Between article and topic: news events as level of analysis and their computational identification

Trilling, D.; van Hoof, M.

DOI
10.1080/21670811.2020.1839352

Publication date
2020

Document Version
Final published version

Published in
Digital Journalism

Digital Journalism

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rdij20

Between Article and Topic: News Events as Level of Analysis and Their Computational Identification

Damian Trilling & Marieke van Hoof

To cite this article: Damian Trilling & Marieke van Hoof (2020) Between Article and Topic: News Events as Level of Analysis and Their Computational Identification, Digital Journalism, 8:10, 1317-1337, DOI: 10.1080/21670811.2020.1839352

To link to this article: https://doi.org/10.1080/21670811.2020.1839352

DIGITAL JOURNALISM
2020, VOL. 8, NO. 10, 1317–1337
https://doi.org/10.1080/21670811.2020.1839352

Routledge
Taylor & Francis Group

ORIGINAL ARTICLE

OPEN ACCESS Check for updates

Between Article and Topic: News Events as Level of Analysis and Their Computational Identification

Damian Trilling and Marieke van Hoof

Department of Communication Science, Amsterdam School of Communication Research, University of Amsterdam, Amsterdam, Netherlands

ABSTRACT

When comparing media coverage or analysing which content people are exposed to, researchers need to abstract from individual articles. At the same time, aggregating them into broad topics or issues often is too coarse and loses nuance. Both theoretically and methodologically, the analysis of an appropriate intermediate level of aggregation is underdeveloped. This article advances research in various areas of journalism studies by developing a theoretical argument for introducing the "news event" as a level of analysis. Based on this, we discuss several computational approaches to empirically detect such news events in large corpora of news coverage in an unsupervised manner. We provide two approaches: One based on traditional tf·idf based cosine similarities, and one that relies on word embeddings, in particular the softcosine measure. Both methods, combined with a network clustering algorithm, perform very well in detecting news events. We apply this method in a case study of 45k news articles from different outlets, in which we show that different news outlets have distinct profiles in the events they cover.

KEYWORDS

News events; text mining; word embeddings; unsupervised machine learning; network clustering

Introduction

Modern democracies are unthinkable without journalism. One of its tasks is the *dissemination of information* (e.g. Beam, Weaver, and Brownlee 2009). Hence, in well-functioning media systems, the sets of events covered by different outlets should not be disjoint but display a considerable overlap. *Routine reporting* ensures that information about routine events (debates, elections, cultural or sportive events) are disseminated to a wide audience. As journalists across contexts tend to agree on which events are newsworthy (e.g. Harcup and O'Neill 2017), also exclusive stories by one outlet, are ultimately reported on by other outlets as well. For instance, when *investigative reporting* (see, e.g. De Burgh 2008) by one outlet discovers a scandal, competitors will further propagate the information by paraphrasing the original article or by writing

CONTACT Damian Trilling d.c.trilling@uva.nl

Supplemental data for this article can be accessed here.

This article has been corrected with minor changes. These changes do not impact the academic content of the article.

follow-up stories. Take the Watergate scandal: The Washington Post uncovered it, but its immense political significance made all other media report on it as well – which, in fact, is necessary to make investigative reporting effective and exert political pressure. We see that, both in routine reporting and with major scoops, one *news event* is often covered by multiple articles in multiple outlets, to a potentially very different extent and choosing different words. This makes the task to identify articles covering the same event a challenging one. Yet, to systematically compare media coverage, we need to do so.

Shifting our focus from news production to news consumption, we face the same challenge: Tracking data or digital trace data may give us very fine-grained information on which specific articles someone has read, but to compare consumption patterns or study exposure effects, we somehow have to group related articles. We argue that for both theoretical and methodological reasons, we often need to do so on the level of what we refer to as *news events*. Many theories of democracy assume a citizenry that has a more or less shared awareness of relevant events happening (Ferree et al. 2002; Strömbäck 2005). This is even true for models that acknowledge that most citizens are actually rather uninvolved in politics. Here, "rather than try to follow everything, the monitorial citizen scans the environment for *events* that require responses" (Zaller 2003, p. 118; emphasis ours). Yet, most research has investigated the existence of a "common core of issues" (e.g. Chaffee and Metzger 2001; Geiß et al. 2018), whereas we argue that it might actually be at least as informative to investigate a "common core of events" that people are aware of. In fact, in line with the monitorial citizen model, it has been shown that in periods of high political activity, the events that citizens read about are more closely aligned with the events journalists deem most important (Boczkowski and Mitchelstein 2013). News events, in our understanding, are more specific than a topic or an issue, but less specific than an individual article. A news website can choose to devote zero, one, or more articles to one event. A competitor's website might make different choices, while still covering the same event.

To understand and analyse the role of the media in a given democratic society, we need to solve a central problem: *How can we identify and distinguish news events in media coverage?*

We propose a conceptualization of *news events* and a method to detect such news events. To the best of our knowledge, a combination of a theoretical conceptualization with a computational operationalization has not been developed in earlier research on news events. For the sake of simplicity, we assume written news and refer to the individual news item as an "article". Our argumentation can be extended to audiovisual news, but the specific method is focussed on textual content.

Theoretical Background and Related Work

We proceed in three steps. First, we discuss research areas where our proposed method may be particularly relevant: agenda setting research, media hype research, and audience research. Then, we develop a working definition of news events and

review possible methods used to detect news events. After that, we develop our own methodological approach and apply it to a first case study.

Why Do We Need to Study News Events?

News Events and Agenda Setting

Agenda-setting researching studies how the agendas of the public, media, and policy makers interact. Agendas are "issues or *events* that are viewed at a point in time as ranked in a hierarchy of importance" (Rogers and Dearing 1988, p. 556; emphasis ours). For instance, McCombs and Reynolds (1985, p. 12) write: "The New York Times frequently plays the role as intermedia agenda-setter because appearance on the front page of the Times can legitimize a topic as newsworthy".

Notwithstanding the large number of empirical studies that operationalize agendas as prioritized lists of *issues* or "most important problems" (e.g. Iyengar and Simon 1993), we follow Rogers and Dearing (1988) who maintain that "events are specific components of issues" (p. 566) and hence are crucial to focus on when moving towards more fine-grained agenda-setting research. Additionally, it has been argued that related events play a crucial role in promoting an issue on the agenda (see, for instance, the literature summaries in Walgrave and Van Aelst (2006) and Liu, Lindquist, and Vedlitz (2011)). Interestingly, already in 1954, Larsen and Hill (1954) studied empirically the "diffusion of a news event", namely via which media members of different communities heard about the death of a senator. Developing an automated approach to identify news events will allow us to conduct such and similar studies on a much larger scale.

However, we need to add a cautionary note here. Even though causal inferences from content analysis data are always problematic, when newspapers were issued only once a day, agenda-setting studies could reasonably argue that a time lag implied an agenda-setting influence. But notwithstanding some timestamp-based *online* intermedia agenda-setting studies (Haim, Weimann, and Brosius 2018), the fact that contemporary online outlets may publish on the same event within minutes or hours makes it much harder to draw such inferences.

While many agenda-setting studies relied on rather broad categories (see, e.g. Baumgartner, Green-Pedersen, and Jones 2006), recent societal issues may not fit into these. For instance, studies on the agenda-setting power of so-called "fake news" highlighted that a more fine-grained approach is needed, in which coverage of one specific event — be it a real one or a fabricated one — can be traced across outlets (Van Hoof 2019; Vargo, Guo, and Amazeen 2018).

