From newsworthiness to shareworthiness

How to predict news sharing based on article characteristics

Trilling, D.; Tolochko, P.; Burscher, B.

DOI
10.1177/1077699016654682

Publication date
2017

Document Version
Final published version

Published in
Journal & Mass Communication Quarterly

Journalism & Mass Communication Quarterly
2017, Vol. 94(1) 38–60

DOI: 10.1177/1077699016654682
http://journals.sagepub.com/home/jmq

From Newsworthiness to Shareworthiness: How to Predict News Sharing Based on Article Characteristics

Damian Trilling[1], Petro Tolochko[2], and Björn Burscher[1]

Abstract

People increasingly visit online news sites not directly, but by following links on social network sites. Drawing on news value theory and integrating theories about online identities and self-representation, we develop a concept of shareworthiness, with which we seek to understand how the number of shares an article receives on such sites can be predicted. Findings suggest that traditional criteria of newsworthiness indeed play a role in predicting the number of shares, and that further development of a theory of shareworthiness based on the foundations of newsworthiness can offer fruitful insights in news dissemination processes.

Keywords

news dissemination, news sharing, news values, social network sites, online news

Patterns of news consumption are changing drastically. Only two decades ago, the probability of reading a specific news article was mainly a function of being a regular reader of the outlet in which it appeared. Nowadays, this relationship is less straightforward. Journalists are no longer the sole gatekeepers who determine which news one is exposed to, because a second filter layer has emerged: citizens who spread the news by sharing links on social media—or decide not to do so. Even before the rise of social network sites, Bruns (2005) observed this phenomenon in the context of collaborative online news sites and coined the term *gatewatching*: "the observation of the output

[1]University of Amsterdam, The Netherlands
[2]University of Vienna, Austria

Corresponding Author:
Damian Trilling, Afdeling Communicatiewetenschap, Universiteit van Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands.
Email: d.c.trilling@uva.nl

gates of news publications and other sources, in order to identify important material as it becomes available" (p. 17). We argue that the act of news sharing can be described as a new phenomenon that lies somewhere in between news creation and news reception. This article aims to understand the extent to which old explanations like news value theory can be applied to this phenomenon.

Various studies have shown that users receive various gratifications when sharing news (e.g., Ma, Lee, & Goh, 2011). Yet, comparatively little is known about *which* news is actually disseminated via social media—a gap that is especially problematic given the core importance of news exposure for the study of political communication (e.g., Chaffee & Metzger, 2001). We therefore focus on identifying content characteristics of news that predict the frequency with which an article is shared on social network sites.

In his theory of structuration, Giddens (1984) has pointed out that behavior can differ strongly between individuals, but is nevertheless predictable in the aggregate. Drawing on news value theory (Eilders, 2006; Kepplinger, 2008; O'Neill & Harcup, 2009), we argue in a similar vein that, although an individual's news sharing behavior will depend on his personality or her interests, there are structural factors of news content that predict its likelihood of being shared. Yet, it would be naïve to assume that theories from the age of mass media could be applied in a network society (e.g., Van Dijk, 2006) without any modification. For example, as social media are used to shape online identities (e.g., Pempek, Yermolayeva, & Calvert, 2009; Turkle, 1995), articles that jeopardize the users' self-representation by, for example, being overtly controversial might be shared less (e.g., Boczkowski & Mitchelstein, 2012). This article therefore revisits classic news values and takes a first step to extend the concept of *newsworthiness* to one of *shareworthiness*. It aims to assess how far a theoretical framework that has proven to be useful to explain news selection by journalists and by the public has to be adapted if one assumption (in this case, the notion that the roles of gatekeeper and recipient are distinct) does not hold any more. After theoretically conceptualizing shareworthiness, we answer, based on the analysis of 132,682 Dutch news articles, the following question: *How can news sharing on social media be predicted by characteristics of the news article?*

Theoretical Background and Related Research

Accessing news directly from news outlets has never been the only way of doing so. As early as in 1944, Lazarsfeld, Berelson, and Gaudet suggested a two-step flow of communication, in which so-called opinion leaders would spread the news they got from the media to their friends. Interpersonal talk, it seems, plays a big role in the dissemination of news (e.g., Brundidge, 2010; Schmitt-Beck, 2003). In reviewing the literature on news dissemination, Weeks and Holbert (2013) summarized,

> In theory, if people encounter a piece of news that is personally meaningful, they will look to share or talk about it with others, through either conversation or the use of communication technologies. Information sharing is especially likely in situations where

> people have strong social networks and when the content is interesting, helpful, or emotionally arousing. (p. 215)

Thus, in both offline and online settings, there seem to be common factors that determine the shareworthiness of a news story.

Nevertheless, traditional ways to disseminate news, from face-to-face conversations to telephone calls or written messages, are not only costly and time-consuming, but, in most cases, also incapable of reaching more than a handful of persons. Sharing a link to a news article on a social network site, in contrast, requires only a minimum of effort, and is—depending on privacy settings and characteristics of the site—capable of reaching a large to virtually unlimited audience. This also differs from earlier forms of news sharing via websites (e.g., Baym & Shah, 2011) or email (e.g., Berger & Milkman, 2012; Boczkowski & Mitchelstein, 2012)—techniques that lack an integrated functionality for redistribution to a potentially large and undefined audience (retweeting, resharing). This makes these forms of news dissemination much less powerful than social network services. In particular, one may speculate that news sharing is especially effective on Twitter. Facebook, in contrast, has a much more restrictive design that assumes reciprocity in friendship relations and uses a sophisticated (and opaque) algorithm to filter the users' news feeds. Other platforms for news sharing, like del.icio.us or reddit, do not play a significant role in the context of Dutch news, which is why we do not take them into further consideration in this article.

For the case of Twitter, Kwak, Lee, Park, and Moon (2010) showed that already a handful of hops in a retweet chain are enough to reach a substantial audience. Once a tweet is retweeted for the first time, the information disseminates very fast, and saturation is usually reached within 1 day (Castillo, El-Haddad, Pfeffer, & Stempeck, 2014; Kwak et al., 2010; Lerman, Ghosh, & Rey, 2010). Although there are some who say that information cascades are inherently unpredictable (see Salganik, Dodds, & Watts, 2006), others have shown that both user characteristics and message characteristics can predict parameters like speed, range, and scale of a cascade (Cheng, Adamic, Dow, Kleinberg, & Leskovec, 2014; Yang & Counts, 2010; see also Stieglitz & Dang-Xuan, 2012).

There are two complementary approaches to the study of news sharing. The first one relies on experiments (or, occasionally, on surveys) and provides valuable insights into the role of individual traits, attitudes, and habits. Often, a uses-and-gratifications perspective is used (e.g., Lee & Ma, 2012; Ma et al., 2011, 2014), but also concepts as various as opinion leadership, perceived informational utility, partisanship, or diffusion of information theory are used (e.g., Bobkowski, 2015; Ma et al., 2014; Weeks & Holbert, 2013). For example, in two experiments, Bobkowski (2015) investigated how perceived information utility of an article and opinion leadership of an individual can explain news sharing. Ecological validity is a limiting factor: Asking people how likely it is they would share some stimulus hardly mimics a real-life situation, and it is a strong assumption to make that such reported sharing intentions really translate into real-world behavior.

Therefore, approaches that seek to explain sharing with experiments are complemented by content-analytical approaches that investigate sharing as a function of message characteristics, which comes at the expense of the lack of experimental rigor, but can make use of unobtrusively collected real-world data in a nonartificial setting. One example is a study by Boczkowski and Mitchelstein (2012), who showed that potentially controversial news are shared mostly during election times. Berger and Milkman (2012) showed that among other factors, emotional language, but also positivity, increases news sharing. Both studies research sharing via email—thus, sharing with a very limited and clearly defined number of addressees. Shareworthiness, as we define it, refers to sharing with what Schmidt (2014) called "personal publics": a more open and less well-defined group of addressees than possible via personal email.

