$  is marginal negative externality when all news organizations share all stories. When all organizations share M stories, one organization's marginal profit is  $\frac{T}{MK} Np - c - \frac{T}{MK^2} Np$ . The first two terms are the direct effect from sharing the marginal link, and the third term is the indirect marginal loss from lowered probability for a news link to clicked due to one more link on the social media. The indirect effect incurs to all other news organizations as well as the given organization. Because there are (K - 1) other organizations, this marginal externality adds up to  $\frac{T}{MK} \frac{(K-1)}{K} Np$ . Then, the condition for full sharing equilibrium in Proposition 1 is equivalent to the condition where this marginal negative externality at full sharing is larger than the marginal cost of sharing an additional link, c. In other words, full link sharing is SNE.

Now the condition for three possible SNE can be characterized in a two-dimensional space generated by the marginal profit - cost ratio, Np/c, and the number of story – attention ratio, M/T as in Figure 3.1.



Figure 3.1: Conditions for SNE.

#### **3.4.1** Comparative Statics

**Proposition 2.** (Excessive sharing hypothesis) Firms share no fewer stories than the first best under SNE.

*Proof.* To replicate the first best (FB) choice of news organizations, suppose all news organizations collude. Then, the colluding entity's problem is simply:

$$\max_{0 \le x \le KM} \left[ \min\left\{\frac{T}{x}, 1\right\} Np - c \right] x, \tag{3.2}$$

where x denotes the total number of shared news links. It will share no stories if the marginal cost is so high that any sharing is not profitable  $(Np \leq c)$ .

Thus, now suppose Np > c. If  $KM \leq T$ , the colluding firm will share all KM stories as in the competition case because the profit on the marginal link shared is Np - c in this case, which is assumed to be positive.

Suppose KM > T. First consider marginal profit at x > T. Because the profit maximization problem becomes  $\max_{T < x < MK} (\frac{T}{x}Np - c)x$ , the marginal profit is -c < 0. In other words, any choice in (T, MK] cannot be optimal. However, the marginal profit in  $x \leq T$  is Np - c > 0. Thus, sharing T stories is optimal.

In sum, each firm under collusion simply shares M if  $MK \leq T$ , and T/K if MK > Tas long as Np > c. This is no greater than the SNE choice for M if  $MK \leq T$ , and M or  $\frac{Np}{c} \frac{T}{K} \frac{(K-1)}{K} \text{ if } MK > T. \text{ The latter is equal or greater than the former under the assumption,}$  $\frac{Np}{c} \geq \frac{(K-1)}{K}, \text{ we made.}$ 

Let us call each firm's choice under the collusion  $x^{FB}$ . Then, Proposition 2 is summarized as  $x^* \ge x^{FB}$ .

# 3.5 Discussion of the Conceptual Model and An Empirical Hypothesis

The excessive sharing result follows from the fact that the competition for limited attention corresponds to the *the tragedy of the commons* (Hardin, 1968; Ostrom, 2015) where each economic agent's choice exerts *negative externality* to other agents' decision. If firm i shares one more news link, it puts more load on news readers' limited attention when there are sufficiently many news links on social media. Because of the additional link on social media, news readers' distraction affects not only the probability that firm i's link is clicked, but also the probability that all other firms' links are clicked. On the other hand, firm i gets benefit from sharing more links because the firm increases its share of attention relative to other news organizations. The negative externality from sharing more link causes a *game* situation in which each company needs to be *strategic* against each other. Because one company's profit depends on other companies' strategies as well as its own strategy, the company needs to take account of decisions that other firms would made to optimize their strategy for profit. The equilibrium number of shared links,  $x^*$  is the result of this strategic concern.

The standard result of the tragedy of the commons game states that users' attention as a *public resource* (Anderson & De Palma, 2009) will be overly occupied relative to the optimal level because an individual company interested only in its own profit would not care about the negative externality to other companies. The excessive link sharing I proved corresponds to this general conclusion of the tragedy of the commons game. Further, this result implies that competition reinforces information overload on social media.

An outcome of the model in which the amount of shared news links is a strategic decision constitutes a testable hypothesis. If the link sharing were not a strategic concern, the proportion of news stories shared by a news organization will not be associated with the number of other organizations' news stories. If news organizations do not perceive competition as a meaningful concern, the *perceived* probability that a shared news story is clicked will increase in the number of shared news links independent of other news organizations' decisions. Then, the profit maximization problem should be  $\max_{0 \le x_i \le M} [\min\{\frac{T}{x_i}, 1\}Np - c]x_i$ . By the same logic as the proof of Proposition 2, the optimal choice should be all news stories or the maximum number that their readers can process, i.e. M or T. Thus, the proportion of shared stories will be 1 or T/M, which does not depend on other news organizations' choices. On the other hand, if news link sharing is strategic as in the previous model, the proportion is  $x^*/M = \frac{Np}{c} \frac{(K-1)}{K} \frac{1}{\phi} = \frac{Np}{c} \frac{(K-1)}{K} \frac{T}{MK}$ , which depends on the total number of stories all organizations published. In other words, the more other news organizations publish, the fewer of their own news stories a news organization shares. This hypothesis will be tested using a Web scraped data set described in Chapter 4.

Although the model in this chapter illustrates one strategic aspect of the amount of news link sharing, one should note that the model is based on stylized assumptions. Above all, this model ignores horizontal and vertical news qualities. Thus, the simply quantitative reactions to the number of other organizations' shared news links is unlikely to be close to the full description of competition among news organizations on social media. For example, how is a firms selection of news topics influenced by the topics offered by its competitors? Are there other strategic tools that news organizations can use on social media than simply whether to share a decision? In Chapter 6 and 7, using text analysis techniques, I will take on some of these questions about the qualitative aspects of news link sharing on Twitter.

#### CHAPTER 4

# AUTOMATED DATA COLLECTION AND COMPUTATIONAL ANALYSIS OF NEWS DISTRIBUTION ON TWITTER

In the earlier chapters, I proposed potential social media strategies stemming from competition among news organizations on social media. However, the empirical examination of the existence of such strategies poses methodological challenges, which are mainly involved with scale of the data. For example, to test whether a news organization reacts to the number of news stories published by other organizations, I need to have a measure for the total number of news stories. To test the selective news link sharing, I need to have a sample of published news stories which 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. It raises a challenge because a researcher has to analyze the data first to get a good sample.

Even though automated data collection allows for bypassing such a sampling challenge by collecting all the data available online, the sheer amount data also prevents the traditional hand-coding approach. For example, the dataset used for this dissertation includes more than 400K news stories and more than 150K tweets with embedded news links. Further, my analyses involve comparisons of the pairs of news stories and tweets, as well as separate stories and tweets on their own. Thus, to answer the questions posed in this dissertation calls for computational text analysis methods as well as automated data collection.

In this chapter, I will explain the techniques I adopted for the data collection and the analyses. In addition to the utility for this study, the Web scraping software built on an opensource database, *Media Cloud* can be generally used for further public monitoring of new organizations' behaviors. More generally, I apply the analytical tools from machine learning and sentiment analysis fields as computational translations of media-related social scientific research questions. These techniques are valuable because they help expand the frontier of answerable questions in social science, and extend a one-shot observation to long-term monitoring of social phenomena under the evolutionary process.

# 4.1 Automated Data Collection

#### 4.1.1 Data Collecting Software

The data were automatically collected by a system of software I developed for this study with Python language. The system has three main components to collect news stories and tweets with embedded hyperlinks to them, which were required for the subsequent analyses. First, the *news scraper* collects news stories published on each news organization's website. This software is built on *Media Cloud*, an open sources database developed and maintained by the Harvard Berkman Klein Center<sup>1</sup>. *Media Cloud* monitors the RSS feeds of approximately 60,000 news organizations around the world every thirty minutes, and stores meta-information of the news stories fed via RSS. To my knowledge, this is the largest set of data regarding news stories published online. Because Media Cloud does not provide content of news stories, the news scraper opens all the URLs in the list from Media Cloud and downloads the news texts.

Second, the *Twitter scraper* collects tweets from each news organization's official Twitter accounts. When news organizations have more than one account, Twitter scraper downloaded data from all the accounts.<sup>2</sup> The Twitter scraper collects text in tweets, embedded URLs, and the tweeted date via Twitter REST API. Third, to identify which news stories were shared on Twitter by news organizations, the *URL matcher* matches URLs from news dataset with URLs from the Twitter dataset. The URL matcher conducts multiple steps of URL preprocessing because URLs are often shortened to save space within a tweet, or modified to have them contain useful information. Thus, the pre-processing includes URL unshortening and query term parsing. Additionally, it parses out news titles or news IDs to maximize

<sup>&</sup>lt;sup>1</sup>Available at https://mediacloud.org

<sup>&</sup>lt;sup>2</sup>See Appendix B to see the list of the included accounts.



Figure 4.1: A computational method to identify shared and unshared news on Twitter.

performance of the matching task. This data collection system is visualized in Figure 4.1. A news story was recorded as shared when it was shared *at least* once.

#### 4.1.2 Description of Data

News organizations in the dataset 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 US news websites except the BBC, the Guardian, Reuters and Al Jazeera, which potentially have a 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 (4 organizations). This process resulted in 84 news organizations. These news organizations are categorized for convenience as in Table 4.1.

The dataset includes regional media and online-only news websites. I included the online-

National	Regional	Magazine	Online
ABC News	The Arizona Republic	Economist	AlterNet
Al Jazeera	Baltimore Sun	Forbes	Bipartisan Report
BBC	Boston Herald	New York Magazine	Breitbart
Bloomberg	Chicago Sun time	The Atlantic	BuzzFeed News
CBS News	Dallas News	The Christian Science Monitor	Common Dreams
CNBC	Denver Post	The New Republic	Conservative Tribune
Fox News	Des Moines Register	Time	$\operatorname{FactCheck}$
MSNBC	Detroit News		${ m FiveThirtyEight}$
NBC News	Houston Chronicle		Free Beacon
New York Times	Miami Heralds		Gateway Pundit
NPR	New York Post		Huffington Post
PBS Newshour	Newsday		IBTimes
Reuters	NJ.com		Infowars
The Guardian	Orange County Register		$\operatorname{Inquisitr}$
USA Today	Oregonian		Media Matters
Wall Street Journal	Orlando Sentinel		Politico
Washington Post	Seattle Times		$\operatorname{PoliticusUSA}$
	Sun-Sentinel		
	The Indiana Star		$\operatorname{Slate}$
	The Kansas City Star		Talking Points Memo
	The Mercury News		The Conversation
	The Philadelphia Inquirer		The Daily Beast
	The Press Democrat		The Daily Caller
	The Sacramento Bee		The Hill
	The Stranger		The Intercept
	The Tennessean		The Root
	Washington Times		townhall.com
			Vanity Fair
			Vice
			Vox
			Washington Examiner
			Western Journalism
			Westword

Table 4.1: The list of news organizations included in the dataset.

only news websites to see if the traditional news organizations and the online-only outlets adopt different social media strategies. The data were collected between November 20, 2017/11/20 and January 1, 2018, and January 9, 2018 and January 28, 2018 for sixty-three days. The data collection includes a gap week during which data were not collected because the software was mistakenly turned off. This data collection process resulted in 435,355 news stories published by 84 news organizations. Among them, 152,555 new stories were shared, resulting in a 35.04% sharing proportion. The number of published news stories by news organizations varied from 21 (InfoWars) to 30, 341 (Washington Times). This variation



Figure 4.2: The average proportion of news sharing of each news organization against the average number of news stories daily published with a smoothed trend. Each dot represents a news organization.

stems from a difference of their media types, from websites that stem from opinion blogs, such as Infowars or townhall.com, to magazines such as Economist or Time. On the other hand, daily news outlets with blogs as a part of their websites, such as Washington Post, or regional papers partnering with other news outlets, such as the Seattle Times publish far more news stories on their websites. The proportion of shared news stories also varied from 7.122% (the Seattle Times) to 95.16% (Economist). This variation depends on how many news stories they publish on their own. Magazines such as Economist or Time 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 Seattle Times, tend to share less. This relationship is visualized in Figure 4.2.

The non-parametrically smoothed trend calculated using the Local Polynomial Regression Fitting (Cleveland et al., 1992) appears to show a negative trend (the blue line in Figure 4.2). However, there is a possibility that the outliers that publish more than 200 news stories are driving the negative trend. These outliers turn out not to impact the general findings. I



Figure 4.3: Daily Number of Published Online News Stories. The data weren't collected during the period between the blue dotted lines.

will discuss the robustness result in Chapter 5.

Figure 4.3 shows the daily pattern of the number of published online news stories by the 84 news organizations. It shows a clear fluctuation where the number of published news stories is lower on weekends and holidays. That is, in addition to the lower number of published news items on Saturday and Sunday, the figure also shows that news organizations publish much less news on Thanksgiving (November 23, 2017), Christmas, and New Year's. 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 recovery after the new year began. The graphical observation shows that the daily pattern is consistent with weekly and seasonal work patterns, and provides a useful face validity of the data.

### 4.2 Computational Text Analysis

To analyze news organizations' potential selective news link sharing and the extra-framing imposed in the process of news paraphrasing, I adopt computational text analysis techniques



Figure 4.4: Text pre-processing procedure.

from machine learning and sentiment analysis. To apply these methods to text data, news texts should be processed into a machine-readable form. This section will discuss the preprocessing first and then explain the text analysis techniques.

### 4.2.1 Text Preprocessing

The richness of written and spoken natural language make a statistical analysis difficult. To simplify text data to a form that is statistically analyzable and, yet not to the point at which interesting variation in the data vanishes, researchers often apply standard preprocessing procedures to given corpora (Manning et al., 2008). Figure 4.4 summarizes the pre-processing procedures I apply to news stories and social media posts for this study.

The removal of capitalization, punctuation and stop words rids texts of less informative data, and reduces the dimensionality of data. In a final form of text data, each word or phrase stands for a single variable. This can cause dimensionality problems for a statistical analysis because there must be many words (i.e. variables in a text analysis) relative to the number of documents (i.e. observations). Removing uninformative words mitigates this problem at the expense of losing minimal variation in the data. Stemming refers to a procedure that achieves the base form of a word. This procedure also reduces the dimensionality of the data, but removes only the functional parts of a word. This implies an assumption that the functions of words, such as class or grammatical number, are not important for the purpose of analysis. I apply standard Porter stemming algorithm (Porter, 1980).

Discarding word orders is a simplification that assumes a word order is uninformative. Although this is evidently a false assumption, discarding word orders is a standard practice because incorporating word orders produces little enhancement in model performance relative to the computational cost this process saves (Manning et al., 2008). The simplification results in the *bag-of-words* model whereby an author of a text picks multiple words from an identical and independent word distribution to compose the text. The final outcome of the pre-processing procedures is a *word-document matrix*; a column equals words, and a row equals documents. Entries of the word-document matrix stand for the count of phrases used in a document. A researcher uses this matrix just like a data matrix is used in standard statistical methods.

#### 4.2.2 Selective News Link Sharing: Structural Topic Model

To see if news organizations tend to share more popular topics on Twitter, I use a machine learning algorithm, Structural Topic Model (Roberts et al., 2016) that applies to natural language – i.e. news text in this dissertation. The flip side of the research question, "What type of news is likely to be shared on social media?", is "what type of news is *not* likely to be shared?". This means that, to know about news organizations' selective link sharing on news content, a researcher needs to compare shared news and unshared news. Compared with many previous works, which focus only on social media posts, the requirement for the comparison between two sets of news stories dramatically increases the volume of data that needs to be analyzed. First, a researcher needs to look at news stories that contain much longer text than social media posts. Also, one needs to look at both shared and unshared news stories.

Manually reading all the news articles to decide what types of news are more likely to be shared is an implausible task. To illustrate this, suppose that a news organization daily publishes one hundred news stories, and a researcher wants to analyze news stories published by twenty news organizations for a week (say six days excluding Sunday). The researcher needs then to analyze 12,000 news articles, and discover different patterns between shared news and unshared news.

Structural Topic Model (STM) is an algorithm that permits automated analysis for such an inquiry. This statistical algorithm simultaneously discovers latent topics in news stories and identifies topics and phrases statistically associated with the link sharing decision. In other words, to circumvent the implausible manual approach, by applying STM, I replace the research question with a computational task that identifies topics correlated with the link sharing decision.

STM is a recent extension of topic models (Blei et al., 2003; Blei & Lafferty, 2006; Blei, 2012) designed to identify latent topics from the co-occurrence of words. STM extends topic models by embedding regression models within them. That is, STM simultaneously find topics and regresses the topics to other observable variables. This allows for testing the association between latent topics and the link sharing probability, which will be the main interest of Chapter 6.

In the class of topic models, a topic is operationalized as the latent distribution of words (or phrases depending on text models). To find the latent distributions, a topic model assumes a text generating model, in which an author chooses a topic from a multinomial distribution, and words (or phrases) from another multinomial distribution given the topic. Based on this text generating model, the model is estimated by Bayesian methods. Depending on how to specify the prior belief about the probability that a topic is drawn, topics can have only negative covariance between one another (Dirichlet distribution; Latent Dirichlet Allocation), or can have both positive and negative covariance (logistic normal distribution; Correlated Topic Model).

STM is built on the correlated topic model Roberts et al. (2013, 2014); Lucas et al. (2015). The word probability given to a topic is assumed to be determined by document covariates (being shared in my case) in the prior belief. This specification allows for both a identification of topics and the regression to find topics related to the sharing decision as a covariate.

I use STM with modifications to the model in regard to the news link sharing context. In particular, I specify a news text generating process that depends on topics and the link sharing decision. Formally, the generative process for each news article that I assumed for a STM model is:

1. Forming attention to topics: A news organization forms an attention  $(\vec{\theta}_d)$  to topics for an article d, which is determined by an organization, the news sharing decision, and the interaction between the two, which are included in  $X_d$ .

$$\vec{\theta_d}|X_d, \Sigma \sim LogisticNormal(\mu = X_d\gamma, \Sigma)$$
(4.1)

2. Forming a word probability given a topic: A news organization forms a probability to use words  $(\beta_{d,k})$  for the article d and given a topic k, which is in turn determined by the baseline weight (m), the deviation by a topic k,  $\kappa_k$ , the deviation by sharing decision g,  $\kappa_g$  and the interaction between the two  $\kappa_i$ .

$$\beta_{d,k} \propto \exp(m + \kappa_k + \kappa_{g_d} + \kappa_{i=g_d}) \tag{4.2}$$

- 3. For each word (n) in a news article  $(d), n \in 1, \ldots, N_d$ :
  - Decision on news topic: A news organization draws word's topic based on the attention to topics formed in 4.1.

$$z_{d,n} | \vec{\theta}_d \sim Multinomial(\vec{\theta})$$



Figure 4.5: A graphical representation of sharing-dependent topic model.

• Decision on words: Conditional on the topic chosen, a news organization draws an observed word from that topic according to the word probability formed in 4.2.

$$w_{w,n}|z_{d,n},\beta_{d,k=z} \sim Multinomial(\beta_{d,k=z})$$

This formal specification, also visualized as a graphical model in Figure 4.5, allows us to identify discriminating words that control topics. The vanilla topic model without covariates is conditional only on the first and last strata (attention to topics and word use probability) in Figure 4.5 to find latent topics. On the other hand, the two strata in the structural topic model are in turn conditional on the news article specific link sharing decision (shared?) in two ways. First, the attention to a topic depends on the sharing decision, which (1) in Figure 4.5 represents. This corresponds to equation 4.1 in the formal specification above. If  $\gamma$  in equation 4.1 is positively large for a topic k, a news article with the topic k is likely to be shared. Second, word frequencies depend on the sharing decision, which (2) in Figure 4.5 represents.

Lastly, the structural topic models adopt the same Laplace prior to the word frequency coefficients. Thus, the structural topic model adopts the same automatic variable selection technique as in the recent discriminating words algorithms Monroe et al. (2008); Taddy (2013); Mitra & Gilbert (2014).

In sum, the structural topic model provides a computational way to identify several associations between news text and the sharing decision. First, it identifies news topics and how they are associated with the sharing decision. Second, it identifies discriminating words after controlling the topics. Lastly, it also identifies discriminating words within each topic. Further, this is a way to translate my research question – how do news organizations condition their link sharing on news content? – to a computationally answerable question – what types of latent topics are more likely to be shared?

