

Finding the news and mapping the links: a case study of hypertextuality in Dutch-language health news websites

Joyce Stroobant

To cite this article: Joyce Stroobant (2018): Finding the news and mapping the links: a case study of hypertextuality in Dutch-language health news websites, Information, Communication & Society, DOI: [10.1080/1369118X.2018.1477971](https://doi.org/10.1080/1369118X.2018.1477971)

To link to this article: <https://doi.org/10.1080/1369118X.2018.1477971>



© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 05 Jun 2018.



Submit your article to this journal [↗](#)



Article views: 29



View related articles [↗](#)



View Crossmark data [↗](#)

Finding the news and mapping the links: a case study of hypertextuality in Dutch-language health news websites

Joyce Stroobant 

Department of Communication Sciences, Ghent University, Ghent, Belgium

ABSTRACT

This study considers hyperlinks as digital navigational cues that can guide users through the increasingly complex and vast online health information landscape in order to examine how hypertextuality at both search engines and health news websites mediates access to further health-related information. This is important because online news media are frequently used and convenient sources for health information. The methodology unfolds in two steps. First, an environmental scan of search engine result pages for the term ‘health news’ was conducted. Second, an automated quantitative content analysis ($N=5428$) of external hyperlinks found on three types of health news websites, i.e., net-native, mixed and legacy news brands, was performed. Most importantly, this study challenges the dominant internal-external distinction by introducing a systematic distinction between genuine external hyperlinks and pseudo-external hyperlinks when comparing various types of online health news. Net-native news websites provide more hyperlinks to thematically related information than legacy news websites with print origins. The latter often include pseudo-external hyperlinks to thematically unrelated, but organizationally affiliated websites, thus favoring financial relationships over thematic coherence as an incentive to link.

ARTICLE HISTORY

Received 5 December 2017
Accepted 10 May 2018

KEYWORDS

Digital journalism; health information seeking behavior; hypertext; search engine; transparency

Introduction

The Internet is a frequently used and convenient source for health information (McDaid & Park, 2010). People seek online health information, for example, to find reassurance, to self-diagnose, or to seek general lifestyle advice (McDaid & Park, 2010; Powell, Inglis, Ronnie, & Large, 2011). Nevertheless, it has become increasingly difficult - both for health professionals and lay people - to locate reliable online health sources (Higgins, Sixsmith, Barry, & Domegan, 2011; Lee, Hoti, Hughes, & Emmerton, 2014; Macias, Lee, & Cunningham, 2017).

Contrary to the pre-Internet era when the production and distribution of medical knowledge was restricted to medical elites, the Internet provides an open platform for a wide heterogeneity of sources to independently produce and distribute health information

CONTACT Joyce Stroobant  joyce.stroobant@ugent.be  Department of Communication Sciences, Korte Meer 7-9-11, B-9000, Ghent University, Ghent, Belgium

© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

(Clarke, Shim, Mamo, Fosket, & Fishman, 2010, p. 72). The line between lay and expert knowledge, and commercial and non-commercial information is blurring as diverse organizations with varying agendas such as patient organizations, universities, federal health institutes, news media, pharmaceutical companies, physicians' personal webpages, individual patients through social networking or blogging platforms, contribute to the proliferation of online health information (Clarke et al., 2010; Hu & Shyam Sundar, 2010).

Given the exponential growth of online information, locating, editing, enriching and organizing relevant high-quality information on a specific topic into a coherent whole has become an important addition to online journalists' skillset (Bakker, 2014; Cui & Liu, 2017). Hypertextuality constitutes the most obvious means for this type of digital content curation – sometimes also pejoratively referred to as aggregation (Bardoel & Deuze, 2001; Cui & Liu, 2017). By inserting hyperlinks to source, structure and contextualize health information, news consumers can click on links and navigate through the complex online health information landscape, thus engaging in a process metaphorically referred to as 'way-finding' (Pearson & Kosicki, 2016). In this sense, online journalists' role has transformed from watchdogs to 'guidedogs' (Bardoel & Deuze, 2001, p. 94) and from gatekeepers to 'gatewatchers' who collect or aggregate news rather than produce it (Bruns, 2005, p. 17).

