

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/304670935>

A Gatekeeper among Gatekeepers: News agency influence in print and online newspapers in the Netherlands

Article in *Journalism Studies* · June 2016

Impact Factor: 0.54 · DOI: 10.1080/1461670X.2016.1190663

READS

14

4 authors:



[Kasper Welbers](#)

VU University Amsterdam

5 PUBLICATIONS 34 CITATIONS

SEE PROFILE



[Wouter Van Atteveldt](#)

VU University Amsterdam

36 PUBLICATIONS 305 CITATIONS

SEE PROFILE



[Jan Kleinnijenhuis](#)

VU University Amsterdam

62 PUBLICATIONS 776 CITATIONS

SEE PROFILE



[Nel Ruigrok](#)

Erasmus University Rotterdam

28 PUBLICATIONS 153 CITATIONS

SEE PROFILE

A gatekeeper among gatekeepers: The impact of a single news agency on political news in print and online newspapers in the Netherlands.

Kasper Welbers, Wouter van Atteveldt & Jan Kleinnijenhuis
VU University Amsterdam, the Netherlands

Nel Ruigrok
Erasmus University Rotterdam, the Netherlands

Abstract

This paper investigates the influence of news agency ANP on the coverage and diversity of political news in Dutch national newspapers. Using computational text analysis, we analyzed the influence on print newspapers across three years (1996, 2008 and 2013) and compared influence on print and online newspapers in 2013. Results indicate that the influence of ANP on print newspapers only increased slightly. Online newspapers, however, depend heavily on ANP, which causes their content to be highly similar. We draw conclusions pertaining to the gatekeeping role of news agencies in the digital age in general, and in the context of the Netherlands in particular. Additionally, we demonstrate that techniques from the field of information retrieval can be used to perform these analyses on a large scale. Our scripts and instructions are provided online to stimulate the use of these techniques in communication studies.

This is the pre-print version of the manuscript published by Taylor & Francis in *Journalism Studies* (online first, 30 June 2016). The published manuscript is available online from <http://www.tandfonline.com/doi/full/10.1080/1461670X.2016.1190663#.V3o0wUK1i1M>

Introduction

There is a conundrum regarding the diversity of news in the digital age: in spite of the larger variety of news publishers, there does not appear to be a larger variety of news (Boczkowski and De Santos, 2007). One of the explanations for this phenomenon is that a large part of the news across a wide range of news publishers can be traced back to the same news agencies, or wire services. These news agencies thereby act as powerful *gatekeepers*: their choices in gathering, filtering and shaping news messages affect the information input of many news publishers, and thereby indirectly have a substantial impact on the information input of citizens (Shoemaker and Vos, 2009).

As professional news brokers, news agencies offer news publishers a relatively cheap, reliable and fast supply of information. Although this can be a boon for the availability of affordable news content, it also restricts the diversity of news in society to the gatekeeping choices of one or several news agencies. This diversity is crucial in a democratic society and a core element in communications policy, because it helps distinguish facts from falsehoods, and ensures that the diversity of viewpoints in society is represented ([Van Cuilenburg, 2007](#); [Napoli, 1999](#)). Accordingly, the prominent role of news agencies in the contemporary media landscape “raises issues for news diversity and free speech” ([Johnston and Forde, 2011](#)).

In this paper we investigate whether the gatekeeping influence of news agencies has increased in the digital age. Specifically, we analyze to what extent the print and online editions of five Dutch national newspapers rely on *ANP*, which is the largest and currently only national news agency in the Netherlands. We herein focus on political news coverage, for which the diversity of information has the most direct implications for the democratic process. Our analysis consists of three parts.

First, we analyze whether *ANP*'s influence on print newspapers has increased over time. Studies suggest that newspapers are becoming more dependent on news agencies ([Lewis et al., 2008](#); [Paterson, 2005](#)) which has mainly been attributed to economic cutbacks in newspapers ([Frijters and Velamuri, 2010](#)). However, more research is needed on the influence of individual news agencies within specific national contexts (see e.g., [Johnston and Forde, 2009, 2011](#)). In the Netherlands, we observe that *ANP* is also struggling with economic cutbacks, and has difficulties offering its clients the exclusivity of information they desire ([Reijnders, 2009](#); [Van Vulpen, 2010](#)). Thus, it cannot be taken for granted that its influence has increased. To test this, we analyze its influence on the print newspapers in three disjoint years (1996, 2008 & 2013) covering two periods of economic turmoil: the economic recession and financial crisis that emerged in 2000 and 2008, respectively.

Second, we compare the influence of *ANP* on the print and online editions of newspapers in the first half of 2013. Online news publishers seem to rely more on news agencies, mainly because of the difficulty of making profit from online news and the quick paced 24/7 news cycle ([Klinenberg, 2005](#); [Johnston and Forde, 2011](#)). We could not find studies that compared print and online newspapers, which is an interesting comparison because it provides ‘a window into how new journalistic forms emerge in the context of existing ones’ ([Boczkowski and De Santos, 2007, 167-168](#)). Furthermore, we found no figures on just how quick paced this news cycle is. We address this gap by measuring the average time for an online newspaper to adopt a news agency article. If this time is indeed short, it puts additional weight to the argument that time pressure is an important reason for online newspapers to rely on news agencies. In addition, this information is relevant for studies that aim to model intermedia dynamics.

Finally, we investigate how the shared dependence of Dutch national newspapers on *ANP* affects the diversity of news content. This consequence is often assumed, but not analyzed empirically. Whether this is the case depends on the range of information offered by *ANP*, the extent to which newspapers rely on *ANP*, and the news selection choices of the newspapers. We measure how often different newspapers are influenced by the same *ANP* articles.

We performed a computational text analysis, based on techniques from the fields of information retrieval (IR) and natural language processing (NLP). This allows us to measure

the cross-time similarity of news content across news organizations at the level of events and identical phrases. The use of IR and NLP techniques to measure document similarity is well established ([Salton and Harman, 2003](#); [Bagga and Baldwin, 1998](#)), and similar techniques have on several occasions been used in communication research ([Landauer and Dumais, 1997](#); [Van Atteveldt, 2008](#), e.g.). Yet, the use of document similarity measures appears to have received little attention in communication studies as a tool to measure interactions between news organizations. We discuss the advantages of this approach compared to other approaches, and facilitate its application in communication studies by providing scripts and instructions to apply it using the open-source statistical package R.

The gatekeeping role of news agencies

The influence of news agencies on society can be conceptualized in terms of gatekeeping, which Shoemaker and Vos define broadly as “the process of culling and crafting countless bits of information into the limited number of messages that reach people each day” (2009, 1). The people that perform this culling and crafting are referred to as gatekeepers. By controlling society’s supply of news, gatekeepers have a strong influence on society’s perception of relevant developments and the interpretation of these developments.

The more news publishers rely on a news agency as a source of news, the more influence the news agency has as a gatekeeper ([McNelly, 1959](#)). News publishers might filter, re-interpret and add new elements to the messages they obtain from news agencies, but the news agencies largely determine the agenda. This raises several concerns. For one, it has been argued that this can harm the quality of news, because journalists often blindly rely on the facts presented in news agency messages, but news agencies do not always uphold journalistic standards for checking sources ([Davies, 2008](#); [Forde and Johnston, 2013](#)). Another concern is that the shared reliance of newspapers on the same news agency can harm the diversity of news content. The importance of diversity in the news has been well established in the communication literature, and is a core element in communications policy ([Van Cuilenburg, 2007](#); [Napoli, 1999](#)). In the Netherlands, where one news agency is dominant and used by almost all major newspapers, diversity could indeed be in peril.