A number of scholars have studied the sharing of partisan information on social network sites (An, Cha, Gummadi, & Crowcroft, 2011; An, Quercia, & Crowcroft, 2014; Morgan, Shafiq, & Lampe, 2013). Our study adopts a different but complementary approach by investigating the sharing of all kinds of news from journalistic nonpartisan outlets. This is not only novel but also appropriate as it has been shown that people share content from ideologically very different news outlets at the same time (Barbera, Jost, Nagler, Tucker, & Bonneau, 2015; Morgan et al., 2013). In the following section, we therefore lay out our concept of shareworthiness of news, drawing heavily on the idea of newsworthiness based on news value research.

News Values as Predictors of News Sharing

To conceptualize shareworthiness, we depart from the notion of newsworthiness—a concept frequently applied in journalism research and in journalism practice. In particular, we follow an approach taken by *news value* research, as one might expect that articles that are more relevant from a journalistic point of view also get more attention from people who spread journalistic products. The idea that there are some inherent factors to a news item that determine its newsworthiness can be traced back to the classic studies by Östgaard (1965) and Galtung and Ruge (1965), who suggested that some empirically determinable news factors determine the news value of some information. Although the question of how many and which factors have to be distinguished remains subject to debate (e.g., Eilders, 2006; Harcup & O'Neill, 2001), the concept itself has been proven useful in a number of empirical studies (see Kepplinger, 2008; O'Neill & Harcup, 2009). Eilders (2006) argued that these factors guide not only journalists' news selection, but also the selection of news by the audience. In our case, this suggests that users' decision whether to share an article is guided by these news factors as well. And indeed, very recently, it has been suggested to use news value theory to explain engagement with online news: Weber (2014) used the concept of newsworthiness to predict commenting on news articles, and Ziegele, Breiner, and Quiring (2014) used it to explain the amount of interaction within such comments.

It lies beyond the scope of this article to test all of the news factors that have been suggested by various authors. We therefore limit our investigation to a selection of factors that have repeatedly proven to be relevant and can be measured using automated content analysis methods (see Boumans & Trilling, 2016). In the following, we first develop a series of hypotheses and research questions, before we bring them together in a research question that asks about the relative strength of these factors, which will further our understanding of what drives news sharing.

One of the factors that has frequently been shown to be of substantial influence is *proximity*. Several, partly overlapping, definitions and operationalizations have been used, but the general finding is that geographical proximity (e.g., domestic issues) and the involvement of elite nations that are considered culturally proximate (e.g., from a Western perspective, the United States) increase the news value of a story (see the overview by Eilders, 2006). Their involvement also increases engagement with an online news story (Weber, 2014). Involvement is not necessarily equal to location: A U.S. bombing in Syria takes place in a non-Western country, but nevertheless prominently involves a Western actor.

We hypothesize the following:

H1a: Domestic issues increase the number of shares an article receives.
H1b: The lower the geographical distance to the closest country involved, the higher the number of shares an article receives.

Although geographical distance, especially in Europe, can coincide with cultural distance, this is by no means necessarily the case. For example, for Australians, their geographical proximity to Asia could easily be overshadowed by their cultural proximity to the United States or the United Kingdom. As this is—maybe to a lesser extent—also likely to be the case in Europe, with the Netherlands being culturally close to, for example, the United States, we also include the following hypothesis:

H1c: Issues involving Western countries increase the number of shares an article receives.

Most news value studies include a factor called *controversy* or *conflict* (Eilders, 2006). If there is no disagreement on how to evaluate a given issue, then the issue is unlikely to be newsworthy. Rather than wanting to hear about what is consensus in society, both news consumers and journalists seem to demand news that contain some kind of conflict (Trussler & Soroka, 2014; Van Dalen, 2012). We hypothesize the following:

H2: The presence of conflict increases the number of shares an article receives.

Of course, not only news about hard conflict is shared. For instance, Harcup and O'Neill (2001) suggested a news value called *entertainment*. One might even state that there is a visible trend toward an emphasis on entertainment within journalism (for a broader theoretical perspective, see Brants & Van Praag, 2006). In fact, a large share of online news use does not fall into the category of hard news, but is related to softer categories like entertainment or other nonhard news topics (Tewksbury, 2003). We extrapolate these arguments about a preference for entertainment stories to, more generally, stories that contain elements that can be summarized under the umbrella term of *human interest* and expect the following:

> **H3:** The presence of a human interest angle increases the number of shares an article receives.

In line with the old cliché that only bad news is good news, research has consistently found that *negativity* is a factor with substantial influence (Eilders, 2006; Harcup & O'Neill, 2001; Shoemaker & Cohen, 2006). Moreover, it has been suggested that negativity seems to thrive on social media when political topics are discussed (e.g., Trilling, 2015). As we will outline below, also very positive news can be shareworthy, and it is an open question whether negativity or positivity is of higher importance.

In fact, notwithstanding the newsworthiness of negativity, positivity is also regarded as a news value (e.g., Harcup & O'Neill, 2001). It is also conceivable that a news story contains *both* a strongly positive and strongly negative angle. For example, a story about the suffering of refugees could contain paragraphs about sympathetic reactions or help that is provided. Negativity and positivity, thus, can be seen as orthogonal concepts rather than as two values of one concept. Positivity might be of special interest regarding the shareworthiness of online content: When a person shares content on social media, the nature of the shared content reflects on the individual's identity. Unlike a journalist, who acts in a professional role, and unlike an individual's decision to *read* a news piece, the individual's decision to share it can be seen as a part of a manifestation of their online identity. Already in the first years of the existence of the Internet and a decade before Twitter and Facebook were founded, Turkle (1995) described how people can shape online identities that do not necessarily have to correspond with their real-life identity. She refers to homepages as "virtual home, like a real one, [that] is furnished with objects you buy, build, or receive as gifts" (Turkle, 1995, p. 259). These objects can be all types of resources, especially links to other sites. Applying this metaphor to today's social network sites, one can consider someone's profile on such a site as being a home furnished with status updates, pictures, and links. Just as the furnishing of a house reflects the inhabitant's identity, the links reflect on the online identity of the social network user. It makes sense to see identities as something people are constructing continuously, as an ongoing "project of the self" (Giddens, 1991). Empirical studies have adopted the notion that social media users actively engage in constructing representations of their identities by posting and sharing content (Pempek et al., 2009), and the links they share are thought to reveal much about them (Dominick, 1999). Such a constructed image is obviously intended to be positive (see, for example, Zhao, Grasmuck, & Martin, 2008).

There is some empirical evidence for the assumption that people are more likely to associate themselves with positive news content. Berger and Milkman (2012) found that positive news articles tend to be emailed more often than articles containing negative emotions, and Stieglitz and Dang-Xuan (2012) showed that political tweets with a positive tone are shared more often.

Summarizing the arguments so far, we expect the following:

> **H4:** A positive tone increases the number of shares an article receives relative to a neutral tone.

H5: A negative tone increases the number of shares an article receives relative to a neutral tone.

It is a well-known fact that exclusiveness is an important currency among media, and that having a scoop can be highly prestigious for media and journalists (e.g., Fengler & Russ-Mohl, 2008). Also among users who share news, one may reasonably assume that they want to be the ones that point others to new and interesting content, not just reiterating something their personal public already knows. Conversely, Knobloch-Westerwick, Sharma, Hansen, and Alter (2005) discussed whether exposure to online news can be explained by a bandwagon effect. Although they conceded that at first glance, a bandwagon effect where attention attracts more attention seems logical, they argued that this effect can be overshadowed by counteracting forces, especially people seeking for distinctiveness and uniqueness.

Obviously, a news story cannot be exclusively published by one outlet if it is distributed to several outlets at the same time via a news agency. On the contrary, content stemming from a news agency is likely to be encountered via such a large number of channels that it seems unlikely that one feels the need to additionally share it. Although not being written by a news agency may be a rather crude indicator for exclusiveness, we ask the following research question:

RQ1: How does being written by a news agency influence the number of shares an article receives?