#### 4.2.3 Paraphrasing for Twitter: Measuring Sentiment with Dictionaries

To analyze how news organizations paraphrase their news stories for social media, I adopt the dictionary based word-counting method that has become a primary way to measure emotion or sentiment in computational text analysis works. In particular, by counting emotion words in news stories and tweets, I test whether news organizations add emotional framing as they paraphrase their news stories. For example, if news organizations want to make news stories look sensational to attract social media users' attention, the paraphrases will contain words with strong sentiment (negative or positive) compared to news stories.

Identifying words likely to appear on a paraphrase can be efficiently conducted by computational algorithms that count words. Although they haven't been applied to news organziations' paraphrasing for social media, dictionary based counting words has been widely adopted to study other online news contexts. For example, Horne & Adali (2017) counted words related to sentiments as categorized in pre-existing dictionaries to measure sentiment in Reddit users' news paraphrase. Also, Berger & Milkman (2012) and Kim (2015) applied a similar method to understand the relationship between sentiment in news stories and online news popularity. I will take a similar word counting approach to identify kinds of phrases likely to appear in social media posts of news organizations.

A caveat in the previous literature that analyzed news paraphrasing is that most of the

literature analyzed only paraphrased text. However, this approach makes it hard to know whether sentiment in a news tweet came from the original news story or from the news organization's strategy. In an exceptional work, Andrew (2007) compared headlines in major Canadian newspapers with actual news content during the 2004 Canadian federal election to focus on news organizations' paraphrasing choices. With this approach, he revealed that the Canadian newspapers choose issue words rather than horserace words (i.e. words that describe election as a game), and more non-neutral (biased) words in news headlines, compared to original news stories. In a similar way, my word counting approach will include a comparison between social media posts and original news texts to distinguish news organizations' paraphrasing choices from news content.

There are many available dictionaries developed in the context of different fields. These include General Inquirer Database (GI) from political science Stone et al. (1966), DICTION from communication Hart (2001), Linguistic Inquiry and Word Count (LIWC) Pennebaker et al., the Regressive Imagery Dictionary (RID) Martindale (1975) and TAS/C Mergenthaler (1996) from psychology, Affective Norms for English Words (ANEW) from behavioral science Bradley & Lang (1999), Dictionary of Affect in Language (DAL) from literature Whissell & Dewson (1986), WordNet-Affect (WNA) from linguistics Strapparava et al. (2004), and Pointwise Mutual Information wordlist (PMI) from computational linguistics Turney & Littman (2003). Most recently, by synthesizing existing dictionaries, Young & Soroka (2012) developed a dictionary called Lexicoder Sentiment Dictionary (LSD) tailored to political news.

In Chapter 7, I will apply three popular dictionaries to the data; Affin, LSD and LIWC. Affin and LSD are used because they are customized for two specific information formats of interest in this section – tweets and news stories. In other words, this choice is intended to check the robustness of results across different dictionaries designed for different contexts. LIWC is chosen because it contains the most extensive words and word categories, and has been by far the most often used for academic research. I specifically used the latest version of LIWC 2015.

#### CHAPTER 5

### A SIMPLE TEST OF THE STRATEGIC LINK SHARING MODEL

## 5.1 Introduction

It is not feasible to test empirically the excessive link sharing hypothesis from the economic model in Chapter 3 because the *first best* choice is unobservable from the data. However, whether news organizations take competition for limited attention into consideration for the number of shared news stories is testable. As illustrated at the end of Chapter 3, if the conceptual model is modified so that news organizations do not consider the competition, the proportion of shared news stories as their optimal choice should not react to the number of stories other organizations publish. On the other hand, according to the original model with competition, the proportion of shared news stories in the equilibrium is a decreasing function of the total number of news stories published by all news organizations. Specifically, the proportion of shared news stories is decided as  $\frac{Np}{c}\frac{K-1}{K}\frac{1}{\phi} = \frac{Np}{c}\frac{K-1}{K}\frac{T}{MK}$ . In other words, if there are more news stories compared to social media users' attention, the proportion decreases. Therefore, the direction of news organizations' reaction to the total number of news stories is indicative of whether news organizations consider competition as they share news links. This comparison provides a testable hypothesis: news organizations will share a lesser proportion of published news stories when all news organizations publish more stories, if they consider the competition for limited attention. In this chapter, I will test this hypothesis by analyzing news organizations' daily news link sharing choices on Twitter.

# 5.2 Analysis

### 5.2.1 Empirical Models

Whether a news organization considers the competition or not, the theoretical model predicts that an organization reacts to the number of its own news stories. However, the model does not say how differently they will react to their own and the total news because all stories published by all organizations were assumed to have the same quality in the model. In practice, there is a good reason for the conjecture that news organizations more sensitively react to the number of its own news. Unlike the model in Chapter 3, each news organization is likely to have a number of *captured* readers who prefer the specific organization to others. In this case, the captured readers' attention will be more responsive to the amount of information from the preferred news organization than those from others. For example, whether Breitbart publishes a lot of news would not matter to far-left readers. Because those captured readers are more likely to incur profit for the preferred news organizations, organizations will also care specifically about the number of their own news links shared on social media in order to not overload attention of the captured readers as well as the total amount.

Figure 4.2 visualizes the relationship between the proportion of shared news stories and the number of published stories on Twitter, broken down by news organizations. There are a few news organizations that publish as many as 500 daily stories because some regional papers publish stories originally published by other outlets with the same owner or by those with whom they partner. These news organizations tend to share fewer of their partners' stories through their Twitter account, presumably because they prioritize their own stories. This pattern from partnership may make the potential negative association between the number of published news stories and the shared proportion look stronger than it really is. Thus, I repeatedly conduct analyses with data without those outliers.

The main dataset described in Chapter 4 is used for the regression analysis. Each news

story in the dataset is tagged with whether it was shared by a news organization on Twitter. By counting stories shared by each news organization everyday, I calculated the daily proportion of shared news. Thus, the dependent variable is the proportion (percentage) of news stories shared by a news organization on each day, and the independent variables are the number of its own news stories on each day, and the total number of news stories published by all 84 news organizations included in the dataset on each day. The impact of the total news implies a news organization's strategic situation in the sense that one's action depends on others. I also included types of news organizations, as in Table 4.1, to see if the different types tend to adopt different news link sharing strategies, and the day of the week to control an apparent weekly cycle.

In addition to the basic linear regression model (Model 1), I was able to take advantage of the panel structure of the dataset. In particular, Model 2 is a fixed effect model that controls each news organization's unobservable tendency regarding the news link sharing proportion by looking at the variation *within* each news organization. In this case, however, the impact of the types of news organizations cannot be examined because the types are an invariant variable. Model 3 and Model 4 are replications of the two models without the outliers.

### 5.2.2 Results

Table 5.1 shows results of the regression. Model 1 shows that a news organization reduces the proportion of news links it shares on Twitter responding to both the number of its own news stories and the total number of news stories published by all news organizations. If a news organization publishes one more news story, then it will share 0.0869% fewer news stories (SE 0.0048). If there is one more news story published in the market, a news organization will share 0.0011% *fewer* news stories (SE 0.0003). This result confirms the hypothesis from the model that news organizations are strategic about social media users' limited attention. The impact of the number of the total news stories is not conflated with a reaction to the number of a news organization's own news stories because the negative impact of the total

Variable	Model 1	Model 2	Model 3	Model 4
Intercept	79.9512***	2.1521*	$81.0988^{***}$	2.2692*
_	(3.2813)	(0.9231)	(3.3776)	(0.9568)
Own News	-0.0869 * * *	-0.0365***	-0.0964***	-0.0307*
	(0.0048)	(0.0095)	(0.0060)	(0.0119)
Total News	-0.0011**	-0.0017***	$-0.0011^{***}$	-0.0019***
	(0.0003)	(0.0003)	(0.0003)	(0.0.0003)
Tuesday	0.8413	0.6911	0.8468	0.6868
	(1.7496)	(1.3602)	(1.8092)	(1.4092)
Wednesday	0.2331	0.1827	0.2228	0.1747
	(1.7711)	(1.3762)	(1.8313)	(1.4257)
Thursday	-0.3278	-0.4177	-0.3735	-0.4245
	(1.7608)	(1.3687)	(1.8210)	(1.4182)
Friday	-1.3547	-1.4576	-1.4386	-1.5114
	(1.7165)	(1.3353)	(1.7751)	(1.3836)
Saturday	-8.1999***	-8.4452***	-8.6746	-8.9201 ***
	(2.0238)	(1.5682)	(2.0948)	(1.6261)
Sunday	-6.5905 * *	-6.4503 ***	-6.9454 **	$-6.7895^{***}$
	(2.0424)	(1.5804)	(2.1143)	(1.6389)
National	-10.3072***		-10.9206 ***	
	(1.9874)		(2.0395)	
Online	-12.4145***		-12.5414***	
	(1.8272)		(1.8545)	
Regional	-15.8720***		-15.1256 ***	
	(1.8649)		(1.8998)	
R	0.1140	0.0204	0.0946	0.0200
Ν	$4,\!532$	$4,\!532$	4,361	4,361
Fixed Effect	No	Yes	No	Yes
Outliers	Yes	Yes	No	No

Table 5.1: Regression of the proportion of news stories shared by news organizations on Twitter.

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

news stories is observed even when controlling out the number of the organization's own news stories.

Although the total number of news stories seems to have a relatively small impact as expected – smaller than the impact of their own news, it can be translated into a significant impact in a realistic situation. For example, possibly because of a significant event, suppose 100 news organizations published ten more news stories. The impact of their own news is only translated into 0.869% fewer news links. If a news organization was publishing 100 news stories, ten more news stories simply cancel out this impact, leaving the number of news links almost the same. However, ten more news stories also means 1,000 more news stories on the market in this scenario. Its impact is translated into 1% fewer news link sharing according to the estimated results. This impact would become even larger if news organizations consider competition with other news sources, and non-news-specific sources discussing a popular news issue are excluded from the data.

This negative impact of the total news stories is robust even when controlling the idiosyncratic characteristics of news organizations using the fixed effect estimation (Model 2). While the estimated impact of the number of their own news became smaller (-0.0365; SE 0.0095), the estimated impact of the total number of news becomes even stronger (-0.0017; SE 0.0003) than it was in Model 1. Thus, the fixed effect model provides a stronger support for the strategic link sharing hypothesis.

To see if the organizations that publish many stories on their websites drive the results, I excluded those outliers on the far right side of Figure 4.2. In particular, I excluded Reuters, The Seattle Times, USA Today, Washington Post and Washington Times – those publishing more than 200 stories daily. Models 3 and 4 in Table 5.1 are the reiterated versions of Models 1 and 2. However, the results without the outliers are hardly distinguishable from the earlier results. The only significant difference is that the impact of the amount of own news estimated by fixed effect becomes less statistically significant when data without the outliers are used (from P-value < 0.001 to P-value < 0.05). However, the estimated impact of the total news on the market stays as statistically significant as that in Model 2, and its effect size became slightly larger (from -0.0017 to -0.0019). Thus, the estimated results without the outliers are the same or slightly closer to the strategic situation between news organizations that the conceptual model tried to capture.

The pattern from each day of the week is as expected; news organizations share a significantly smaller proportion of news stories on Twitter during the weekends (about 8% less on Saturday and about 6.5% less on Sunday). This result is very similar in Model 3.

Finally, different types of news organizations predict significantly different proportions of news stories shared on Twitter, according to Model 1. The baseline case here is magazines, which publish a significantly smaller proportion of news stories compared to daily news outlets. Thus, they tend to share a large portion of news stories (potentially suffering less of the information overload problem). Among other types of news outlets, national media tend to share a larger fraction of their news stories followed by online news outlets. However, the difference between these two groups was not large ( $-10.3072 \approx -12.4145$ ). Indeed, the difference between the two coefficient was not statistically significant by a linear hypothesis test (p = 0.13). On the other hand, regional media share a significantly smaller proportion of their news (-15.8720; SE 1.8649) than both national and online media do. This is also confirmed by linear hypothesis testing (p < 0.01 and p < 0.001 respectively). The estimated differences in proportions of shared news stories between different types of organizations in Model 3 were very similar to those of Model 1.

This result has two potential explanations. First, because regional news organizations often partner with national news organizations, they may not share many of stories from the partners. Second, because financial suffering prevails among regional media, they may be less engaged in social media strategies. Indeed, it is known that national news organizations are engaged in sophisticated social media strategies and adjust themselves with the new online environment (NYT, The Guardian, BBC) as much as online only ventures (Kew, 2016, August 17; Rowan, 2014, January 2; Newman, 2011) whereas regional media keep downsizing their organizations. However, there may be a counter-argument to the second explanation where those less engaged in social media strategies leads to more shared news because they resort to an old practice to automatically feed their news stories through social media accounts.

# 5.3 Discussion

A simplest, but intuitive guess about news organizations' social media strategy would be sharing as many stories as they believe readers can process regardless of other organizations' choices. However, if news organizations care about the chance that their news links are clicked, and competition with other media for users' limited attention, there is an incentive to adjust how much information they want to disseminate as formalized in the theoretical model in Chapter 3. Older evidence revealed that tweets from news organizations were automatically fed as in RSS (Palser, 2009; Armstrong & Gao, 2010). However, more recently, news organizations are increasingly adopting manual customization of news tweets (Newman, 2011), hiring data scientists to discover optimal social media strategies (Rowan, 2014, January 2), and developing social media monitoring platforms (Diakopoulos, 2017).

The empirical results in this chapter support the hypothesis that news organizations indeed care about users' information overload and what other organizations do as they disseminate news stories on Twitter, hence *strategic*. As the economic model predicts, news organizations shrink the proportion of news stories they disseminate via Twitter when the whole market publishes more news stories. They also reduce the proportion when they publish more on their own, which implies that they also care about captured consumers who prefer their stories. However, this result does not contradict the idea of strategic link sharing. News organizations' reactions to the total amount of news published in the market was statistically significant even when controlling the reaction to their owns. The impact of the strategic choice was even stronger than the reaction to their own news in a realistic situation.

Although the simple analysis that focuses only on the quantitative aspect of link sharing on Twitter omits many qualitative aspects of potential strategies, it clarifies that news dissemination through Twitter is significantly governed by strategic concerns. This raises a concern about whether the strategic link sharing on social media will provide the same sufficient, fair, and diverse news information that can inform society as we expected from traditional journalistic norms. This market driven news dissemination can still support the desirable provision of news if organizations' strategies are diverse enough. But, if they are highly patterned, the immigration toward Twitter and Facebook for news consumption can restrict or bias news perception. Thus, to what extent news information disseminated via social media deviates from the traditional editorial decisions is an important empirical question to gauge the significance of the new information mediating process newly imposed by social media.

Neither the game theoretic model nor the empirical results addresses *how* news organizations reflect other news organizations' behaviors because both explain only outcomes of the interaction rather than how they get there. One realistic scenario is that news organizations preemptively shrink the proportion of shared news when dominantly popular news breaks out to mitigate potential information overload. In this case, news organizations may want to share less about less important news topics. This scenario specifically evokes a possibility that strategic news link sharing causes a concentration on certain news topics on social media. The potential concentration issue is particularly important because it can reduce the diversity of news information on social media. The empirical assessment of the distribution of shared news topics is the subject of the next chapter.

#### CHAPTER 6

## SELECTIVE NEW LINK SHARING ON TWITTER: A STRUCTURAL TOPIC MODEL

# 6.1 Introduction

The economic model suggested in Chapter 3 implies that news organizations cannot individually escape from the equilibrium where they excessively share many stories. Although the limited attention as a public resource provides a theoretical basis that predicts the quantitative choice of organizations as an outcome of competition, the excessive sharing equilibrium is built on the simplifying assumptions made for the model, such as identical news information across different organizations and stories. In reality, factors other than how many news stories are published must affect a news organization's decision on whether to share. As one such possibility, in this chapter, I investigate *selective news link sharing* – how news organizations condition the link sharing decision on news content. Applying a structural topic model (STM) recently developed at the intersection of political science, statistics and machine learning (Roberts et al., 2016) to the Web scraped news stories and tweets, I will show which news topics are more likely to be more visible than others on news organizations' Twitter accounts.

This study is particularly meaningful for the potential impact of social media as a news platform on news diversity because types of news stories shared on Twitter will influence the set of information to which users have access. In other words, news organizations' selective link sharing is potentially a new form of *gatekeeping* (Dimitrova et al., 2003) to the extent that it systematically deviates from traditional editorial decisions.
# 6.2 Motivation and Related Works

Unlike the simplified model suggested in Chapter 3, news quality consists of a variety of dimensions; there is a distinction between factual news and commentary. Some news stories are in-depth whereas others are short reports. Different news stories may contain different viewpoints about the same public issue. A single news report may be about a variety of public issues. Moreover, news readers' preference for each dimension of news quality is diverse as well. Thus, each news organization aims at different groups of news readers; to use an economic term, news is a *vertically and horizontally differentiated good* (Gentzkow & Shapiro, 2010). A natural economic behavior of a firm that sells differentiated goods is to condition its supply on the quality of the goods (Anderson et al., 1992). This is because the quality affects consumers' valuation of the good, which in turn determines demand the firm faces. Thus, as far as firm's behavior depends on demand, it also depends on the quality of the product.

To translate this argument into the context of this study, news organizations will condition news link sharing on news content. However, it is hard to incorporate the multidimensional quality differentiation into a mathematical model; such a model will have too many parameters structured in a complicated way to be solved to produce predictable hypotheses. Instead, one can empirically explore how the news link sharing decision differs depending on a variety news qualities to identify the selective news link sharing. Although drawing predictions from a formal economic model is not feasible, findings from the online journalism literature provide some conjectures about news types more or less likely to be disseminated via social media. In this section, I will review this literature to complement the standard economic theory with what previous findings about news organizations' online news choices and their social media postings predict about their selective news link sharing on social media as a news distribution platform.

## 6.2.1 Social Media as A News Distribution Platform

Studies about news organizations' tweets hint at what types of news stories news organizations share on Twitter. Initially, media scholars expected that the Web would afford news organizations a channel for the mutual interaction between media and audience (Chan-Olmsted & Park, 2000), and would be an effective promotion tool to attract younger audiences who do not regularly access traditional media (Palser, 2009; Chan-Olmsted et al., 2013). However, evidence that news organizations are using these opportunities are rather weak. For example, Greer & Ferguson (2011) analyzed tweets from 488 local TV stations, and found out that only 23.3% of 455 commercial TV stations tweets for interaction with news readers. Similarly, Meyer & Tang (2015) recently hand-coded 4,507 tweets from 60 local news organizations (TV and newspaper), and only 7.4% of the tweets from local television stations and 11.6% from local newspapers were intended for interaction. Cleary et al. (2015) drew 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 & Gao (2010) concluded that traditional news organizations generally extend their conventions to Twitter rather than adopting a customized strategy.

Traditional news organizations rather use social media as an additional news distribution platform. Greer & Ferguson (2011) found that 94.9% of the commercial TV stations tweets to disseminate their news articles whereas only 17.6% tweet to promote their programs. Similarly, Meyer & Tang (2015) reported that 94.4% of the tweets from TV stations and 96.3% from newspapers are for news link sharing. The news distribution through social media is a profitable tactic. Hong (2012) found out that the social media use and the number of tweets by the news organizations induce more traffic toward their news websites. Therefore, news link sharing would be the main concern when news companies contemplate their social media use.

Previous studies have found some evidence that news organizations are strategic in dis-

seminating their stories via social media. Armstrong & Gao (2010) speculated that the abundance of links to 'sensationalistic' news with topics, such as crime and many life-style news links by local media compared to national and regional media are a result of a consideration for the targeted demand. Although news news link sharing seemed automatic for many organizations that considered the same paraphrasing for social media as headlines from original stories (Palser, 2009; Armstrong & Gao, 2010), Armstrong and Gao also found that some organizations add news leads whereas other news organizations post only headlines. Further, more recent studies found that news organizations are adopting manual curation optimized for social media. According to Newman (2011), the BBC switched from automatic feed to manual choice and editing optimized for Twitter in 2011, and the number of followers has doubled since then. Recognizing this potential, Facebook published a guideline for strategic news posts based on their own user news engagement studies.<sup>1</sup>

## 6.2.2 News Organizations' Online News Choice

Although studies about how news organizations, as firms, choose a subset of published news stories to disseminate via social media are rare, their traditional gatekeeping practice provides some clue. The selective news link sharing decision for social media resembles the traditional gatekeeping function of journalism in that it decides the salience of a subset of information. One can even find a primitive form of selective news link sharing from the original conception of "gatekeeper" in the literature; As news organizations choose news likely to draw more attention or to be more newsworthy on social media, in the past a wireeditor selected stories from a mass of wire copies from wire services (White, 1950). The news bulletin is another old practice even closer to news link selection. Whereas the traditionally understood gatekeeping is a process through which journalists selected news information out of *raw* information, both the news link sharing and the news bulletin give more salience to a

 $<sup>\</sup>label{eq:linear} $$^1$ https://www.facebook.com/notes/facebook-journalists/study-how-people-are-engaging-journalists-on-facebook-best-practices/245775148767840$ 

subset of *published* news (Bruns, 2011). Indeed, in the online context, Dimitrova et al. (2003) showed that news organizations use hyperlinks as a gatekeeping tool to control out-traffic to information sources.