Newspapers and news agencies

In 2008, [Davies \(2008\)](#) released the book *Flat Earth News*, in which he painted an alarming picture of the increased reliance of UK newspapers on news agencies. Based on a study by [Lewis et al. \(2008\)](#), he claimed that no less than 70% of the news stories in the five most prestigious Fleet Street titles (i.e. popular London newspapers) were direct copies (30%) or rewrites (19%) of news agency articles, or at least contained elements (21%) from them ([Davies, 2008, 74](#)). The concern that newspapers have become too dependent on news agencies is also shared in other countries, for instance in Australia ([Johnston and Forde, 2009, 2011](#)) and the Netherlands ([Scholten and Ruigrok, 2009](#)).

The increased reliance on news agency copy has mainly been attributed to economic cutbacks ([Frijters and Velamuri, 2010](#)). Journalists are pressured to spend less time gathering and investigating information, and instead just “take it off the wires and knock it into shape” (journalist quoted in [Davies, 2008, 75](#)). A survey of UK journalists showed that journalists indeed ‘felt that the pressure to produce a high number of stories daily has

intensified, and that this increased their reliance on recycling material rather than reporting independently’ Lewis et al. (2008, 4).

The reliance on news agencies appears to be even stronger for online editions of newspapers (Johnston and Forde, 2009), which is alarming given the increasing popularity of online news consumption. The literature offers two main explanations for this difference. Firstly, economic constraints are higher for online newspapers due to the difficulty of making money from online news. Many users are not willing to pay for online content (Chyi, 2005). Online newspapers therefore often rely solely on advertising, believing that ‘the revenue they could gain from content charging would be less than what they would lose in advertising’ (Herbert and Thurman, 2007, 213). Some newspapers have experimented with paywalls, but this was often not a viable business model (Arrese, 2015)—through recently there have been more successful cases such as the New York Times (see e.g., Cook and Attari, 2012). Given these economic difficulties, the reliance on news agency copy is likely to be high to reduce expenses.

Secondly, the influence of news agencies on online newspapers is boosted by the speed of the online news cycle. Online news can be published 24/7, which has created ‘an informational environment in which there is always breaking news to produce, consume, and—for reporters and their subjects—react against’ (Klinenberg, 2005, 54). Johnston and Forde (2011) argue that this acceleration leads to ‘an even greater reliance on news agency copy than perhaps at any other time in news media history’ (Johnston and Forde, 2011, 195-196).

News agencies in the Netherlands

In this paper we focus on a single news agency, the *Algemeen Nederlands Persbureau* (ANP). ANP was founded in 1934 by the Association of the Dutch Daily Press (De Nederlandse Dagbladpers) as a joint effort of the national newspapers to create a quick and independent source of news facts. Since it became a private limited company in 2001, newspaper publishers gradually pulled out. Since 2010 ANP is owned by the investment company V-Ventures. (Rutten and Slot, 2011)

The disjunction of ANP and the newspaper publishers opened up the market for new competition. In 2001 the news agency *Novum* was founded. Together with *GPD*—which was founded in 1936, and mainly provided news for regional newspapers—there were now three national news agencies. This competition eventually proved fatal. *GPD* ended its long history of service in 2013 after it lost an important client. *Novum* was taken over by ANP in 2014.

This shows that, even if newspapers have become more dependent on external news gatherers, the digital age is certainly not a golden age for news agencies. One of the problems is that digital technology has made it much easier for news publishers to monitor and use news agency content without paying for it. Rutten and Slot (2011) report that an interviewee from a news organization estimated that with the use of digital tools such as webcrawlers and rss-feeds about 90% of the ANP news could also be tracked without an ANP subscription. Copyright law provides limited protection against this indirect use of news agency content due to the *press exception*—an exception in Dutch copyright law that allows news organizations to use each others news, at least in terms of bare facts (Guibault,

2012). This greatly harms the value of news agency subscriptions, which depend on the exclusivity of information.

Together with competition from *Novum*, this caused significant economic cutbacks for *ANP*, due to which a large part of the workforce was fired after 2009 (Rutten and Slot, 2011; Ebisch, 2012). Despite these developments, *ANP* remained to be the largest news agency in the Netherlands during our study, and is currently the only national news agency. Except for *NRC Handelsblad* in 2013, all national newspapers subscribed to *ANP* during the years analyzed in this paper.

The influence of *ANP* on Dutch newspapers

Studies that looked for traces of news agency copy in Dutch national newspapers confirmed that news agencies are an important source of information (Heijmans et al., 2009; Scholten and Ruigrok, 2009). Scholten and Ruigrok (2009) focused specifically on *ANP*, and found that for nine prominent Dutch national newspapers in 2008, on average 27.6% of the articles were copies or rewrites of *ANP* content.

Scholten and Ruigrok (2009) also found some evidence that the influence of *ANP* increased between 2006 and 2008. Other than that, there have been no longitudinal studies that compare the influence of *ANP* over time. Based on the theory that the influence of news agencies has increased due to economical cutbacks, we expect that this influence has increased more over a longer period of time. Newspaper companies in the Netherlands experienced two substantial economic cutbacks during the last two decades (Bakker and Scholten, 2011). One is the economic recession that started in 2000, and the other is the financial crisis of 2008. Both sales and advertising incomes decreased, and a large part of the workforce was fired, especially between 2008 and 2011. To analyze the impact of these developments, we compare the influence of *ANP* on print newspapers across these periods, focusing on three years: 1996, 2008 and 2013.

H1: The influence of *ANP* on political news in the Dutch national print newspapers increased between 1998 and 2013.

Prior studies did not measure the influence of *ANP* on the online editions of newspapers. Also, while the literature suggests that the influence of news agencies is higher for online news (Johnston and Forde, 2009; Paterson, 2005), this has not yet been compared for the print and online editions of the same newspaper. This comparison is interesting because it focuses our lens on the difference between print and online news production, with relatively little interference from other differences.

As discussed, economic constraints are often higher for online editions of newspapers, due to the problem of generating profit from online news (Chyi, 2005; Mitchelstein and Boczkowski, 2009). In the Netherlands this is the case as well. In 2012, Christian van Thillo, CEO of media company *De Persgroep*, stated that the free model—generating income through advertising only—does not work (Van Soest, 2012). Despite announcements of experiments with paywalls (Van der Laan, 2013), no lasting solutions appear to have been found, and during the period in which we analyzed the online newspapers (the first half of 2013) the free model was still used. This struggle to make online newspapers profitable, together with the theory that reliance on news agency copy is higher due to the 24/7 online news cycle (Johnston and Forde, 2011), leads to the following hypothesis.

H2: The influence of *ANP* on political news in online editions of Dutch national newspapers is stronger than the influence on the print editions.

To look closer into the influence of the 24/7 news cycle, we also measure the time it takes for online newspapers to respond to *ANP* publications. We did not find prior research that measured this, but the theory suggests that online newspapers will copy news agency items as quickly as possible. We pose the following question.

RQ1: What is the average time for online newspapers to adopt an *ANP* article?

If news publishers are influenced by the same news agency, this potentially affects the diversity of news content. Whether this is the case depends on the amount of news supplied by the news agency, the amount of news required by the newspaper, and the newspaper's news selection criteria. Put simply, if the news agency publishes sufficient news items, and newspapers select different items to cover, then their shared reliance on the news agency does not affect the diversity of news. We thus pose the following question.

RQ2: To what extent are newspapers influenced by the same *ANP* articles?