News Events and Media Hypes

The "event" is also a central term in Vasterman's (2005) work on news waves and media hypes. Building on Kepplinger and Habermeier (1995), Vasterman identifies so-called "key events" that trigger media coverage – very much in line with what the events-as-part-of-issues perspective on agenda setting would also expect. In what he calls a media-hype, media are then "making the news instead of reporting events by: reporting comparable incidents and linking them to the key event; reporting thematically related news such as features, analyses and opinions" (p. 516). He distinguishes between genuine events ("like

violent incidents or court convictions", p. 528) that are happening in the real world, independently of news coverage and other events, such as "an interview, a speech, an official warning (regarding health risks) or, as often happens in scandals, a startling disclosure by investigative reporter" (p. 514). Importantly, key events can be of either category.

Media hype researchers typically define the event in advance (e.g. Hellsten and Vasileiadou 2015), while we assume that events are not known a priori. If we were to distinguish between between staged "pseudo-events" (Boorstin 1987) and genuine events, one could argue that many pseudo-events are, in fact, known a priori. Yet, if we follow the distinction between routine events, serendipity, scandals, and accidents (e.g. Molotch and Lester 1974), only the routine events can be reasonably assumed to be known by an observer. Still, when analysing a corpus of historical data, all types of events can, in theory, be said to be "known" – yet, compiling a comprehensive list of them seems infeasible except for datasets of trivial size.

Also, we would consider an interview that accompanies a news story as belonging to the same event, whereas Vasterman (2005) would see both articles as belonging to the same "news wave", but the interview being a separate event.

News Events and Audience Research

Research on news audiences has for a long time investigated how exposure to various media relates to their audience's knowledge of current events, largely relying on survey data (e.g., Oeldorf-Hirsch 2018; Schoenbach, de Waal, and Lauf 2005).

Yet, audience research is facing large transformations, moving away from survey data and towards the use of tracking data and digital trace data. This abundance of data poses new challenges in how to abstract and aggregate. When dealing with data on the browsing histories of respondents (for a data collection tool, see Menchen-Trevino 2016), we need to settle on a level of analysis on which the abundance of URLs that people visited can be aggregated. Aggregating on the domain level shows which combinations of outlets people visit (e.g. Mukerjee, Majó-Vázquez, and González-Bailón 2018). This allows to study, for instance, whether the audiences of extremist and mainstream media overlap, but does not answer the question whether there is a fragmentation of the public sphere (see, e.g. Marcinkowski 2008) such that some groups of people are not aware of the same events happening in society as others. To do this, we need to aggregate on the event level.

The importance of this comes hand-in-hand with a development referred to as "unbundling" (e.g. Trilling 2019) or "atomization" (Bruns 2018), a move towards very individual news diets of separate articles from different outlets. If we still want to answer questions about people's knowledge on current events as a function of their media use, we cannot use an aggregation of the coverage of "their" newspaper as a proxy of their news exposure. Instead, we need to link the small pieces they consume to the events they cover. Crucially, it is not sufficient any more to this *for every outlet*, but *for every individual user*, which stresses the need for automatization. There is a pressing need for such studies: The unbundling of news consumption has explicitly been linked to a persumed (but not empirically confirmed) influence on the diversity of news events on users' agendas (Moeller et al. 2016).

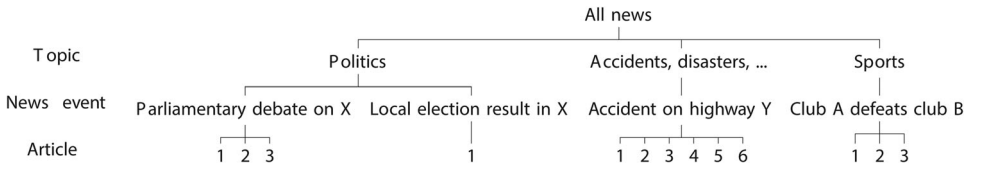


Figure 1. News events can be covered by one or more articles in one or more outlets, but relate to one specific and identifiable event and are thus much more fine-grained than news topics or news categories.

How Can We Define News Events?

While both news users and journalists will have an intuitive understanding of an "event", it is surprisingly hard to pinpoint a definition. Even seemingly simple characteristics of an event, such as its scope, its beginning and end, are hard to define. For instance, a public clash between two politicians of the same party probably is a newsworthy event. But is the speech that the first politician gives one event and the reaction of the second another one? Or is neither, and is the occasion where it happened (let's say, a party congress) the event? The decision how the stream of what's happening is chunked into meaningful "stories" is largely contingent, and thus, news can be seen as "a socially determined construction of reality" (Staab 1990, p. 428). Hence, Molotch and Lester (1974) distinguish between socially constructed "events" and their underlying "happenings".

Consequently, any attempt to determine the one-and-only event classification of what happens on a given day is doomed to fail. Yet, unless one denies the existence of any physical reality, one must acknowledge the existence of happenings that spark media attention: the final match of a world championship, a sudden invasion of one country by another, the outbreak of a pandemic. Unfortunately, distinguishing between underlying happenings and socially constructed events seems hard if not impossible in content-analytical approaches. Molotch and Lester (1974) classify events depending on whether the underlying happening was intentional or not, and who "promoted" the event. But to investigate such differences, we first need to take a step back and be able to identify events.

The discursive construction of news events implies that an event is not necessarily confined to the boundaries of a single article. If we want to determine, for instance, in how far news outlets publish about the same news event as others, we are not interested in an individual news article, but in all articles that cover such a *news event*. We therefore preliminary define news events as follows:

> News events are specific events that lead to news coverage, such as a specific debate on a specific day in a specific parliament, a specific accident, or a specific football match. They can be covered by one or more articles in one or more outlets, but relate to one specific and identifiable event and are thus much more fine-grained than news topics, issues, or news categories.

Figure 1 offers a graphical illustration of an example of this definition.

Note the hierarchical structure in Figure 1. Another example of this hierarchy is provided by Yang et al. (1999), who write: "An event [...] is different from a topic [...]. An event identifies something (non-trivial) happening in a certain place at a certain time. For example, USAir-427 crash is an event but not a topic, and 'airplane accidents'

is a topic but not an event." (p. 34). Clearly, the topic here can contain multiple specific accidents. As we will discuss later, our model could be extended to more levels of hierarchy. For instance, "airplane accidents" could be a subtopic of the broader topic "accidents and disasters".

Our conceptualization is notably broader than in some other fields. For instance, Schrodt's work on event data in political science (e.g. Schrodt, Davis, and Weddle 1994) has focussed on parsing sentence structures to extract information, but is essentially dictionary-based: events are defined by a set of pre-defined verbs and actors that are grammatically related to them. This considerably narrows down what can be detected as an event. More recenty, the GDELT Project defined an event as "capturing two actors and the action performed by Actor1 upon Actor2" (Schrodt and Leetaru 2013, p. 41). It is focussed on international politics, but used in communication science as well (see Hopp et al. 2019). As our definition is much less restricted, GDELT events fit our concept of a news event, but the reverse is not necessarily true.

Also Boorstin's (1987) work is based on a narrower conceptualization, as it mainly focuses on "pseudo-events" – planned events that are created with the very purpose of being reported about, such as press conferences, speeches, celebrations, etc. Next to these, we also include what Molotch and Lester (1974) would describe as events based on unintentionally accomplished happenings.