Recent discussions about news aggregating services show that what happens on those platforms imposes information constraints on news readers on top of traditional journalism practice. When news readers rely on separate platforms, such as Google News or Yahoo News, to reduce information to the amount that they can intelligently process, the platform can have a strong power to decide the salience of news information. In this vein, Carlson (2007) argues that Google's choice of news has a different motivation from traditional media's, and that the move of the gatekeeping power to the aggregators can ultimately undermines the traditional values of journalism. More recently, the discussion about a filter bubble (Pariser, 2011) warns that the algorithmic recommendation of news stories can disturb the traditional process through which legacy journalism provides quality information about public issues because news readers fall into a feedback loop whereby news readers become trapped by the algorithms in what they are estimated to prefer.

Although media theories have expanded the gatekeeper concept to capture multiple gates within an organization that happens during the editorial process and the inter-media gatekeeping (Shoemaker & Vos, 2009). The selective link sharing on social media is fully captured by neither framework because gatekeeping on social media happens *within* an organization, but also *between* multi-platforms. Indeed, if news organizations selectively share news links on social media, this suggests that the gatekeeping process within an organization can be governed by different socio-technical constraints depending on platforms used for news dissemination.

Although research about multi-platform gatekeeping within an organization is limited, there are comparisons between audience's news choices on online platforms and journalists' hints at social media users' news preferences. For example, Zubiaga (2013) compares the New York Times' top news stories as an editorial decision with stories that are the most popular on Twitter and Facebook. Analyzing more than 56K news stories, he found that professional editors tend to select hard news such as US politics and international relations whereas social media users like relatively soft news such as science, fashion and technology. If news organizations optimize their link selection for the number of clicks or virality on social media, users' orientation toward soft news will manifest itself in news organizations' choices for social media as well.

Recent evidence from the field suggests that decisions involved with news dissemination via social media is relatively autonomous from the traditional editorial process. According to social media specialists' testimonies to the American Press Institute (Elizabeth, 2017, November 14), the duties and responsibilities regarding social media posting are fractured rather than being tied to the centralized editorial desk decisions. Although they vary across organizations, social media strategies are perceived and organized as business strategies rather than journalism practices. In 2015, Michael Roston, a senior staff editor social media at The New York Times, wrote that The Times' social media desks' editors joined the Audience Development department, which focuses on search engine optimizations, analytics and growth (Roston, 2015, January 22). As a result, many social media specialists feel that their job does not reside at the core of tasks of news organizations. "My work in social media always felt somewhat removed from the rest of editorial; like the writers were operating independently and my job was to catch and promote as much of their content as I could. We didn't feel integrated" (Elizabeth, 2017, November 14).

This tendency is escalated by the wide-spread use of audience metric techniques for social media optimization. What to post and when to post online increasingly depend on the reaction of audiences. For many social media specialists, the basic routine that dominates their job is "post and count" (Elizabeth, 2017, November 14). Using an ethnographical method, Tandoc Jr (2014) found that use of audience metric and recommendation algorithms indeed results in a new drive toward an additional dimension of gatekeeping for online journalism. Selection and deselection of news stories on organizations' news dissemination platforms are

quickly decided according to outcomes of real-time audience metrics or even of automated algorithms. To make this process even more efficient, leading news organizations have been developing social media monitoring and decision-aid platforms on their own (Diakopoulos, 2017). The degree to which social media related decisions are integrated into editorial varies. Although it is partially influenced by how well a news organization does in the market, but it's not definite. For example, whereas The New York Times' case is leaning toward a business orientation of social media strategies, Elizabeth (2017, November 14) found that the Washington Post takes a different path to encourage social media specialists and news rooms to work together. In some regional newspaper companies, a very limited number of staff, even a single person, decides the social media posting. But sometimes it is entirely dispersed among individual reporters without aid of social media specialists, mainly because of the tight financial constraint. Thus, although it is almost certain that selective link sharing will deviate from the traditional gatekeeping decisions, its degree for individual organizations is hard to predict beforehand.

All in all, previous studies imply that news organizations are likely to be engaged in a gatekeeping task for social media that is separate from the traditional editorial process, and that the social media gatekeeping strategy is likely to be governed by a perceived mechanism by which news stories would go viral on social media. Thus, news organizations are likely to share news stories with specific components, such as topics that are known to contribute to virality. Previous studies about news virality have identified such news components that help virality: Evidence generally shows that emotional arousal (intensity) is associated with virality on social media (Berger & Milkman, 2012; Stieglitz & Dang-Xuan, 2013). Focusing on news topics, García-Perdomo et al. (2017) found that conflict / controversy, human interest and odd news categories are more likely to be shared by Twitter users. Thus, assuming that news organizations are adapting to such news demand on Twitter, one can hypothesize that news organizations will share news topics likely to contain high emotional arousal, controversy, human interest, or oddity. Further, as the emerging online news organizations

have bigger stakes in social media to disseminate their stories, this pattern will be more evident for them. I will test these hypotheses by analyzing in the next section which news links each news organization shared on its Twitter account.

# 6.3 Computational Identification of Selective News Link Sharing

Identifying the selective news link sharing on Twitter requires a two-step analysis to find out (a) what a news story is about and (b) how the news content is associated with the link sharing decision. Because of the difficulty to achieve a representative sample of published news and the long list of news organizations, the amount of data involved with this analysis became enormous. In this chapter, I adopt a machine learning technique called Structural Topic Model (STM) (Roberts et al., 2016) to analyze more than 40K web-scraped online news stories from 84 outlets. STM will simultaneously identify news *topic* – a computational operationalization of news content – and find the relationship between the news topics and observable variables of interest, such as which organization published the news story and whether the organization shared the news on Twitter in this case.<sup>2</sup>

## 6.3.1 News Topic Categorization

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 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. The challenge is, slightly different, but conceptually similar to the traditional question for manual content analysis: are the

<sup>&</sup>lt;sup>2</sup>See Chapter 4 for a detailed description of the method.

discovered topics based on the number of topics assumption 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. If the number of topics is too low, on the other hand, a researcher is likely to miss interesting variations across different news contents.

Researchers have suggested multiple indices to compare performance of different topic number assumptions from the perspective of the statistical model comparison. For example, Blei et al. (2003) suggested the held-out log-likelihood approach. This approach compares likelihoods from held-out data – as in the common cross-validation – using estimated parameters based on different choices of topic number. Other scholars suggested an approach closer to human intuition of topics; if a topic is well-defined, words that characterize a topic should appear frequently in the same document (coherence) (Mimno et al., 2011; Newman et al., 2010), and words that characterize different topics should not be in the same document frequently (exclusivity) (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 the choices of human coders (Chang et al., 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. After all, the number of topics tends to be up for researchers' careful choices, based on their research questions. In this case, the interpretation of topic models is "What's the best categorization of news topics given the number of topics?"

For this study, I chose thirty topics. The main reason for this choice is 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 great interest to this study. However, if the model too finely categorizes news topics 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, I found that STM begins to identify news about specific regions with thirty-two topics from this study's dataset. Thus, I decided to limit the number of topics to thirty so that the model can identify more broadly labeled regional topics.

Table 6.1 summarizes the identified topics and words most likely to appear in a given topic. I labeled each topic by inspecting the list of the associated words and randomly chosen news stories among those estimated to be most likely to be about the specific topic. For most cases, the associated words provide a coherent interpretation about each topic. For example, the output is precise enough to distinguish different sports: the Sports/Basketball topic is associated with words such as '3-pointer', '3-point' and 'rebound' whereas 'nfc', 'touchdown', 'nfl' 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 and Politics/Congress - from the Regional Politics topic, which is associated with words, such as 'mayor', 'counti', 'neighborhood', and names of regional communities. There are a few cases in which the news format or functionality dominate news content, such as Subscription, Advertising and French (language). Although these topics are of less interest in this study, identifying them was unavoidable because news websites include a few native advertising or advertising for its own, and Media Cloud does not distinguish them. However, STM's successful detection of non-news messages helped narrow my focus on actual news stories of my interest for this study. Moreover, these categories are not frequent in the dataset, and are less likely to contain substantive news content overlapped with other news topics because most have non-news related functions (e.g. subscription and adverstising) or characteristics (e.g. French language).

Figure 6.1 visualizes 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 Style, Crime, and sports related topics occupy top ranks of the list. However, there is a good amount of *hard news* with topics related to national politics, economy and international relations as well. Although there is only topic directly related to regional issues, Regional, the sports-related topics and topics, such as Education, Transportation and Crime are likely to be more related to regional issues.

## 6.3.2 Overall Patterns of Selective Link Sharing

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, and (c) interactions between the two. T estimated coefficients for the covariates mean (a) how much more often a given topic is *shared overall* compared to the topic's proportion among the total news stories, (b) how much more often a given topic is *published* by a news organization compared to other organizations, and (c) how much more often a given topic is *shared by a news organization* compared to other organizations. In other words, this regression allows detecting both topics more or less likely to be shared overall, and topics that are more or less likely to be shared by specific news organizations.

Figure 6.2 presents the overall propensity. The zero in the middle is the topic-wise baseline proportion among all the published news stories, which in fact varies across different topics. The points represent how the proportion deviates from the baseline when only looking at the shared topics. Thus, it can be interpreted as a relative propensity of topics to be shared. In what follows, ABC was chosen as a baseline to visualize how different organizations' selective link sharing deviates from national, traditional, and (relatively) neutral media outlet. The topics are sorted from the most common to the rarest.

Table 6.1: Words likely to appear given each topic.

Label	Words with High Probability
Life Style/Interview	yeah, youv, im, ive, your, weird, oh, mayb, stuff, shes, funni, guy, crazi, guess, laugh, youll, ok, hey, id, somehodi
Crime	polic, custodi, arrest, sheriff, prison, feloni, jail, detain, sentenc, murder, prosecutor, juri, patrol indict, prosecut, convict, testifi, plead, shoot, fied
${ m Sp}{ m orts}/{ m Basket}{ m ball}$	3-pointer, 3-point, rebound, turnov, basket, streak, nba, halftim, foul, overtim, score, basketbal,
m Economy/Tech	softwar, retail, compani, inc, googl, batteri, appl, buyer, user, maker, innov, manufactur, sale,
$\rm Social/Conflict$	democraci, planet, reader, earth, radic, mainstream, cultur, phrase, context, truth, racist,
International/Middle East	jerusalem, syria, saudi, islam, isra, embassi, arab, israel, refuge, iraq, muslim, iran, minist,
Legal/Regulation	lawsuit, regulatori, court, suprem, constitut, regul, commiss, legislatur, lawyer, ban, legal,
Economy/Finance	bit coin, currenc, inflat, economist, index, investor, stock, billion, growth, profit, equiti, market,
UK	strategist, forecast, fike, ouyer, sector, asset, price, fivest princ, queen, marriag, marri, royal, bbc, caption, divorc, harri, copyright, kate, pregnant, wed, givifiniand bitthday us, ches, extress london babi
$\operatorname{Politics}/\operatorname{President}$	mueller, flynn, fbi, hillari, clinton, russia, trump, russian, presidenti, judiciari, counsel, obama,
Updates	p.m, a.m, vs, dec, pic.twitter.com, t.co, jan, 8, 9, 7, aug, feb, nov, 6, pm, 2017, nashvill, decemb 4 oct
$\operatorname{Sp}\operatorname{orts}/\operatorname{Foot}\operatorname{ball}$	nfc, touchown, vike, nfl, lineback, postseason, quarterback, cornerback, playoff, yard, offsea-
$\operatorname{Sp}orts/\operatorname{Other}$	basebal, tiger, leagu, cup, soccer, championship, tournament, player, stadium, roster, sport, offseason, premier, squad, golf, olymp, captain, athlet, australian, club
$\operatorname{Ent}\operatorname{ert}\operatorname{ain}\operatorname{ment}$	album, disney, music, song, theater, comedi, movi, studio, netflix, singer, film, sing, episod, artist hand festiv funni actor premier viewer
Social/Gender	franken, inappropri, harass, lawsuit, resign, alleg, complaint, misconduct, settlement, sexual, accus, kiss, investig, staffer, applog, ethic, inquiri, employe, uncomfort, workplac
$\operatorname{Regional}/\operatorname{Politics}$	firefight, evacu, homeless, santa, shelter, mayor, counti, neighborhood, francisco, flame, diego, orang downtown smoke oakland los hurrican flee rent fire
Weather	temperatur, snow, outdoor, fish, tree, bird, salt, winter, weather, inch, ice, hike, rain, hotel, ocean, mountain, denver, tourist, sea, forest
Transportation	passeng, airlin, airport, flight, navi, plane, port, crash, vehicl, pilot, transport, crew, highway, driver, patrol, truck, traffic, rescu, accid, ship
Health	diseas, patient, clinic, medicin, cancer, addict, drug, medic, health, doctor, brain, alcohol, dr, hospit, nurs, surgeri, depress, studi, suicid, treatment
Subscription	subscrib, newslett, editor, inbox, advertis, reader, click, robot, pleas, jersey, york, verifi, cabl, privaci, occasion, commentari, columnist, publish, editori, manhattan
Tax	deduct, tax, trillion, taxpay, debt, fiscal, insur, economist, revenu, corpor, premium, incom, hike, provis, enrol, deficit, wage, treasuri, reform, repeal
Social Media	youtub, user, instagram, facebook, googl, app, platform, internet, onlin, blog, fake, web, video, social, mail, camera, audio, footag, content, compu
Immigration	daca, bipartisan, shutdown, mcconnel, schumer, immigr, deport, legisl, congression, gop, senat, republican, sen, democrat, congress, capitol, lawmak, short-term, mitch, rep
Education	student, teacher, educ, campus, graduat, school, nonprofit, teach, enrol, donat, academ, balti- mor, district, donor, colleg, grade, fundrais, chariti, volunt, taught
$\operatorname{Sp}\operatorname{orts}/\operatorname{College}$	ncaa, freshman, coach, footbal, iowa, championship, espn, sophomor, bowl, athlet, tennesse, mississippi, michigan, ohio, miami, basketbal, alabama, teammat, georgia, oklahoma
$\operatorname{Politics}/\operatorname{Congress}$	voter, ballot, moor, roy, alabama, elect, gop, nomine, republican, poll, strategist, democrat, re-elect, nomin, senat, presidenti, legislatur, vote, jone, doug
Me Too	weinstein, women, sex, hollywood, femal, harvey, male, rape, gender, actress, woman, nbc, gay, workplac, men, sexual, rose, charli, uncomfort, discrimin
International/Far~East	korea, missil, korean, nuclear, china, chines, sanction, russia, japan, moscow, russian, diplomat, kim, u.n, ministri, regim, un, olymp, weapon, deleg
French	n, la, de, e, dalla, h, el, del, franc, b, m, v, p, co, l, f, s, r, fort, k
Advertising	emb, U, copi, cup, code, paus, bowl, video, email, link, courtesi, kid, super, sale, camera, recipi, tip, mom, tast, facebook

	REGIONAL: firefight, evacu, homeless     WEATHER: temperatur, snow, fish     TRANSPORTATION: passeng airlin flight					
	HEALTH: patient, SUBSCRIPTION:	diseas, medicin newslett, subscrib	inbox			
	TAX: deduct, tax, t SOCIAL MEDIA: y	trillion outub, user, instag	jram			
	EDUCATION: da	ent, teacher, educ	coach			
	<ul> <li>POLITICS/CONGRE</li> </ul>	ESS: voter, ballot,	, coach moor			
	<ul> <li>ME TOO: weinstein, INTERNATIONAL/FAR</li> </ul>	hollywood, gende EAST: korea, mis	r ssil, korean			
	FRENCH: dalla, n, co ADVERTISING: x, emb.	0				
	, _ ,					
		I				
0.00	0.05	0.10	0.15			
Expected Topic Proportions						

Figure 6.1: A bar graph for estimated overall proportions of 30 topics. The unit is fraction.



Figure 6.2: Relative propensity to be shared on social media of 30 topics. The propensity is measured as a fractional deviation of the proportion of a news topic among shared stories from that among published stories. The topics are ordered by the frequency on news websites. The propensity to be shared and frequency on websites are not notably associated.

There does not seem to be a clear correlation between the frequency of a news topic and the topic-wise propensity to be shared. If there were such a correlation, the points in Figure 6.2 should have been roughly laid on a line, which does not seem to be the case. For example, whereas Social/Conflict and International/Middle East topics are relatively common, they are less common among the shared topics. On the other hand, social media users will see Sports/College topic with a higher chance than when they find news from websites. This does not necessarily mean that social media users will see more Sports/College topic than Economy/Tech topic because the results show only deviations from the published proportion rather than actual proportions on Twitter. For example, because there are many more news stories about Economy/Tech published on websites, even if news organizations are slightly less likely to share them on Twitter, there are likely to be more Economy/Tech news stories than Sports/College ones. What the results do mean is that a news reader is less likely to see Economy/Tech news from Twitter than from news websites.

The observation that the chance of a topic being shared on Twitter does not match with its frequency on news websites implies that the publication decision and the link sharing decision do not necessarily follow the same logic. A common knowledge about a publication decision is that a formal editorial process within a news organization determines what to communicate to readers according to newsworthiness. This decision is governed by traditional journalism norms (Hermida, 2010). On the other hand, recent observations about news organizations' Social Media Optimization (SMO) seems to be more governed by economic incentives (Newman, 2011; Diakopoulos, 2017). Thus, the difference between the patterns on websites and on their social media accounts are likely to represent different sets of information given to news readers, which stem from different governing principles. If news readers' migration to social media remains consistent in the future, the news choices made by economic incentives may significantly impact what news readers see as news.

Types of news more likely to be shared seem to confirm the conjecture that the sharing propensity is governed by economic incentives rather than newsworthiness. The topics with higher propensity to be shared are ones traditionally categorized as 'soft' topics, such as Crime, Sports/ Basketball and Sports/Football. Indeed, among eleven topics that have a larger proportion on Twitter than on news websites (topics on the right of the dotted line in Figure 6.2), the only news topic that is difficult to categorize as soft news was Regional Politics. On the other hand, all other topics traditionally categorized as 'hard' topics, such as International/Middle East, Economy/Finance and Politics/President, have a lower proportion on social media than on websites (topics on the left of the dotted line). If the common wisdom that readers tend to prefer soft news to hard news still holds for Twitter, news organizations' incentives to share news links likely to be clicked may have created this sharing pattern – Economically driven gatekeeping for social media pushes consumable news toward soft topics.

## 6.3.3 Difference Across Organizations

Although the previous analysis of overall news link sharing reveals a general tendency for news links shared on Twitter to be biased toward soft news, the pattern may differ depending on the strategies of individual news organizations. Previous literature has shown that news publishing patterns, which is traditionally conceived as governed by values and conventions of journalism, hinge on news organizations' different incentives. For example, Lischka (2014) illustrates how different media ownership structures and media modes lead to different volumes of news topics, different tones, and different news interpretations. If social media strategies are more subject to economic incentives than the traditional editorial decisions, the differences of news choices across news organizations will be even more conspicuous than the news publishing pattern. Although little is known about news topics shared on social media, there is evidence that different types of news organizations – e.g. public vs. commercial TV stations and TV stations vs. newspapers – tweet in distinct patterns (Greer & Ferguson, 2011; Meyer & Tang, 2015). Further, as illustrated in the overall pattern in the previous section, there is no guarantee that the news publishing pattern and the conditional news link sharing pattern are analogous.