Data

We collected the news articles of the print and online editions of five national newspapers from the Netherlands: *De Telegraaf*, *Algemeen Dagblad*, *De Volkskrant*, *Trouw* and *NRC Handelsblad*. Print and online news consumption are high in the Netherlands compared to other countries ([Bakker and Scholten, 2011](#); Eurobarometer, 2012). A representative sample of the market research agency GfK in July 2012 ($n = 1805$) shows that 73.7% attended to at least one print newspaper over the course of one week, and 46.4% attended to one of the print newspapers in this study. In the same week, 57.6% visited a news website at least once, and 31.9% visited the online editions of the newspapers in this study.

For the news agency (*ANP*) and the print newspapers we gathered all articles for three disjoint years: 1996, 2008 and 2013. For the online newspapers we gathered all articles for the first half of 2013. For the print edition of *De Telegraaf* the data for 1996 was unavailable. Therefore, we only used *De Telegraaf* for the comparison of print and online news in 2013. The set of news agency articles also contained different versions of the same article due to updates. We filtered out all updates, since we are interested in the time at which an event is first covered.

To focus the analysis on political news, we used a search query to select only news articles that mention Dutch political parties. We also included *ANP* articles that did not match the search query, but that addressed the same event as a newspaper article that did. We found that when *ANP* articles mention a minister or undersecretary the political party is often not mentioned, but in the newspaper this is added. Also, newspapers often added quotes or statements from politicians to add a political context to event coverage.

In total 848,479 news articles were collected, of which 59,687 were selected as political news and used in the analysis. The average number of articles per medium per period is presented in Table 1. There is a notable increase in *ANP* articles between 1996 and 2008. This is consistent with Rutten and Slot (2011), who mentioned that in the 5 years leading

Table 1
Average number of news articles per day

	All news			Political news		
	1996	2008	2013	1996	2008	2013
News agency						
ANP	164.2	246.2	202.9	6.8	14.4	10.9
Print newspapers						
Algemeen Dagblad	152.1	133.7	141.9	8.1	6.2	7.6
De Volkskrant	109.3	121.3	115.7	9.2	11.8	8.8
Trouw	105.9	113.8	92.4	10.2	10.1	9.5
De Telegraaf			184			11.7
NRC Handelsblad	118.2	111.6	100.6	9.0	14.5	13.1
Online newspapers						
Algemeen Dagblad			219.8			9.9
De Volkskrant			179.4			12.6
Trouw			123.5			11.4
De Telegraaf			193.5			10.2
NRC Handelsblad			29.6			4.4

Note. Online newspapers based on first half of 2013 only

up to 2011 *ANP* produced about 40% more output than before. The decrease between 2008 and 2013 is likely related to economic cutbacks (Ebisch, 2012; Rutten and Slot, 2011). Also notable is the low amount of news article on the website of *NRC Handelsblad*. Unlike most online newspapers, *NRC Handelsblad* mainly provides longer background stories instead of short news updates.

Measuring news agency influence

Ideally, one would be able to learn about the influence of a news agency on a newspaper article from explicit source references, such as author information or hyperlinks (see e.g., Johnston and Forde, 2009; Meraz, 2009). The problem is that this information is often inaccurate. Lewis et al. (2008) studied 5 UK newspapers, and found that while only 1 percent of news articles mention news agencies as a source, more than half of the news could actually be traced back to news agencies, with about 30% being near exact copies. Print newspapers in particular are often reluctant to mention news agencies, presumably because this ‘dilutes the authority of a newspaper’ (Matheson, 2004, 458). In the Netherlands, the number of articles in print newspapers that can be traced back to a news agency also appears to be higher than the number of articles that explicitly refers to *ANP* (Scholten and Ruijgrok, 2009).

Meraz (2011) used time-series analysis as an alternative to hyperlink analysis as a method to analyze the intermedia influence of blogs and newspapers (also see [Hollanders and Vliegthart, 2008](#); [Vliegthart and Walgrave, 2008](#)). Influence is then measured in terms of Granger (1988) causality, or predictive causality; as the extent to which the attention for a news item in one medium can be predicted based on the recent attention for

this item in another medium.

Similarly, to analyze to what extent newspapers adopt stories from news agencies, we want to measure whether newspaper coverage of specific events can be traced back to prior news agency coverage of these events. This introduces two complications. The first is content-analysis. To code all news items at the level of events would require an enormous effort. All news items would have to be coded inductively, and coders would need to be able to distinguish a huge amount of codes (one for each event). The second complication is that this data cannot be analyzed with common time-series models. Time-series analysis requires repeated measurements over time¹, but media attention for specific events generally only lasts one or a few days.

We therefore use an alternative approach. Using a document similarity measure, we measure for each newspaper article how similar it is to recent news agency articles (the similarity measure is discussed in the next section). This type of approach was also used by Scholten and Ruigrok (2009) and Paterson (2005), who measured the similarity of news articles as the percentage of overlapping word n-grams (i.e. sets of n consecutive words). Influence is then measured based on the extent to which newspapers contain literal copy from news agencies, which is akin to scanning for plagiarism.

An important difference of our approach is that we used a different similarity measure, to account for the fact that news agency influence does not necessarily result in literal copy. To measure influence on a more subtle level, we calculated the similarity of news articles based on the most distinguishing nouns and proper names in the lead of the article. The resulting similarity score indicates whether articles address the same event. We use this to measure the influence of *ANP* on a newspaper as the proportion of newspaper articles for which the event was earlier covered by *ANP*.

A general limitation of content-analysis based approaches for measuring influence in news diffusion is that content similarities can also result from journalists using the same sources. Thus, even if traces of news agency copy are found in an article, this does not prove that the article would not have been published if the news agency rejected it. This is particularly the case for press releases and public relations material, which journalists can often easily obtain without relying on a news agency.

Nevertheless, previous studies show that journalists do often rely on news agencies also for press releases and public relations material (Forde and Johnston, 2013; Lewis et al., 2008). Even if a news agency is not the only possible source of certain information for a journalists, they do make this information more accessible and lend legitimacy to it (Forde and Johnston, 2013). Accordingly, notwithstanding the aforementioned limitation, traces of news agency content in newspaper articles provide useful insight in the gatekeeping influence of news agencies, as also demonstrated below in our validity tests.

Measuring document similarity

To measure the similarity of documents we use the vector space model (VSM) approach (Salton et al., 1975; Salton and Harman, 2003). The first step of this approach is to decide which elements of documents are used to represent them as vectors. We used a bag-

¹Meraz, for instance, restricted her analysis to “hot-button” events that generated “notable discussion across all blogs and media” (2011, 181).

of-words approach, which means that we only look at word-occurrence. More specifically, we only look at the nouns and proper names, which we extracted using a natural language processing techniques called part-of-speech tagging². Proper names refer to unique, named entities, such as specific people, organizations and locations, and thus contain much information to distinguish events in news content. We also used normal nouns, because news articles also often describe events using unnamed actors and things, such as in the sentence: "a young man stole a bicycle". To ignore different word forms (e.g., singular versus plural) we used lemmatization to reduce words to their morphological root form.

To focus on the main event of a news article, we only used the headline and first 5 sentences, based on the domain knowledge that newspaper articles generally have an inverted pyramid structure—the who, what and where are immediately introduced ([Knobloch et al., 2004](#)). We did not delete low-frequency words because these can be very informative about specific events. Of the high-frequency words, that occurred in more than 1% of all articles, we manually deleted words that were not informative about events, such as temporal location expressions (e.g., yesterday) and author information.