Instead, our conceptualization of news events is more similar to the notion of news story chains by Nicholls and Bright (2019), who "define news 'story chains' as events or single issues which receive repeated coverage in the news media through a series of initial articles and followup pieces" (p. 44). The differences are subtle; nevertheless, they result in a different analytical and interpretative lens. First, the notion of a chain evokes the image of a strict temporal order. However, it may not always be meaningful to say "who was first", and links between articles may be more complex: If two outlets A and B publish two articles A1 and B1 about the same event on day 1, B publishes a follow-up B2 on day 2 (but A does not), and both A and B publish articles A3 and B3 on day 3: how would the chain look like? What is the predecessor of A3? Is it B2 or A1? And is A1 or B1 the "initial article"? Depending on the specific research question, either the temporal notion of a story chain or our more general notion of a news event may be more useful. Second, Nicholls and Bright (2019) are interested in "repeated coverage" only, while here again our scope is wider: a standalone article that covers one event covered by no one else is within the scope of our definition of a news event, but not within the scope of the definition of a news story chain.

For instance, studies that are interested in the diversity of events that are present in a given subset of the data (such as coverage by a specific outlet, or the composition of an individual's news diet (see, for instance, Moeller et al. 2016)) are not interested in temporal ordering, but need to allow for single-article events.

Still, the notion of a news story chain and a news event have more commonalities than differences. Most notably, they introduce a more fine-grained level of analysis than the often-used topic: "News stories are conceptually distinct from news 'topics,' which we define as thematic news areas which also receive repeated coverage but which naturally encompass multiple events, and whose time span is much longer" (Nicholls and Bright 2019, p. 44). Given a large-enough corpus, we would assume that

one news event often (but not always!) triggers multiple articles. For instance, Buhl, Günther, and Quandt (2018) grouped 1,919 online news reports into 131 events, and Nicholls and Bright (2019) automatically identified 5,753 "story chains" in 39,558 articles.

One could also conceive of a situation in which one article covers multiple events; for instance an analytical background piece reflecting on several events. For now, we will focus on the typical case, in which an article (mainly) reports on one event.

A second consideration is how events are grouped into topics or issues. Again, we can ask whether an event can belong to multiple issues. For instance, in the Comparative Agendas Project (Baumgartner, Green-Pedersen, and Jones 2006), topics are coded in a hierarchical manner, with broader topics having sub-topics, and one article can be assigned to multiple ones. In contrast, many analyses ultimately focus on the "main topic", substantiating the view that a news event can be assigned to one topic.

A third consideration is how news events are related to each other. One can treat them as unrelated, for instance to study *how many* different events are covered by which outlets. But one could also treat them as related via their parents in the tree in Figure 1, adding information about "how different" events are. Additionally, and as we will discuss as a suggestion for future work in our conclusion, one might consider adding more hierarchical relationships to the model to allow for sub-topics and/or sub-events. Finally, some have proposed to talk about *serial* events or *linked* events when, for instance, analysing follow-up news coverage (see, e.g. Geiß 2018). Generalizing this idea, one could also think of a network approach to model the relationship between events, just as network approaches have been proposed to model associations between issues in agenda setting theory (Guo 2013).

Automated Approaches to the Analysis of News Events

To determine which articles cover the same event, we need pairwise comparisons between individual articles. As the number of comparisons increases exponentially with the number of articles, manual comparisons are infeasible.

As we assume the number of news events to be large and unknown, rule-based and supervised methods are not appropriate for our purpose. Also, many unsupervised methods like topic models or *k*-means clustering require specifying the number of clusters (thus, events) in advance. We, in contrast, need an unsupervised method that is able to cluster news articles into an undefined number of news events. It also needs to scale nicely to (very) large datasets.

Prior research has largely focussed on *literal* overlap using the Levenshtein distance, cosine similarities of word count or tf·idf representations, (Boumans et al. 2018; Welbers et al. 2018), or the BM25F score (a measure similar to tf·idf scores) and the proportion of common keywords (Nicholls and Bright 2019). These methods assume that to describe the same event, one essentially needs the same words: "an unusual place name […] would be a strong indicator that these two articles are part of the same story" (Nicholls and Bright 2019, p. 48). While this indeed is true for examples involving such unique identifiers, the general assumption has recently been challenged. In particular, standard overlap methods do not take into account that texts can be *semantically similar*, i.e. the same news event can

be described in different words. To illustrate, the sentences '*Obama speaks* to the *media* in *Illinois*' and 'The *President greets* the *press* in *Chicago*' share none of the same words (excluding stop words) resulting into a zero similarity score in classic similarity measurements. However, it is clear that these sentences describe the same news event (Kusner et al. 2015).

This can be solved with so-called word embedding models, in which words are represented in a high-dimensional vector space where more similar words are closer together (Mikolov et al. 2013). In a classic bag-of-words (BOW) model, a specific word in document A can only be either identical or not to a specific word in document B; but if we represent each word by a high-dimensional vector, we can say *how* identical on a continuous scale the two words (and documents) are.

Kusner et al. (2015) Word Mover's Distance (WMD) utilizes such a word embeddings model to compute the cumulative semantic "travel cost" between two documents. Even though Kusner et al.'s "the president greets the press"-example has become a widely cited example, and in spite of its obvious relationship to news events, we are not aware of any empirical study that actually used their WMD technique to explicitly detect news events.

Another approach to compare documents based on word embeddings is the soft cosine measure (SCM), introduced by Sidorov et al. (2014). When two words are completely unrelated, the soft cosine is identical to standard cosine similarity. SCM has been shown to be considerably faster than WMD while showing almost no loss in precision (Novotný 2018). Again, to the best of our knowledge, no prior news event research has used SCM so far.

After calculating document similarities with any of these approaches, one needs to determine which documents "belong together". Traditionally, two articles with a similarity score above a certain threshold are considered to belong to the same group (e.g. Boumans et al. 2018; Welbers et al. 2018). Nicholls and Bright (2019) argue that the threshold approach does not work well in cases where documents can belong to multiple groups and develop a novel approach using network partitioning techniques to identify the boundaries of their news story chains.

We will combine a network-approach with a word-embedding approach to identify news events. We are interested to see how a word-embedding-based approach compares to a traditional literal-overlap approach. How does the strength of the latter ("unique events are characterized by unique words like place names") compare to the strength of former ("unique events can still be described with different words, especially synonyms and near-synonyms")? In how far do these strengths outweigh their associated weaknesses (missing relevant articles because of different wording, and considering different entities as identical that are only similar, respectively)?

A First Application: Coverage of News Events in Dutch Media

As a first application and test case for our proposed approach, we use half a year of coverage in three Dutch media outlets, which we describe below. We do so mainly for illustrative purposes, and acknowledge that a dedicated study would be necessary to give a fuller account. As we have seen, scholars are concerned that a lack of overlap of covered events may be detrimental to democratic discourse (e.g. Moeller et al.

2016; Schoenbach, de Waal, and Lauf 2005). Therefore, we will show how our method can be used to answer the following questions:

RQ1 To what extent do the events covered overlap between outlets?

RQ2 (a) Which kind of events enjoy the largest overlap, and (b) which kind of events are exclusive to specific outlets?

For pragmatic reasons, we operationalize the "kind" of event here as the overarching topic to which it belongs, as this could be easily determined. Crucially, though, one could use any other available feature of the events (e.g. whether they were genuine or pseudo-events).