Thus, there are two questions regarding the potentially varying selective news link sharing strategies across different news organizations: (a) Do the likelihoods that a certain topic is shared differ across different organizations? (b) Does the gap between news publishing and news link sharing patterns differ across different organizations? In the regression in the STM model, coefficients for news organization indicators represent how much more of a news topic a news organization publishes than the baseline (ABC). And coefficients for the interaction terms between the link sharing decision dummy and the news organization indicators represent how much the proportion of a topic among shared news stories deviates from the proportion among all the published stories. Thus, the first set of coefficients shows the publishing pattern whereas the second set of coefficients shows how the link sharing deviates from the publishing pattern. To illustrate how different types of news organizations adopt different link sharing strategies, I will visualize these coefficients broken down by news topics with clear patterns.

News publishing patterns accord with organization type. Estimated proportions of published news topics accord with commonsense patterns; national media and online media publish more about national politics – e.g. Politics/President and Politics/Congress whereas regional media publish more about topics that are likely to be preferable to regional audience – e.g. Regional Politics, Sports/Basketball, Sports/College, Sports/Football. Figure 6.3 and 6.4 present a typical example. The transparent bars mean how much proportion of a published topic for each organization deviates from the baseline (ABC) whereas the solid bars mean how much proportion of a tweeted topic for each organization deviates from the baseline deviates from the given organization's publishing pattern. Thus, the transparent and solid bars can be interpreted as each new organization's propensity to *publish* a given news topic on its website and the propensity to *share* links about a given topic on its Twitter account respectively. The green bar means that a news website has a paywall, and the red bar means no paywall. In Figure



Figure 6.3: News publication and link sharing patterns by news organizations: Politics/President topic.

6.3, the national and online media are more likely to publish about Poltics/President topics than the regional media. On the other hand, the regional media are much more likely to publish about Sports/Basketball topics (Figure 6.4).

Although online media's favored topics resemble those of the national media, online media tend to publish more about social issues, such as Immigration (Figure 6.5), Social Media, and Social/Gender, whereas national media tend to publish more about international topics – e.g. International/Middle East (Figure 6.6) and International/Far East (Figure 6.7).

A reversionary strategy is common for favored topics. 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 of a particular news topic when they publish a lot about the topic. For example, Figure 6.5 shows that national and online media frequently write about Immigration, but share less on Twitter. On the other hand, Figure 6.8 shows that many regional media publish a lot about Sports/Football topic, but they tend to share less of their links on Twitter. A natural out-



Figure 6.4: News publication and link sharing patterns by news organizations: Sports/Bas-ketball topic.



Figure 6.5: News publication and link sharing patterns by news organizations: Immigration topic.



Figure 6.6: News publication and link sharing patterns by news organizations: International/Middle East topic.



Figure 6.7: News publication and link sharing patterns by news organizations: International/Far East topic.

# <figure><figure>

Figure 6.8: News publication and link sharing patterns by news organizations: Sports/Football topic.

Shared

Paywall

No

Yes

Published

Coef

come 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 do not want to be too specialized in a given topic and reach a broader audience base. For example, although the Wall Street Journal publishes more about finance than others do, to overwhelm general social media users with the professional analyses of finance may not be a profitable approach. 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 Wall Street Journal obviously publishes 25% points more about Economy/Finance topic than the baseline ABC in the data, but it shares 16% points fewer news links about the same topic on Twitter compared to what it published. A similar explanation can be applied to regional media's reversionary strategy for the Sports/Football topic (Figure 6.8). Although the regional media's specialty is in regional news, and they are likely to expect regional football fans who visit their websites, they may have different incentives on Twitter – to reach out to broader social media users. If



Figure 6.9: News publication and link sharing patterns by news organizations: Crime.

this is the case, regional media are unlikely to share too much about regional football games. 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.

The opposite direction of the reversionary strategy is also possible. In other words, 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 Crime Topic (Figure 6.9) whereas regional media tend to do so with Transportation topic (Figure 6.10).

A strategic retreat is another common strategy for non-specialized topics for regional media. Unlike specialized news topics in which news organizations often counterbalance with less link sharing, news organizations tend to share even fewer non-specialized topics. For example, regional media publish less about International/Middle East topics, and share even less than that on Twitter (Figure 6.6). They also behave in a similar manner with the Politics/President topic (Figure 6.3). This is understandable because if a news

#### Transportation



Figure 6.10: News publication and link sharing patterns by news organizations: Transportation.

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 not likely to attract either the specific target group or the broad news demand on social media.

On the other hand, there was no evidence that other media types also adopt strategic retreat. This is a more predictable result for national media because they tend to have asymmetrically larger reporting infrastructure compared to regional media. In other words, there may not be many news topics in which national media conceive a lack of expertise to the level where they want to adopt strategic retreat. The reason online media, which often have as limited a reporting infrastructure as regional media do, do not adopt the strategic retreat is murkier. One possible explanation is that online news organizations supplement news they cannot directly cover with the reporting of other news organizations, with or even without legitimate partnerships (Weber & Monge, 2011). In this case, they may not particularly conceive their lack of expertise in a certain news topic.

#### Politics/Congress



Figure 6.11: News publication and link sharing patterns by news organizations: Politics/-Congress topic.

A concentration strategy is used for specialized and popular topics. There are some hot topics about which news organizations publish more and share more links. For both national and online media, Politics/Congress is such a hot topic (Figure 6.11). On the other hand, Social/Gender is a hot topic for online media, but not for national media (Figure 6.12). This difference seems related to the earlier point that online media's specialty is more in social issues than national media's.

In the meantime, regional media's hot topics are quite disconnected from national/online media. Figure 6.9 and 6.4 show that regional media tend to both publish and share more of Crime and Basketball topics. Again, these are topics on which regional media have traditionally focused, and which their regional readers are likely to expect from them.

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 news



Figure 6.12: News publication and link sharing patterns by news organizations: Social/Gender topic.

organization's relative specialty (comparative advantage) – e.g. Immigration for national media and Sports/Football for regional media – but the topic's popularity was modest. 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 6.2. 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 great 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 regarding hot topics. Regional media retreat from a few topics that are not their traditional focus of coverage. It was harder to detect discernible patterns from

	Reversion	Retreat	Concentration	
Magazine	e   International/Far East			
	Health			
	Weather			
	President (4)			
National	Crime		$\operatorname{Politics}/\operatorname{Congress}$	
	Finance		m Legal/Regulation~(2)	
	Politics/President			
	Weather			
	Immigration			
	Social/Gender			
	Tax			
	International/Middle East			
- ··	Sports/Others (9)		<b>P.11.</b> (2)	
Online	Crime		Politics/Congress	
	Life/Interview		Social/Gender	
	Politics/President		Legal/Regulation $(3)$	
	Weather			
	Immigration			
D	$\begin{bmatrix} Me & 100 & (6) \\ Cr & arta & (Fr & ath a H) \end{bmatrix}$	Internetional/Middle East	Crante /Deplethell	
Regional	Sports/Football	International/Middle East	Sports/Basketball	
	Education	Politics/President Social/Identity Conflict (2)	Lonomy/lechnology	
	Sports (Others	Social/Identity Connict (5)	Opdate (5)	
	Bagional Delition			
	L oral / Dorulation			
	$\begin{array}{c} \text{Legal/Regulation} \\ \text{Transportation} (7) \end{array}$			

Table 6.2: Summary of the selective news link sharing strategies. Numbers in parenthesis are number of topics categorized as a given strategy.

magazines because fewer media are included in the dataset compared to other types.

# 6.4 Discussion

The findings from the pooled data show the link sharing propensity of a news topic does not accord with the frequencies among all the published stories, confirming that news organizations' link sharing is governed by a different logic from the one that governs traditional gatekeeping decisions. Traditional editorial decisions are thought to be governed by journalistic norms (Hermida, 2010). Instead, there has been a concern that economic incentives drive traditional reporting toward popularity (Lischka, 2014). However, the even distribution between hard news topics, such as Social/Identity Conflict, International/Middle East and Legal/Regulation, and soft news topics, such as Life/Interview, Crime, and Sports/Basketball, illustrates that the market drive is not influential enough to banish the hard news on news websites. However, the deviation of the selective link sharing for Twitter – e.g. high sharing propensity of sports topics vs. low sharing propensity of politics / international relation related topics – is more consistent with the widely expected outcomes of market drive in journalism's gatekeeping function. The conjecture that gatekeeping on Twitter is more market driven than the traditional one is also supported by recent observations from within news organizations. Though the outcomes of the news link selection have similar modality to those of the traditional gatekeeping, it is not always an official part of the editorial decisions, and an independent social media team decides the link selection in many organizations (Rowan, 2014, January 2; Tandoc Jr, 2014; Elizabeth, 2017, November 14). The logic behind this decision resembles that of the earlier SEO practices that mainly aim at maximizing the popularity of a news link. However, the selective news link sharing is a more explicit form of gatekeeping than SEO in that news organizations have full control over which news links should be *invisible* on their social media accounts (Tandoc Jr, 2014). Thus, the selective link sharing is likely to provide a powerful momentum toward popularity-biased gatekeeping on the top of the traditional gatekeeping.

Although the pooled analysis of the topic-wise link sharing propensity indicates that the selective link sharing carries out the longstanding concerns about the bias toward soft news on Twitter, such a conclusion may still be hasty, because of the dominance of specific types of organizations in the data. For instance, because there are many regional media in the dataset (Table 4.1), their strategies might dominate the overall patterns.

The organization-level analyses reveal more details of selective link sharing. The finding that news organizations' link choice differs across organizations shows that their criteria for selective link sharing cannot be reduced to the a one-dimensional degree of popularity. The observation that the same types of news organizations occasionally adopt a coherent selective link sharing strategy implies that inherent characteristics of an organization, as well as popularity as an external factor, are also a significant determinant 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 naming appropriate; national media significantly publish more about international topics and national politics, and regional media publish more about sports and regional issues. I use the term "specialty" to mean "what a news organization is good at." It is related to news organizations' capabilities, such as reporters' training levels at different expertise, the number of reporters, the expertise of editorial staffs, connections to certain news sources, etc. Although these capabilities are changeable, they will characterize news organizations at least in a short time span because they tend to co-evolve slowly with readers' long-term expectations.

Overall, specific selective link sharing strategies seem to be determined by the intersection of the popularity of a news topic and a news organization's specialized and *non*-popular topics strategy from this cross was the *reversionary strategy* for specialized and *non*-popular topics and the *concentration strategy* for specialized and popular topics. Whereas the reversionary strategy acts as a *homogenizing momentum* for non-popular topics in the sense that different news organizations will share with similar proportions of a given topic under this strategy, the concentration strategy can be diversifying because different types of news organizations will share more about their specialized hot topics. Additionally, non-specialized and unpopular topics will be less visible by the *strategic retreat*. These predictions are summarized in Figure 6.13. The overall result implies that *news can be diverse (specialized) on social media only for popular topics*. To take a notable scenario based on the findings, national media share more about popular, national politics, and regional media share more about regional sports.

On the contrary, news organizations are less likely to share specialized, but less popular topics by the reversionary strategy. Those topics are Immigration, International/Far East topics for national/online media, and Sports/Football and Education topics for regional news. Less popular, specialized topics may include topics that are complicated enough to require professional treatment to cover but do not stimulate interest from 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



Figure 6.13: 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.

may mean a socio-technical momentum which weakens the role of journalism that informs public decisions. A comparison between civic topics falling into the concentration and reversionary strategies somewhat supports this conjecture. Topics such as Politics/Congress or Social/Gender are indeed important, but vulnerable to sensational conflictive framing whereas Immigration and Education lack two opposing sides fighting against one another. 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.

The momentum, which social media as a news distribution platform generates does not seem favorable to regional media. What they concentrate on are likely to be regional topics, but mostly confined to local sports. Many authors recently discuss the potential of the regional media's revival from a hyperlocal approach (Metzgar et al., 2011; Lowrey, 2012). Yet, regional media's selective link sharing strategy shows that they focus on local sports, which broader local audiences expect, refrain from local civic issues, and give up on national issues for social media. For social media to help hyperlocal news, local media should be able to reasonably believe that there is a broad audience who would rather pay attention to those issues on social media than other local issues. This does not seem an easy task.

Selective news link sharing as a reaction to their environment has an implication for the design of information curating algorithms on social media. The conclusion that news organizations' selective link sharing is an outcome of their adaptation to inherent and external situations implies that they will also adapt to potential changes in news demand on social media caused by changes in an algorithm. This adaptation process is likely to become quicker and smoother, determined by the data scientists they hire (Rowan, 2014, January 2) and the social media monitoring platforms they adopt (Diakopoulos, 2017). However, quick adaptation is likely to make news organizations vulnerable because their demand depends on how the platforms' decision changes users' news reading behavior. Based on the two-sided market theory, economists point out that the dependence on monopolized platforms may suppress independent content providers, as notably in the Microsoft case (Evans, 2003). Indeed, there is evidence that Facebook's algorithm change greatly impacted news organizations' online revenue (Brown, 2018, April 18). Therefore, the discussion of a desirable algorithm design should include consideration about the potential strategic adaptation of news organizations as well.

Finally, the results found in this chapter may be subject to the topic of popularity cycle although the dataset includes a fair amount of time – sixty-three days. For example, STM identifies Me Too topic as separate from Social/Gender topic. Thus, verifying the conclusions 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 an infrastructure for the constant monitoring of news organizations' selective link sharing strategy.

#### CHAPTER 7

## NEWS PARAPHRASING AND SENTIMENT

# 7.1 Introduction: News Paraphrasing as a Strategy

Thus far, I have considered only the 'whether to share' decision as a news organization's choice in its social media strategy. However, that is not the only possible choice a news organization can make when it disseminates news links on social media. In this chapter, I focus on how the organization paraphrases a news story for a social media post.

To incorporate the multiplicity of news organizations' choices, my simplified conceptual model in Chapter 3 can be extended to a more realistic problem as follows:

$$\max_{\boldsymbol{x_i}} \tau_i(\boldsymbol{x_i}, \boldsymbol{x_{-i}}) N p_i - c_i(\boldsymbol{x_i})$$
(7.1)

Now  $\boldsymbol{x_i}$  is a vector that contains news organization's multiple strategic choices on social media. And the function  $\tau_i(\cdot)$  is a function that maps the choices to the probability that a news organization, *i*'s news story is read by social media users. As in 3.1, the probability is also determined by other new organizations' choices,  $\boldsymbol{x_{-i}}$ . Cost of social media strategies  $c_i(\cdot)$  is allowed to differ for each firm in this model, which implies asymmetric equilibria. However, unless one has a strong simplifying belief about what the plausible choice set is and what the functions  $\tau_i(\cdot)$  and  $c_i(\cdot)$  look like, it is hard to set up a solvable formal model based on the equation 7.1. Instead, in this chapter, I again empirically explore news organizations' additional choices – news paraphrasing by analyzing tweets embedded in news links.

There is evidence that news organizations are strategically paraphrasing their news stories for social media posts. Although news link sharing is deemed automatic for some organizations, considering the same text in social media posts was used as headlines for original stories until a few years ago (Palser, 2009; Armstrong & Gao, 2010), Armstrong and Gao observed some differences across organizations. For example, some organizations add a part of a lead to a social media post whereas other news organizations use only headlines. More recent reports show that news organizations take one step further to adopt a manual curation process optimized for social media. According to Newman (2011), in 2011, the BBC switched from the automatic feed to the manual choice and editing optimized for Twitter, and as a result, the number of followers doubled. Recognizing the potential of social media optimization for news, Facebook published a guideline for strategic news posts based on their own user-news engagement studies.<sup>1</sup>

The recent surge of online-only emerging media makes the manual social media optimization more common because they depend heavily on the viral dissemination of their news stories as a major means of news distribution. According to a recent Wired article, Buzzfeed relies on machine learning techniques and A/B testing to optimize itsr social media strategies for virality (Rowan, 2014, January 2). Similarly, Breitbart's co-founder Larry Solov recently said that the company has a separate social media team and that they are very careful about how they portray their stories on social media (Kew, 2016, August 17). Indeed, Reis et al. (2015) found that 51% of tweets were phrased differently from online headlines in 5,182 tweets they collected from four major news organizations' accounts in 2014 (The New York Times, BBC, Reuters and Daily Mail).

In this chapter, I focus on whether and how news organizations add sentiment as they paraphrase their news stories for Twitter. Although the literature on news popularity has generally reached a consensus that negative framing for news contents and headlines increases news popularity or virality on social media, researchers do not often distinguish between negative framing from a social media strategy and negative framing in the original story. However, the distinction is necessary to learn whether a social media strategy indeed adds negativity to news framing separate from the traditional reporting. Thus, in this chapter, I compare the sentiment of news tweets with that of the original news stories. I conjecture

 $<sup>\</sup>label{eq:linear} ^{1} https://www.facebook.com/notes/facebook-journalists/study-how-people-are-engaging-journalists-on-facebook-best-practices/245775148767840$ 

that news organizations consistently add *extra-negative* framing as they paraphrase original stories because social media strategies are more economically driven than the traditional editorial decisions. As multiple experiment studies have exihibited, negative news framing increases readers' engagement and changes opinions (Price et al., 1997; Zillmann et al., 2004; Trussler & Soroka, 2014). Thus, the extra-negativity as an outcome of social media strategies may have a significant impact on news readers on social media.

# 7.2 Related Works

# 7.2.1 Paraphrasing News on Social Media

Text is one of the most salient components of a social media post that contains a major portion of the information (Hu & Liu, 2012). And how news content is paraphrased largely affects the popularity of a news story on social media (Horne & Adali, 2017). Whereas many social media posts lack visual components, virtually all of them contain some elements of text. As social media is a text media above all things, news texts in social media have been often analyzed (Kwak et al., 2010; Newman, 2011). Text in social media posts can have different topics (Zhao et al., 2011) or sentiments (Hansen et al., 2011) from other texts because a length limit is imposed by social media users' cognitive capacity<sup>2</sup> or by a technology – e.g. 280 character limits on Twitter.

Observations that text on a social media post affects the popularity of a news article, and that the text is not a mere summary of news content, imply an incentive for a news link sharer to compose a paraphrasing text in a way that increases the popularity of the social media post with the news link. Previous studies discovered a few text features on social media posts with a news link were positively associated with their popularity. Hansen et al. (2011) found that negative sentiment in a social media post with a news link was more

<sup>&</sup>lt;sup>2</sup>The online advertising literature consistently reports the negative correlation between length of advertising message and the click-through rate (Baltas, 2003; Robinson et al., 2007). In the context of social media, De Vries et al. (2012) found brand posts with lengthy texts on Facebook has a smaller number of likes from users.

likely to be retweeted whereas positive sentiment increases a personal tweet's popularity. Further, news sharers on social media often aim at popularity when they paraphrase news stories. Horne & Adali (2017) found that many news sharers on Reddit paraphrase news stories in different ways from original news headlines. Like Hansen et al. (2011), Horne and Adali also found that negative sentiment and more emotional words in a paraphrase predict greater popularity of a news story on Reddit. Moreover, in their analysis, a paraphrase of news content predicted the news link's popularity with more accuracy than the news content itself. In other words, the association between a news paraphrase and news popularity on Reddit was not necessarily determined by the original news content. This distinguishable and significant power of news paraphrasing on social media is likely to provide an incentive for journalists to make strategic choices with it to attract more attention.

Being strategic for paraphrasing is not a new task for news organizations. For a traditional newspaper report, journalists must compose headlines that summarize news stories they publish. However, the headline is not only a mere summarization of a news content, but also the main 'hook' of a news story to readers (Molek-Kozakowska, 2013) similar to news paraphrasing for social media posts. In the same vein, Bell (1991) listed the main purposes of headlines as a) summarizing, b) framing and c) attracting. In other words, how the content is summarized in headlines influences news readers' choice to actually read the story. This similarity between headlines and news paraphrasing for social media implies that journalists' choices made for headlines that have been observed in previous studies can inform us about potential choices made for social media posts.

Although components in main news texts that increase news demand are still notable in the studies of headlines, many researchers focus on sensationalism in headlines. For example, Tenenboim & Cohen (2015) conducted a content analysis on headlines of 15,431 online news stories published on an Islaeli website, and found out sensational topics and curiosity arousing elements in headlines lead to more clicks. Also applying qualitative linguistic analysis, Molek-Kozakowska (2013) showed that sensationalism in headlines has multiple dimensions: illocutions (aims of journalists), themes, narrative, evaluation (value judgment) and proximity. The prevalence of sensationalism in news headline also predicts strong sentiment in news paraphrasing. According to Uribe & Gunter (2007)'s empirical work based on a hand-coding approach, sensational news about politics and crime issues tend to contain strong sentiment.