Exclusiveness can also be inferred from the topic of an article. An exclusive article, after all, either reveals novel facts about a given topic or introduces a new topic to the debate. In the case of the latter, one could expect that an article that covers a topic that is *not* widely discussed in other articles receives more shares, as it offers something new. In contrast, in the case of a widely discussed topic, the shares will be distributed among several similar articles, lowering the number of shares of each individual one. We can therefore assume that when one assigns a "topic popularity score" to each article, which reflects on how far its contents overlap with topics that are popular in the whole news corpus at the same time, a higher topic popularity score will lead to fewer shares. Of course, this is no rule without exceptions: There will be some topics that are dominating the media agenda and will also receive many shares. That might, for instance, be the case during election campaigns—a time in which, as Boczkowski and Mitchelstein (2012) have shown, sharing behavior deviates from routine periods anyway. We examine whether exclusiveness of a news story to be a positive predictor of sharing:

RQ2: How does the frequency with which the specific topic of an article is covered in the news in general at the same time influence the shares a specific article receives?

Summarizing the Concept of Shareworthiness

As we have shown, although details may vary, there is some general agreement that journalists (and the public as well) share some general ideas about what constitutes a newsworthy story. However, to be able to apply the concept of newsworthiness to explain news sharing, one needs to extend and modify it. In particular, we investigate the role of a number of factors that we expect to explain shareworthiness: (a) geographical distance, (b) cultural distance, (c) negativity, (d) positivity, (e) conflict, (f) human interest, and (g) exclusiveness. We do not want to claim that this list is exhaustive, but we believe that it can serve as a starting point for developing a framework of shareworthiness. In particular, it has also been suggested that messages with a high practical utility are shared more frequently (Berger & Milkman, 2012; Bobkowski, 2015). Although this is an important additional factor to further investigate in future research, it lies beyond the scope of this article due to practical reasons: The size of our data set requires automated coding, and to our best knowledge, no method to automatically code this feature has been tested and described in the literature yet.

Although we have argued that we expect the seven factors we mentioned above to influence the number of shares, confirming these hypotheses would not lead to a full understanding of shareworthiness. Rather, it is important to assess their relative importance and to tease out whether, for example, positivity or negativity plays a bigger role, or whether conflict or human interest is more important. We therefore ask the following:

RQ3: What is the relative importance of the factors mentioned above?

Method

To test our conceptualization of shareworthiness, we analyzed the sharing of news articles originating from six major Dutch news sites: websites of the three nationwide quality newspapers (nrc.nl, trouw.nl, volkskrant.nl), an Amsterdam-based local newspaper (parool.nl), a popular newspaper (ad.nl), and the largest online news site (nu.nl), which is operated by a large publishing house, but has no offline counterpart. The Netherlands enjoys one of the highest Internet penetration rates in the world (International Telecommunications Union, 2012) and is in the top three of European countries with the highest share of social media users (Centraal Bureau voor de Statistiek [Statistics Netherlands], 2013). This makes the country a good case for studying social media news sharing as a mass phenomenon, as one can safely assume that news sharing has passed the stage where it was only practiced by early adopters. In contrast to media from English-speaking countries, Dutch news outlets can be assumed to have a very limited audience outside of the Netherlands. This is a big advantage and allowed us to study news sharing among the intended audience, without having to take into account possible distortions caused by factors relevant only in some other country.

Table 1. Sample Description (N = 132,682).

News site	Description	n
ad.nl	Popular newspaper	45,525
nrc.nl	Quality newspaper	6,267
nu.nl	Online-only news site	5,721
parool.nl	Amsterdam newspaper	24,370
trouw.nl	Quality newspaper	19,299
volkskrant.nl	Quality newspaper	31,500

Note. n = Number of articles included in the model.

Data

Following the method suggested by Trilling (2014), we subscribed to the main RSS feed of each of the sites under study to retrieve the news articles published on these sites between January and August 2014. To this end, we wrote a Python program that queried the RSS feed once an hour, added the RSS items of each new article to a table, and additionally downloaded the whole article by following the link provided by the RSS feed. As the different number of articles retrieved shows (Table 1), the policies of the sites differ in terms of inclusiveness of the feeds. Although in general all articles published are included, some of the sites do not include each and every article. However, as our interest does not lie in the direct comparison of these sites, we do not see this as problematic for our purposes.

In the period under study, we collected 139,132 articles. A small number of the cases were not available because of downloading errors, reducing the sample size to 135,871. Because we used coverage in the week before and after an article was published to determine a topic popularity score, we could not estimate this variable for the first and last week of the data collection, leaving us with 132,712 cases for inclusion in the regression model. A tiny share of 30 cases caused errors while we tried to estimate their sentiment, which gives a final N of 132,682 cases in our models.

Our data approximately follow a count data distribution (Figures 1 and 2), which is why we decided to estimate negative binomial regression models. Count data distributions consist of positive integers only and are right-skewed. The number of social media reactions is a typical example of such a distribution (Saxton & Waters, 2014). Although the simplest count model is a Poisson regression, in our case (as in the comparable study by Saxton & Waters, 2014), the standard deviations of the dependent variables are much higher than the mean (see Table 2), which means that a negative binomial regression is more appropriate (e.g., Gardner, Mulvey, & Shaw, 1995).[1]

Dependent Variables

We retrieved the *number of tweets* and the *number of Facebook interactions* by querying the respective application programming interfaces (APIs). This took some time due to rate limitations and was done between October 3 and 16, 2014. By introducing a time lag of more than a month between the last day of retrieving the news articles and the first day

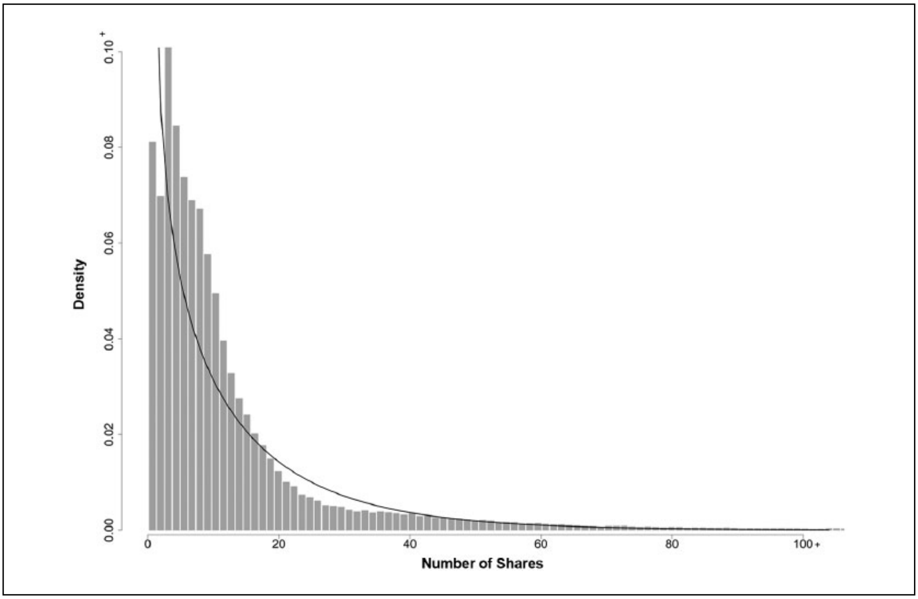

Figure 1. Distribution of Twitter shares.
Note. Histogram with fitted negative binomial distribution. For better readability, the graph is cropped at 100 shares.