Next, we weight the vectors. Turney et al. explains that “The idea of weighting is to give more weight to surprising events and less weight to expected events”, which is important because “surprising events, if shared by two vectors, are more discriminative of the similarity between the vectors than less surprising events” (2010, 156). Thus, we want to give more weight to rare words than common words. For this we use the term-frequency inverse document frequency (tf.idf), which is a classic weighting scheme and recommended standard in information retrieval ([Sparck Jones, 1972](#); [Monroe et al., 2008](#)).

The similarity of documents can now be measured based on how close they are together in the vector space. A common measure used in information retrieval is the cosine of the angle between vectors. Since there are no negative values in our document vectors, the cosine similarity measure ranges from 0 (zero similarity) to 1 (identical).

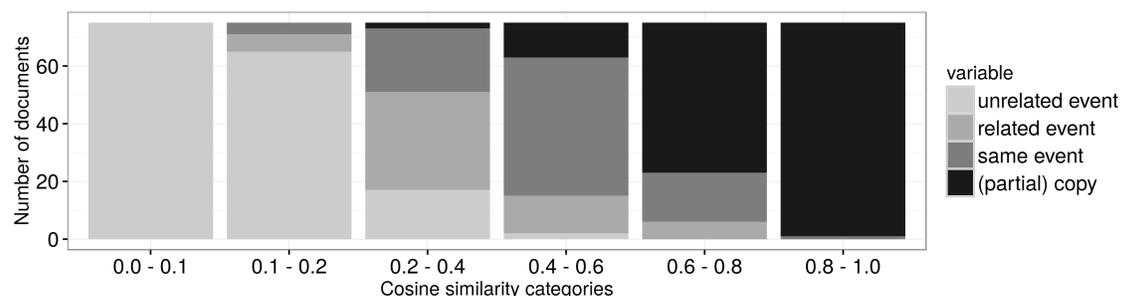
Method of analysis

To determine whether a newspaper article can be traced back to a news agency article, we compared each newspaper article to all news agency articles that were published within 2 days before the newspaper article. If this similarity score is higher than a certain threshold (determined in the validity section) then the news agency article is considered an influence.

The two day time-frame was used because we assume that if a newspaper is influenced by a news agency article this happens on a short term—which we also demonstrate for online news. For print newspapers, we took into account that the *ANP* article had to be published before the newspaper is pressed, which is midnight for most newspapers, and in the morning for the afternoon newspaper *NRC Handelsblad*. For online newspapers, we did not impose a similar publication delay, because we also found exact copies of *ANP* articles that were published simultaneously by *ANP* and the online newspaper. Furthermore, we subtracted one hour from the *ANP* publication time because we found that some articles that are

²For the natural language processing techniques used in this paper we used the *Frog* software ([Van den Bosch et al., 2007](#)), a free to use memory-based morphosyntactic tagger and parser for the Dutch language. Similar software is also freely available for other languages, such as CoreNLP for English ([Manning et al., 2014](#)).

Figure 1. Document similarity scores versus manual codings of document similarity.



certainly *ANP* copies—they were identical, and some also credited *ANP*—were published before the *ANP* publication time in our data³.

Validity and similarity threshold

To determine whether documents address the same event, we need to decide on a threshold for the similarity score. There is no logical default for this threshold, and the most useful threshold—that is best at measuring what we want to measure—will differ depending on the data and research question. To determine the most useful threshold for our study, we performed two tests to measure the validity of our similarity measure at various thresholds.

By validity we mean the extent to which the results of the computational approach correspond to a gold standard (i.e. results that are assumed to be correct). For the first gold standard, we drew 6 samples of 75 pairs of newspaper and *ANP* documents with different levels of similarity. These document pairs were manually coded by a coder who did not see the similarity scores. The coder had to select from one of three options: the documents address unrelated events, different but related events, or the same event. If documents address the same event, the coder also coded whether the documents are (partial) copies, which was assisted by highlighting identical seven-word phrases. Although for this study we are not interested in articles with different but related events, we added this category for additional insight in the performance of the similarity measure.

The results are presented in Figure 1. We see that document pairs with similarity scores above 0.4 very often address the same event, and above 0.6 are often (partial) copies. Similarity scores below 0.2 generally indicate that documents address different events, and between 0.2 and 0.4 the results are more ambiguous, with many documents addressing different but related events.

To determine the most suitable threshold, we calculated the precision and recall scores at different levels of similarity. Precision is the proportion of pairs with a similarity score above the threshold that are actually similar (based on the gold standard). Recall is the proportion of actually similar pairs that have a similarity score above the threshold. The performance of the computational approach is only good if both scores are high, which can be measured as their harmonic mean, called the F1 score. These results are presented in

³Based on the validity tests using explicit source references in Trouw as a gold standard, we verified that subtracting one hour from the *ANP* publication time increases the recall.

Figure 2. Precision, Recall, F1 and Cohen’s Kappa for two gold standards.

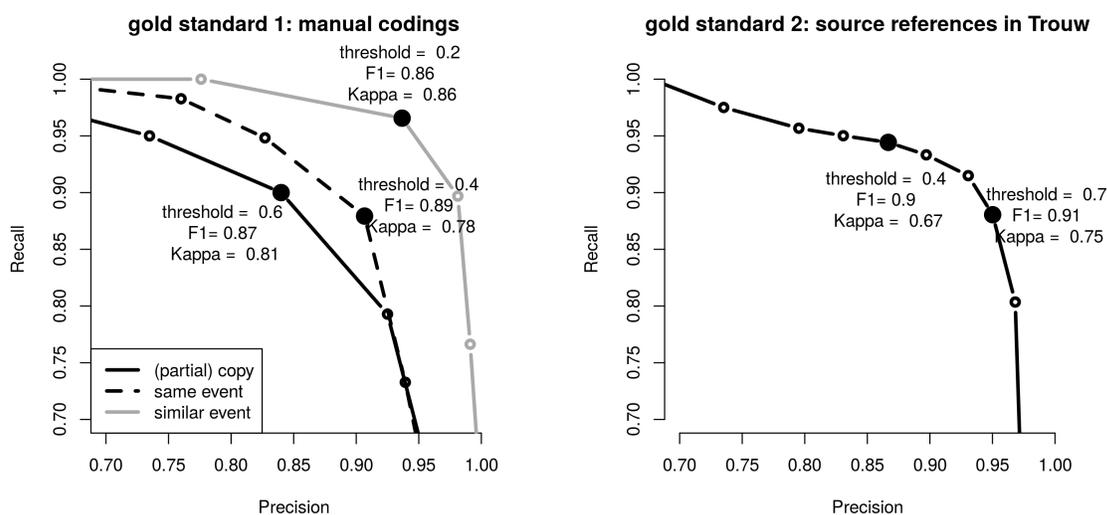


Figure 2. For events the F1 score is highest at a threshold of 0.4 (F1 = 0.89, Kappa = 0.78). For reference, we added Cohen’s Kappa, which is a common inter-coder reliability measure⁴. Both values indicate that the computational measurement of events is good. The measurement of (partial) copy is also good at a threshold of 0.7 (F1 = 0.87, Kappa = 0.82). For reference, a threshold of 0.2 would provide a good measurement of whether documents at least address related events (F1 = 0.86, F1 = 0.85).