Methods

Data and Resources

We used a large corpus ($N = 45k$) of Dutch news articles published between 26-11-2018 and 26-05-2019, consisting of articles from ad.nl (popular newspaper with regional editions), volkskrant.nl (national quality newspaper), and nu.nl (large online-only news site). The data are a subset of a larger project, in which news articles are continously gathered using RSS feeds and web scraping (Trilling et al. 2018). Custom parsers extract the plain text of the article body (length: $M = 1608$ characters ($SD = 1552$)).

We furthermore use the Amsterdam Embedding Model (AEM), which has been specifically developed for news-related text and has been shown to outperform competing models in tasks like topic classification (Kroon et al. 2019). We preprocessed our data in the same way that the training data for the word embedding model were preprocessed: we removed punctuation and lowercased the texts.

Similarity Calculation

A naïve approach to obtain similarity scores would be to compare all articles with each other, for instance by multiplying a matrix with all article representations with its own transpose. Both for efficiency reasons and theoretical reasons, we instead narrowed down the set of candidate articles that may be considered belonging to the same event. Conceptually, it seems wrong to consider an article that was published – say – a month after another article to be part of the same event.

Castillo et al. (2014) have shown that after three days, the interest for news-related articles vanishes; and by extension, it seems unreasonable that a news outlet will still cover a more than three-day old event. Following Nicholls and Bright (2019), we assume news events in principle take place within a window of three days. For any given article, we look whether there are similar articles on the same day, one day after, and two days after. This does *not* imply that the maximum duration of the event is three days: The maximum distance of three days does not refer to the first, but to the last article of an event. Hence, the maximum time span is unlimited, but as soon as one day is followed by two days without any coverage about the event, we consider the event closed. This *chaining* that can occur is both a curse and a blessing: If we use too low similarity thresholds, it may happen that some article *always* has a

similar-enough article in the next one or two days, leading to an "event" spanning months, and in which the content drifts and the last article has no similarity at all with the first one. But if we use a strict maximum of three days from the *first* day of an event and not allow for chaining, we are too strict: Yang et al. (1999) even go as far as maintaining that coverage over an event can last several weeks.

We slightly modified the three-day threshold to account for weekends. Traditionally, Saturday newspapers are much thicker than weekday editions, and no newspaper appears on Sunday in the Netherlands. Even though the shift towards online news has made this distinction less clear, the general pattern still holds true, also because much online news content is in fact produced by the same newsrooms as for the print editions. We therefore assume a six-day week in which Saturday and Sunday are collapsed.

(Soft)cosine similarity was computed using the implementation provided by the gensim package (Rehurek and Sojka 2010). To give more weight to infrequent words, which are likely to characterize a given event, we applied tf·idf weighing (see also Yang et al. 1999) and discarded all words that occurred in more than 50% of the articles. As a side benefit, the latter greatly diminishes the memory resources needed and sped up the calculations significantly. We also removed all words that occurred two times or less in the entire corpus.

Partitioning into Events

We constructed a network with each article as a node, and each similarity score as an edge weight. Obviously, most articles are *not* similar to each other. To enable efficient calculations and to remove the arbitrary difference between articles that we did not even compare (because of a >3 days date difference) and unrelated articles with an edge weight close to 0, we decided to not store any edge with weight $< .2$, a value below which it is very unlikely that articles address the same event (Welbers et al. 2018).

We partitioned the graph using the Leiden algorithm (Traag, Waltman, and van Eck 2019), an extension of the well-known Louvain algorithm (Blondel et al. 2008). In particular, we used the Surprise method (Traag, Aldecoa, and Delvenne 2015), which, compared to the often used Modularity method, is more suitable for a large number of very small communities. A comparison of both methods in our data set confirmed that Modularity generally performed well, but additionally created a few too-large partitions with hundreds of articles.

We also considered the use of a hierarchical clustering technique used by Nicholls and Bright (2019), Infomap (Edler, Bohlin, and Rosvall 2017). Infomap creates subclusters as long as the links between the smaller groups are stronger than those between random articles within the bigger group.The results were almost identical (see Online Appendix). Consistent

Table 1. Descriptives for different threshold/similarity combinations.

	cosine					softcosine				
	0.2	0.3	0.4	0.5	0.6	0.2	0.3	0.4	0.5	0.6
mean	2.03	1.58	1.35	1.21	1.12	6.78	2.89	1.88	1.51	1.27
std	3.48	2.00	1.22	0.71	0.45	30.41	10.04	4.27	2.27	1.07
max	88	53	41	21	15	551	367	161	70	30
single-art. events	15626	21854	27135	32232	36348	4262	11043	18305	24337	30700
multi-art. events	6685	6777	6241	5165	3899	2460	4736	5961	5940	5257

with our results, using Infomap to identify news events requires a relatively high threshold; without it, it will rather identify news story chains (Nicholls and Bright 2019).

Manual Annotation for Evaluation Purposes

We manually annotated 100 randomly chosen events per model (https://github.com/damian0604/newsevents/tree/master/data/evaluation). With on average numbers of articles per event between 1.12 and 6.78 (Table 1), per model, the number of annotated articles is in the order of magnitude of hundreds. For each article within an event, we annotated whether it belonged to the main event (i.e. the event that most articles in the cluster described).

Results

Cosine versus Softcosine

As Figure 2 shows, the softcosine approach is able to match more articles per event than the cosine approach. For instance, using a threshold of 0.4, the cosine approach identifies 27,135 single-article events and 6,241 multiple-article events ($M = 1.35$, $SD = 1.22$), whereas the softcosine approach only leaves 18, 305 single-article events, and merges the rest into 5,961 multiple-article events ($M = 1.88$, $SD = 4.27$). This raises the question whether the higher degree of "merging" by the softcosine approach is justified or not. We therefore evaluated whether the identified events indeed consisted of articles pertaining to one event.

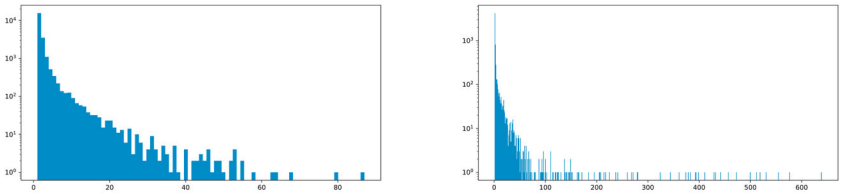
The evaluation of six models with thresholds 0.4, 0.5, and 0.6 for both cosine and softcosine similarity, points out that both cosine and softcosine score high on precision[1] (Table 2). Even the share of news events that were entirely correct is high across most thresholds.

Even at lower thresholds, the cosine method is able to group related articles. Even though in some cases it is not up to the standard of our definition of a news event (in which most follow-up developments would count as separate events), by far most articles are at least related at the "story chain"-level.[2] Although the cosine method enjoys relatively high precision across thresholds, it is more conservative: it creates smaller events and leaves more articles unrelated.

In contrast, softcosine is able to create bigger news events and leaves less single-article events. However, its precision is generally lower, especially at lower thresholds. We may already conclude that for the softcosine method, we need to set a higher threshold than necessary in the cosine approach.

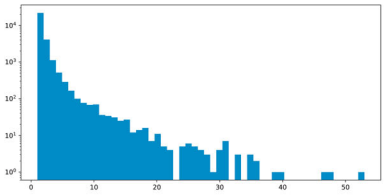
Though softcosine similarity is able to compensate for the fact that journalists use different words to describe the same events, it results in matches between articles that are similar in topic but are distinct event. For instance, we observed how the softcosine approach merged two articles about two unrelated events involving Puma and Nike, while the cosine method did not fall into this trap.