## 7.2.2 Negative Framing in Online News

In their seminal study, McCombs et al. (1997) acknowledged the importance of sentiment in shaping how to think about news. By connecting social issues with positive/negative/neutral sentiment, news reporting informs readers' judgment about real-world issues (McComas & Shanahan, 1999). Research has particularly focused on negativity in news framing. Trussler & Soroka (2014) hypothesized that news readers have a psychological preference for negative news framing because negativity is further from humans' innately positive expectations, and is thus considered a signal for more useful information (Kahneman, 1979). Indeed, Trussler & Soroka (2014) found that politically news readers prefer negative news framing under an experiment condition. They suggest this result as evidence for the demand side explanation of the prevalence of negativity in political reporting. Built on this assessment, researchers have applied automated measurements of sentiment in news stories. For example, using human coders and multiple lexicons, Young & Soroka (2012) found that news framing is consistently biased toward negativity, particularly for crime and foreign policy topics.

Negative summarization has been shown to have a strong impact on news perception. By manipulating headlines and leads, Price et al. (1997) found that experiment subjects' opinion and emotional valence on funding to public universities change depending on how the same information is framed (conflict/human interest/consequence). Further they found that the changed opinion had an impact on decision making about the policy. With a similar experiment design, Zillmann et al. (2004) showed that experiment subjects spend more time reading news stories when they are framed as conflict or victims' agony rather than misfortune or economic loss. Analyzing an aggregate time series of public opinion and sentiment in news on economic issues, Soroka (2006) argues that negative sentiment contained in news stories largely affects opinion change not only because media coverage tends to be negative rather than positive, but also because negative information tends to have a greater impact on the impressions of news readers. These studies imply that news paraphrasing, such as social media feeds has the power of extra-framing, on top of framing in a main body of news stories.

Although news organizations' paraphrasing on social media has not been directly analyzed, findings from previous works on news demand show that news sentiment and popularity are closely related. First, media studies generally agree that the negative sentiment in a news story is positively associated with a news demand in a traditional news reading setting. Newhagen (1998) conducted an experiment in which subjects watch news images that evoke negative emotion (anger, fear and disgust). He found that images that evoke anger most increase subjects' approach to the images, followed by images that induce fear and images that induce disgust. He also found that the images that induce were the most memorable. He explains this result with a biological definition of anger, which is the physiological result of territorial violation. This result implies that a certain emotion in a news story can induce news readers to "approach" more, which may in turn affect demand for the news. Indeed, Trussler & Soroka (2014) found that a politically interested person is likely to read a political news story with a negative tone and that frames politics as a game rather than informative news that uses an eye-tracking experiment.

Yet, some researchers point out that the generality of the negative framing should be understood with care. In particular, the popularity of such a strategy and its effectiveness can differ depending on news topics. For example, focusing only on health related news, Kim (2015) found that only positive sentiment in health news is positively associated with its likelihood to be read and shared on social media. Similar results were found for news paraphrasing. By measuring the sentiment scores (valence) of 69,907 top news headlines from four major news websites (BBC News, Daily Mail, the New York Times and Reuters), Reis et al. (2015) found that the majority of the headlines has negative sentiment overall, but the level of the dominance varies, depending on news topics. For example, news stories falling under a topic category, "world," most frequently have negative headlines. However, stories about sports or science & technology tend to have neutral headlines compared to other topics. In the context of my study, this implies that the negative framing for news and social media will differ across news organizations depending on the news topics they mostly cover.

Further, the potential negativity from news paraphrasing is likely to be clearer on social media, where economic incentives to elicit more clicks dominates, compared to traditional headline making, which is a part of traditional editorial decisions. In traditional news consumption where news stories are consumed as a bundle (e.g. a newspaper or TV news program), headlines signal news content to compete with other articles within the same bundle. Thus, the headlines had to be written to show a story's relevance to a reader (Dor, 2003). However in the online environment, news readers typically navigate across different news sources. As a result, attracting news readers' attention with a headline became a more challenging task because it also needs to take account of competition with other sources beyond the given bundle. In this circumstance, news organizations have a strong incentive to signal their stories' appeal to readers. Moreover on social media, this signaling incentive is likely to be even larger because users' attention is easily distracted by a mixture of various types of information. Indeed, researchers have detected the proliferation of the use of sensational and particularly negative expressions in online news headlines for *click-bait*. For example, Chakraborty et al. (2016) and Potthast et al. (2016) showed sentiment polarity (valence) can be used as a predictor of an automated click-bait detection algorithm.

All in all, news organizations are likely to frame their news negatively to attract readers. News paraphrasing for social media will be more negative because it is more governed by the economic motivation compared to traditional editorial decisions. Thus, I hypothesize that news organizations apply *extra-negative framing* to news as they paraphrase for social
media. Testing this hypothesis requires the comparison of the sentiment of news tweets with the sentiment of the original news stories. In this chapter, I will conduct the comparison using three different dictionaries – Affin, LSD and LIWC. Further, the literature implies that different types of news organizations that cover different topics are likely to show different levels of negative framing. I will test this conjecture by comparing the extra-negative framing across different types of organizations.

## 7.3 Results

In this section, I present results from the dictionary-based sentiment analysis. Measured sentiments of news content that each news organization publicizes on its news website and Twitter account will reveal that the extra-negativity from paraphrasing for tweets is consistent across different organizations. Subsequently, using an inferential analysis, I will discuss how the extra-negativity differs depending on types of news organizations. All the analyses in this section were iterated using the three different dictionaries – Affin, LSD and LIWC – to validate the results. Results omitted to avoid repetition are included in Appendix D.

### 7.3.1 Negativity of News by Organizational Types

The previous literature has found generally that negativity in a news story tends to increase its demand, and news organizations take advantage of this tendency to attract more readers. This implies that tweets about news stories are likely to have a systematic bias toward negative framing. Although results in the context of online news consumption are yet limited and more debatable, there is much evidence that negativity in social media posts attracts more attention. Further, I conjecture that news organizations' social media strategies are largely governed by economic incentives stemming from competition for limited attention. If this conjecture holds, paraphrases of news stories for Twitter are likely to contain more negative sentiment compared to original stories, which are presumably governed more by journalism norms.



Figure 7.1: Average sentiment scores of news stories and tweets: LIWC dictionary.

Figure 7.1 presents results of the sentiment analysis based on the LIWC dictionary broken down by news organizations. The red bar means the average sentiment scores of news stories and the blue bar represents the average sentiment scores of tweets with embedded news links to the original news stories. The score 50 indicates neutral sentiment for LIWC. The results show that the tone of news stories is indeed consistently negative across different news organizations. In particular, whereas regional news media occasionally have positive tones, on average, most of other types of news media have negative tones. Moreover, this result is robust across all three dictionaries.<sup>3</sup> Excluding regional media, the media with positive tones include; Forbes, CNBC, USA Today, Inquisitr and Vanity Fair. These seem related to a specific focus of media, such as finance or entertainment, and the nonpartisan orientation it pursues.

To check if this pattern holds consistently across different dictionaries, Figure 7.2 visualizes the average news sentiment of each type of news organizations using three dictionaries,

<sup>&</sup>lt;sup>3</sup>Results based on Affin and LSD dictionaries are Figure D.1 and Figure D.2 in Appendix D.



Figure 7.2: Sentiment level of news stories by news organization type.

Affin, LSD and LICW. Indeed, national and online news media have the most negative tones whereas regional media have the least negative tones regardless of dictionaries. The ranks between national and online media differ depending on the dictionary, but the difference between the two is much smaller than the difference with the other two types. The only significant difference across different dictionaries was that the analysis based on Affin results in positive average tones of magazines and regional papers rather than negative.

### 7.3.2 Strength of Extra-Negative Framing

The news dissemination through social media may raise yet another layer of news negativity issue – the extra-negativity of news paraphrasing for social media. To many social media users who look only at the news paraphrases without clicking the link, the potential extranegativity coupled with insufficient information about the news topic may greatly affect users' news perception. If the social media strategy is largely governed by economic incentives compared to news publishing, the extra-negativity on Twitter is likely to occur even for regional media, which had moderate sentiment on their websites.

Again, Figure 7.1 shows the negativity of tweets (blue bars) measured by the LIWC

dictionary.<sup>4</sup> For most news organizations, blue bars are lower than red bars, which means that tweets have more negative sentiment than the original news stories. However, there are some exceptions, particularly among online media such as Breitbart, Conservative Tribune, and Western Journalism, which are all known to be conservative partisan news outlets. A notable aspect of these exceptions is that these news organizations tend to have even more negative tones on their original news stories than others. Thus, the lack of extra-negativity may be only because their original news information is already very negative, and not because they do not frame tweets negatively.

A simple regression analysis can extend this visual analysis based on averages to a news story level analysis. In addition, the regression allows for explicitly revealing whether different types of news organizations add the extra-framing to a different extent, which is not easy to demonstrate with a visual analysis. The regression model I ran has the sentiment gap between a tweet and an original news article as a dependent variable, which represents the extra-negativity a news organization adds as it paraphrases its own stories for a tweet. The gap is expected to be "positive" with the extra-negativity. Having the sentiment of the original article as a control variable allows for distinguishing the extra-negative framing from negativity in an original story. The unit of analysis is each article-tweet pair.

Table 7.1 presents the regression results. The results appear to be consistent with the expectations; more negative news stories have less extra-negativity for tweet paraphrasing because they are already framed negatively. Although Affin measure results in the opposite sign (-0.0446), it was not statistically significant. Also, it should be noted that the  $R^2$  is extremely low for the Affin-based and LSD-based regressions although LSD resulted in a significant finding. The main reason for this is that they pick up less variation from the gap between a news story and a tweet because they include relative limited number of words in the dictionaries compared to LIWC. Although the size of the dictionaries is large enough to pick up the variation from aggregated text, they tend to miss information when they are

 $<sup>^{4}</sup>$ Again, results consistent with those based on other dictionaries are reported in Figure D.1 and D.2 in Appendix C.

Variable	Affin	LSD	LIWC
Intercept	$0.4771^{***}$	$0.2575^{***}$	-15.3137***
	(0.0293)	(0.0124)	(0.3809)
News Sentiment	-0.0446	$0.0751^{***}$	$0.5853^{***}$
	(0.0082)	(0.0078)	(0.0028)
National	-0.0928***	-0.0462***	$-2.5787^{***}$
	(0.0310)	(0.0132)	(0.3825)
Online	-0.0514**	-0.0287*	$-2.6715^{***}$
	(0.0327)	(0.0138)	(0.4007)
Regional	$-0.1601^{***}$	-0.0739***	-2.0643***
	(0.0311)	(0.0132)	(0.3815)
R2	0.0015	0.0017	0.2110
Ν	$74,\!931$	$90,\!593$	$166,\!159$

Table 7.1: Linear regression of sentiment gap on news sentiment and news organizations type.

to detect small information from a gap between a single news story-tweet pair. Thus, the regression based on LIWC seems more reliable for this analysis.

Even after controlling the sentiment of the original stories, regional news media tend to add more negativity as they paraphrase their news stories for tweets than national and online media (-2.0643 > -2.5787 > -2.6715).<sup>5</sup> However, the national media and online media show a similar level of extra-negative framing. This means that although magazines and regional media are relatively positive on their websites, they catch up with online and national media on Twitter by adding more negativity on news paraphrasing. In other words, news paraphrasing for Twitter negatively homogenizes different types of media. This is visualized in Figure 7.3 where the sentiments are not too different compared to Figure 7.2.

### 7.4 Discussion

News paraphrasing has always been an essential part of the news making process, and its framing impact has been proven to be significant. Although headline making – traditional paraphrasing – has long been accused of being sensational, it is a part of the traditional

<sup>5</sup>Note that this result means the baseline category, magazine, has the largest gap, which is expected to be positive with the extra-negativity.



Figure 7.3: Sentiment level of tweets by news organization type.



Figure 7.4: Additional negative sentiment to tweets by news organization type.

editorial decision, likely to be governed by journalistic norms. On the other hand, recent testimonies have shown that a similar task of paraphrasing for social media is often outside of newsroom Elizabeth (2017, November 14); Roston (2015, January 22), and largely relies on numbers from audience metric and automated algorithms (Tandoc Jr, 2014, 2016). Considering the observation that social media users often read only the paraphrased news without clicking embedded links (Horne & Adali, 2017), it is reasonable to worry about the impact of the additional mechanism whereby news organization can impose news framing on social media, and monitor it as a quasi-editorial decision. By comparing sentiments of tweets by different types of news organizations and original news stories, I found evidence that the incentives from the demand of social media users led organizations to place more negative framing on news compared to the original stories in this chapter.

The reported results are generally consistent with conjectures I made based upon the literature. (a) Online news stories are negatively framed overall, and (b) regional media likely to cover less controversial issues and frame news in a significantly less negative way compared to national and online media. (c) Social media posts are likely to be more driven by audience's reception than journalistic norms and are consistently more negatively framed than original stories across different outlets. (d) In addition, regional media significantly catch up to other types of media with extra-negative framing for Twitter.

The negative news framing of national/online media in the results has the potential to affect the news perception of online news readers, as scholars have pointed out (Newhagen, 1998). Although regional media supplies less negative framing, it might be limited to certain news topics, and many report evidence of the rapid decline of regional/local journalism in the rise of online journalism (Curran, 2010; Nielsen, 2015). Although some see potential from the Internet to rebuild a strong link between news readers and the local community (Lowrey, 2012), recent evidence shows that social media tends to function as a momentum toward the centrality of national legacy media or nation-wide emergent forms of online-only media (Hodson & Lindgren, 2017). If regional media is indeed losing the competition for limited attention on social media, users' news perception may become negatively biased as an incidental impact of the negativity of national/online media observed.

Given that social media users read only paraphrased news most of time without clicking embedded news links, the extra-negative framing for social media exacerbates such impact on news perception. Furthermore, news organizations' sensitivity to market demand on social media, enhanced by data-driven social media monitoring technologies (Diakopoulos, 2017), creates a condition vulnerable to a feedback loop spiraling toward negative news framing on social media. The negativity in news paraphrasing induces negative news perception, which in turn potentially makes social media users demand more negative framing.

The observation that regional papers are catching up is consistent with the conjecture that news paraphrasing for social media is more driven by the audience's reception compared to traditional editorial decisions. If journalists' framing for news stories is more governed by their traditional decision about newsworthiness, regional news has a reason to have moderate sentiment from their less-controversial regional topics than national and online media. On the other hand, if paraphrasing is governed more by profitability from social media users' clicks and engagement, there is no reason that social media posts from regional media are less negatively framed than those from other types of media. In this case, the gap between the two sentiments (i.e. the extra-negative framing) should be much larger for regional media as observed in the presented analysis.

Although news readers are increasingly susceptible to such market driven negative framing on social media, they have been relatively free from normative discussions, such as an evaluation of its impact and what we should do about it as a society. Although sensationalism of traditional headlines has frequently raised researchers' concern (Chakraborty et al., 2016; Molek-Kozakowska, 2013; Blom & Hansen, 2015), the traditional decision could still resort to the possibility of self-regulation based on journalists' values from their long-term training and experience as journalists. However, it is harder to expect that news paraphrasing for social media will be controlled under a similar self-regulation because social media strategies are more likely to be subject to economic principles. If the dominance of audience's reception for news paraphrasing is the case, then the resulting paraphrases are likely to be homogenized to reflect social media users' news demand. Surely, it is still possible that diversity in social media users' news demand will induce a diversity of paraphrases. However, the empirical results in this chapter show that framing from news paraphrasing on Twitter is currently undergoing a rather homogenizing momentum, disconnected from original news stories. This raises a need for the public monitoring of paraphrasing on social media as news organizations' quasi-editorial decisions.

Regarding public monitoring of social media, the reported results suggest that LIWC shows a better performance than the other two dictionaries. The main reason seems that a generally purposed LIWC dictionary contains a larger set of words and finer-ground sentiment scores associated with the words whereas LSD and Affin are designed for a specific domain – LSD for news and Affin for tweets – and contain a much shorter list of words. LSD and Affin contain enough information to pick up average sentiment from large sample of text as I confirmed from the consistency of news sentiment across different dictionaries. However, they do not seem to have sufficient information to detect limited information from the sentiment gap between a tweet and a news story. Interestingly, this conclusion somewhat contradicts the previous dictionary-based sentiment analysis that called for a need for public monitoring. For example, Young & Soroka (2012) argued that monitoring news needs a news-customized dictionary. However, the present setting that compares news to tweets results in the opposite conclusion, i.e. that we need a general dictionary. It seems that there is a trade-off between generalizability and customization. To increase the precision of sentiment information from one (each) type of text, it is better to customize. However, to compare two different types of text, such as a news story and a tweet, it is better to have a general dictionary.

### **CHAPTER 8**

## CONCLUSION: SOCIAL MEDIA STRATEGIES AS QUASI-EDITORIAL DECISIONS

Although Internet technologies provide strong momentum toward "flatter" communication in a sense that users, who used to be the mere audience of information, can bring their voices into journalists' work via the new vehicles (Singer, 2009), the news making - distribution process is still *structured* by powerful actors. Information platform owners, such as Facebook and Google, now have significant control over the priority and relevance of information for individual users (Pariser, 2011). News organizations, albeit weakened, retain a privileged position over nonprofessional journalists to access public information because of their organized reporting and editorial infrastructure, financial capital, and historically accumulated connections to information sources. Journalism is a network composed of multiple human actors and technologies (Turner, 2005). Yet, it is a network structured by powerful actors who possess the means of production and distribution.

As an analysis of one kind of such powerful actors, I investigated news organizations' social media strategies as an emergent quasi-editorial decision, focusing on news dissemination via Twitter. I consider the news dissemination strategies on social media as a form of the news organizations' adaptation to socio-technical traits of the new news distribution platform. Drawing on recent reports that news organizations' social media strategies is disconnected from traditional editorial decisions it terms of personnel, organizational structure and decision processes Roston (2015, January 22); Elizabeth (2017, November 14); Tandoc Jr (2014, 2016), I tested whether this disconnection indeed produces difference in visible news topics and framing between news websites and organizations' Twitter accounts. Based on this comparison, I argued that the different mechanisms whereby the quasi-editorial decisions for social media are made can impose an additional layer of information-mediating process, relatively free from journalistic norms and routines, as readers find news information from

social media more often.

Monitoring such a quasi-editorial decision is not a simple task. The social media strategies, as an adaptation process, are still in flux partially because of the inertia in the older regime of journalism, and partially because of the ceaselessly changing technologies. Furthermore, my research questions, which involve two different forms of text, news and tweets from a rapidly growing number of news organizations, make the manual analysis hardly feasible. Fortunately, the complicated and everyday mutating online information mediating process is recorded on the Internet in an almost complete form. Thus, I proposed a computational system of data collection and automated text analysis for the constant monitoring of news organizations' social media strategies. This approach will expand our frontier of "monitoring citizenship" (Schudson, 1998) over the public information flow toward communication that is happening via social media, an emergent news distribution platform.

## 8.1 News Link Sharing as a Strategic Choice

A fundamental question before gauging news organizations' social media strategies includes, *Is there a reason for a news organization to be strategic on social media?* I use the term, "strategic," based on the economic definition, to refer to a situation where every actor's decision depends on one another's. In Chapter 3, I found a reason for a positive answer to this question from social media users' limited attention. In the economic model I suggested, the scarcity of users' attention capacity relative to the volume of information propagated on social media creates competition between news organizations. This model implies that users' attention can be conceived as a *public resource*, which is a source of economic benefit that anyone can access to exploit, but whose amount is limited (Ostrom, 2015). As most public resources, the outcome of my model reveals that one news organization's attempt to capture users' attention undermines a chance for other organizations to do so. In other words, sharing more news links to attract social media users' click eats up users' attention, resulting in less opportunity for the links of other news organizations to be in users' awareness. This competition resembles a situation in which everyone is shouting to be heard. Because I shout, a person next to me also has to shout to be heard (Anderson & De Palma, 2013). Because people shout, I have to shout even louder in turn. By similar logic, a news organization should *strategically* decide how many news links it is going to share considering how many *others* would share. Although recent media studies have focused on the impact of information overload, as a kind of bounded-rationality, on individual information choice behavior (Liang & Fu, 2016), competition for a limited attention model shows that news organizations' adaptation to such human boundedness can also give rise to new communication patterns in an emergent media environment.