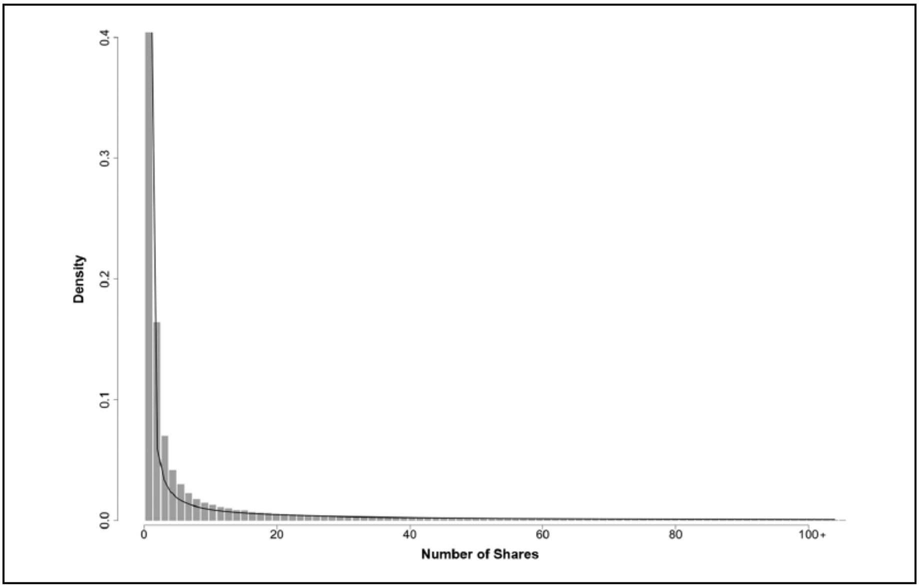

Figure 2. Distribution of Facebook shares.
Note. Histogram with fitted negative binomial distribution. For better readability, the graph is cropped at 100 shares.

Table 2. Descriptive Statistics of All Variables in the Data Set.

Variable	*n*	*M*	*SD*	Minimum	Maximum
Dependent variables					
Facebook interactions	132,682	49.25	590.91	0	79,975
Twitter shares	132,682	11.95	33.79	0	4,235
Independent variables					
Domestic topic	132,682	0.59	0.49	0	1
Geographical distance (in kilometers)	68,290	2,926	3,419	0	18,552
Cultural distance (Western = 1)	68,290	0.68	0.46	0	1
Negativity	132,682	2.85	0.85	1	5
Conflict	132,682	0.59	0.49	0	1
Human interest	132,682	0.84	0.37	0	1
Positivity	132,682	1.87	0.97	1	5
Press agency	132,682	0.50	0.50	0	1
Topic popularity score	132,682	0.07	0.11	0	1
Control variables					
Length (in 1,000 characters)	132,682	1.50	1.74	0	70.66
Topic: Defense and foreign affairs	132,682	0.14	0.34	0	1
Topic: Political system	132,682	0.07	0.26	0	1
Topic: Economic policy	132,682	0.05	0.22	0	1
Topic: Social affairs	132,682	0.05	0.22	0	1
Topic: Law and order	132,682	0.14	0.34	0	1
Topic: Infrastructure	132,682	0.05	0.22	0	1
Topic: Science	132,682	0.01	0.10	0	1
Topic: Culture	132,682	0.09	0.29	0	1
Topic: Weather	132,682	0.01	0.09	0	1
Topic: Sports	132,682	0.26	0.44	0	1

of querying the APIs, we as far as possible reduced the likelihood that people were still sharing the articles. In the case of Twitter, the number can straightforwardly be interpreted as the number of shares, as each tweet that includes a link to the article is counted. For Facebook, this was not possible, because the number provided by the API is calculated as the sum of the number of likes of the URL, the number of shares including copy/pasting a link to Facebook manually, and the number of likes and comments on stories on Facebook about the URL. In that sense, Facebook's measurement of sharing is artificially inflated by including not directly sharing-related activities. On the contrary, one could argue that these activities indirectly contribute to sharing, as an article that enjoys more activity whatsoever is more likely to be displayed in users' news feeds.

Independent Variables

We wrote a series of Python programs to parse the articles and retrieve the relevant information for each independent variable. We applied a supervised machine learning (SML) classifier to code whether an article refers to a *domestic* or an *international* *issue* (**H1a**). SML is a technique in which a computer learns from a set of human-coded training documents to predict content-analytical variables in texts automatically (e.g., Grimmer & Stewart, 2013; Russell & Norvig, 2002). We used a classifier that was trained on a representative collection of Dutch news articles (for a detailed description of the classifier, see Burscher, Vliegenthart, & de Vreese, 2015). The classifier has an accuracy of 87% (area under the curve [AUC] = 0.82).[2] The scikit-learn machine learning library (Pedregosa et al., 2011) was used for training and testing the classifier.

Geographical distance (**H1b**) is determined as follows. First, we parsed the names of all countries and capitals in each article. Then, we calculated the distance between each capital (or the capital of each country) referred to in the article and Amsterdam, the capital of the Netherlands. Distances were obtained via the World Distance Calculator (GlobeFeed, n.d.). Per article, we included the geographical distance of only one country in the data set, the one with the smallest distance to the Netherlands. We subsequently created a series of dummy variables (<500 km, 501-1,000 km, 1,001-2,000 km, 2,001-5,000 km, 5,001-10,000 km, >10,000 km), with 0 km (= within the Netherlands) as reference category.

Cultural proximity (**H1c**) is based on Huntington's (1996) *Clash of Civilizations*. Cultural proximity is a binary variable with 1 indicating the Western civilization and 0 indicating non-Western civilizations.

To infer whether an article reflects *conflict* (**H2**) and whether it has a *human interest angle* (**H3**), we based ourselves on the conceptualization and operationalization of conflict frames and human interest frames provided by Semetko and Valkenburg (2000). Specifically, indicators of disagreement between individuals or groups and the mentioning of several opposing positions were seen as indicators of conflict. Providing human examples and faces, the use of vignettes that generate strong feelings, and the mentioning of how an issue affects individuals or groups were seen as indicators of human interest. We applied two SML classifiers, which were developed based on this operationalization. The procedure is similar to the one used for coding the domestic versus international scope of articles. For a detailed description of both classifiers, we refer to Burscher, Odijk, Vliegenthart, de Rijke, and de Vreese (2014). The coding accuracy is equal to 80% (AUC = 0.78) for conflict and 79% (AUC = 0.78) for human interest.

Using the SentiStrength algorithm (Thelwall, Buckley, Paltoglou, Cai, & Kappas, 2010), we determined the *positivity* (**H4**) and the *negativity* (**H5**) of each article. We reversed the negativity scale, so that both scales range from 1 (*not positive/not negative*) to 5 (*very positive/very negative*). A very unemotional article will score low on both scales. Although it is possible that an article contains a high amount of positivity and negativity on both scales at the same time, this is not too common, as is illustrated by the low correlation of r = .16. The empirical observation that there is no negative correlation between the two is also in line with our theoretical argument that negativity and positivity are two rather independent factors. The SentiStrength algorithm has an accuracy of 96.9% when one deems a 1-point difference on the 5-point scale acceptable (Thelwall et al., 2010). Although the accuracy has been formally evaluated only for the original English-language version of SentiStrength, we believe that even if the Dutch version was less accurate, it still yields largely valid results.

Whether an article was *written by a news agency* (**RQ1**) was determined by parsing the byline. In case the initials of a wire service or news agency were included, we coded the article as written by a news agency.

To determine the popularity of an article's topic at the time when the article was published (**RQ2**), we applied the following procedure. The basic idea behind this procedure is that the nouns in an article's title indicate the topic of that article. Therefore, comparing the distribution of nouns in an article's title with the distribution of such nouns in the titles of all articles published within a certain time period indicates the popularity of an article's topic in that time period. To extract the nouns, we used the Frog Dutch morphosyntactic analyzer and dependency parser (Van den Bosch, Busser, Canisius, & Daelemans, 2007).

First, we split the total period that our study covers into a set of 2-week periods. For each 2-week period $p \in P$, we created a vector $\mathbf{V}$ containing a popularity weight W for each unique noun $n \in N$ in the titles of each article $a \in A$ that was published in the period, $\mathbf{V}_p = (W_{n1}, W_{n2}, \ldots, W_{nN})$. As our study covers 32 weeks, this resulted in 16 popularity weight vectors. In each vector, the popularity weight for each noun is equal to the standardized count of articles in the period containing the noun in its title. Each vector thus represents the popularity of topics within the news coverage of a 2-week period.