This confirms our initial expectation: The softcosine approach seems to be successful in finding potentially related articles that otherwise could not be found because of their different word choice. On the other hand, as the example illustrates, such

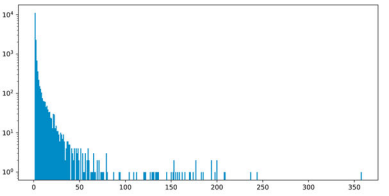


(a) cosine, 0.2

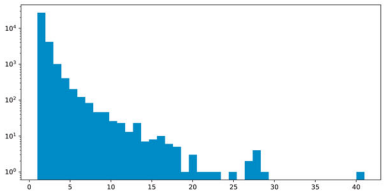
(b) softcosine, 0.2



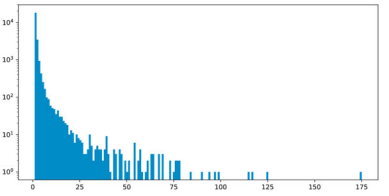
(c) cosine, 0.3



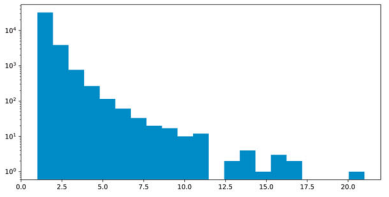
(d) softcosine, 0.3



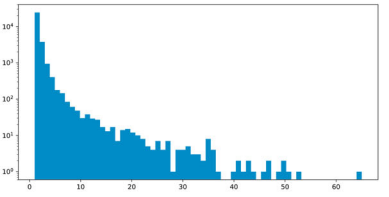
(e) cosine, 0.4



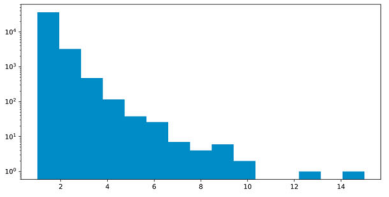
(f) softcosine, 0.4



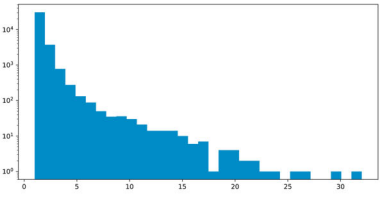
(g) cosine, 0.5



(h) softcosine, 0.5



(i) cosine, 0.6



(j) softcosine, 0.6

Figure 2. Histograms (logarithmic scale) of the number of articles per event using different similarity functions and thresholds.

Table 2. Precision for different threshold/similarity combinations.

Similarity method	Threshold	Prec. 1 (%)	Prec. 2 (%)	TP/AP
cosine	0.4	74	88.52	223/268
cosine	0.5	78	89.02	217/253
cosine	0.6	89	94.39	204/225
softcosine	0.4	56	76.20	234/521
softcosine	0.5	65	81.77	236/379
softcosine	0.6	75	86.92	222/289

Note. Precision 1: The percentage of news events that are entirely clustered correctly. *Precision 2:* The percentage of news articles that are correctly clustered. *AP* (all positives) is the number of articles that are assigned to an event in the sample; hence, the maximum number of true positives that can be achieved.

potentially related articles may, upon closer inspection, turn out to be *related* in some sense, yet *not about the same event*.

We therefore need to ask: Does the softcosine approach lead to a problematic false positive-rate? The comparatively high maximum number of articles per event may suggest that. Yet, this is not the case: In the softcosine-0.6 model, the largest event consists of 35 articles. A manual inspection confirmed that all of them are related to a soccer match between Liverpool and Manchester. If we have a look at all events consisting of more than 20 articles, a similar picture arises. These 12 events span a diverse range of domains (sports, weather, an environmental disaster, …), but only very few of these articles can be regarded as misclassified: in one of these 12 events, celebrity news are merged together with a sports event.

Selecting a Threshold

In contrast to earlier work (Welbers et al. 2018), our experiments suggest that a threshold of .2 is too low to capture *events* rather than broader issues or topics. As we see in Table 1, while a threshold of 0.2 may still work for a tf·idf based cosine similarity score, it is clearly leading to too many false positives when employing softcosine. Out of 45k articles, only 4,262 (thus, $<$10%) were not grouped, but the price we pay for this is too high: An event generating 551 articles is clearly implausible, and also mean and standard deviation seem too high, given that we study only three news sources (see also Figure 2). A qualitative examination of a sample of detected events confirmed this. In addition, it is highly plausible that at least some niche events are genuinely covered by only one article, and hence, we do not want the number of one-article events to be too close to zero.

More specifically, while there are some models and thresholds that are clearly inferior, there is not necessarily one "best" model. Researchers have to make a well-informed trade-off between more conservative models that offer a high precision (but miss some articles and some events) and more encompassing models that recover more links articles and events at the expense of a slightly lower precision.

Answers to the Research Questions

RQ1 asked to which extent the events overlap between outlets. It seems to be surprising at first sight that the overlap is comparatively little (Figure 3). On closer inspection, though, this makes sense, as the answer to RQ2 shows. RQ2 asked: (a) Which kind of

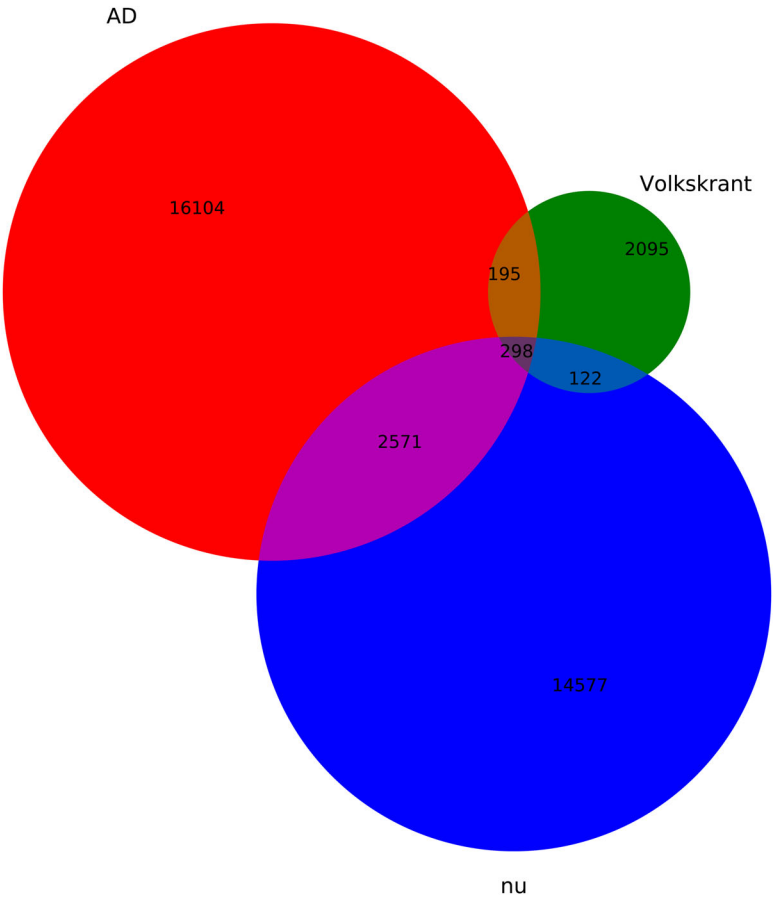


Figure 3. Overlap in event coverage based on the the softcosine measure and a threshold of 0.6.

events enjoy the largest overlap, and (b) which kind of events are exclusive to specific outlets? To answer these questions, we compared the events that are covered by only one outlet with those that were covered by all.