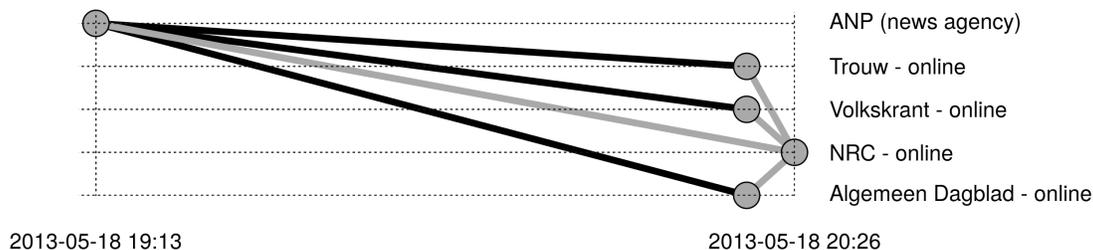
The first gold standard shows how the computational similarity score relates to a human interpretation of similarity, but does not show how good this enables us to measure whether a newspaper article is actually based on an ANP article. To test this, we used the data for the online newspaper *Trouw*. *Trouw* appears to be reliably consistent in crediting sources, and 66% of their articles explicitly credits *ANP*. Thus, we use this explicit reference to *ANP* as our second gold standard.

The results are presented in the right graph of Figure 2. Here we see that the results are good when using a threshold of 0.4 (F1 = 0.90, Kappa = 0.67) and even better when using a threshold of 0.7 (F1 = 0.91, Kappa = 0.75). This indicates that *Trouw* articles that credit *ANP* as a source are often (partial) copies. Interestingly, precision is clearly lower when using a threshold of 0.4, which indicates that there are quite some *Trouw* articles that address an event that was earlier covered by *ANP*, but that does not explicitly credit *ANP* as a source. As discussed in the section on measuring news agency influence, it is still possible that in these cases *ANP* was a source, or at least a factor in the news selection process. We found some indication of this: *Trouw* often covered these articles very shortly after they were published by *ANP*, as we show in the results section. In addition, we also found some articles that were (partial) copies of *ANP* articles but did not credit *ANP*, indicating that source references in *Trouw* are not 100% reliable—which also means that the precision of our measurement is in truth higher.

In summary, the first validity test verifies that the similarity score is a valid measure-

⁴The Kappa and the F1 scores across thresholds were highly correlated (Pearson correlation = 0.95).

Figure 3. News articles per organization (y-axis) over time (x-axis) that address the same event (grey lines) or are partially identical (black lines)



ment of whether a newspaper article contains an event or textual passages that previously occurred in a news agency article. The second test verifies that this measurement corresponds to the actual use of *ANP* as a source. The 0.4 threshold appears to be the best measure for whether articles address the same event, and the 0.7 threshold indicates that one article is the (partial) copy of the other. We report our results using both thresholds, as two complementing measures of influence.

Results

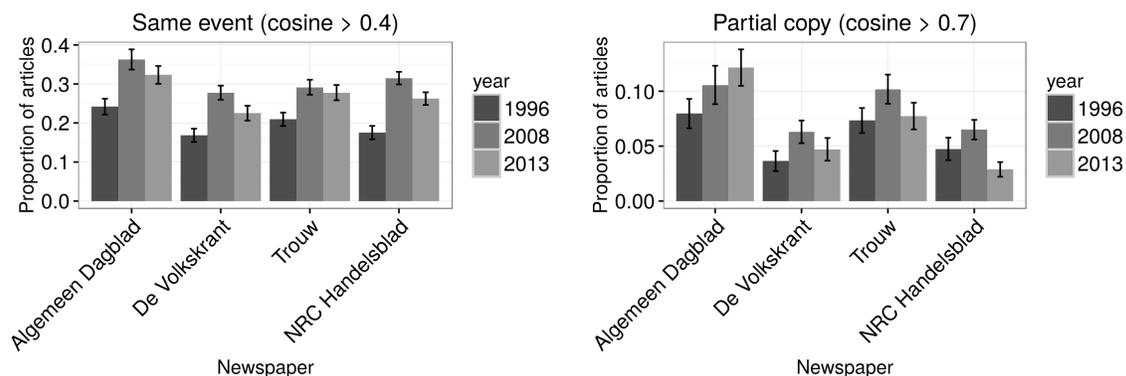
As a demonstration of how we use document similarity data to investigate our hypotheses, Figure 3 shows a network of news articles that address the same event. The x-axis represents time, and the y-axis distinguishes the newspapers and news agency. Nodes represent news articles, and edges represent similarity⁵. Grey lines indicate that articles address the same event ($\text{cosine} > 0.4$) and black lines indicate that they are partially identical ($\text{cosine} > 0.7$).

In this example *ANP* breaks a story, and an hour later three online newspapers copy it. Briefly after that, the online edition of *NRC Handelsblad* publishes an article covering the same event but not a partial copy (indicated by the grey line). Interesting to see is that during the time of this particular example, *NRC Handelsblad* did not have an *ANP* subscription, meaning that they could not legally have copied the *ANP* article. The timing of the publication suggests *NRC Handelsblad* covered an event after learning about it from the other newspapers.

The network contains all the information we need for our analyses. The influence of *ANP* on a newspaper is measured as the percentage of newspaper articles with an edge towards a recent *ANP* article (H1 & H2). The reaction time is calculated based on the distance of similar articles on the x-axis (RQ1). And the extent to which newspapers are influenced by the same news agency articles is measured as the number of different newspapers that are similar to the same news agency article (RQ2 & H3).

⁵Note that for this network we also looked at inter-newspaper similarities to get a better feel of our data, but for the analysis we are only interested in edges from newspapers to the news agency.

Figure 4. Proportion of articles that can be traced back to ANP per newspaper per year.



Note: Error bars represent confidence intervals, based on a binomial distribution with a 95% confidence level.

Influence of ANP on print newspapers over time

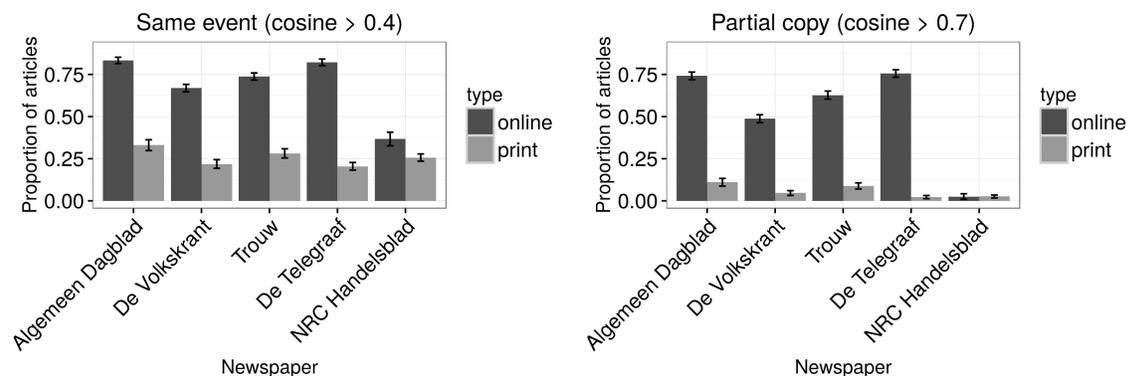
The left chart in Figure 4 presents the influence of ANP on the print newspapers in 1996, 2008 and 2013, measured as the proportion of newspaper articles of which the event can be traced back to a news agency article. Overall we see that this lies around 31%, ranging from 29% to 36%. In comparison to Scholten and Ruigrok (2009), who found an average of 27.6% ANP based articles across newspapers in 2008, our results are slightly higher for that year. This is likely because we look for similarity in terms of events, whereas Scholten and Ruigrok focused on literal quotes. Recall that newspaper *De Telegraaf* is missing from this analysis, because data for 1996 was not available.