In order to explore hypertextual information pathways laid out by health news websites, a quantitative content analysis of external hyperlinks on nine Dutch-language health news websites will be conducted. Since online news media occupy a central position in the online health information landscape (Groselj, 2014), health news websites present interesting case studies for assessing how hypertext mediates between news audiences and health information. Not only are news media important sources for health information, individual news articles often trigger reader curiosity thus inciting additional health-information seeking (Tang & Lee, 2006). The presence of links in articles could therefore eliminate the cognitive burden on health-information seekers caused by information overload and frustration due to the inability to effectively cope with this overload (Macias et al., 2017).

Literature review

The meaning of hyperlinks in journalism

Theoretically, three main functions are attributed to hypertextuality in journalism (De Maeyer, 2012). First, hyperlinks function as a transparent sourcing mechanism by providing direct access to raw source material thus revealing otherwise covert news sourcing practices (Napoli, 2008). This function is also referred to as hyperlinks' citational function (Ryfe, Mensing, & Kelley, 2016). Second, hyperlinks to sources also enhance the perceived credibility of certain types of articles as well as readers' inclinations to seek further information (Borah, 2014; Chung, Nam, & Stefanone, 2012). Third, hyperlinks to webpages containing opposing views or different interpretations generate greater diversity of opinions in news, leading to what Gans (2011) calls 'multiperspectival news'. Ultimately, in all three of these functions, hyperlinks are metaphorical signposts that can guide news

consumers through the complex digital health information landscape (Clarke et al., 2010; Hu & Shyam Sundar, 2010; Pearson & Kosicki, 2016).

Empirical hyperlinking studies in journalism extend over three successive research traditions that roughly correspond to three theoretical perspectives (De Maeyer, 2017; Doherty, 2014), i.e., the technological-reductionist normative, technological-reductionist empirical and technological-constructivist perspective (Vobič, 2014, p. 258).

First, from the technological-reductionist normative perspective, inspired by early technological optimism surrounding the popularization and commercialization of Internet browsers in the early 1990s, scholars were preoccupied with the question of whether hyperlinks, alongside other specific features of online news such as interactivity and multi-mediality, were in fact present in online news (e.g., Barnhurst, 2002; Kenney, Gorelik, & Mwangi, 2000; O'Sullivan, 2005; Paulussen, 2004; Tankard & Ban, 1998). In these early studies, online news was reduced to, and described by, the presence or absence of technological features against a backdrop of what online journalism – in theory – could and should look like (Deuze, 2003). The overall conclusion here was that online news made scant use of new technologies and therefore did not live up to its full potential.

The second wave of hyperlinking research is characterized by a strong emphasis on the distinction between internal and external linking (Dimitrova, Connolly-Ahren, Williams, Kaid, & Reid, 2003; Engebretsen, 2006; Oblak, 2005). Internal hyperlinks that keep news consumers within the boundaries of the website are the most common type of hyperlinks on news websites. Internal hyperlinks are widely used because, contrary to external hyperlinks, they keep visitors within the same website thus safeguarding traffic and advertising revenue (Chang, Southwell, Lee, & Hong, 2012). News websites' internal linking preference is dubbed 'jurisdictional protectionist' (Chang et al., 2012), 'gated cybercommunity' (Tremayne, 2005) or 'walled garden' phenomenon (Napoli, 2008, p. 64). Since external hyperlinks, which are not associated with financial considerations, are valued more than internal ones, the second wave of research rather bleakly (and uniformly) concluded that online journalism had failed to adopt the technical affordances of the web to enhance news content (De Maeyer, 2017).

The underlying theoretical assumption here is that internal and external hypertextuality contribute in fundamentally different ways to the news content (Bardoel & Deuze, 2001; Deuze, 2003). While external hyperlinks to other offsite webpages open up the original news content, e.g., by linking to original sources in a direct and transparent manner thus boosting transparency, credibility and diversity (De Maeyer, 2012), internal hyperlinks are seen as embodiments of the commercialization and commodification of news rather than as an emanation of journalism's public service value to sustain open and democratic debate. As argued by Deuze (2003, p. 212) internal on-site hyperlinks could 'lead to a downward spiral of content' that 'tells the end-user that the "worldwide" web does not exist'.