As with most public resources, the predicted equilibrium of the model is that the limited attention of social media users is over-exploited. Roughly, news organizations will share too many news links to beat other organizations' links. Because everyone tries to beat one another, there will be so many news links on social media that users will be able to process only a small portion of the shared news links. However, if news organizations cooperated jointly to maximize their profits rather than compete, they would disseminate fewer links to save users' attention.

This result has an implication on the discussion about public accountability of social media platforms, which was dramatically escalated by the Cambridge Analytica case, in which the private information of users that was retained by Facebook was politically abused. The following controversy as to responsibility of Facebook and other Internet platforms has reached an agreement that the platform owners' self-prescription will not fundamentally resolve the risk of similar troubles (Vasu et al., 2018). However, given that social media owners have been considered technology companies relatively free from social accountability, it is difficult to establish a grounding principle that allows for public engagement with the information platforms.

The conclusion of my model that users' attention is a public resource can offer such a principle. As Ostrom (2015) argues, governance of a public resource susceptible to overexploitation is a part of the key infrastructure for success and sustainability of different levels of communities. Because democracy desperately needs public attention to civic issues, as Berger (2011) points out, and if attention can be perceived as a public resource, companies that profit from (re)distributing users' attention would have social responsibility for the commercial use of the resource. This argument may be used to warrant publicly governing online platforms to prevent overuse or abuse of attention – the scarce and valuable resource of society. However, although the natural time is limited, human attention may be extended by technologies, such as new devices and information curation algorithms, partially provided by the platform companies. Thus, to what extent users' attention is a limited resource is an empirical question that calls for further research.

Whether attention is over-exploited is hard to test because the firms' collectively optimal use of attention – the theoretical comparison target – is not observable from the data. Yet, the model still produces a testable prediction: "News organizations will reduce the proportion of news links they share on social media as more stories are published by all organizations." This prediction resonates with the proposition that news link sharing is 'strategic' in the sense that one organization's decision relies on others. The regression results in Chapter 5 based on news links shared on Twitter are consistent with this prediction. They are also in accord with recent reports that news organizations have hired social media specialists and adopted social media monitoring platforms to optimize their social media strategies to the reactions of both users and competitors (Kew, 2016, August 17; Rowan, 2014, January 2; Diakopoulos, 2017). As technologies they adopt for such monitoring develop, sensitivity of their reaction is expected to increase in the future.

Realistically, news organizations' strategic choice may not be confined to decisions about the number of news link they will share. For example, news organizations may want to concentrate on some popular news topics, when they shrink the proportion of shared news as a reaction to the large volume of published stories, as the model predicts. Instead of selecting news links to share, they may want to adopt different content and signaling this decision by manipulating text on social media posts. However, neither the model nor the simple regression about the aggregated proportion of shared news links allows for investigating those qualitative dimensions of social media strategies. This challenge raises a need to take a deeper look at the relationship between news content and news organizations' choices on social media.

## 8.2 Selective News Link Sharing as Gatekeeping

A news story, as an information good, has multiple qualitative characteristics, such as length, topic, factuality, opinion, etc. Basic economic theory dictates that a news organization, like an economic entity, naturally conditions its strategy about these characteristics to maximize profits. In the context of news distribution on social media, this economic prediction should manifest itself as *selective news link sharing* – conditioning the link sharing decision on news content. Whether a news organization shares a specific news link on social media has a potentially significant impact on the visibility of a news story. Users will be more likely to be aware of the news story, and have easier access to it when there is a hyperlink to it. In this sense, hyperlinks can function as a new form of information 'gates.' (Dimitrova et al., 2003).

The outcome of the link sharing decision – differentiated accessibility to news – is close to the traditional gatekeeping by journalists. Both tasks decide accessibility to certain information by controlling its visibility on a medium. However, principles governing the two decisions may be different. Traditional gatekeeping has been arguably conceived as governed by journalistic norms. Relying on journalism norms and values, journalists carefully decide what information is worth citizen's knowing. Years of training for job ethics based on professionalism and accumulated experience for objective reporting inform 'newsworthiness.' On the other hand, news organizations' competition for users' attention implies that news link sharing is more likely to be governed by economic incentives; the more likely users are generating advertising profit, the better. According to recent reports, the decision making for social media posting is often separate from traditional editorial decisions organizationally and personnelwise, and is often more attached to a management side within an organization (Roston, 2015, January 22; Elizabeth, 2017, November 14). This tendency is accelerated because social media strategies are determined by audience metrics and automatic algorithms (Tandoc Jr, 2014, 2016; Diakopoulos, 2017). Given that gatekeeping is a critical role of journalism for citizens' informed decisions, the dominance of economic value in deciding accessibility to news can pose a challenge to the right to be informed. With a computational text analysis technique, Structural Topic Model, I tested whether news content often distributed via Twitter is indeed significantly different from the traditional gatekeeping discussed in Chapter 6.

The overall comparison between proportions of published topics and shared topics provides evidence that traditional gatekeeping and the selective link sharing do indeed follow different logic. Whereas Crime and Sports/Basketball topics were frequently published and even more likely to be distributed via Twitter, Social/Identity Conflict and Middle East topics are also often published but less likely to be shared by organizations. In general, soft news topics, such as sports and crime are more likely to be shared by news organizations than hard news topics related to politics and international relations although many hard news topics are among the most frequent topics on the news websites. This result indicates that the popular conjecture that commercialized media drive news toward human interest rather than newsworthiness is crystallized more visibly on social media than on news websites. If decision makers for link sharing consider expected profit to be more important than newsworthiness as editors presumably do, this discrepancy makes sense. But it is not yet definite whether the new gatekeeping on social media implies a different regime of the information mediating process for news readers. It raises a question, "Are we informed enough with the new way of deciding the visibility of news?" This is not only a normative question, but also an empirical one because, even if an individual organization's link sharing has a strong bias in an undesirable way, the collection of news organizations may still provide a good mix of

news on social media.

To take a deeper look at how news organizations as a collection selectively share news links, I also broke down their selective link sharing by individual organizations and broader categories of organization types. The outcome revealed different strategies of selective link sharing. News organizations share more news on certain hot topics (concentration strategy), less on other less popular topics, but they published much (reversionary strategy). The latter strategy characterizes the discrepancy between traditional gatekeeping and selective link sharing. Even though a news organization considers a certain topic to be important in their editorial decision, thus they became 'specialized' throughout their history, they would not share much about it on social media because popularity in the short term dominates social media strategies. Thus, I suggested that specialty and popularity as two major determinants of selective link sharing. Specialty is connected to what news organizations consider themselves to be as a group of journalists whereas popularity is more related to short-term profitability.

Topics on which news organizations concentrate and from which they refrain differ across different types of news organizations. National and online media concentrate on particular political topics that can be readily depicted as a horse-fighting whereas online media also concentrate on Social/Gender topic. However, they refrained from other hard topics, such as Middle East, Immigration and Finance. The finding that these types of organizations are highly likely to share a few hard topics somewhat mitigates the earlier concern that the link sharing is biased toward soft news. However, the shared hard news topics are still focused on specific topics that appeal to the popular preference of a polarized perception about public issues. In that sense, the market-driven news distribution via social media seems to bias information available for informed political thinking.

Regional media's hot topics were sports, mostly Basketball, presumably due to the specific period of data collection. On the other hand, they refrain from regional issues such as education, regional politics and transportation even though they are 'specialized' in these topics. This may result in an unfortunate situation where regional papers are mostly conceived as information sources for regional sports as opposed to the vision of hyper-locality as regional media's viable surviving strategy (Metzgar et al., 2011). Moreover, regional media retreat from many topics they are not specialized in, such as international relations. If shrinking toward a regional sport media does not provide enough demand to restore their business, this outcome does not seem favorable for the future viability of their business.

The concentration on hot topics does not seem favorable to the pluralist ideal of diversitybased informed decisions for democracy (Dahl, 2005; Walker, 1991). Because overall popularity is the only profit source on social media, this outcome seems almost inevitable. Further, it is not an outcome a single organization can 'decide' to overcome. The findings imply that the improvement of social media design should be based on news organizations' economic reactions to changes to enhance news organizations' incentives toward more diverse link sharing. This approach will require further in-depth monitoring of the behaviors of news organizations and the interactions between news organizations and platforms.

## 8.3 Extra-negative Framing for Paraphrases

News organizations have more potential choices than whether to share when they distribute news via social media. In Chapter 7, I focused on text in a tweet occuring with news links. Text is one of the most salient form of an organizations' choice, not only because it comes almost certainly with news links on social media, but also because our existing knowledge about traditional headlines and online political comments dictates that news paraphrasing has significant impacts on readers' news perception. In particular, I focused on the sentiment that the paraphrasing for a tweet adds to the frames of an original news story. The literature generally agrees that negative sentiment both increases readers' attention to news (McCombs et al., 1997; Trussler & Soroka, 2014) and changes reactions to news, such as reading time, attitude, opinion, etc. (Price et al., 1997; Zillmann et al., 2004; Soroka, 2006; Newhagen, 1998). As a result, news stories are consistently negatively framed (Young & Soroka, 2012).

However, the difference in the degree of negative framing across different topics (Young & Soroka, 2012) predicts that different types of news organizations with specific focuses – 'specialty' to use the term in Chapter 6 – will have different levels of the negative framing. For example, regional media that covers regional issues, such as regional politics, education, and transportation are likely to frame news less negatively than national / online media, which covers more partisan issues. Yet, social media strategies where the audience's reception is expected to govern may put a different layer of sentiment on news because the original news story is paraphrased for a social media post. In other words, the attention drawing function of negative framing is expected to prevail on social media.

The dictionary-based sentiment analysis of news and Twitter data confirms that news stories are consistently negatively framed in news organizations. In addition, regional media conveyed less negative sentiment through news stories than other types of news organizations. Further, paraphrasing on Twitter showed even more negativity than original news stories overall regardless of the type of organization. Although these results have a significant implication on readers' news perception in their own right, given the waning industry of regional media in terms of online presence, a separate analysis of a single type of text cannot distinguish between negativity in news stories and negativity from paraphrasing strategy.

To see systematically the source of the negativity observed on news paraphrases, I also conducted an inferential analysis that regresses the negativity gap between a story and a news paraphrase on original news sentiment and types of organizations. The results indicate that regional media catch up to other types by adding even more negativity on news paraphrases for tweets. Contrasting with the significantly low level of negativity in news stories from regional media, this finding provides additional significant indication that social media strategies are governed by a different logic than that governing traditional journalism practices. In other words, news organizations seem to paraphrase news mostly to attract users rather than to fairly represent the news content. In parallel with selective link sharing, the overall results in Chapter 7 imply that social media strategies that are largely driven by audience's reception outside newsroom will greatly affect "how readers think about news" as well as "what to think about." Again, a scalable monitoring system for news strategies is called for to decide whether we will be satisfied with the emerging quasi-editorial process to guarantee well-informed decision making for society.

## 8.4 Moving Forward

In this dissertation, I investigated news organizations' social media strategies to disseminate news stories as potential quasi-editorial decisions. In Chapters 3 and 5, I showed that news organizations' behavior on social media is governed by economic motivation. I used an theoretical modeling and an empirical analysis of news organizations' quantitative link sharing decisions against the decisions of others on Twitter. These results, combined with the observation that news organizations' tasks on social media have forms and functions similar to those of their traditional editorial decisions, such as gatekeeping and framing, pose the major theoretical and practical concern of this dissertation about whether economically-driven quasi-editorial decisions are significantly different from traditional ones. Using computational methods, I illustrated that the gatekeeping and framing for Twitter indeed generate significantly different patterns. They were biased toward popularity on social media, resulting in concentration on specific topics and additionally negative framing.

Although this dissertation illustrated the emergence of a new information filter on Twitter, for two reasons, it is too early to say that the new filter illuminates a different regime of journalism. First, although Twitter is one of two social media through which most news organizations regularly share news links, Twitter accounts for a small portion of news reading through social media. I believe that the general conclusion that social media as a news distribution platform create an additional information mediating process, can be generalized to Facebook. Facebook also retains the characteristics that I argue are a main driving force of distinctive news dissemination strategies – users' limited attention and virality. However, the specific patterns of the strategies might be different because Facebook also has different traits than Twitter, such as a more active algorithmic curation, relatively closed networks, and a larger text limit. Thus, applying similar analyses to the Facebook case will help draw a more complete picture of the emergent information mediating mechanism that exists on multiple platforms. Second, we do not know completely significance of the impact the new filter has on users because user behaviors are mostly omitted from the analyses in this dissertation. With multiple platforms designed for information sharing, news organizations are no longer an exclusive source of news for most audiences. Thus, we need a further study that can test whether news organizations' strategies indeed impact news perceptions. This type of future study cannot be based on small samples either. To gauge the impact of news organizations' social media strategies, one needs to measure how often audiences are exposed to such strategies, which can hardly be measured relying on the memory of the audience. A possible path is to connect news readers' log data and a survey on their perceptions.

From the perspective of policy, news organizations' social media strategy is a moving target. Because news organizations are likely to react to audience's reception, their strategies must be changed to environmental variables that affect users' news demands, just as platforms' algorithms and news organizations' business models, change. Thus, to decide whether the emerging filter indeed calls for intervention, we need a constant monitoring of it. The methods to link different sources of data and machine learning techniques that I suggested can extend existing open source platforms with a similar goal, such as MediaCloud, to make the monitoring systems more specifically geared toward meaningful questions. Further, the empirical findings in this dissertation only describe outcomes of organizations' social media strategies, but mechanisms whereby use of audience metrics and organizational disconnectedness of social media specialists lead to specific strategies still remain to be explained. Understanding "how they get there" will help us predict news organizations' adaptation to future environmental changes, which will inform platforms' better algorithm design and potential policy intervention. Although the framing study in this dissertation focuses only on sentiment, studies of news paraphrasing in the traditional media environment encompass a much wider conception of framing. In general, a framing concept includes both the selection of ideas (Entman, 1993) and a narration of the selected ideas (Gamson & Modigliani, 1994). Although previous attempts to computationally operationalize framing concepts have applied topic models to news text (Boydstun et al., 2013), they fail to differentiate between news topics and framing because the choice of algorithms is not based on the theories. I believe explicit modeling of the word selection process and sequential word choice, complying with the original idea of the concept rather than applying a ready-made algorithm, is a more plausible way to the framing concept. Newly developed algorithms can also be a valuable extension of the public monitoring system for news organizations' social media strategies.

APPENDICES

### APPENDIX A

## **PROOFS OF LEMMAS AND PROPOSITIONS IN CHAPTER 3**

## A.1 Proof of Proposition 1

If  $KM \leq T$ , we know that full sharing characterizes the equilibrium as long as Np > cbecause the profit on the marginal link shared is Np - c in this case, which assumed to be positive.

If KM > T, the marginal profit can be lower than Np - c. Let us consider an *unconstrained* version of equation 3.1 without the upper bound for the probability for a link to be clicked at one:

$$\max_{0 \le x_i \le M} E[\pi_i(x)] = \left[\frac{T}{\sum_j x_j} Np - c\right] x_i \tag{A.1}$$

Suppose that all other news organizations except the firm, i, are choosing the SNE choice,  $x^*$ . Then the maximization problem becomes:

$$\max_{0 \le x_i \le M} \left[ \frac{T}{(K-1)x^* + x_i} Np - c \right] x_i$$

The first order condition of this problem is<sup>1</sup>:

$$TNp\frac{(K-1)x^*}{[(K-1)x^*+x_i]^2} - c = 0$$

In an SNE,  $x_i = x^*$ . Thus,

<sup>&</sup>lt;sup>1</sup>The second order condition holds because the derivative of the marginal profit is  $-2MT(K-1)Np/[M(K-1)+x_i]^3$ , which is negative as far as  $x_i$  is nonnegative.

$$TNp \frac{(K-1)x^*}{K^2 x^{*2}} = c$$
$$\iff x^* = \frac{Np}{c} \frac{T(K-1)}{K^2}$$
(A.2)

Note that, when  $x_i = x^*$  for all *i*, each firm's profit is  $c\frac{K}{K-1}$ , which is strictly positive. Thus,  $x^*$  is not constrained by the requirement that profit be non-negative.

Now by comparing unconstrained optimal choice  $x^*$  with the number of news stories an individual organization published, M, we can find the range for full sharing and no sharing equilibrium. That is, if  $x^* \ge M$ , news organizations cannot increase the number of news link shared on social media more than M even though the marginal profit is positive. Hence, full sharing is an SNE in this case. Rearranging the inequality, we get  $\phi = \frac{MK}{T} \ge \frac{Np}{c} \frac{K-1}{K}$  for full sharing condition. Partial sharing will be sustained as an SNE otherwise.

#### APPENDIX B

### DETAILS OF DATA COLLECTION

## **B.1** Official Twitter Accounts

News organizations often maintain more than one "official" Twitter accounts. For example, *Time* only has one official account whereas *Washington Post* has Washington Post, Washington Post World, Washington Post Sports, and so on. To take account of this, the Twitter scraper collects tweets from all the *sub-accounts*. However, there are cases where the distinction between official accounts and those maintained by individual journalists is unclear. To make a distinction, I coded the Twitter accounts according to following steps: (a) Search accounts with a news organization's name (not the description), (b) Choose accounts with the organization's name in the account name (acronyms are accounted, e.g. WP = Washington Post) and (c) If an account description states that the account is maintained by individual reporters, the account was excluded.

These steps aim to count only social media strategies come out of the organization level rather than a journalist level or some bureau (e.g. Economy or International).

## **B.2** How Scraper Works

The news scraper opens all the URL from MediaCloud and extracts news text. The text extraction is conducted by a popular open-source detector Goose, which has been an originally Java based open-source program, and now it is implemented with Scala and Python. Goose uses rules to search through HTML tags to detect text. The news scraper also uses three strategies to extract published date. The first priority is to parse the text and match date with regular expressions. If it doesn't work, the web crawler extracts dates from URL. If that doesn't work either, the news scraper looks for meta data that contains date. To see these algorithms work well enough, I sampled 10 news stories from each news organiza-

Organization	Twitter	Organization	Twitter
CNN	CNN	Washington Post	washingtonpost
	$\operatorname{cnnbrk}$		$\operatorname{postpolitics}$
	cnni		PostWorldNews
	<b>CNNPolitics</b>		$\operatorname{postlive}$
New York Time	$\operatorname{nytimes}$		WashPostDC
	nytimesbooks		wapodesign
The Guardian	guardian		WashPostPR
	guardiannews		PostVideo
	GanPolitics		wasnpostneip Dest Greenbies
	Guardianfilm		PostSports
	guardianopinion		PostStyle
	guardianmusic		wpmagazine
	GuardianTravel		WaPoFood
	guardianculture		WaPoTravel
	guardianstyle		post lead
	$\operatorname{guardianlife}$	BBC	BBCWorld
	guardian_sport		BBC
	guardiantech		BBCBreaking
	SocietyGuardian		BBCTechnology
	guardianscience		BBCLondonNews
	GuardianData		BBCBolitics
	GuardianEdu		bbchealth
	GuardianMTN		BBCBD
	guardianweekend		BBCScienceNews
	GuardianFashion		BBCNewsEnts
	${ m guardianstage}$	Forbes	Forbes
	${ m guardiang2}$		$\mathbf{For} \mathbf{besTech}$
	$\operatorname{guardianeco}$		$\operatorname{ForbesLife}$
	guardianworld	Fox News	FoxNews
	GdnVoluntary		FoxSports
	GuardianGDP	Uuffington Post	FUAIV
	a offilms ndmusic	numington i ost	HuffPostPol
	GuardianBooks	Bloomberg	husiness
	guardianreview	Dicomperg	technology
	guardiancities		markets
	GdnSocialCare		bpolitics
	GuardianJobs		luxury
USA Today	usatoday		economics
	USATODAYmoney	Wall Street Journal	WSJ
	usatodaytech		WSJPolitics
	usatodaysports		WSJSports WSIM-slats
	US ATODA Vhoslth		WSJMarkets WSJAcie
	USATODAT lleann USATOpinion		WSJASIa WS Jopinion
	usatodavhss		WS.Ibusiness
	usatodayvideo		WSJeurope
	usatodaystyle		WSJIndia
	usatodaylife		WSJMoneyBeat
	usatodaytravel		WSJNY
	${ m usatodaymags}$		WSJbreakingnews
	${ m usatodaymlb}$		WSJecon
	usatodaynba		WSJtech
	USATODAYeats	D	WSJatrica
	usatodayweather	neuters	Reuters ReutersW1-1
	usatouayfin usatsima		Reuters World
	USATODAVRooks		ReutersOpinion
	COLLI O DULL DOORS		Terrerochmon

Table B.1: List of Twitter accounts included in the dataset.