If we look at the results using a cosine threshold of 0.7, we zoom in on newspaper articles that are likely to be copies or rewrites of ANP articles. The average percentage drops to about 9%, and differs more strongly across newspapers. Most noticeable is a sharp decrease in influence on *NRC Handelsblad* in 2013. This makes sense, because in 2010 this newspaper broke of their contract with ANP, meaning that it can no longer publish literal copies of ANP content. The articles that do match an ANP article at this level of similarity are mainly short articles that contain the same quotes from politicians as reported by ANP, meaning that its probable but not certain that ANP is a (indirect) source.

Interestingly, even though *NRC Handelsblad* is no longer subscribed to ANP in 2013, many of the events it covered can be traced back to ANP, as seen in the results using the 0.4 threshold. To some extent, this can simply be the result of coincidence: ANP publishes faster than *NRC Handelsblad*, so if they independently cover the same event then ANP is faster to report it. But, it should also be taken into account that *NRC Handelsblad* can indirectly rely on ANP by monitoring news publishers that do have a subscription (which, as discussed, is also legal to do due to the press exception in copyright law). Also, journalists tend to monitor the work of their colleagues to gather information and to confirm their own sense of news (Gans, 1979). Since *NRC Handelsblad* is an afternoon newspaper, this can include news from the morning newspapers.

Looking at the changes over time, there is a clear increase between 1993 and 2008 in all four newspapers in the proportion of articles that is traced back to ANP. To account

Figure 5. Proportion of print and online articles that can be traced back to ANP in the first half of 2013



Note: Error bars represent confidence intervals, based on a binomial distribution with a 95% confidence level.

for coincidental differences (though essentially we analyse the whole population) we also calculated whether the differences between proportions are significant based on a binomial distribution. This is the case for both measurements of influence: in terms of events ($p < 0.001$) and in terms of partial copy ($p < 0.01$).

Between 2008 and 2013 we did not see an increase, and in some cases even a significant decrease. This could be related to economic pressures, as discussed in the section on news agencies in the Netherlands. Note, for instance, that the number of political ANP articles decreased during this period. In combination with many free online sources that depend almost exclusively on ANP for information, this harms the exclusivity of ANP. *NRC Handelsblad* stated this as a main reason for breaking off their contract with ANP (Van Vulpen, 2010). Other newspapers might have responded by looking for more alternative sources of information.

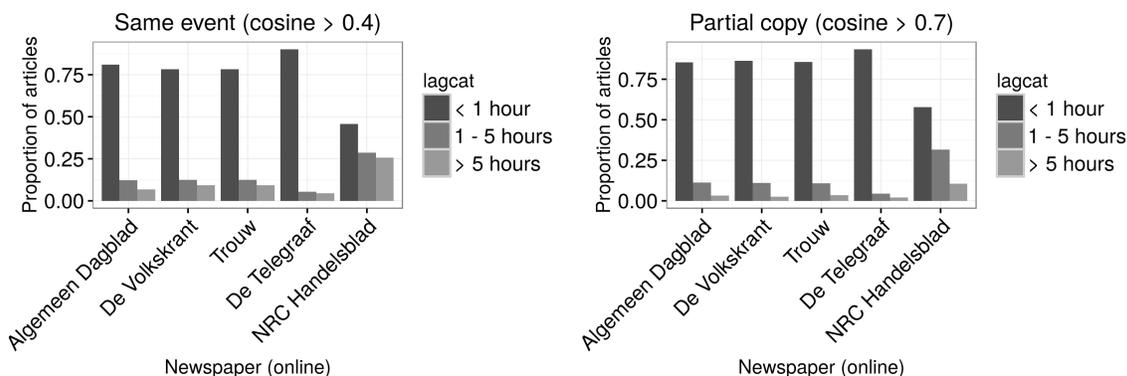
If we compare 1996 to 2013 for the measurement based on similar events, we still see a significant increase in all newspapers ($p < 0.01$). For the measurement based on partial copy this is also the case for *Algemeen Dagblad* ($p < 0.001$). Thus, we still find evidence of an increase of ANP influence between 1993 and 2013, based on which we accept H1.

Influence of ANP on online newspapers

Figure 5 presents the influence of ANP on the print and online newspapers in the first half of 2013. According to Tony Gillies, Editor-in-chief of the Australian Associated Press (AAP), the online editions of Australian newspapers used about 3 to 4 times as much AAP content compared to their print counterparts (quoted in Johnston and Forde, 2011, 198). Our results suggest that in the Netherlands the differences are roughly of the same magnitude. Also, looking at the results for similarity scores above 0.7, we see that the websites often publish (near) exact copies of ANP articles. The only exception is *NRC Handelsblad*, but this makes sense since it was not subscribed to ANP in 2013. These results provide strong support for H2: news agency reliance is stronger for online newspapers.

The online editions of newspapers in the Netherlands thus depend heavily on ANP.

Figure 6. Time between an online newspaper article and the news agency article to which it can be traced back



One of the main explanations for this is that online newspapers require a steady supply of information because of the quick paced online news cycle. We investigate this by looking at the time it takes for online newspapers to respond to *ANP* publications. If a newspaper article matched with multiple *ANP* articles—for instance, if multiple *ANP* articles cover the same event—then only the strongest match was used to calculate the time difference.

The results are presented in Figure 6. As explained, we subtracted one hour from the *ANP* publication time because we found that some articles that are certainly *ANP* copies were published before the *ANP* publication. For this analysis we used the original publication times, and if the newspaper article was published (within one hour) before *ANP*, then the response time was set to 0.

The results clearly show that online newspapers most often adopt an *ANP* article within one hour—at least 75%, except for *NRC Handelsblad*. For partial copies this was even above 85%. For all newspapers combined, the average response time (RQ1), measured as the median⁶, is 14 minutes for same event articles and 12 minutes for partial copies. Note that this includes the articles from the *ANP* buzzfeed service, which are published simultaneously with *ANP*. Overall, this clearly demonstrates the role of *ANP* in the quick paced online news cycle in the Netherlands.

Interestingly, while *NRC Handelsblad* was not subscribed to *ANP*; the average response time for the same event articles (36% of *NRC* articles) was still only 70 minutes, and almost half (46%) of the articles was covered within one hour after *ANP*. This suggests that even if *ANP* only had an indirect influence, news travels fast.

Homogeneity in adopting *ANP* articles

Based on the previous results regarding the influence of *ANP* on online newspapers, it is likely that online newspapers often cover the same events based on their shared use of *ANP* as a source. But, it is still possible that different newspapers are influenced by different *ANP* articles. To investigate this, we analyzed what proportion of a newspaper’s articles can be traced back to the same *ANP* articles as another newspaper’s articles. For the sake of parsimony, we only report the results for the analysis at the level of events.