Nevertheless, an excellent longitudinal study of hyperlinks at four leading Swedish newspapers found that external hyperlinks often 'linked to other *businesses* within the same parent corporation and also to *business* collaborations with external partners' (Karls-son, Clerwall, & Örnebring, 2015, p. 860, *italics mine*). Contrary to Deuze (2003), Karlsson et al. (2015) suggest that offsite external hyperlinks may be governed by protectionist marketing incentives, and may therefore be similar to internal hyperlinks. In order to survive the transition from print to digital publishing, newspaper industries have adopted new

business models that rely on streams of revenue unrelated to the journalistic activities of the organization such as partnerships with web shops, telecommunications and other services (Villi & Hayashi, 2015). Therefore, despite the rise of external links observed by Karlsson et al. (2015), the question remains whether external links to affiliated websites add value to the original content in terms of source transparency, credibility and diversity.

In order to meet the methodological challenge posed by external links to affiliated businesses, De Maeyer (2013), in her dissertation on the use of hyperlinks in French-speaking Belgium, applies a three-way distinction for classifying hyperlink destinations, i.e., alongside internal and external hyperlinks, she adds 'fake external links' (translated from French, 2013, p. 167). Hyperlinks can therefore be classified as internal when they refer to another page on the same website, as pseudo-external (or 'fake external') when they refer to affiliated websites, and as genuinely external when they refer to websites from other organizations.

Starting in the 2010s, a third wave of research recognizes the overall dominance of internal hyperlinks in news, but differentiates between linking behavior in different types of online journalism (Sjøvaag, Moe, & Stavelin, 2012). Progressive genuine external hyperlinking is found most in citizen journalism (Paulussen & D'heer, 2013), blogs (Coddington, 2012) and alternative net-native explanatory news websites (Cui & Liu, 2017). Rather than considering hyperlinking as an entirely novel practice, these studies identify it as a process of normalization that is shaped by existing professional journalistic norms which differ according to the type of news outlet (Coddington, 2014). The use of hyperlinks, in other words, varies according to 'the socialization of the various media types in the profession of journalism' (Cui & Liu, 2017, p. 853).

Health news websites in the online health information landscape and the centrality of search engines

Groselj (2014) who analyzed 641 websites about the top 10 searched health topics according to search engine *Yahoo!* in 2010 (i.e., breast cancer, depression, diabetes, fibromyalgia, gall bladder, herpes, HIV, Lupus, pregnancy, shingles), found that news websites constitute the second largest group of websites providing health information (21%). Yet, since news is event-driven, some topics receive more attention from news sites than others. HIV (45%) and depression (40%) receive most attention, while pregnancy (4%) and fibromyalgia (6%) are covered less. News websites share this second position with informational portals that provide systematic health information (21%). Homepages with the purpose of representing the owner and his activities (mainly non-profit, e.g., physicians, patient organizations or government) (30%) are ranked first (Groselj, 2014). The remaining 48% of the online health information landscape is occupied by eight other types of websites (Groselj, 2014).

Typically, search engines feature as a trusted and frequently used starting point for seeking online health-information (Lee et al., 2014; Powell et al., 2011). Similarly, search engines are identified as important starting points for finding online news (Newman, Fletcher, Levy, & Nielsen, 2016; Ørmen, 2016). In response to this increased use of search engines, news organizations are concerned with search engine optimization (SEO) in order to give their news content maximal online visibility (Giomelakis & Veglis, 2016). A combination of SEO and a large archive of online news articles explain why news

websites are among the top results when using search engines for health information. Nevertheless, while algorithmic gatekeeping has been criticized for being impersonal and for returning irrelevant results (Powell et al., 2011), public trust in these algorithmic gatekeeping processes remains high (Newman et al., 2016, pp. 112–114).