Organization	Twitter	Organization	Twitter
CNBC	CNBC		ReutersChina
	CNBCnow		reuterspictures
	CNBCTopStories		ReutersUK
	cnbcafrica		ReutersLive
	CNBCPolitics		ReutersUS
	CNBCi		ReutersIndia
New York Post	nypost		ReutersTech
	nypostsports		ReutersAfrica
	nypmetro		ReutersSports
	nypostbiz		ReutersPolls
	NYPfashion		ReutersPolitics
US News	usnews		ReutersShowBiz
	USNewsEducation		specialreports
	${ m USNewsHealth}$	NBC News	NBCNews
The Atlantic	TheAtlantic		NBC
	TheAtlNews	The Hill	${ m thehill}$
	TheAtlEducation		TheHillOpinion
	TheAtlPoliltics		hilltransport
	TheAtlPhoto		The Hill Events
	TheAtlVideo	$\operatorname{Time}$	TIME
	TheAtlTech	CBS News	CBSNews
	TheAtlGlobal		CBS
	TheAtlCulture		CBSTweet
	${ m TheAtlHealth}$	ABC News	ABC
$\operatorname{Economist}$	The Economist		ABCNetwork
	${\tt EconomistEvents}$		abcnews
	$\operatorname{EconAmericas}$		ABCPolitics
	E conWhichMBA	The Daily Beast	${ m thedailybeast}$
	ECONdailycharts	Chicago Tribune	chicagotribune
	$\mathbf{E}\mathbf{conEurope}$		ChiTribEnt
	$\operatorname{EconUS}$		${ m chicago\_homes}$
	$\operatorname{EconAsia}$		ChiTribBiz
	${f EconCulture}$		ChiTribBooks
	$\operatorname{EconEconomics}$		$\operatorname{ChiTribuneAuto}$
	EconBizFin		$\operatorname{ChiTribFood}$
	$\operatorname{EconSciTech}$		ChiTribPhoto
The Mercury News	mercnews	Al Jazeera	AlJazeera
Nj.com	n j dot com		AJEnglish
	$\mathrm{HSS}\mathrm{portsNJ}$		AlJazeera_World
	NJ_Sports		AJENews
	NJcomsomerset		AJEVideos
	NJentertainment		AJEWeather
	Monmouth_NJ	The Daily Caller	DailyCaller
	NJ_Politics		TheDC_Opinion
	Bergen_NJ		TheDCSports
	njerseypolitics		TheDCPolitics
	NJ_Morris		thedctechnews
	NYJetsNews	The Root	TheRoot
Q.,	butler_nj	Miami Herald	MiamiHerald
Seattle 11mes	seattletimes		HeraldSports
	SealimesJobs		WitamineraldFood
	seatimespreps		HeraldOpEd
	Sea 1 imes Photo		neraldbusiness
	Seatimesoiz	The Dhile delphic In	witamineraldLive
	SeatimesSports	i në Finladelphia inquirer	pninyinquirer
	SeaTimesOpinion	Inquisitr	${ m theinquisitr}$

Table B.1 (cont'd)

Organization	Twitter	Organization	Twitter
Washington Examiner	dcexaminer		InquisitrSports
The Conversation	ConversationUS		${ m IQShowbiz}$
Washington Times	WashTimes		${ m InquisitrLife}$
	wtimespolitics		InquisitrGaming
	WashTimesLocal		InquisitrFunny
Detroit Free Press	freep		InquisitrHealth
o :	freepsports		iqcontest
Oregonian	Oregonian		Inquisitr world
	OregonianBiz		Inquisitr lech
	besports olive	The Secremente Ree	
	insports_onve	The Sacramento Dee	SacDee_news
	OregonianStump	Baltimoro Sun	baltimoresup
Dallas Nows	dallasnews	Dattinore 5 un	baltennarte
Danas news	DallasISD Nows		BaltSunVid
	dmn_collogos		Balt Sun Hoalth
Denver Post	umn_coneges		Danounneann
	denverpost		BaltSunSports
	DenverPostPicks	PBS Newshour	NewsHour
	denverpolitics	1.0	PBS
	denveropinion		NewsHourWorld
	DPostSports	Sun Sentinel	SunSentinel
	DenverPostBrk		PhotoSSentinel
	dpcommunity		SFLEventsForYou
	DPRockies		SSCourts
	denverbusiness		
	denverpostlite	The Press Democrat	NorthBayNews
	PostBroncos	Newsday	newsday
Orange County Register	ocregister		NewsdayHSsports
AlterNet	AlterNet		NewsdaySports
The Kansas City Star	$\operatorname{KCStar}$		${ m NewsdayBiz}$
	m KCStar HS		${ m NewsdayHealth}$
Orlando Sentinel	${ m or } { m landos entinel}$		NewsdayOpinion
	OSCrime		NewsdayEnt
	OSPhoto	Boston Herald	bostonherald
	OSLakeCounty	${ m npr.org/programs}$	NPR
	orlandosports		nprpolitics
~	OSentinelBiz		nprscience
Chicago Sun Times	suntimes		nprbusiness
	suntimes_hoops		nprclassical
	suntimes_preps	Slate	Slate
	CST breaking		SlateMoneybox
	CS reditorials	Dinanican Danant	Slatev Ideo Dinantisanism
	Sun Limes CHI	Western Lournalism	Bipartisanism West Iournalism
	summes_sports	Talling Doints Momo	westJournansm TDM
Detroit News	dotroitnows	Informars	infowars
Detroit ivews	DetNewsOpinion	FiveThirtyFight	FiveThirtyFight
The Stranger	TheStrange	New York Megazine	NVMao
The New Republic	NewBenublic	Media Matters	mmfa
The Indiana Star	industar	Free Beacon	FreeBeacon
The Christian Science Monitor	csmonitor	The Intercept	theintercept
The emission science monitor	csmonitorphoto	Vanity Fair	VanityFair
Common Dreams	commondreams	Vice	VICE
The Tennessean	Tennessean	, 200	vicedocs
Ins Ionnoboun	TNMusicNews		VICESports

Table B.1 (cont'd)

Organization	Twitter	Organization	Twitter
Buzz Feed News Conservative Tribune Gateway Pundit	BuzzFeed BuzzFeedNews BuzzFeedFood BuzzFeedBooks buzzfeedPol BuzzFeedPol BuzzFeedQuiz BuzzFeedQuiz BuzzFeedAnimals BuzzFeedCeleb fwd BuzzFeedSports BuzzFeedSports BuzzFeedGeeky BuzzFeedGeeky BuzzFeedEnt conserv_tribune gatewaypundit	FactCheck IBTimes townhall.com Des Moines Register Westword Politico Breitbart PoliticusUSA MSNBC Vox	VICESportsNZ VICERECORDS vicetech VICE_NZ vice_money factcheckdotorg IBTimes townhallcom DMRegister RegisterVisuals DenverWestword politico BreitbartNews politicususa MSNBC voxdotcom

Table B.1 (cont'd)

tions, and manually see if the text and date extraction worked well. I only included news organizations for which the extraction algorithm worked well with 9 out 10 news stories.

When there's paywall on a news organization's website, I use Selenium to scrape news text. Since Selenium allows automate searching through a Web browser, it allows for logging in as if a user manually type in an account name and a password, and extract text from the displayed text. Although many websites have paywalls, most of them had a partial paywall in a sense that, when a user uses URL from RSS feed, they show the full text. Thus, the Selenium-dependent scraper was only used to the New York Time, Wall Street Journal, Washington Post and Economist.

## B.3 Twitter Scraper

To scrape news companies' tweets, I use Twitter API's *GET statusesuser\_timeline* method that returns up to 3,200 the most recent tweets from a user's account. Scraping news organization's all tweets is straightforward because the timeline returned by the method is equivalent to what a user sees on screen.<sup>1</sup>

 $<sup>^{1} \</sup>rm https://dev.twitter.com/rest/reference/get/statuses/user_{t}imeline$ 



Figure B.1: An example of shortened news URL in a tweet.

## B.4 URL Matcher

To relate news link embedded in tweets to original news articles, I also develop a URL matcher. This software works in three steps. The URL matcher extracts a URL from tweets scraped by the Twitter scraper. This task is straightforward because an embedded URL is tagged in a JSON form as Twitter API provides. Then it finds an original form of the URL if it is shortened. Most of URLs embedded in social media posts are shortened as in Figure B.1 so that it would not take up much space or the character limit. The original forms were identified by sending HTTP request using each shortened URLs. Since URLs often contain query terms attached to use them for a variety of functionality, such as the traffic source identification, the URL matcher parses out. As a final normalization process, the URL matcher only extracts 'path' part that represents unique location within a website to minimize the false negative cases from the matching. Due to a popular search engine optimization (SEO) strategy, majority of the paths contain some form of news headlines. However, sometimes they are numeric news IDs. After the normalization process, a news story is labeled as 'shared' when the normalized URL from a news organization's Twitter account.

## APPENDIX C

## ADDITIONAL STM GRAPHS AND FULL OUTPUT

# C.1 Additional Graphs



Figure C.1: News publication and link sharing patterns by news organizations: Life/Interview topic.



Figure C.2: News publication and link sharing patterns by news organizations: Economy/Technology East topic.



Figure C.3: News publication and link sharing patterns by news organizations: Social/Identity Conflict topic.



Figure C.4: News publication and link sharing patterns by news organizations: Legal/Regulation topic.



Figure C.5: News publication and link sharing patterns by news organizations: Economy/-Finance topic.



Figure C.6: News publication and link sharing patterns by news organizations: UK topic.



Figure C.7: News publication and link sharing patterns by news organizations: Update topic.



Figure C.8: News publication and link sharing patterns by news organizations: Sports/Others topic.



Figure C.9: News publication and link sharing patterns by news organizations: Entertainment topic



Figure C.10: News publication and link sharing patterns by news organizations: Regional Politics topic.



Figure C.11: News publication and link sharing patterns by news organizations: Weather topic.



Figure C.12: News publication and link sharing patterns by news organizations: Health topic.


Figure C.13: News publication and link sharing patterns by news organizations: Subscription topic.



Figure C.14: News publication and link sharing patterns by news organizations: ITax topic.



Figure C.15: News publication and link sharing patterns by news organizations: Social Media topic.



Figure C.16: News publication and link sharing patterns by news organizations: Education topic.



Figure C.17: News publication and link sharing patterns by news organizations: Sports/-College topic.



Figure C.18: News publication and link sharing patterns by news organizations: Me Too topic.



Figure C.19: News publication and link sharing patterns by news organizations: French topic.



Figure C.20: News publication and link sharing patterns by news organizations: Advertising topic.

## APPENDIX D

## SENTIMENT ANALYSIS OUTPUTS BASED ON LSD AND AFINN



Figure D.1: Average sentiment scores of news stories and tweets: AFINN dictionary. Zero score indicates neutral sentiment.



Figure D.2: Average sentiment scores of news stories and tweets: LSD dictionary. Zero score indicates neutral sentiment.

BIBLIOGRAPHY

## BIBLIOGRAPHY

- An, Jisun, Meeyoung Cha, Krishna Gummadi & Jon Crowcroft. 2011. Media landscape in twitter: A world of new conventions and political diversity. In *Proceedings of icwsm 2011*, 18–25. AAAI.
- An, Jisun, Daniele Quercia, Meeyoung Cha, Krishna Gummadi & Jon Crowcroft. 2014. Sharing political news: the balancing act of intimacy and socialization in selective exposure. *EPJ Data Science* 3(1). 12.
- Anderson, Chris. 2007. The long tail: how endless choice is creating unlimited demand. Random House.
- Anderson, Simon P & André De Palma. 2009. Information congestion. The RAND Journal of Economics 40(4). 688–709.
- Anderson, Simon P & André De Palma. 2012. Competition for attention in the information (overload) age. The RAND Journal of Economics 43(1). 1–25.
- Anderson, Simon P & André De Palma. 2013. Shouting to be heard in advertising. Management Science 59(7). 1545–1556.
- Anderson, Simon P, Andre De Palma & Jacques François Thisse. 1992. Discrete choice theory of product differentiation. MIT press.
- Andrew, Blake C. 2007. Media-generated shortcuts: Do newspaper headlines present another roadblock for low-information rationality? *Harvard International Journal of Press/Politics* 12(2). 24–43.
- Armstrong, Cory L & Fangfang Gao. 2010. Now tweet this how news organizations use twitter. *Electronic News* 4(4). 218–235.
- Bakshy, Eytan, Solomon Messing & Lada A Adamic. 2015. Exposure to ideologically diverse news and opinion on facebook. *Science* 348(6239). 1130–1132.
- Baltas, George. 2003. Determinants of internet advertising effectiveness: an empirical study. International Journal of Market Research 45(4). 505–515.
- Barzilai-Nahon, Karine. 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.
- Bell, Allan. 1991. The language of news media. Blackwell Oxford.
- Benkler, Yochai, Robert Faris, Robertsm Hal & Ethan Zuckerman. 2017. Study: Breitbartled right-wing media ecosystem altered broader media agenda. Tech. rep. Columbia Journalism Review. https://www.cjr.org/analysis/breitbart-media-trump-harvard-study.php.

- Bennett, Shea. 2013, July 13. Social media overload how much information do we process each day? Social Times. http://www.adweek.com/socialtimes/social-media-overload/ 488800.
- Bennett, W Lance & Shanto Iyengar. 2008. A new era of minimal effects? the changing foundations of political communication. *Journal of Communication* 58(4). 707–731.
- Berger, Ben. 2011. Attention deficit democracy: The paradox of civic engagement. Princeton University Press.
- Berger, Jonah & Katherine L Milkman. 2012. What makes online content viral? Journal of marketing research 49(2). 192–205.
- Bischof, Jonathan & Edoardo M Airoldi. 2012. Summarizing topical content with word frequency and exclusivity. In Proceedings of the 29th international conference on machine learning (icml-12), 201-208.
- Blei, David & John Lafferty. 2006. Correlated topic models. Advances in neural information processing systems 18. 147–154.
- Blei, David M. 2012. Probabilistic topic models. Communications of the ACM 55(4). 77-84.
- Blei, David M, Andrew Y Ng & Michael I Jordan. 2003. Latent dirichlet allocation. the Journal of machine Learning research 3. 993–1022.
- Blom, Jonas Nygaard & Kenneth Reinecke Hansen. 2015. Click bait: Forward-reference as lure in online news headlines. *Journal of Pragmatics* 76. 87–100.
- Boydstun, Amber E, Justin H Gross, Philip Resnik & Noah A Smith. 2013. Identifying media frames and frame dynamics within and across policy issues. In *New directions in analyzing text as data workshop, london,*.
- Bozdag, Engin & Jeroen van den Hoven. 2015. Breaking the filter bubble: democracy and design. *Ethics and Information Technology* 17(4). 249–265.
- Bradley, Margaret M & Peter J Lang. 1999. Affective norms for english words (anew): Instruction manual and affective ratings. Tech. rep. Technical report C-1, the center for research in psychophysiology, University of Florida.
- Brown, Pete. 2018, April 18. Facebook struggles to promote 'meaningful interactions' for local publishers, data shows. *Columbia Journalism Review*. https://www.cjr.org/tow\_center/facebook-local-news.php.
- Bruns, Axel. 2011. Gatekeeping, gatewatching, real-time feedback: new challenges for journalism. *Brazilian Journalism Research* 7(2). 117–136.
- Carlson, Matt. 2007. Order versus access: News search engines and the challenge to traditional journalistic roles. *Media, Culture & Society* 29(6). 1014–1030.

- Chakraborty, Abhijnan, Bhargavi Paranjape, Sourya Kakarla & Niloy Ganguly. 2016. Stop clickbait: Detecting and preventing clickbaits in online news media. In 2016 ieee/acm international conference on advances in social networks analysis and mining (asonam), 9-16. IEEE.
- Chan-Olmsted, Sylvia, Hyejoon Rim & Amy Zerba. 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.
- Chan-Olmsted, Sylvia M & Jung Suk Park. 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.
- Chang, Jonathan, Sean Gerrish, Chong Wang, Jordan L Boyd-Graber & David M Blei. 2009. Reading tea leaves: How humans interpret topic models. In Advances in neural information processing systems, 288–296.
- Cleary, Johanna, Eisa al Nashmi, Terry Bloom & Michael North. 2015. Valuing twitter: Organizational and individual representations at cnn international. *Digital Journalism* 3(6). 908–924.
- Cleveland, William S, E Grosse & WM Shyu. 1992. Local regression models. Statistical models in S, edited by John M. Chambers and Trevor J. Hastie 309-376.
- Cook, Clare & Esa Sirkkunen. 2013. What's in a niche? exploring the business model of online journalism. *Journal of Media Business Studies* 10(4). 63–82.
- Cottle, Simon. 2007. Ethnography and news production: New (s) developments in the field. Sociology Compass 1(1). 1–16.
- Curran, James. 2010. The future of journalism. Journalism studies 11(4). 464–476.
- Dahl, Robert A. 2005. Who governs?: Democracy and power in an american city. Yale University Press.
- De Vries, Lisette, Sonja Gensler & Peter SH Leeflang. 2012. Popularity of brand posts on brand fan pages: An investigation of the effects of social media marketing. *Journal of Interactive Marketing* 26(2). 83–91.
- Deuze, Mark. 2008. Understanding journalism as newswork: How it changes, and how it remains the same. Westminster Papers in Communication and Culture 5(2). 4–24.
- Diakopoulos, Nicholas. 2017. The routledge companion to digital journalism studies chap. Computational journalism and the emergence of news platforms, 176–184. New York, NY: Routledge.
- Dick, Murray. 2011. Search engine optimisation in uk news production. *Journalism Practice* 5(4). 462–477.

- Dimitrova, Daniela V, Colleen Connolly-Ahern, Andrew Paul Williams, Lynda Lee Kaid & Amanda Reid. 2003. Hyperlinking as gatekeeping: Online newspaper coverage of the execution of an american terrorist. *Journalism Studies* 4(3). 401–414.
- Domingo, David & Chris Paterson. 2011. Making online news (volume 2). Newsroom Ethnography in the Second Decade of Internet Journalism, New York: Peter Lang.
- Dor, Daniel. 2003. On newspaper headlines as relevance optimizers. *Journal of Pragmatics* 35(5). 695–721.
- Elizabeth, Jane. 2017, November 14. Social media teams today: A summary of what we learned. *American Press Institute*. https://www.americanpressinstitute.org/publications/reports/strategy-studies/social-media-teams-today/.
- Entman, Robert M. 1993. Framing: Toward clarification of a fractured paradigm. *Journal* of communication 43(4). 51–58.
- Evans, David S. 2003. The antitrust economics of multi-sided platform markets. Yale Journal on Regulation 20. 325–381.
- Faris, Robert, Hal Roberts, Bruce Etling, Nikki Bourassa, Ethan Zuckerman & Yochai Benkler. 2017. Partisanship, propaganda, and disinformation: Online media and the 2016 us presidential election. Tech. rep. Berkman Klein Center.
- Gamson, William A & Andre Modigliani. 1994. The changing culture of affirmative action. Equal employment opportunity: labor market discrimination and public policy 373–394.
- García-Perdomo, Víctor, Ramón Salaverría, Danielle K Kilgo & Summer Harlow. 2017. 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* 1–22.
- Gentzkow, Matthew, Edward L Glaeser & Claudia Goldin. 2006. The rise of the fourth estate. how newspapers became informative and why it mattered. In *Corruption and* reform: Lessons from america's economic history, 187–230. University of Chicago Press.
- Gentzkow, Matthew & Jesse M Shapiro. 2010. What drives media slant? evidence from us daily newspapers. *Econometrica* 78(1). 35–71.
- Giomelakis, Dimitrios & Andreas Veglis. 2015. Employing search engine optimization techniques in online news articles. *Studies in Media and Communication* 3(1). 22–33.
- Granovetter, Mark S. 1973. The strength of weak ties. American journal of sociology 78(6). 1360–1380.
- Greer, Clark F & Douglas A Ferguson. 2011. Using twitter for promotion and branding: A content analysis of local television twitter sites. *Journal of Broadcasting & Electronic Media* 55(2). 198-214.
- Grueskin, Bill, Ava Seave & Lucas Graves. 2011. The story so far: What we know about the business of digital journalism. Columbia University Press.