We decided to use the softcosine model with a threshold of 0.6, because it identifies more multiple-article events and misses less true positives than the standard cosine measure, be it at the price of a lower but still very high precision, compared to the standard cosine models with a threshold of 0.5 or 0.6. Other trade-offs are possible.

First, we used a classifier that was pre-trained on Dutch news to classify our news events into four topics: Business, Politics, Entertainment and Other (Vermeer 2018). We use this classification to illustrate that our event clustering makes sense, and depending on the substantive research question a researcher is interested in, one may use very different – and more fine-grained – approaches. However, on our focus here is on the detection of events (i.e. the first step in a research pipeline), not their classification according to, for instance, specific event types (a possible second step). Figure 4 points out that AD and nu.nl are more likely to write about entertainment news events that other outlets do not

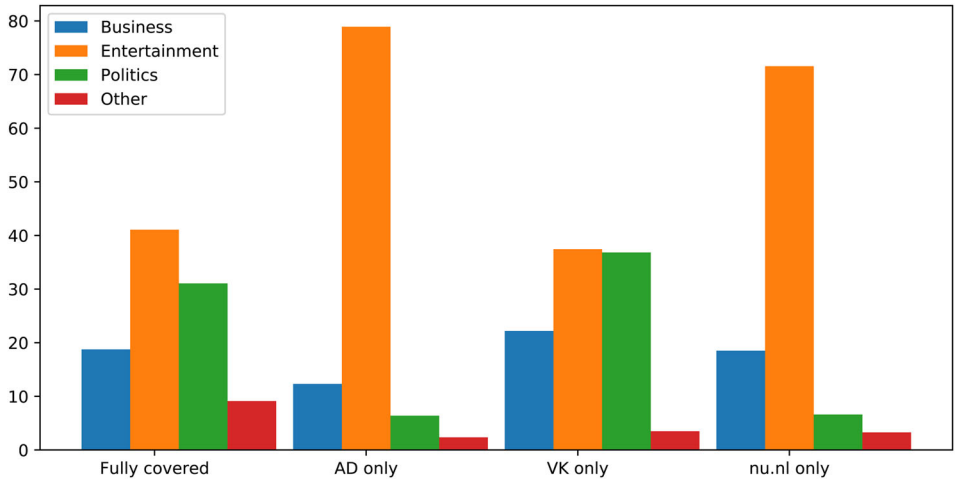


Figure 4. Distribution of topics in fully covered, AD-only, Volkskrant-only and nu.nl-only news events.

cover. In their exclusive news events, AD and nu.nl are relatively less likely to write about exclusive political events than Volkskrant (be it not in absolute numbers). This very much fits what one would expect from these outlets, as Volkskrant is considered a quality newspaper, while AD is a popular newspaper with regional editions, and nu.nl an online-only outlet aiming at a very broad audience.

To confirm this picture in a more fine-grained way, we followed a suggestion by Rayson and Garside (2000) and determined the most characteristic words for events covered by one outlet only versus fully covered events by calculating their loglikelihood based on their observed and expected frequencies. In addition, and to ease interpretation, we did not only consider unigrams, but also bigrams. Inspecting the 20 most characteristic words for each outlet, we see that most of these words are sport-related in the case of AD, entertainment and crime-related in the case of nu.nl, and politics-related in the case of VK, even though the differences are less clear-cut for the last outlet (see Online Appendix).

Conclusion and Discussion

Our aim was twofold: First, we offered a theoretical conceptualization of what we call a *news event* and argued what this level of analysis can offer for agenda setting, media hypes, and audience research. Second, we explored computational approaches to detecting such events and to cluster articles accordingly.

We showed that approaches based on literal overlap (such as the cosine similarity of tf·idf representations) perform surprisingly well to detect news events. However, they seem to be overly conservative and miss some articles that belong to an event, due to – for instance – differences in phrasing. Softcosine approaches can solve these issue due to the use of word embeddings. But this comes at a comparatively high cost by also merging articles that are about similar, but not identical events. Which approach is preferable may depend on the application. In particular, one may expect

that the softcosine based approach is less sensitive to different ways of framing the same issue. In a highly partisan media landscape, it may be that one event is reported on in such a different way that document-similarity based approaches fail to recognize them. We suggest that future research should tease out whether the strengths of different approaches can be combined to reach the best possible robustness. This seems especially relevant when our event clustering is used as a first step in a pipeline, in which the second step consists of an analysis of different perspectives on or framing of the same event.

Regardless of whether one chooses the cosine or softcosine approach, our experiments demonstrate that these approaches work well and can be applied in a very efficient way by limiting the number of comparisons that need to be made. Compared to a naïve compare-everything-with-everything approach, our approach based on a three-day-moving window is orders of magnitude faster and renders the analysis of long-term datasets possible, as the number of comparisons only increases linearly instead of exponentially. It also leads to a lower amount of false positives as it prevents grouping articles about similar but temporally distinct events.

We have shown that the combination of the Leiden network clustering algorithm with (soft)cosine based similarity metrics, applied to a moving window of news articles, can be a feasible approach. We suggest, though, to further explore alternatives and to systematically compare them. In particular, it may be interesting to see whether our approach can be further refined by taking more meta-data into account.

In order to do so, though, we need to systematically think about the right evaluation metrics. For instance, given that the population of all events and the number of articles per event are unknown, we can calculate a measure of precision by checking how many of the articles ascribed to an event indeed are about the event; but we cannot calculate a measure of recall that tells us whether we found all articles about the event, or whether we indeed found all events. Some (Nicholls and Bright 2019; Welbers et al. 2018) have randomly drawn pairs of articles and compared manual and automated assessment of whether they cover the same event. While this allows the calculation of both precision and recall, it requires a lot of pairs to be evaluated (as most randomly drawn pairs will not be about the same event) and – in our view – answers a different question than the questions that should be of core interest: (1) How many articles ascribed to an event indeed belong to it?; (2) How many articles were not assigned to any event though they do belong to a multi-article event?; and (3) How many events did we miss? Future research needs to assess what the best feasible evaluation criteria for news event detection are.

We finally demonstrated how our method can be used in a small case study. This suggested that overlap between different online news outlets might actually be rather low if it comes to the specific events covered, and probably lower than an analysis on the issue or topic level would suggest. Yet, the differences made immediately sense once we looked in more detail into which events differed in their coverage: Both a top-down topic classification and a bottom-up inspection of most over-represented uni- and bigrams showed that the quality paper covered political events that the other outlets missed, the popular newspaper covered additional sports events, and the online-only news site offered crime and entertainment events that the other outlets did not cover. While the question of what the best way to categorize detected events is beyond the scope of this article, we

presented the necessary first step that needs to be taken to do so. Future research could explore, for instance, natural language generation techniques to create short summaries of the events, and/or cluster the events into topics.

Towards a Hierarchical and Dynamic Understanding of News events

We started from the observation that a topic can contain multiple events which can be covered by multiple articles each. But the hierarchy could also contain more than three layers. For instance, depending on the researcher's goal, it may be useful to further split events into sub-events.[3]

Conceptually, this could offer a better lens to understand how coverage puts different emphases on different sub-events within an event; or it could help explain dynamics of spin-off coverage where a sub-event generates follow-up coverage instead of the event that was newsworthy in the first place. In our example in Figure 1, a player getting seriously injured during the match (sub-event) may have more long-term consequences than the original event (the match). The current coronavirus crisis offers a prominent example of how entities from different layers may "promote" in the hierarchy. If we would have conducted our analysis in December 2019, we would probably have identified the outbreak of a new disease in Wuhan as a news event. An overarching topic like "Covid-19" did not exist back then. Yet, it is one of the most prominent topics in the news coverage of 2020.