Research questions

Despite their wide use and perceived trustworthiness, search engines are also criticized for returning hyperlinks to irrelevant and even untrustworthy websites, especially in the context of health information (Lee et al., 2014). Relevant health news websites are, for the purpose of this study, defined as websites that provide news about health-related topics, produced by professional journalists, on a regular daily or weekly basis. Given the possible effects of the quality of online health information for public health, the following research question is proposed:

RQ1: Does a basic search in three popular search engines return relevant results for the query ‘gezondheidsnieuws’ (Eng. ‘health news’)?

Secondly, since news websites occupy a prominent position in the online health information landscape (Groselj, 2014), it is important to consider hyperlinks on those news websites as navigational objects that have the potential to mediate access to additional health information. Hyperlinks potentially add value to the original journalistic content if they link out to thematically related content, i.e., further information, source material, original scientific studies, etc. However, hyperlinking behavior on news websites is often motivated by financial considerations rather than journalistic ideals as envisioned in the first wave of hyperlinking research. The discord between commercial and journalistic incentives for using hyperlinks is often measured by quantifying the ratio of internal and external hyperlinks within websites. Despite the relative dominance of internal linking on news websites in general (Karlsson et al., 2015), previous research shows that the occurrence of internal and external hyperlinks varies according to news websites’ socialization in the principles of institutional journalism (Cui & Liu, 2017). New entrants to the field, such as blogs, citizen journalism and net-native explanatory news websites, are more likely to use external hyperlinks than news websites of legacy media brands (Coddington, 2012; Cui & Liu, 2017; Paulussen & D’heer, 2013; Sjøvaag et al., 2012).

Recent empirical hyperlinking studies, however, suggest that external hyperlinks should not be treated monolithically (De Maeyer, 2013; Karlsson et al., 2015). Besides genuine external hyperlinks, so-called pseudo-external hyperlinks, which are technically external, but similar to internal hyperlinks because they link to domains pertaining to the overarching media conglomerate, are introduced. To my knowledge, no research has been conducted on the occurrence of pseudo-external hyperlinks on different types of news websites. Drawing on the arguments made above, the following research questions are put forward:

RQ2: To what extent can technically external hyperlinks on different types of health news websites be categorized as pseudo-external hyperlinks?

RQ3: To what extent do external hyperlinks on different types of health news websites refer to health-related content?

In an attempt to uncover whether pseudo-external hyperlinks are, like internal hyperlinks, commercially inspired, this study will measure whether hyperlinks to health-related content are either genuinely external or pseudo-external. Finally, in order to get a more fine-grained assessment of how hyperlinks mediate access to further health-related content, this paper will also identify the originators of the linked content, e.g., non-profits, patient organizations, pharmaceutical industries, etc.

RQ4a: To what extent do pseudo-external hyperlinks refer to health-related content compared to genuine external hyperlinks?

RQ4b: If health related content is linked to, who is the originator of that content?

Method

Environmental scan to identify health news websites

In order to identify health news websites as they might be encountered by a news audience that increasingly relies on search engines to locate news (Newman et al., 2016), a two-step environmental scan of search engine result pages was conducted (Bowler, Hong, & He, 2011). First, a keyword search was performed for the term ‘gezondheidsnieuws’ (Eng. ‘health news’) in three leading search engines (i.e., *Bing*, *Google* and *Yahoo!*). Since this research is situated in Flanders, the Dutch-speaking part of Belgium, the search queries were also performed in Dutch. Secondly, all URLs on the first five search engine result pages (henceforth, ‘SERPs’) were manually collected and coded using SPSS. Collection of URLs was limited to the first five SERPs as a measure of practicality. *Google* returned more search results per SERP due to the inclusion of sponsored results, i.e., 16 rather than 10–12, hence the slight overweight of *Google* search results in the sample. Yet, given *Google*’s overwhelming domination of the search engine market (i.e., in Europe, mobile and desktop searches combined, *Google* holds 91.78% of the market), this was not deemed problematic (StatCounter, 2017).