- Hansen, Lars Kai, Adam Arvidsson, Finn Årup Nielsen, Elanor Colleoni & Michael Etter. 2011. Good friends, bad news-affect and virality in twitter. In *Future information* technology, 34–43. Springer.
- Hardin, Garrett. 1968. The tragedy of the commons. Science 162(3859). 1243–1248.
- Hart, Roderick P. 2001. Redeveloping diction: theoretical considerations. Theory, Method, and Practice in Computer Content Analysis 16. 43-60.
- Hermida, Alfred. 2010. Twittering the news: The emergence of ambient journalism. Journalism Practice 4(3). 297–308.
- Hermida, Alfred, Fred Fletcher, Darryl Korell & Donna Logan. 2012. Share, like, recommend: Decoding the social media news consumer. *Journalism Studies* 13(5-6). 815–824.
- Hodson, Jaigris & April Lindgren. 2017. Is no election news good news?: The 2015 canadian election and locally relevant news on twitter. In Proceedings of the 8th international conference on social media & society, 39. ACM.
- Hong, Sounman. 2012. Online news on twitter: Newspapers' social media adoption and their online readership. *Information Economics and Policy* 24(1). 69–74.
- Horne, Benjamin D & Sibel Adali. 2017. The impact of crowds on news engagement: A reddit case study. *arXiv preprint arXiv:1703.10570*.
- Hu, Xia & Huan Liu. 2012. Text analytics in social media. In *Mining text data*, 385–414. Springer.
- Ingram, Mathew. 2018, January 22. Facebook's latest changes will probably make misinformation worse. Columbia Journalism Review. https://www.cjr.org/innovations/facebooknewsfeed-changes.php.
- Kahneman, Daniel. 1979. Prospect theory: An analysis of decisions under risk. *Econometrica* 47(2). 263–292.
- Kaplan, Sarah & Rebecca Henderson. 2005. Inertia and incentives: Bridging organizational economics and organizational theory. *Organization Science* 16(5). 509–521.
- Kew, Ben. 2016, August 17. Digiday: Breitbart 'dominating' social media and 'punches above its weight'. Breitbart. http://www.breitbart.com/big-journalism/2016/08/17/digidaybreitbart-dominating-social-media-and-punches-above-its-weight/.
- Kim, Hyun Suk. 2015. Attracting views and going viral: How message features and newssharing channels affect health news diffusion. *Journal of Communication* 65(3). 512–534.
- Kiousis, Spiro & Matt Ragas. 2015. Implications of third-level agenda building for public relations and strategic communication. The Power of Information Networks: New Directions for Agenda Setting 8. 161–174.

- Kitchin, Rob. 2016. Thinking critically about and researching algorithms. Information, Communication & Society 20(1). 1–16.
- Kleinberg, Jon M. 1999. Authoritative sources in a hyperlinked environment. Journal of the ACM (JACM) 46(5). 604–632.
- Kroll, Joshua A, Joanna Huey, Solon Barocas, Edward W Felten, Joel R Reidenberg, David G Robinson & Harlan Yu. 2016. Accountable Algorithms. University of Pennsylvania Law Review 165. 1–60.
- Kwak, Haewoon, Changhyun Lee, Hosung Park & Sue Moon. 2010. What is twitter, a social network or a news media? In *Proceedings of the 19th international conference on world wide web*, 591-600. ACM.
- Lasorsa, Dominic L, Seth C Lewis & Avery E Holton. 2012. Normalizing twitter: Journalism practice in an emerging communication space. *Journalism studies* 13(1). 19–36.
- Lawrence, Regina G, Logan Molyneux, Mark Coddington & Avery Holton. 2014. Tweeting conventions: Political journalists' use of twitter to cover the 2012 presidential campaign. *Journalism Studies* 15(6). 789–806.
- Lerman, Kristina & Rumi Ghosh. 2010. Information contagion: An empirical study of the spread of news on digg and twitter social networks. *Icwsm* 10. 90–97.
- Liang, Hai & King-wa Fu. 2016. Information overload, similarity, and redundancy: unsubscribing information sources on twitter. *Journal of Computer-Mediated Communication* 22(1). 1–17.
- Lischka, Juliane 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.
- Lowrey, Wilson. 2009. Institutional roadblocks: Assessing journalism's response to changing audiences. Journalism and citizenship new agendas in communication 44-67.
- Lowrey, Wilson. 2012. Journalism innovation and the ecology of news production: Institutional tendencies. Journalism & Communication Monographs 14(4). 214–287.
- Lowrey, Wilson. 2017. The emergence and development of news fact-checking sites: Institutional logics and population ecology. *Journalism Studies* 18(3). 376–394.
- Lucas, Christopher, Richard Nielsen, Margaret Roberts, Brandon Stewart, Alex Storer & Dustin Tingley. 2015. Computer assisted text analysis for comparative politics. *Political Analysis* 23(2). 254–277.
- Lucas Jr, Robert E. 1976. Econometric policy evaluation: A critique. In *Carnegie-rochester* conference series on public policy, vol. 1, 19–46. Elsevier.
- Manning, Christopher D, Prabhakar Raghavan, Hinrich Schütze et al. 2008. Introduction to information retrieval, vol. 1 1. Cambridge university press Cambridge.

- Manovich, Lev. 2012. Data stream, database, timeline: The forms of social media. Software Studies Initiative 27. http://lab.softwarestudies.com/2012/10/data-stream-database-timeline-new.html.
- Martindale, Colin. 1975. Romantic progression: The psychology of literary history. Hemisphere publishing corporation.
- McComas, Katherine & James Shanahan. 1999. Telling stories about global climate change: Measuring the impact of narratives on issue cycles. *Communication Research* 26(1). 30–57.
- McCombs, Maxwell. 2014. Setting the agenda: Mass media and public opinion. John Wiley & Sons.
- McCombs, Maxwell, Juan Pablo Llamas, Esteban Lopez-Escobar & Federico Rey. 1997. Candidate images in spanish elections: Second-level agenda-setting effects. Journalism & Mass Communication Quarterly 74(4). 703-717.
- Meraz, Sharon. 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.
- Meraz, Sharon & Zizi Papacharissi. 2013. Networked gatekeeping and networked framing on # egypt. The international journal of press/politics 18(2). 138–166.
- Mergenthaler, Erhard. 1996. Emotion-abstraction patterns in verbatim protocols: A new way of describing psychotherapeutic processes. *Journal of consulting and clinical psychology* 64(6). 1306–1315.
- Metzgar, Emily T, David D Kurpius & Karen M Rowley. 2011. Defining hyperlocal media: Proposing a framework for discussion. New Media & Society 13(5). 772–787.
- Meyer, Kelly M & Tang Tang. 2015. # socialjournalism: Local news media on twitter. International Journal on Media Management 17(4). 241-257.
- Mimno, David, Hanna M Wallach, Edmund Talley, Miriam Leenders & Andrew McCallum. 2011. Optimizing semantic coherence in topic models. In Proceedings of the conference on empirical methods in natural language processing, 262–272. Association for Computational Linguistics.
- Mitchell, Amy, Mark Jurkowitz & Kenneth Olmstead. 2014. Social, search and direct pathways to digital news. Tech. rep. Pew Research Center. http://www.journalism.org/ 2014/03/13/social-search-direct/.
- Mitra, Tanushree & Eric Gilbert. 2014. The language that gets people to give: Phrases that predict success on kickstarter. In *Proceedings of the 17th acm conference on computer supported cooperative work & social computing*, 49–61. ACM.
- Molek-Kozakowska, Katarzyna. 2013. Towards a pragma-linguistic framework for the study of sensationalism in news headlines. *Discourse & Communication* 7(2). 173–197.

- Monroe, Burt L, Michael P Colaresi & Kevin M Quinn. 2008. Fightin'words: Lexical feature selection and evaluation for identifying the content of political conflict. *Political Analysis* 16(4). 372–403.
- Mourão, Rachel Reis. 2015. The boys on the timeline: Political journalists' use of twitter for building interpretive communities. *Journalism* 16(8). 1107–1123.
- Newhagen, John E. 1998. Tv news images that induce anger, fear, and disgust: Effects on approach-avoidance and memory. *Journal of Broadcasting & Electronic Media* 42(2). 265–276.
- Newman, David, Jey Han Lau, Karl Grieser & Timothy Baldwin. 2010. Automatic evaluation of topic coherence. In Human language technologies: The 2010 annual conference of the north american chapter of the association for computational linguistics, 100–108. Association for Computational Linguistics.
- Newman, Nic. 2011. Mainstream media and the distribution of news in the age of social discovery. Tech. rep. Reuters Institute for the Study of Journalism, University of Oxford.
- Newman, Nic, Richard Fletcher, David A. L. Levy & Rasmus Kleis Nielsen. 2016. Reuter institute digital news report 2016. Tech. rep. Reuters Institute for the Study of Journalism, University of Oxford. http://www.digitalnewsreport.org/.
- Nicholls, T, Nabeelah Shabbir & Rasmus Kleis Nielsen. 2016. Digital-born news media in europe. Tech. rep. Reuters Institute for the Study of Journalism, University of Oxford.
- Nielsen, Rasmus Kleis. 2015. Local journalism: The decline of newspapers and the rise of digital media. IB Tauris.
- Ostrom, Elinor. 2015. Governing the commons. Cambridge university press.
- Palser, Barb. 2009. Hitting the tweet spot: news outlets should use twitter to reach elusive and valuable audiences. *American Journalism Review* 31(2). 54–55.
- Pariser, Eli. 2011. The filter bubble: How the new personalized web is changing what we read and how we think. Penguin.
- Parmelee, John H. 2013. Political journalists and twitter: Influences on norms and practices. Journal of Media Practice 14(4). 291–305.
- Parveen, Farzana, Noor Ismawati Jaafar & Sulaiman Ainin. 2015. Social media usage and organizational performance: Reflections of malaysian social media managers. *Telematics* and Informatics 32(1), 67–78.
- Pennebaker, James W, Martha E Francis & Roger J Booth. ???? Linguistic inquiry and word count: Liwc 2001. Mahway: Lawrence Erlbaum Associates.

Porter, Martin F. 1980. An algorithm for suffix stripping. *Program* 14(3). 130–137.

- Potthast, Martin, Sebastian Köpsel, Benno Stein & Matthias Hagen. 2016. Clickbait detection. In European conference on information retrieval, 810-817. Springer.
- Price, Vincent, David Tewksbury & Elizabeth Powers. 1997. Switching trains of thought: The impact of news frames on readers' cognitive responses. *Communication research* 24(5). 481–506.
- Purcell, Kristen, Lee Rainie, Tom Rosenstiel & Kenneth Olmstead. 2010. Understanding the participatory news consumer. Tech. rep. Pew Research Center. http://www.pewinternet. org/2010/03/01/part-5-news-gets-personal-social-and-participatory/.
- Rauch, Jennifer. 2015. Exploring the alternative-mainstream dialectic: What "alternative media" means to a hybrid audience. *Communication, Culture & Critique* 8(1). 124–143.
- Reis, Julio, Fabricio Benevenuto, P Vaz de Melo, Raquel Prates, Haewoon Kwak & Jisun An. 2015. Breaking the news: First impressions matter on online news. In Proceedings of the international conference on weblogs and social media 2015, 357-366.
- Roberts, Margaret E, Brandon M Stewart & Edoardo M Airoldi. 2016. A model of text for experimentation in the social sciences. *Journal of the American Statistical Association* 111(515). 988–1003.
- Roberts, Margaret E, Brandon M Stewart, Dustin Tingley, Edoardo M Airoldi et al. 2013. The structural topic model and applied social science. In Advances in neural information processing systems workshop on topic models: Computation, application, and evaluation, .
- Roberts, Margaret E, Brandon M Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson & David G Rand. 2014. Structural topic models for open-ended survey responses. *American Journal of Political Science* 58(4). 1064–1082.
- Robinson, Helen, Anna Wysocka & Chris Hand. 2007. Internet advertising effectiveness: the effect of design on click-through rates for banner ads. *International Journal of Advertising* 26(4). 527–541.
- Roston, Michael. 2015, January 22. Don't try too hard to please twitter and other lessons from the new york times' social media desk. 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, David. 2014, January 2. How buzzfeed mastered social sharing to become a media giant for a new era. *Wired*. http://www.wired.co.uk/article/buzzfeed.
- Russell Neuman, W, Lauren Guggenheim, S Mo Jang & Soo Young Bae. 2014. The dynamics of public attention: Agenda-setting theory meets big data. *Journal of Communication* 64(2). 193–214.

- Sandvig, Christian, Kevin Hamilton, Karrie Karahalios & Cedric Langbort. 2016. Automation, algorithms, and politics when the algorithm itself is a racist: Diagnosing ethical harm in the basic components of software. *International Journal of Communication* 10. 4882–4890.
- Scheufele, Dietram A. 1999. Framing as a theory of media effects. *Journal of communication* 49(1). 103–122.
- Scheufele, Dietram A. 2000. Agenda-setting, priming, and framing revisited: Another look at cognitive effects of political communication. *Mass Communication & Society* 3(2-3). 297–316.
- Schmierbach, Mike & Anne Oeldorf-Hirsch. 2012. A little bird told me, so i didn't believe it: Twitter, credibility, and issue perceptions. *Communication Quarterly* 60(3). 317–337.
- Schudson, Michael. 1998. The good citizen: A history of american civic life. Free Press.
- Shaw, Donald Lewis & Maxwell E McCombs. 1977. The emergence of american political issues: The agenda-setting function of the press. West Group.
- Shearer, Elisa & Jeffrey Gottfried. 2017. News use across social media platforms 2017. Tech. rep. Pew Research Center. http://www.journalism.org/2017/09/07/news-use-acrosssocial-media-platforms-2017/.
- Shoemaker, Pamela J & Timothy Vos. 2009. *Gatekeeping theory*. Routledge.
- Singer, Jane. 2009. Journalism in the network. Routledge.
- Singer, Jane B. 2005. The political j-blogger: 'normalizing'a new media form to fit old norms and practices. *Journalism* 6(2). 173–198.
- Sjøvaag, Helle & Eirik Stavelin. 2012. Web media and the quantitative content analysis: Methodological challenges in measuring online news content. *Convergence* 18(2). 215–229.
- Soroka, Stuart N. 2006. Good news and bad news: Asymmetric responses to economic information. *Journal of Politics* 68(2). 372–385.
- Starbird, Kate. 2017. Examining the alternative media ecosystem through the production of alternative narratives of mass shooting events on twitter. In *Icwsm*, 230–239.
- Stieglitz, Stefan & Linh Dang-Xuan. 2013. Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior. Journal of Management Information Systems 29(4). 217–248.
- Stigler, George J. 1961. The economics of information. *Journal of political economy* 69(3). 213–225.
- Stone, Philip J, Dexter C Dunphy & Marshall S Smith. 1966. The general inquirer: A computer approach to content analysis. MIT press.

- Strapparava, Carlo, Alessandro Valitutti et al. 2004. Wordnet affect: an affective extension of wordnet. In *Lrec*, vol. 4, 1083–1086. Citeseer.
- Taddy, Matt. 2013. Multinomial inverse regression for text analysis. *Journal of the American Statistical Association* 108(503). 755–770.
- Tandoc Jr, Edson C. 2014. Journalism is twerking? how web analytics is changing the process of gatekeeping. New Media & Society 16(4). 559–575.
- Tandoc Jr, Edson C. 2016. Journalistic autonomy and web analytics. The Routledge Companion to Digital Journalism Studies 293–301.
- Tenenboim, Ori & Akiba A Cohen. 2015. What prompts users to click and comment: A longitudinal study of online news. *Journalism* 16(2). 198–217.
- Thomson, Robert, Naoya Ito, Hinako Suda, Fangyu Lin, Yafei Liu, Ryo Hayasaka, Ryuzo Isochi & Zian Wang. 2012. Trusting tweets: The fukushima disaster and information source credibility on twitter. In *Proceedings of the 9th international iscram conference*, 1–10.
- Trussler, Marc & Stuart Soroka. 2014. Consumer demand for cynical and negative news frames. The International Journal of Press/Politics 19(3). 360-379.
- Tsimonis, Georgios & Sergios Dimitriadis. 2014. Brand strategies in social media. Marketing Intelligence & Planning 32(3). 328–344.
- Turner, Fred. 2005. Actor-networking the news. Social Epistemology 19(4). 321–324.
- Turney, Peter D & Michael L Littman. 2003. Measuring praise and criticism: Inference of semantic orientation from association. ACM Transactions on Information Systems (TOIS) 21(4). 315–346.
- Uribe, Rodrigo & Barrie Gunter. 2007. Aresensational'news stories more likely to trigger viewers' emotions than non-sensational news stories? a content analysis of british tv news. European Journal of Communication 22(2). 207–228.
- Vargo, Chris J, Lei Guo & Michelle A Amazeen. 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.
- Vargo, Chris J, Lei Guo, Maxwell McCombs & Donald L Shaw. 2014. Network issue agendas on twitter during the 2012 us presidential election. *Journal of Communication* 64(2). 296–316.
- Vasu, Norman, Benjamin Ang, Shashi Jayakumar Terri-Anne-Teo, Muhammad Faizal & Juhi Ahuja. 2018. Fake news: National security in the post-truth era. Policy Report, Singapore: CENS-S. Rajaratnam School of International Studies 25.
- Walker, Jack L. 1991. Mobilizing interest groups in america: Patrons, professions, and social movements. University of Michigan Press.

- Wallace, Julian. 2018. Modelling contemporary gatekeeping: The rise of individuals, algorithms and platforms in digital news dissemination. *Digital Journalism* 6(3). 274–293.
- Warzel, Charlie. 2018, January 12. Facebook couldn't handle news. maybe it never wanted to. *BuzzFeed News*. https://www.buzzfeed.com/charliewarzel/facebooks-news-feed-change-is-a-retreat-from-the-public.
- Weber, Matthew S & Peter Monge. 2011. The flow of digital news in a network of sources, authorities, and hubs. *Journal of Communication* 61(6). 1062–1081.
- Weltevrede, Esther, Anne Helmond & Carolin Gerlitz. 2014. The politics of real-time: A device perspective on social media platforms and search engines. *Theory, Culture & Society* 31(6). 125–150.
- Whissell, Cynthia M & Michael RJ Dewson. 1986. A dictionary of affect in language: Iii. analysis of two biblical and two secular passages. *Perceptual and Motor Skills* 62(1). 127–132.
- White, David Manning. 1950. The "gate keeper": A case study in the selection of news. Journalism Bulletin 27(4). 383–390.
- Yankelovich, Daniel. 1991. Coming to public judgment: Making democracy work in a complex world. Syracuse University Press.
- Young, Lori & Stuart Soroka. 2012. Affective news: The automated coding of sentiment in political texts. *Political Communication* 29(2). 205–231.
- Zhao, Wayne Xin, Jing Jiang, Jianshu Weng, Jing He, Ee-Peng Lim, Hongfei Yan & Xiaoming Li. 2011. Comparing twitter and traditional media using topic models. In *European* conference on information retrieval, 338–349. Springer.
- Zhu, Jian-Hua, David Weaver, Ven-Hwei Lo, Chongshan Chen & Wei Wu. 1997. Individual, organizational, and societal influences on media role perceptions: A comparative study of journalists in china, taiwan, and the united states. *Journalism & Mass Communication* Quarterly 74(1). 84–96.
- Zillmann, Dolf, Lei Chen, Silvia Knobloch & Coy Callison. 2004. Effects of lead framing on selective exposure to internet news reports. *Communication Research* 31(1). 58–81.
- Zubiaga, Arkaitz. 2013. Newspaper editors vs the crowd: on the appropriateness of front page news selection. In Proceedings of the 22nd international conference on world wide web, 879–880. ACM.