Then, we computed a topic popularity score S for each article in the data set. This score is equal to the mean of popularity weights of each unique noun $k \in K$ in the article, where popularity weights were derived from the popularity weight vector of the time period in which the article was published. More formally, we computed $S_a = 1/K \sum_{i=1}^{K} W_{ki}$.

Control Variables

It seems obvious that different topics receive a different number of shares, and this has also been demonstrated empirically (Berger & Milkman, 2012; Wu, Hofman, Mason, & Watts, 2011). At the same time, we have little theoretical interest in the differences between each of the possible categories, which is why we include *main topic* as a control variable. To determine it, we used a machine learning classifier (see Burscher et al., 2015 for details about the classifier and the conceptualization of the topics). It can distinguish between 19 topics as defined by the Comparative Agendas Project (Baumgartner, Green-Pedersen, & Jones, 2006). The coding performance of the classifier is equal to F1 = .71.[3]

To limit the number of control variables, we aggregated the 19 original topics into 10 topic categories: defense and foreign affairs, political system, economic policy, social affairs and education, law and order, infrastructure, science and technology, culture and entertainment, weather and disasters, and sports.

As it is possible that the popularity of one or both of the network sites decreases or increases over time, we also control for the number of *days since beginning of the data collection*. Furthermore, we include the *length of the article*, measured in characters, as it might be the case that short news stories are shared differently than longer ones (Berger & Milkman, 2012). We centered the variable around its mean before entering it into the model. To account for the different intercepts between *news sites* (some are more popular than others), we inserted dummy variables for each site, except for parool.nl, the site with the least shares, which served as the reference category.

Results

Our hypotheses have to be tested for different social media. Therefore, we first determined whether they were used in a comparable fashion: Sharing on Facebook and on Twitter is correlated, but not identical, $r_{\ln(\text{twit}+1),\ln(\text{fb}+1)} = .44$. A first descriptive look at the Twitter sharing data (see Table 2 and Figures 1 and 2) already suggested that sharing behavior indeed takes place in a manner that resembles our theoretical expectations: Although 98.9% of the articles received less than 100 shares on Twitter, some received more than 4,000. No shares are received by 8.1%, and 72.6% received 10 shares or less. Although *zero* shares is seldom, by far most articles receive a very small number of shares—while some articles do receive great attention. Not surprisingly, these data roughly follow a *count distribution*, or, more precisely, a *negative binomial distribution* (see Figures 1 and 2). Facebook interactions show a much bigger spread. On one hand, 40.4% (n = 53,614) of the articles did not receive a single interaction; on the other hand, the three most popular items received 48,689, 53,844, and 79,975 interactions.

To find out which characteristics of a news article drive sharing, we entered all predictors in a negative binomial regression model. The incidence rate ratios (IRRs) given in Table 3 are straightforward to interpret: When the independent variable increases by one unit, the expected count of the dependent variable has to be multiplied by the IRR. For example, an IRR of 0.7 means that a one-unit increase leads to only 70% of expected shares, whereas an IRR of 1.3 can be interpreted as 130% of expected shares.

H1a predicted a higher amount of attention for domestic issues. This hypothesis receives support: According to our model, articles covering domestic issues are shared 1.29 as many times as nondomestic issues on Twitter, and 1.80 as many times as nondomestic issues on Facebook. This is in line with the finding that news stories that take place in the Netherlands receive more Twitter shares than those that take place in geographically distant places (**H1b**). Cultural distance does (**H1c**) matter as well: Our analysis suggests that stories about non-Western countries receive only 0.83 (Twitter) and 0.69 (Facebook) times as many shares as the baseline of Western countries. The conclusion seems clear: Whatever measure we use, a topic that is closer to home (because it is inherently domestic, because of geographical distance, or, in a more abstract sense, because it involves another Western and thus culturally similar country) is shared more often. Comparing the effect sizes, it is also safe to say that whether or not a topic involves the Netherlands is more important than how far away a country is exactly.

Compared with proximity, both the presence of conflict (**H2**) and human interest (**H3**) are less important. On Twitter, articles with a conflict angle are shared 1.11 times as often as articles without any conflict angle; the same is true on Facebook (1.09),

Table 3. Negative Binomial Regressions Predicting the Number of Shares on Twitter and Facebook (N = 132,682).

	Twitter	Facebook
Controls		
ad.nl	3.950*** [3.886, 4.015]	8.419*** [8.097, 8.752]
nrc.nl	8.184*** [7.972, 8.402]	13.721*** [12.821, 14.694]
nu.nl	15.865*** [15.439, 16.305]	62.082*** [57.894, 66.526]
trouw.nl	1.743*** [1.710, 1.777]	0.856*** [0.818, 0.897]
volkskrant.nl	2.364*** [2.320, 2.409]	1.081** [1.032, 1.132]
Days since t_0	0.999*** [0.999, 1.000]	1.002*** [1.002, 1.003]
Length (in 1,000 characters)	1.166*** [1.160, 1.172]	1.278*** [1.259, 1.297]
Topic: Defense and foreign affairs	0.803*** [0.786, 0.821]	0.662*** [0.627, 0.698]
Topic: Political system	0.992 [0.967, 1.016]	0.796*** [0.749, 0.847]
Topic: Economic policy	1.008 [0.980, 1.035]	0.634*** [0.593, 0.679]
Topic: Social affairs and education	1.412*** [1.375, 1.450]	1.435*** [1.341, 1.536]
Topic: Law and order	0.871*** [0.853, 0.889]	0.639*** [0.608, 0.672]
Topic: Infrastructure	1.101*** [1.072, 1.132]	0.964 [0.901, 1.032]
Topic: Science and technology	1.128*** [1.069, 1.190]	2.053*** [1.800, 2.352]
Topic: Culture and entertainment	1.136*** [1.110, 1.162]	1.543*** [1.456, 1.636]
Topic: Weather and disasters	0.786*** [0.741, 0.835]	1.627*** [1.412, 1.885]
Topic: Sports	0.638*** [0.625, 0.650]	0.367*** [0.350, 0.384]
Expected predictors of shareworthiness		
Domestic topic	1.286*** [1.269, 1.302]	1.797*** [1.741, 1.854]
Geographical distance: outside the Netherlands, but <500 km	0.760*** [0.738, 0.784]	0.545*** [0.507, 0.587]
Geographical distance: <1,000 km	0.765*** [0.744, 0.788]	0.591*** [0.551, 0.635]
Geographical distance: <2,000 km	0.794*** [0.773, 0.816]	0.729*** [0.683, 0.779]
Geographical distance: <5,000 km	0.839*** [0.817, 0.861]	0.734*** [0.687, 0.784]
Geographical distance: <10,000 km	0.829** [0.807, 0.851]	0.718*** [0.674, 0.766]
Cultural distance: Non-Western country	0.831* [0.813, 0.849]	0.690 *** [0.654, 0.728]
Distance: NA	0.771*** [0.758, 0.783]	0.640*** [0.614, 0.667]
Conflict	1.105*** [1.091, 1.119]	1.093*** [1.062, 1.126]
Human interest	1.003 [0.988, 1.018]	1.333*** [1.285, 1.383]
Positivity	1.043*** [1.037, 1.049]	1.171*** [1.153, 1.190]
Negativity	1.026*** [1.019, 1.033]	1.080*** [1.062, 1.099]
Press agency	0.666*** [0.657, 0.675]	0.276*** [0.267, 0.285]
Topic popularity score	0.745*** [0.710, 0.783]	2.443*** [2.151, 2.776]
Nagelkerke pseudo-R^2	.56	.36
Log likelihood	−422,316	−381,921
θ	1.307*** (0.006)	0.187*** (0.001)
AIC	844,695	763,907

Note. IRRs with confidence intervals in brackets. Values < 1 indicate a negative effect, values > 1 indicate a positive effect. AIC = Akaike information criterion; IRRs = incidence rate ratios.
*p < .05. **p < .01. *** p < .001.

which we interpret as a comparatively small, but still substantial effect. In contrast, **H3** is not supported at all in the case of Twitter, and it is probably safe to conclude that the presence of a human interest angle is basically irrelevant here. However, this is different on Facebook, where human interest has a strong influence, and articles may expect the number of shares to go up by a third (they are shared 1.33 as many times as nonhuman interest articles).