⁶The median is more appropriate than the mean given the highly skewed distribution

Figure 7. Proportions of a newspaper's (rows) *ANP* influenced articles that also influenced another newspaper (columns)

Algemeen Dagblad - print	1	0.23	0.27	0.26	0.28	0.58	0.57	0.57	0.51	0.19
De Volkskrant - print	0.33	1	0.41	0.33	0.42	0.64	0.68	0.64	0.56	0.29
Trouw - print	0.25	0.29	1	0.26	0.32	0.62	0.62	0.66	0.58	0.21
De Telegraaf - print	0.29	0.28	0.3	1	0.35	0.58	0.55	0.56	0.53	0.25
NRC Handelsblad - print	0.21	0.24	0.28	0.24	1	0.44	0.45	0.44	0.39	0.27
Algemeen Dagblad - online	0.19	0.17	0.21	0.18	0.2	1	0.96	0.94	0.71	0.17
De Volkskrant - online	0.19	0.18	0.23	0.18	0.21	0.95	1	0.93	0.68	0.17
Trouw - online	0.19	0.17	0.24	0.18	0.21	0.97	0.96	1	0.71	0.17
De Telegraaf - online	0.19	0.16	0.22	0.18	0.2	0.75	0.73	0.74	1	0.17
NRC Handelsblad - online	0.3	0.32	0.33	0.34	0.47	0.68	0.63	0.66	0.58	1
	Algemeen Dagblad - print	De Volkskrant - print	Trouw - print	De Telegraaf - print	NRC Handelsblad - print	Algemeen Dagblad - online	De Volkskrant - online	Trouw - online	De Telegraaf - online	NRC Handelsblad - online

The results are presented in Figure 7. Scores represent proportions for the newspapers in the rows. For example, the cell in the second row, first column indicates that 33% of the articles in the print edition of *De Volkskrant* can be traced back to *ANP* articles that also influenced the print edition of *Algemeen Dagblad*.

These results contain our answer to RQ2, and several findings are particularly interesting. Firstly, we see a clear cluster of strong proportions between the online newspapers, in particular between *Algemeen Dagblad*, *De Volkskrant* and *Trouw*. It is notable that these three newspapers are all owned by *De Nederlandse Persgroep*. In 2011, a central editorial board was formed that would manage the general news for their online editions (Nu, 2011). This partially explains why their similarity in the use of *ANP* articles is higher than 90%. For the other online newspapers the proportions are lower, but still high. Most of the articles of *De Telegraaf* can be traced back to *ANP* articles that also influenced *Algemeen Dagblad* (75%), *De Volkskrant* (74%) and *Trouw* (73%). For *NRC Handelsblad* the proportions for these newspapers are slightly lower (68%, 63% and 66%). Given that except for *NRC Handelsblad* each of these newspapers depends strongly on *ANP*, as seen in Figure 5, we conclude that their shared dependence on *ANP* indeed harms the diversity of their political news coverage.

Secondly, we see that among print newspapers these proportions are clearly lower. Most are below 30%, and the strongest proportions are found for *De Volkskrant* towards *Trouw* (41%) and *NRC Handelsblad* (42%). Overall, this signifies that print newspapers

largely filter *ANP* news in different ways. Thus, the diversity of news in print newspapers suffers less from their shared dependence on *ANP*.

Conclusion

In this paper we analyzed the influence of a single news agency, *ANP*, on political news coverage in print and online newspapers in the Netherlands. The first part of our analysis focused on changes in the influence of *ANP* on political news in print newspapers between 1996, 2008 and 2013. We observed an increase between 1996 and 2013, which can be explained by economic cutbacks that force newspapers to cut back on news gathering expenses. But we also found that its influence decreased between 2008 and 2013, despite economic cutbacks for newspapers within this period. This development makes sense if we take into account how the market for news agencies in the Netherlands has changed in the past two decades. Like newspapers, news agencies are struggling to find a profitable business model in the digital age. Given the apparent role of news agencies in the gatekeeping process, it is important to study this struggle and how it affects their influence on newspapers and other major news publishers.

In the second part of our analysis we compared the influence of *ANP* on political news in print and online newspapers in the first half of 2013. Previous studies suggest that the influence of news agencies on online news is particularly strong, because of the difficulty of making profit from online news and the high speed 24/7 news cycle which demands constant updates of the latest news (Johnston and Forde, 2011; [Frijters and Velamuri, 2010](#); [Klinenberg, 2005](#)). We zoomed in on these two factors by comparing the print and online editions of the same newspapers, and by measuring the average speed with which online newspapers use *ANP* articles.

Our results clearly verify that the online editions depend more on *ANP* than the print editions. For the four newspapers with *ANP* subscriptions we even found that between 50% and 75% of political news consisted of (partial) copies of *ANP* articles. Also, we found strong empirical support for the high speed online news cycle: about 85% of (partial) copies was published within one hour after *ANP*. Note that in addition to theoretical implications, this finding has important methodological implications for time-series studies on the interactions of online news publishers. It underlines the need for models that are able to capture interactions at the level of minutes.

The third part of the analysis focused on how the shared dependence of newspapers on *ANP* affected the diversity of political news across newspapers. We found that print newspapers were often influenced by different *ANP* articles. Online newspapers, however, were often influenced by the same articles, which in combination with their strong dependence on *ANP* copy substantially harms the diversity of their political news coverage.

More generally, we believe that this signifies an important difference in the market logic for print and online news. Whereas diversity is an important area for competition between print newspapers, online diversity appears to be sacrificed for the sake of speed. Other studies already pointed out that in online newsrooms the pressure to be first suppresses the pressure to be right ([Johnston and Forde, 2009](#)). Our study adds that it also suppresses the pressure to be diverse.

Based on these findings, we conclude that the news agency *ANP* has indeed become a more influential gatekeeper regarding political news in the Netherlands. In recent decades,

its influence appears to have increased due to economic cutbacks in newspapers, and even more so as a result of the growing popularity of internet technology as a news medium. Notwithstanding the importance of *ANP* as a news gatherer, this raises concerns for the diversity of news.

It is important to note that we did not investigate the quality of journalistic work within *ANP*, nor did we investigate how well newspaper journalists check the reliability of news agency content. The harm to the the quality of news, as Davies (2008) claimed to observe in the UK, might not apply in the Netherlands. To conclude whether the strong influence of *ANP* also harms the quality of news content, additional studies are required that look into these journalistic practices.

In this paper we used two complementary measures of influence, and each has an important limitation. Regarding the first measure: if an event is first covered by the news agency and later covered by a newspaper then there is not necessarily a causal relation. There can be alternative sources from which a news publisher could have learned about an event, and it is generally not possible to take all possible sources into account. Furthermore, if multiple sources covered the event, it is unclear which—if any—causes the news publisher to cover it. To some extent, we can address this problem with the second measure. That is, by looking for explicit traces of influence found in how the article is written, either by using higher levels of document similarity (as in this paper) or by looking for literal quotes (see e.g., Scholten and Ruigrok, 2009; Paterson, 2005). The limitation of this measure is that influence does not always leave these explicit traces. To our best knowledge, the limitations of both measures cannot be addressed with only content-analysis data. We therefore suggest to use both measures as complementary indications of influence.

We found that the use of document similarity scores can be a powerful approach for tracing informational relations between news organizations on a large scale. In this line of research, we only encountered the use of this approach for the purpose of tracing literal quotes (Scholten and Ruigrok, 2009; Paterson, 2005). We expanded on this approach by using techniques from the fields of information retrieval (IR) and natural language processing (NLP). For future studies we will further explore and improve this approach. Our computer scripts and instructions are available as a package for the open-source statistical software R⁷. We aim to keep developing this package as a free and accessible tool for the analysis of news diffusion patterns.

References

- Arrese, Ángel 2015. [From gratis to paywalls: A brief history of a retro-innovation in the press's business.](#) *Journalism Studies* (ahead-of-print), 1–17.
- Bagga, Amit and Breck Baldwin 1998. [Entity-based cross-document coreferencing using the vector space model.](#) In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics-Volume 1*, pp. 79–85. Association for Computational Linguistics.

⁷The R package is hosted on the Comprehensive R Archive Network (CRAN) under the name RNewsflow. A vignette with detailed instructions is available online: <https://cran.r-project.org/web/packages/RNewsflow/vignettes/RNewsflow.html>. The code is available open-source on GitHub: <https://github.com/kasperwelbers/RNewsflow>.