Besides providing insight in the usefulness of search engines for finding health news (RQ1), results of the environmental scan were used as a starting point for further analysis of the use of hyperlinks on health news websites. In other words, drawing on the environmental scan, the sample of health news websites under scrutiny in this study, includes nine health news websites that pertain to three categories of news websites with varying roots in the institution of journalism¹: (1) legacy media brands (A, B and C), (2) net-native websites by publishers with additional print brands (D, E and F), (3) net-native websites by publishers without print activities (G, H and I) (cf. Table 1).

Researching SERPs, nevertheless, poses several methodological challenges due to their peculiar ontological status as research objects (Ørmen, 2016). Search results are, unlike Tweets or news articles, created in the act of searching and are increasingly personalized to the preferences of the searchers, in this case the researcher. Consequently, the research object does not exist outside the research. Firstly, personalization of search results was avoided by using private (or ‘incognito’) browsing modes. This study used *Google Chrome*’s incognito browser (version 55.0.2883). Private or incognito browsing allows to surf the web without leaving behind any traces of previous browsing history, cookies, passwords, favorites or bookmarks. Secondly, browser add-ons providing services such

Table 1. Overview of seed websites, individual external hyperlinks and distinct websites after page grouping.

Seed website	Number of pages crawled	External hyperlinks (5428)	Number of websites (254)
[A] hln.be/subdirectory ^a <i>De Persgroep</i>	2	40	26
[B] telegraaf.nl/subdirectory ^a <i>Telegraaf Media Groep</i>	23	100	26
[C] knack.be/subdirectory ^a <i>Roularta Media Group</i>	18	167	34
[D] nu.nl/subdirectory <i>SANOMA</i>	21	92	22
[E] gezondnu.nl <i>MixCom Media</i>	319	811	23
[F] gezondheidsnet.nl <i>Senior Publications Nederland</i>	702	721	40
[G] e-gezondheid.be <i>E-Santé</i>	1328	1682	24
[H] gezondheid.be <i>Rossel/Mediahuis</i>	453	504	39
[I] gezondheidsnederland.nl <i>CEBAM</i>	1284	1311	20
Total	4150	5428	254

^aFor legacy websites, only subdirectories related to health were used as starting point for further analysis.

Cursive indicates publisher.

as the blocking of advertisements (e.g., *Adblock Plus*) or the evaluation the safety of search results (e.g., *Web of Trust*) were disabled. While these strategies may effectively overcome some issues of personalization (i.e., the researcher's location could still be determined based on the IP-address which was not anonymized), it may run the risk of becoming 'too artificial and detached from real-world search situations' (Ørmen, 2016, p. 118). Yet, because the aim of this paper is not to make generalizations about the search algorithms of the various search engines, the validity of the search results does not downplay the reliability and usefulness of the collected sample as a basis for analysis in this case study of online health news websites (Karpf, 2012).

Hyperlink mapping of nine health news websites

The collection of the hyperlinks was performed automatically using web-based software VOSON (Ackland, 2011). The software was used to crawl the nine health news websites (called 'seeds') identified through the environmental scan and to harvest all external hyperlinks encountered on the seed page and on pages one click away from the seeds² (cf. Table 1 for an overview of crawled websites). Internal hyperlinks were not included because external hyperlinks, as argued by Deuze (2003), show the greatest theoretical potential added value for journalistic content regarding transparency, diversity and credibility. Additionally, previous research has already convincingly illustrated the overall dominance of internal hyperlinks in news websites (e.g., Karlsson et al., 2015).

After crawling each individual seed, a total dataset of 5428 unique hyperlinks was composed. Yet, rather than coding and visualizing 5428 individual pages, the dataset was cleaned via a page grouping process. This entails that hyperlinks referring to the same website were grouped together. Due to technical errors with the VOSON software related to mismatches between <http://> and newer secure <https://>-protocols (R. Ackland, personal

communication, February 20, 2017), page grouping was performed manually. Crawling news websites can also pose difficulties for crawlers because news websites are exceptionally rich in content and have very extensive sitemaps. After page grouping, the initial 5428 external URLs were reduced to 254 individual websites.