Turning to the tone of the article, we see that positivity (**H4**) has a stronger influence than negativity (**H5**). The hypothesis that positivity increases the number of shares is supported for both platforms: A 1-point increase on the 5-point positivity scale increases the number of shares by a factor of 1.043. This effect is much more pronounced on Facebook, where a 1-point increase of positivity results in 1.171 times as many interactions compared with an article that scores 1 point lower on the positivity scale. But when an article is written 1 point more negatively on the 5-point negativity scale, one can expect it to be shared 1.026 times as often on Twitter and 1.080 times as often on Facebook, compared with an article that scores 1 point less on the negativity scale. Taken together, this means that the effect of positivity is roughly twice as large as the effect of negativity.

The last aspect we were interested in was the effect of the exclusiveness. We asked whether written by a news agency reduces the number of shares of an article (**RQ1**), which can be interpreted as exclusive stories being more shareworthy. This seems to be the case, and the effect seems to be strong: They receive only two thirds of the Twitter shares of other articles, typically those written by journalists of the site itself. An even sharper decline is observed on Facebook, where agency-written articles can expect only 0.28 times as many interactions compared with nonagency articles. In line with this, for Twitter we find a negative effect of the topic popularity score of the topic of an article (**RQ2**). Topics that were very present in the media received less shares than topics that did not belong to the top issues. On Facebook, however, we observe an opposite effect. In combination with the bigger spread in terms of numbers of shares that we discussed when looking at the descriptive statistics, one interpretation would be that sharing on Facebook centers more around few dominant issues, whereas on Twitter there is more variation.

Summarizing the results, **RQ3** can be answered as well: The most shares will be received by an article about the own country (or at least another Western country) and not written by a news agency. Of less, but still considerable importance, is the presence of conflict, while human interest works only on Facebook (where it has a strong influence). Regarding tone, positivity works better than negativity, especially on Facebook. The results regarding the popularity of a topic are inconclusive, as the effect on Twitter and Facebook shares is opposite.

Discussion and Conclusion

In this article, we have shown that the concept of newsworthiness—once developed to explain news selection at the production stage, later also used to explain audience choices—can form a fruitful starting point to develop a concept of shareworthiness. This is of high theoretical importance for better understanding news flows in the 21st century, where the audience themselves play a role in redistributing the content by means of sharing. It has implications for the study of mass communication, journalism, and political communication, as people increasingly get their news via social media.

We investigated which characteristics of a news article predict how often the article is shared on social media. In doing so, we followed recent work by Weber (2014) and Ziegele et al. (2014), who advocated a further development of the concept of newsworthiness for engagement with online news. We found that all of our proposed list of factors that might increase shareworthiness indeed to be predictors of news sharing: geographical distance, cultural distance, negativity, positivity, conflict, human interest (not supported in the Twitter model), and exclusiveness (on Facebook, only supported for the news agency indicator, not for the topic popularity indicator). We interpret our results as encouragement to further refine the concept of shareworthiness, and as evidence that it is more than necessary to integrate and modify different theories from the age of mass media to be able to explain news consumption in today's media landscape.

Although all hypotheses were partly or fully supported, another question is how substantial the effects are. Negative tone, for instance, was much less important than the question whether an article dealt with a domestic issue or whether it was written by a news agency. This can partly be a methodological artifact, as especially the last example can be coded in a very straightforward way without any substantial measurement error, which is much more problematic for abstract concepts like negativity. In this article, it was not our aim to give exact estimates of effect sizes, but rather assess which factors can possibly contribute to the shareworthiness of an article. Further research is needed to investigate the exact role of each factor.

In our definition, we understand shareworthiness as the shareworthiness of *journalistic* content. It goes without saying that also all kinds of other content are shared online, for example, humorous content and so-called memes (Guadagno, Rempala, Murphy, & Okdie, 2013). Practitioners suggest that listicles ("The top 10 . . ."), headlines directly involving the reader ("Are you a . . ."), and, of course, cat pictures receive a lot of shares; but to our best knowledge, no systematic academic studies are available on this. Nevertheless, future research might want to integrate such a perspective and compare the influence of such "superficial" characteristics with the factors we identified. This could have important theoretical and practical implications, as it would allow to predict more accurately how news has to be presented to be shared. It would also be worth investigating in how far the predictors of shareworthiness that we identified translate to nonjournalistic content. We assume that many of our arguments are also valid in other contexts. For instance, whether a negative or a positive tone increases sharing more is also relevant in commercial contexts or in health campaigns.

One obvious drawback of our study is that we could not take user characteristics into account. It makes much sense to expect that personal interests, needs, and perceptions moderate an individual's news sharing (e.g., Bobkowski, 2015). Further research on shareworthiness can therefore profit from the collection of digital trace data on an individual level, for example, by collecting sharing data through an app that also allows surveying the users (e.g., Wells & Thorson, 2015). Another drawback is that we sometimes had to rely on rather crude measures in our automated coding, which could be refined in future research.

Our results also suggest that it might be worth studying which features of social network sites encourage which forms of engaging with the news. We showed that Facebook and Twitter are used in generally similar but sometimes also different ways. This was especially the case when looking at the role the popularity of a news topic played, but also in the fact that many effects were more pronounced on Facebook than on Twitter. Hence, the idea that communication about news and politics on social network sites can be partly explained by the different affordances of these sites (see, for example, Halpern & Gibbs, 2013) should be further explored.

All in all, our study extends our knowledge on how news spreads in an online environment. Journalists no longer have a monopoly on gatekeeping, and media users' role as gatewatcher (Bruns, 2005) becomes more important. We were able to show that this does not result in some kind of unpredictable news flow in which only people's personal interests determine whether they share a news item or not. In contrast, characteristics inherent to the news items seem to be good predictors of sharing. This extends to the topic as well: Domestic issues and, in general, "softer" topics get a bit more attention than political or economic issues—which is in line with the findings of Boczkowski and Mitchelstein (2012). But, although earlier research suggested a much stronger interest for sports news than for more relevant topics (Tewksbury, 2003), social media users seem to care much more about relevant issues than about sports—which, in fact, is good news for democratic discourse. Relatedly, we only found mixed evidence for the hypothesized preference of human interest stories. These observations are interesting regarding the frequently voiced concern that using social media would further the consumption of nonsubstantial news on the expense of hard news, which results in a lack of knowledge about current affairs and which lowers the quality of public discourse. The mixed evidence in our data suggests that the situation is not as bleak as this pessimistic view suggests. This is in line with recent findings from Sweden, where it has been shown that even if a news site provides a lot of human interest stories, these are not necessarily those the readers interact with the most (Larsson, 2016). This illustrates, though, how the functioning of democratic discourse more and more becomes an interplay of the performance of journalism and user interactions. After all, even relatively small differences in sharing behavior can lead to a spiraling process: If, as our data also show, a hardcore politic topic or a foreign topic is slightly less likely to be shared, there will be a few people less who see it as a result of the sharing, out of which slightly less reshare it, and so on. In extreme cases, this could lead to the disappearance of topics (like, for example, foreign news) from the public agenda.

To be able to analyze the prevalence of such processes and their impact on the democratic discourse, we therefore have to improve our understanding of the mechanisms, characteristics, and conditions of news sharing. This is, after all, how a growing part of the population encounters news and political information. Understanding what makes a news item shareworthy is one key element to this.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) declared receipt of the following financial support for the research, authorship, and/or publication of this article: This work was carried out on the Dutch national e-infrastructure with the support of the SURF Foundation.