- Bakker, Piet and Otto Scholten 2011. *Communicatiekaart van Nederland. Overzicht van media en communicatie (achtste geheel herziene druk)*. Kluwer.
- Boczkowski, Pablo J and Martin De Santos 2007. When more media equals less news: Patterns of content homogenization in argentina's leading print and online newspapers. *Political Communication* 24(2), 167–180.
- Chyi, Hsiang Iris 2005. Willingness to pay for online news: An empirical study on the viability of the subscription model. *Journal of Media Economics* 18(2), 131–142.
- Cook, Jonathan E and Shahzeen Z Attari 2012. Paying for what was free: Lessons from the new york times paywall. *Cyberpsychology, Behavior, and Social Networking* 15(12), 682–687.
- Davies, Nick 2008. *Flat Earth news: An award-winning reporter exposes falsehood, distortion and propaganda in the global media*. Random House.
- Ebisch, Bart 2012, Februari. Anp, novum en de digitale paradox.
- Eurobarometer 2012, November. Media use in the European Union. *Eurobarometer*.
- Forde, Susan and Jane Johnston 2013. The news triumvirate: Public relations, wire agencies and online copy. *Journalism Studies* 14(1), 113–129.
- Frijters, Paul and Malathi Velamuri 2010. Is the internet bad news? the online news era and the market for high-quality news. *Review of Network Economics* 9(2), online.
- Gans, H. 1979. *Deciding What's News*. Vintage Books.
- Granger, Clive WJ 1988. Some recent development in a concept of causality. *Journal of Econometrics* 39(1), 199–211.
- Guibault, L 2012. The press exception in the dutch copyright act. In B. Hugenholtz, A. Quaedvlieg, and D. Visser (Eds.), *A century of Dutch copyright law: auteurswet 1912-2012*, pp. 443–475. Amsterdam: deLex.
- Heijmans, Ellen, Kees Buijs, Pytrik Schafraad, Anne Marie Frye, Serena Daalmans, and Daphne Ten Haaf 2009. Nieuwsbronnen en de kwaliteit van de journalistiek. In B. Ummelen (Ed.), *Journalistiek in diskrediet*, pp. 23–40. AMB.
- Herbert, Jack and Neil Thurman 2007. Paid content strategies for news websites: An empirical study of british newspapers' online business models. *Journalism Practice* 1(2), 208–226.
- Hollanders, David and Rens Vliegthart 2008. Telling what yesterday's news might be tomorrow: Modeling media dynamics. *Communications* 33(1), 47–68.
- Johnston, Jane and Susan Forde 2009. 'Not wrong for long': The role and penetration of news wire agencies in the 24/7 landscape. *Global Media Journal-Australian Edition* 3(2), online.

- Johnston, Jane and Susan Forde 2011. The silent partner: News agencies and 21st century news. *International Journal of Communication* 5, 195–214.
- [Klinenberg, Eric 2005. Convergence: News production in a digital age. *The Annals of the American Academy of Political and Social Science* 597\(1\), 48–64.](#)
- [Knobloch, Silvia, Grit Patzig, Anna-Maria Mende, and Matthias Hastall 2004. Affective news effects of discourse structure in narratives on suspense, curiosity, and enjoyment while reading news and novels. *Communication Research* 31\(3\), 259–287.](#)
- [Landauer, Thomas K and Susan T Dumais 1997. A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review* 104\(2\), 211–240.](#)
- Lewis, Justin, Andrew Williams, and Bob Franklin 2008. A compromised fourth estate? UK news journalism, public relations and news sources. *Journalism Studies* 9(1), 1–20.
- Lewis, Justin, Andrew Williams, Robert Franklin, James Thomas, and Nicholas Mosdell 2008. The quality and independence of british journalism: Tracking the changes over 20 years. *MediaWise*.
- [Manning, Christopher D., Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky 2014. The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pp. 55–60.](#)
- [Matheson, Donald 2004. Weblogs and the epistemology of the news: Some trends in online journalism. *New Media & Society* 6\(4\), 443–468.](#)
- [McNelly, John T 1959. Intermediary communicators in the international flow of news. *Journalism & Mass Communication Quarterly* 36\(1\), 23–26.](#)
- [Meraz, Sharon 2009. Is there an elite hold? Traditional media to social media agenda setting influence in blog networks. *Journal of Computer-Mediated Communication* 14\(3\), 682–707.](#)
- [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.](#)
- [Mitchelstein, Eugenia and Pablo J Boczkowski 2009. Between tradition and change: A review of recent research on online news production. *Journalism* 10\(5\), 562–586.](#)
- [Monroe, Burt L, Michael P Colaresi, and Kevin M Quinn 2008. Fightin' words: Lexical feature selection and evaluation for identifying the content of political conflict. *Political Analysis* 16\(4\), 372–403.](#)
- [Napoli, Philip M 1999. Deconstructing the diversity principle. *Journal of Communication* 49\(4\), 7–34.](#)

- Nu 2011, Januari. Een webredactie voor AD, Parool, Trouw en Volkskrant.
- Paterson, Chris 2005. News agency dominance in international news on the internet. In D. Skinner, J. Compton, and M. Gasher (Eds.), *Converging Media, Diverging Politics*, pp. 145–163. Lexington Books.
- Reijnders, Maarten 2009, June. ANP: gratis media zitten niet voor dubbeltje op de eerste rang.
- Rutten, Paul and Mijke Slot 2011. Zijn de persbureaus te verslaan? De positie van de nederlandse persbureaus in de nieuwsketen. *TNO*.
- Salton, Gerard and Donna Harman 2003. *Information retrieval*. John Wiley and Sons.
- Salton, Gerard, Anita Wong, and Chung-Shu Yang 1975. A vector space model for automatic indexing. *Communications of the ACM* 18(11), 613–620.
- Scholten, Otto and Nel Ruigrok 2009. Bronnen in het nieuws. Nieuwsvoorziening landelijke dagbladen steeds afhankelijker van ANP. *Mediamonitor*.
- Shoemaker, Pamela J and Timothy Vos 2009. *Gatekeeping theory*. Routledge.
- Sparck Jones, Karen 1972. A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation* 28(1), 11–21.
- Turney, Peter D, Patrick Pantel, et al. 2010. From frequency to meaning: Vector space models of semantics. *Journal of artificial intelligence research* 37(1), 141–188.
- Van Atteveldt, Wouter 2008. *Semantic Network Analysis: Techniques for Extracting, Representing, and Querying Media Content (dissertation)*. BookSurge.
- Van Cuilenburg, Jan 2007. Media diversity, competition and concentration: Concepts and theories. *Media between Culture and Commerce* 4, 25–54.
- Van den Bosch, A, G.J. Busser, W. Daelemans, and S. Canisius 2007. An efficient memory-based morphosyntactic tagger and parser for dutch. In *Selected Papers of the 17th Computational Linguistics in the Netherlands Meeting*.
- Van der Laan, Servaas 2013, Januari. Nederlandse krantensites verdwijnen achter betaalmuur.
- Van Soest, Thijs 2012, June. Van Thillo: Het gratis content model online heeft zijn langste tijd gehad.
- Van Vulpen, Maarten 2010, December. NRC zegt contract met ANP op.
- Vliegthart, R. and S. Walgrave 2008. The contingency of intermedia agenda-setting. A longitudinal study in Belgium. *Journalism and Mass Communication Quarterly* 85(4), 860–877.