In this attempt to assess how external hyperlinks on health news websites guide users to other health-related information on the net, the 254 identified websites were content analyzed using SPSS and visualized in a networked structure using Gephi. Firstly, it was identified whether hyperlinks were genuinely external or pseudo-external (RQ2). Exhaustive lists of all domains owned by the nine publishers in the sample were composed in order to complete this task. Secondly, the websites were coded according to whether or not they contained health-related content (RQ3, RQ4a). Thirdly, in order to get a more complete picture of the type of information that was linked to, further distinctions relating to the providers of the information were made (RQ4b). This categorization takes into account the heterogeneity that characterizes the contemporary biomedical field (Clarke et al., 2010): (1) policy-makers, (2) government institutions, (3) news websites, (4) social media, (5) industry, (6) sickness funds, (7) consumer organizations, (8) patient support and advocacy groups, (9) academic, (10) associations of health professionals and hospitals, (11) personal or community websites, e.g., blogs, *Wikipedia*, (12) civil society and non-profit health information providers, e.g., *Red Cross*, and a final miscellaneous category 'other', e.g., information about cookies, links to app-stores such as *Google Play* (13). Intra-coder reliability for the coding was measured using Cohen's kappa coefficient and ranged between 0.87 and 0.96 which indicates good agreement (Neuendorf, 2002). Statistical software SPSS was used for cross-tabulations, Chi-square analysis and further post-hoc testing, i.e., pairwise column comparisons and adjusted standardized Pearson residuals (Beasley & Schumacker, 1995).

Results

Finding health news using search engines

Firstly, out of 180 search results retrieved from the first 5 SERPs of 3 popular search engines, only 21 hyperlinks, which led to 9 individual websites, linked to authentic and up-to-date health news websites. Secondly, the most frequent results provided commercially biased health 'news'. In fact, the most popular search result, albeit absent from *Yahoo!*, linked to a network of self-employed pharmacists. This website mainly contains business information for pharmacists, but also has a small category of irregularly updated health news highlighting a certain product or service. Another type of bias was found in self-proclaimed health news websites and blogs containing conspiracy theories. Authors of these websites transparently communicate their intention to act as guard dogs of health. Yet, rooted in negative personal experiences with traditional medicine they harbor a deep mistrust towards governments, pharmaceutical, food and chemical industries. They collect and write stories that confirm these previously held beliefs, e.g., about the dangers of artificial sweeteners, and publish this on their website. This type of bias is called confirmation bias (Leman, 2007).

It is, however, important to distinguish between biased information and biased information that is intentionally deceptive. All three search engines returned links to five

different fake news websites promoting the same contested herbal weight-loss supplement, *Garcinia Cambogia* (Astell, Mathai, & Su, 2013). These websites systematically presented homepages with the header ‘gezondheidsnieuws’ (Eng., ‘health news’) and various other news categories (e.g., mental health, sexual health, fitness, etc.), yet these categories are just a gimmick because they are images rather than links to other news stories. None of the surrounding text provides links to a profile page nor to legitimizing information on other websites despite the presence of the logos of other news brands, i.e., ‘gezondheidsnet’ and ‘telegraaf’, suggesting endorsement by these brands. In fact, these websites did not contain any links except for a link to a webpage where the product that is being appraised can be purchased.

Overall this means that for this sample only 14.4% of search results referred to relevant health news websites (RQ1). Nevertheless, since most user searches are often limited to the first page of search results (Macias et al., 2017), more important than the frequency of a particular search result is its rank. *Google* provides six different relevant health news websites on its first SERP, *Bing* shows five, but *Yahoo!* only shows one relevant search result on the first SERP. If we continue to the second page, *Google* and *Yahoo!* add one more relevant search result, while *Bing* adds two more thus equaling *Google*. Generally speaking in terms of ranks and relevance, *Bing* and *Google* yield the best results (Table 2). All three search engines provided a hyperlink to one of the websites that aggressively promote *Garcinia Cambogia* on the first SERP. *Yahoo!* returned the lowest number of relevant results, but also the lowest number of websites that contain misinformation.