Notes

1. Nevertheless, we estimated Poisson models as well, and indeed their log likelihood was significantly lower than the log likelihood of the negative binomial models: $LL_{TwitterNB} = -422{,}313$, $df = 34$; $LL_{TwitterPoisson} = -1{,}106{,}592$, $df = 33$; $\chi^2 = 1{,}368{,}558$, $p < .001$; and $LL_{FBNB} = -381{,}855$, $df = 34$; $LL_{FBPoisson} = -15{,}455{,}272$, $df = 33$; $\chi^2 = 30{,}146{,}834$, $p < .001$.
2. Area under the curve (AUC) is a commonly used evaluation method for binary choice problems (Sokolova & Lapalme, 2009). A perfect model will score an AUC of 1, whereas random guessing will score an AUC of approximately 0.5. The measure thus allows us to quantify how much better than random the classifier's choices are.
3. The F1 score is equal to the harmonic mean of recall and precision, and is a standard evaluation measure for multiclass supervised machine learning (SML) classification tasks. Due to the aggregation to 10 instead of 19 topics, the performance of the classifier is likely to be slightly better than reported.

References

An, J., Cha, M., Gummadi, K., & Crowcroft, J. (2011). Media landscape in Twitter: A world of new conventions and political diversity. In *Proceedings of the fifth international AAAI conference on Weblogs and Social Media* (pp. 18-25). Retrieved from http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewFile/2825/3283

An, J., Quercia, D., & Crowcroft, J. (2014). Partisan sharing: Facebook evidence and societal consequences. In *Proceedings of the second ACM conference on Online Social Networks* (pp. 13-24). New York, NY: ACM. doi:10.1145/2660460.2660469

Barbera, P., Jost, J. T., Nagler, J., Tucker, J. A., & Bonneau, R. (2015). Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological Science*, *26*(10), 1531-1542.

Baumgartner, F. R., Green-Pedersen, C., & Jones, B. D. (2006). Comparative studies of policy agendas. *Journal of European Public Policy*, *13*, 959-974. doi:10.1080/13501760600923805

Baym, G., & Shah, C. (2011). Circulating struggle. *Information, Communication & Society*, *14*, 1017-1038. doi:10.1080/1369118X.2011.554573

Berger, J., & Milkman, K. L. (2012). What makes online content viral? *Journal of Marketing Research*, *49*, 192-205. doi:10.1509/jmr.10.0353

Bobkowski, P. S. (2015). Sharing the news: Effects of informational utility and opinion leadership on online news sharing. *Journalism & Mass Communication Quarterly*, *92*, 320-345. doi:10.1177/1077699015573194

Boczkowski, P. J., & Mitchelstein, E. (2012). How users take advantage of different forms of interactivity on online news sites: Clicking, e-mailing, and commenting. *Human Communication Research*, *38*, 1-22. doi:10.1111/j.1468-2958.2011.01418.x

Boumans, J. W., & Trilling, D. (2016). Taking stock of the toolkit: An overview of relevant automated content analysis approaches and techniques for digital journalism scholars. *Digital Journalism*, *4*, 8-23. doi:10.1080/21670811.2015.1096598

Brants, K., & Van Praag, P. (2006). Signs of media logic: Half a century of political communication in the Netherlands. *Javnost—The Public*, *13*(1), 25-40.

Brundidge, J. (2010). Political discussion and news use in the contemporary public sphere: The "accessibility" and "traversability" of the Internet. *Javnost—The Public*, *17*(2), 63-82.

Bruns, A. (2005). *Gatewatching: Collaborative online news production*. New York, NY: Peter Lang.

Burscher, B., Odijk, D., Vliegenthart, R., de Rijke, M., & de Vreese, C. H. (2014). Teaching the computer to code frames in news: Comparing two supervised machine learning approaches to frame analysis. *Communication Methods and Measures*, *8*, 190-206. doi:10.1080/19312458.2014.937527

Burscher, B., Vliegenthart, R., & de Vreese, C. H. (2015). Using supervised machine learning to code policy issues: Can classifiers generalize across contexts? *The ANNALS of the American Academy of Political and Social Science*, *659*, 122-131.

Castillo, C., El-Haddad, M., Pfeffer, J., & Stempeck, M. (2014). Characterizing the life cycle of online news stories using social media reactions. In *Proceedings of the 17th ACM conference on Computer Supported Cooperative Work & Social Computing* (pp. 211-223). doi:10.1145/2531602.2531623

Centraal Bureau voor de Statistiek [Statistics Netherlands]. (2013). *Seven in ten internet users active on social media*. Retrieved from http://www.cbs.nl/en-GB/menu/themas/vrije-tijd-cultuur/publicaties/artikelen/archief/2013/2013-3907-wm.htm

Chaffee, S. H., & Metzger, M. J. (2001). The end of mass communication? *Mass Communication and Society*, *4*, 365-379. doi:10.1207/S15327825MCS0404_3

Cheng, J., Adamic, L., Dow, P. A., Kleinberg, J. M., & Leskovec, J. (2014). Can cascades be predicted? In *Proceedings of the 23rd international conference on World Wide Web* (pp. 925-936). New York, NY: ACM. doi:10.1145/2566486.2567997

Dominick, J. R. (1999). Who do you think you are? Personal home pages and self-presentation on the World Wide Web. *Journalism & Mass Communication Quarterly*, *76*, 646-658.

Eilders, C. (2006). News factors and news decisions. Theoretical and methodological advances in Germany. *Communications*, *31*, 5-24. doi:10.1515/COMMUN.2006.002

Fengler, S., & Russ-Mohl, S. (2008). Journalists and the information-attention markets: Towards an economic theory of journalism. *Journalism*, *9*, 667-690. doi:10.1177/1464884908096240

Galtung, J., & Ruge, M. H. (1965). The structure of foreign news. *Journal of Peace Research*, *2*, 64-91.

Gardner, W., Mulvey, E. P., & Shaw, E. C. (1995). Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. *Psychological Bulletin*, *118*, 392-404. doi:10.1037/0033-2909.118.3.392

Giddens, A. (1984). *The constitution of society: Outline of the theory of structuration*. Cambridge, UK: Polity.

Giddens, A. (1991). *Modernity and self-identity: Self and society in the late modern age*. Cambridge, UK: Polity.

GlobeFeed. (n.d.). World Distance Calculator [Web application]. Retrieved from http://distancecalculator.globefeed.com/world_distance_calculator.asp

Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, *21*, 267-297. doi:10.1093/pan/mps028

Guadagno, R. E., Rempala, D. M., Murphy, S., & Okdie, B. M. (2013). What makes a video go viral? An analysis of emotional contagion and Internet memes. *Computers in Human Behavior*, *29*, 2312-2319. doi:10.1016/j.chb.2013.04.016

Halpern, D., & Gibbs, J. (2013). Social media as a catalyst for online deliberation? Exploring the affordances of Facebook and YouTube for political expression. *Computers in Human Behavior*, *29*, 1159-1168. doi:10.1016/j.chb.2012.10.008

Harcup, T., & O'Neill, D. (2001). What is news? Galtung and Ruge revisited. *Journalism Studies*, *2*, 261-280. doi:10.1080/14616700120042114

Huntington, S. (1996). *The clash of civilizations and the remaking of world order*. New York, NY: Simon & Schuster.

International Telecommunications Union. (2012). *Percentage of individuals using the Internet 2000-2012*. Retrieved from http://www.itu.int/en/ITU-D/Statistics/Documents/statistics/2013/Individuals_Internet_2000-2012.xls

Kepplinger, H. M. (2008). *News values*. International Encyclopedia of Communication. Retrieved from http://www.communicationencyclopedia.com/subscriber/tocnode.html?id=g9781405131995_yr21

Knobloch-Westerwick, S., Sharma, N., Hansen, D. L., & Alter, S. (2005). Impact of popularity indications on readers' selective exposure to online news. *Journal of Broadcasting & Electronic Media*, *49*, 296-313. doi:10.1207/s15506878jobem4903_3

Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media? In *19th international conference on World Wide Web* (pp. 591-600). New York, NY: ACM. doi:10.1145/1772690.1772751

Larsson, A. O. (2016). I shared the news today, oh boy. *Journalism Studies*. Advance online publication. doi:10.1080/1461670X.2016.1154797

Lazarsfeld, P. F., Berelson, B., & Gaudet, H. (1944). *The people's choice: How the voter makes up his mind in a presidential campaign*. New York, NY: Columbia University Press.