Hierarchical or even more generalized network approaches that allow an event to be part of multiple topics could be a way to model how what first appears to be a comparatively minor event becomes an overarching topic with many sub-topics and even more events. While this is certainly an extreme example, we hope that future work will systematically think through how such dynamics can be conceptualized and operationalized.

To build on, the Infomap algorithm can provide a hierarchical output, and the Leiden and Louvain algorithms can be run multiple times with a varying resolution parameter. It is also possible to modify the Louvain algorithm to output a full hierarchy (Bonald et al. 2018). Other hierarchical community detection algorithms support the simultaneous identification of multiple resolutions of topics, events, sub-events, etc. (Peixoto 2014). Doing so, however, requires theoretical work first: We need to develop a good definition of what constitutes a sub-event, think of methods to determine the optimal level of granularity, and develop evaluation metrics to to test whether we correctly identified all sub-events without getting unclear or non-sensical subevents.

In conclusion, the conceptual and methodological considerations in this article as well as our first application to a news dataset should rather be seen as the beginning than as the end: as the beginning towards the development of a robust and easy-to-use method to detect news events. In the end, this will greatly improve the way how we can make sense of media content, but also of media use data.

Notes

1. Precision denotes the fraction of articles a method identifies as relevant that indeed are truly relevant. Recall, in contrast, denotes the share of truly relevant articles that were identified as such by the method.

2. For instance, a 10-article news event (cosine, 0.5) consists of several news events: four articles covering the Dutch government's purchase of KLM-Air France shares, two articles on the response of Air France, two articles on the response of Dutch politics, and two articles on the events leading up to the purchase. These are definitely related at the story-level, though not at the news event level.
3. We would like to thank the two anonymous reviewers as well as Christian Baden, Ryan Gallagher, and Felix Victor Münch, all of whom pointed us to these possible hierarchical extensions of our work.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was carried out on the Dutch national e-infrastructure with the support of SURF Cooperative.

ORCID

Damian Trilling https://orcid.org/0000-0002-2586-0352
Marieke van Hoof https://orcid.org/0000-0002-4117-837X

References

Baumgartner, F. R., C. Green-Pedersen, and B. D. Jones. 2006. "Comparative Studies of Policy Agendas." *Journal of European Public Policy* 13 (7): 959–974. doi: 10.1080/13501760600923805

Beam, R. A., D. H. Weaver, and B. J. Brownlee. 2009. "Changes in Professionalism of U.S. journalists in the Turbulent Twenty-First Century." *Journalism & Mass Communication Quarterly* 86 (2): 277–298.

Blondel, V. D., J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. 2008. "Fast Unfolding of Communities in Large Networks." *Journal of Statistical Mechanics: Theory and Experiment* 2008 (10): P10008. doi: 10.1088/1742-5468/2008/10/p10008

Boczkowski, P. J., and E. Mitchelstein. 2013. *The News Gap: When the Information Preferences of the Media and the Public Diverge*. Cambridge, MA: MIT.

Bonald, T., B. Charpentier, A. Galland, and A. Hollocou. 2018. Hierarchical Graph Clustering using Node Pair Sampling. http://arxiv.org/abs/1806.01664

Boorstin, D. J. 1987. *The Image: A Guide to Pseudo-Events in America*. New York, NY: Atheneum.

Boumans, J., D. Trilling, R. Vliegenthart, and H. Boomgaarden. 2018. "The Agency Makes the (Online) News World Go Round: The Impact of News Agency Content on Print and Online News." *International Journal of Communication* 12: 1768–1789.

Bruns, A. 2018. *Gatewatching and News Curation: Journalism, Social Media, and the Public Sphere*. New York, NY: Lang.

Buhl, F., E. Günther, and T. Quandt. 2018. "Observing the Dynamics of the Online News Ecosystem." *Journalism Studies* 19 (1): 79–104. doi: 10.1080/1461670X.2016.1168711

Castillo, C.,. M. El-Haddad, J. Pfeffer, and M. Stempeck. 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 (Cscw '14), 211–223. Baltimore, ML: ACM.

Chaffee, S. H., and M. J. Metzger. 2001. "The End of Mass Communication?" *Mass Communication and Society* 4 (4): 365–379. doi: 10.1207/S15327825MCS0404_3

De Burgh, H. (Ed.). 2008. *Investigative Journalism: Context and Practice*. New York, NY: Routledge.
Edler, D., L. Bohlin, and M. Rosvall. 2017. "Mapping Higher-Order Network Flows in Memory and Multilayer Networks with Infomap." *Algorithms* 10 (4): 112. doi: 3390/a10040
Ferree, M. M., W. A. Gamson, R. Gerhards, and D. Rucht. 2002. "Four Models of the Public Sphere in Modern Democracies." *Theory and Society* 31 (3): 289–324. doi: 10.1023/A:1016284431021
Geiß, S. 2018. "The Dynamics of Media Attention to Issues: Towards Standardizing Measures, Dimensions, and Profiles." In *From Media Hype to Twitter Storm: News Explosions and Their Impact on Issues, Crises and Public Opinion*, edited by P. Vasterman, 83–113. Amsterdam: Amsterdam University Press.
Geiß, S., M. Magin, B. Stark, and P. Jürgens. 2018. "Common Meeting Ground in Gefahr? Selektionslogiken Politischer Informationsquellen Und Ihr Einfluss Auf Die Fragmentierung Individueller Themenhorizonte." *Medien & Kommunikationswissenschaft* 66 (4): 502–525. doi: 10.5771/1615-634X-2018-4-502
Guo, L. 2013. "Toward the Third Level of Agenda Setting Theory: A Network Agenda Setting Model." In *Agenda Setting in a 2.0 World: New Agendas in Communication*, edited by T. Johnson, 112–133. New York, NY: Routledge.
Haim, M., G. Weimann, and H.-B. Brosius. 2018. "Who Sets the Cyber Agenda? Intermedia Agenda-Setting Online: The Case of Edward Snowden's NSA Revelations." *Journal of Computational Social Science* 1 (2): 277–294.
Harcup, T., and D. O. Neill. 2017. "What is News?" *Journalism Studies* 18 (12): 1470–1419. doi: 10.1080/1461670X.2016.1150193
Hellsten, I., and E. Vasileiadou. 2015. "The Creation of the Climategate Hype in Blogs and Newspapers: Mixed Methods Approach." *Internet Research* 25 (4): 589–609.
Hopp, F. R., J. Schaffer, J. T. Fisher, and R. Weber. 2019. "iCoRe: The GDELT Interface for the Advancement of Communication Research." *Computational Communication Research* 1 (1): 13–44.
Iyengar, S., and A. Simon. 1993. "News Coverage of the Gulf Crisis and Public Opinion." *Communication Research* 20 (3): 365–383.
Kepplinger, H. M., and J. Habermeier. 1995. "The Impact of Key Events on the Presentation of Reality." *European Journal of Communication*, *10* (3), 371–390.
Kroon, A., D. Trilling, A. Fokkens, F. Loecherbach, J. Moeller, W. van Atteveldt, and M. van der Velden. 2019. "Deriving Semantics from Dutch Media Corpora: The Amsterdam Word Embedding Model." In Etmaal Van de Communicatiewetenschap, Nijmegen, Netherlands.
Kusner, M. J., Y. Sun, N. I. Kolkin, and K. Q. Weinberger. 2015. "From Word Embeddings to Document Distances." Proceedings of the 32nd International Conference on Machine Learning, 957–966.
Larsen, O. N., and R. J. Hill. 1954. "Mass Media and Interpersonal Communication in the Diffusion of a News Event." *American Sociological Review* 19 (4): 426–433.
Liu, X., E. Lindquist, and A. Vedlitz. 2011. "Explaining Media and Congressional Attention to Global Climate Change, 1969–2005: An Empirical Test of Agenda-Setting Theory." *Political Research Quarterly* 64 (2): 405–419. doi: 10.1177/1065912909346744
Marcinkowski, F. 2008. "Public Sphere, Fragmentation of." In *The International Encyclopedia of Communication*. Chichester, UK: Wiley.
McCombs, M., and A. Reynolds. 1985. "How the News Shapes Our Civic Agenda." In *Media Effects: Advances in Theory and Research*, edited by J. Bryant & M. B. Oliver, 1–16. New York, NY: Routledge.
Menchen-Trevino, E. 2016. *Web Historian: Enabling Multi-Method and Independent Research with Real-World Web Browsing History Data* (Tech. Rep.).
Mikolov, T., I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. 2013. "Distributed Representations of Words and Phrases and Their Compositionality." In Advances in Neural Information Processing Systems, 3111–3119.
Moeller, J., D. Trilling, N. Helberger, K. Irion, and C. H. de Vreese. 2016. aug). "Shrinking Core? Exploring the Differential Agenda Setting Power of Traditional and Personalized News Media." *info* 18 (6): 26–41.
Molotch, H., and M. Lester. 1974. "News as Purposive Behavior: On the Strategic Use of Routine Events, Accidents, and Scandals." *American Sociological Review* 39 (1): 101–112. doi: 10.2307/2094279