Lee, C. S., & Ma, L. (2012). News sharing in social media: The effect of gratifications and prior experience. *Computers in Human Behavior*, *28*, 331-339. doi:10.1016/j.chb.2011.10.002

Lerman, K., Ghosh, R., & Rey, M. (2010). Information contagion: An empirical study of the spread of news on Digg and Twitter social networks. In *Proceedings of the fourth international AAAI conference on Weblogs and Social Media Information* (pp. 90-97). Retrieved from http://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/viewFile/1509/1839

Ma, L., Lee, C. S., & Goh, D. H. (2011). That's news to me: The influence of perceived gratifications and personal experience on news sharing in social media. In *Proceedings of the 11th annual international ACM/IEEE joint conference on Digital Libraries* (pp. 141-144). doi:10.1145/1998076.1998103

Ma, L., Lee, C. S., & Goh, D. H. (2014). Understanding news sharing in social media: An explanation from the diffusion of innovations theory. *Online Information Review*, *38*, 598-615. doi:10.1108/OIR-10-2013-0239

Morgan, J. S., Shafiq, M. Z., & Lampe, C. (2013). Is news sharing on Twitter ideologically biased? In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work* (pp. 887-897). New York, NY: ACM.

O'Neill, D., & Harcup, T. (2009). News values and selectivity. In K. Wahl-Jorgensen & T. Hanitzsch (Eds.), *The handbook of journalism studies* (pp. 161-174). New York, NY: Routledge.

Östgaard, E. (1965). Factors influencing the flow of news. *Journal of Peace Research*, *2*, 39-63.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825-2830.

Pempek, T. A., Yermolayeva, Y. A., & Calvert, S. L. (2009). College students' social networking experiences on Facebook. *Journal of Applied Developmental Psychology*, *30*, 227-238. doi:10.1016/j.appdev.2008.12.010

Russell, S. J., & Norvig, P. (2002). *Artificial intelligence: A modern approach* (3rd international ed.). Harlow, UK: Pearson.

Salganik, M. J., Dodds, P. S., & Watts, D. J. (2006). Experimental study of inequality and cultural market. *Science*, *311*, 854-856.

Saxton, G. D., & Waters, R. D. (2014). What do stakeholders like on Facebook? Examining public reactions to nonprofit organizations' informational, promotional, and community-building messages. *Journal of Public Relations Research*, *26*, 280-299. doi:10.1080/1062726X.2014.908721

Schmidt, J.-H. (2014). Twitter and the rise of personal publics. In K. Weller, A. Bruns, J. Burgess, M. Mahrt, & C. Puschmann (Eds.), *Twitter and society* (pp. 3-14). New York, NY: Peter Lang.

Schmitt-Beck, R. (2003). Mass communication, personal communication and vote choice: The filter hypothesis of media influence in comparative perspective. *British Journal of Political Science*, *33*, 233-259. doi:10.1017/S0007123403000103

Semetko, H., & Valkenburg, P. (2000). Framing European politics: A content analysis of press and television news. *Journal of Communication*, *50*(2), 93-109. doi:10.1111/j.1460-2466.2000.tb02843.x

Shoemaker, P. J., & Cohen, A. (2006). *News around the world: Content, practitioners, and the public*. New York, NY: Routledge.

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, *45*, 427-437. doi:10.1016/j.ipm.2009.03.002

Stieglitz, S., & Dang-Xuan, L. (2012). Political communication and influence through micro-blogging—An empirical analysis of sentiment in Twitter messages and retweet behavior. In *45th Hawaii International Conference on System Sciences* (pp. 3500-3509). doi:10.1109/HICSS.2012.476

Tewksbury, D. (2003). What do Americans really want to know? Tracking the behavior of news readers on the Internet. *Journal of Communication*, *53*, 694-710. doi:10.1093/joc/53.4.694

Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., & Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, *61*, 2544-2558. doi:10.1002/asi.21416

Trilling, D. (2014). Weg vom manuellen Speichern: Automatisierte Datenerhebung bei Onlinemedien [Beyond saving manually: Automated data collection for online media]. In K. Sommer, M. Wettstein, W. Wirth, & J. Matthes (Eds.), *Automatisierung in der Inhaltsanalyse* (pp. 73-89). Cologne, Germany: Herbert von Halem.

Trilling, D. (2015). Two different debates? Investigating the relationship between a political debate on TV and simultaneous comments on Twitter. *Social Science Computer Review*, *33*, 259-276. doi:10.1177/0894439314537886

Trussler, M., & Soroka, S. (2014). Consumer demand for cynical and negative news frames. *The International Journal of Press/Politics*, *19*, 360-379. doi:10.1177/1940161214524832

Turkle, S. (1995). *Life on the screen: Identity in the age of the Internet*. New York, NY: Simon & Schuster.

Van Dalen, A. (2012). Structural bias in cross-national perspective: How political systems and journalism cultures influence government dominance in the news. *The International Journal of Press/Politics*, *17*, 32-55.

Van den Bosch, A., Busser, B., Canisius, S., & Daelemans, W. (2007). An efficient memory-based morphosyntactic tagger and parser for Dutch. *LOT Occasional Series*, *7*, 191-206.

Van Dijk, J. A. (2006). *The network society: Social aspects of new media*. Thousand Oaks, CA: SAGE.

Weber, P. (2014). Discussions in the comments section: Factors influencing participation and interactivity in online newspapers' reader comments. *New Media & Society*, *16*, 941-957. doi:10.1177/1461444813495165

Weeks, B. E., & Holbert, R. L. (2013). Predicting dissemination of news content in social media: A focus on reception, friending, and partisanship. *Journalism & Mass Communication Quarterly*, *90*, 212-232. doi:10.1177/1077699013482906

Wells, C., & Thorson, K. (2015). Combining Big Data and survey techniques to model effects of political content flows in Facebook. *Social Science Computer Review*. Advance online publication. doi:10.1177/0894439315609528

Wu, S., Hofman, J. M., Mason, W. A., & Watts, D. J. (2011). Who says what to whom on Twitter. In *Proceedings of the 20th international conference on World Wide Web* (pp. 705-714). doi:10.1145/1963405.1963504

Yang, J., & Counts, S. (2010). Predicting the speed, scale, and range of information diffusion in Twitter. In *Proceedings of the fourth international AAAI conference on Weblogs and Social Media Predicting* (pp. 355-358). Retrieved from http://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/viewFile/1468@misc/1896

Zhao, S., Grasmuck, S., & Martin, J. (2008). Identity construction on Facebook: Digital empowerment in anchored relationships. *Computers in Human Behavior*, *24*, 1816-1836. doi:10.1016/j.chb.2008.02.012

Ziegele, M., Breiner, T., & Quiring, O. (2014). What creates interactivity in online news discussions? An exploratory analysis of discussion factors in user comments on news items. *Journal of Communication*, *64*, 1111-1138. doi:10.1111/jcom.12123

Author Biographies

Damian Trilling, PhD, is an assistant professor of political communication and journalism at the Department of Communication Science, University of Amsterdam (The Netherlands), where he also received his PhD. He is affiliated with the Amsterdam School of Communication Research. He is interested in questions such as "How does the changing media landscape change political communication and journalism?"; "How do citizens, politicians, and journalists make use of new tools?"; and "What is the impact of these changes?" He is also interested in methodological innovations to study communication in an age where more and more communication is happening digitally. This includes methods of automated content analysis and computational social science approaches in general.

Petro Tolochko, MSc, is a PhD candidate at the Department of Methods in the Social Sciences at the University of Vienna, Austria. He received his MSc at the University of Amsterdam. His research interests include quantitative methods in the social sciences, social networks, quantitative text analysis, and automated language processing in the social sciences.

Björn Burscher, PhD, is a search engineer at Textkernel. He received his PhD from the Department of Communication Science, University of Amsterdam (The Netherlands). His research focuses on the automatization of content analysis for communication research. To this end, he works on machine-learning-based methods for identifying and coding topics and media frames in news coverage.