Mukerjee, S., S. Majó-Vázquez, and S. González-Bailón. 2018. "Networks of Audience Overlap in the Consumption of Digital News." *Journal of Communication* 68 (1): 26–50. doi: 10.1093/joc/jqx007

Nicholls, T., and J. Bright. 2019. "Understanding News Story Chains Using Information Retrieval and Network Clustering Techniques." *Communication Methods and Measures* 13 (1): 43–59. doi: 10.1080/19312458.2018.1536972

Novotný, V. 2018. Implementation notes for the Soft Cosine Measure. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management – CIKM '18*, 1639–1642. New York, New York, USA: ACM Press.

Oeldorf-Hirsch, A. 2018. "The Role of Engagement in Learning from Active and Incidental News Exposure on Social Media." *Mass Communication and Society* 21 (2): 225–247. doi: 10.1080/15205436.2017.1384022

Peixoto, T. P. 2014. The Graph-Tool Python Library. *figshare*. Accessed 10 September 2014. http://figshare.com/articles/graphtool/1164194

Rayson, P., and R. Garside. 2000. "Comparing Corpora Using Frequency Profiling." In Proceedings of the Workshop on Comparing Corpora.

Rehurek, R., and P. Sojka. 2010. "Software Framework for Topic Modelling with Large Corpora." In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, 45–50. Valletta, Malta: ELRA.

Rogers, E. M., and J. W. Dearing. 1988. "Agenda-Setting Research: Where Has It Been, Where is It Going?" *Annals of the International Communication Association* 11 (1): 555–594.

Schoenbach, K., E. de Waal, and E. Lauf. 2005. "Research Note: Online and Print Newspapers: Their Impact on the Extent of the Perceived Public Agenda." *European Journal of Communication* 20 (2): 245–258. doi: 10.1177/0267323105052300

Schrodt, P. A., S. G. Davis, and J. L. Weddle. 1994. "Political Science: KEDS—a Program for the Machine Coding of Event Data." *Social Science Computer Review* 12 (4): 561–587. doi: 10.1177/089443939401200408

Schrodt, P. A., and K. Leetaru. 2013. "GDELT: Global Data on Events, Location and Tone." In International Studies Association Meeting, 1979–2012.

Sidorov, G., A. Gelbukh, H. Gómez-Adorno, and D. Pinto. 2014. "Soft Similarity and Soft Cosine Measure: Similarity of Features in Vector Space Model." *Computación y Sistemas* 18 (3): 491–504.

Staab, J. F. 1990. "The Role of News Factors in News Selection: A Theoretical Reconsideration." *European Journal of Communication* 5 (4): 423–443.

Strömbäck, J. 2005. "In Search of a Standard: Four Models of Democracy and Their Normative Implications for Journalism." *Journalism Studies* 6 (3): 331–345. doi: 10.1080/14616700500131950

Traag, V. A., R. Aldecoa, and J.-C. Delvenne. 2015. "Detecting Communities Using Asymptotical Surprise." *Physical Review E* 92 (2): 22816. doi: 10.1103/physreve.92.022816

Traag, V. A., L. Waltman, and N. J. van Eck. 2019. "From Louvain to Leiden: Guaranteeing Well-Connected Communities." *Scientific Reports* 9 (1): 1–12.

Trilling, D. 2019. "Conceptualizing and Measuring News Exposure as Network of Users and News Items." In *Measuring Media Use and Exposure: Recent Developments and Challenges*, edited by C. Peter, T. K. Naab, & R. Kühne. Köln: Herbert von Halem.

Trilling, D., B. Van De Velde, A. C. Kroon, F. Locherbach, T. Araujo, J. Strycharz, … J. G. Jonkman. 2018. "INCA: Infrastructure for Content Analysis." In 2018 IEEE 14th International Conference on e-Science (e-Science), 329–330.

Van Hoof, M. 2019. *Fake News Contamination? A Study of Agenda Setting Interaction between Fake News and Tradtional Media*. Amsterdam: Universiteit van Amsterdam.

Vargo, C. J., L. Guo, and M. A. Amazeen. 2018. "The Agenda-Setting Power of Fake News: A Big Data Analysis of the Online Media Landscape from 2014 to 2016." *New Media & Society* 20 (5): 2028–2049. doi: 10.1177/1461444817712086

Vasterman, P. L. 2005. "Media-Hype: Self-Reinforcing News Waves, Journalistic Standards and the Construction of Social Problems." *European Journal of Communication* 20 (4): 508–530. doi: 10.1177/0267323105058254

Vermeer, S. 2018. A Supervised Machine Learning Method to Classify Dutch-Language News Items.

Walgrave, S., and P. Van Aelst. 2006. "The Contingency of the Mass Media's Political Agenda Setting Power: Toward a Preliminary Theory." *Journal of Communication* 56 (1): 88–109. doi: 10.1111/j.1460-2466.2006.00005.x

Welbers, K., W. van Atteveldt, J. Kleinnijenhuis, and N. Ruigrok. 2018. "A Gatekeeper among Gatekeepers." *Journalism Studies* 19 (3): 315–333. doi: 10.1080/1461670X.2016.1190663

Yang, Y.,. J. G. Carbonell, R. D. Brown, T. Pierce, B. T. Archibald, and X. Liu. 1999. "Learning Approaches for Detecting and Tracking News Events." *IEEE Intelligent Systems* 14 (4): 32–43.

Zaller, J. 2003. "A New Standard of News Quality: Burglar Alarms for the Monitorial Citizen." *Political Communication* 20 (2): 109–130. doi: 10.1080/10584600390211136