Mapping external hyperlinks in health news websites

Overall, 33.5% of the technically external domains belong to the portfolio of the media conglomerate to which the seed website pertains. As illustrated in Table 3, there is a significant interaction between the occurrence of hyperlinks to pseudo-external domains and the type of website ($\chi^2(2, N = 254) = 90.16, p = .05$). Further pairwise comparisons of column proportions using a *z*-test with Bonferroni adjustments to the alpha-level of 0.05, indicates a statistically significant difference between the legacy news websites and the remaining two categories, but no statistical difference between the occurrence of pseudo-external hyperlinks in net-native and mixed tradition websites. The adjusted standardized Pearson residuals indicate that legacy news websites explain most of the observed

Table 2. Overview of health news websites retrieved via environmental scan per search engine.

Bing	Google	Yahoo!
(2) gezondnu.nl	(3) gezondheid.be	(2) gezondheid.be
(3) nu.nl	(4) hln.be	(11) nu.nl
(4) gezondheidsnawetenschap.be	(5) nu.nl	(40) gezondheidsnet.nl
(5) gezondheidsnet.nl	(6) gezondheidsnet.nl	(44) gezondheidsnet.nl
(8) e-gezondheid.nl	(7) gezondheidsnawetenschap.be	(47) gezondheidsnet.nl
(12) gezondheid.be	(9) knack.be	
(13) knack.be	(15) e-gezondheid.be ^a	
(14) gezondheidsnet.nl	(18) e-gezondheid.be ^a	
(40) telegraaf.nl	(32) e-gezondheid.be ^a	
(48) telegraaf.nl	(49) e-gezondheid.be ^a	
	(65) e-gezondheid.be ^a	

^aIndicates sponsored result.

(#) Bracketed number indicates rank on SERP.

Table 3. Cross-tabulation for type of external hyperlink by type of health news website expressed in % ($N = 254$).

Domain	Website type			Total
	Net-native	Mixed	Legacy	
Pseudo-external	13.3 _a (-4.76)	13.1 _a (-4.84)	72.4 _b (9.5)	33.5
Genuinely external	86.7 _a (4.76)	86.9 _a (4.84)	27.6 _b (-9.5)	66.5

$$\chi^2(2, N = 254) = 90.16, p = .05.$$

Notes: Values in the same row not sharing the same subscript are significantly different at the adjusted alpha-level using the Bonferroni method. Bracketed numbers below column proportions represent the adjusted residuals. All adjusted residuals are significantly different at the adjusted alpha-level using Bonferroni method. Critical z value = 2.64.

variance. In other words, linking behavior of legacy news brands significantly differs from brands that originated online. The website publisher's involvement in other print activities alongside the net-native health news website, has no significant impact. Both net-native health news websites and websites with a mixed tradition link to pseudo-external domains less than legacy news websites.

Even though the crawl was limited to health news, only 52.4% of the linked to domains were health-related. Pairwise comparisons of column proportions indicate a statistically significant difference between legacy websites, one the one hand, and the websites in mixed tradition and net-native websites, on the other hand ($\chi^2(2, N = 254) = 72.05, p = .05$). The adjusted standardized Pearson residuals again demonstrate that legacy news websites explain most of the variance (cf. Table 4). Furthermore, there appears to be a statistically significant difference between the occurrence of health-related hyperlinks and the type of hyperlink. As demonstrated in Table 5, only 24.7% of pseudo-external hyperlinks refer to health-related content, contrary to 66.3% of genuine external hyperlinks.

Of those 52.4% of health-related domains, Table 6 illustrates that almost one in five domains represent other news websites (19.5%). Yet, e-shopping websites (15%) are also frequent. These websites sell products with health-advancing characteristics, but which are not pharmaceuticals. For example, beauty products with nursing aspects, food supplements, ergonomic mattresses to alleviate back pain or organic foods. Other linked to domains pertain to established sources: associations of medical professionals and hospitals (15%), government (14.3%) and academic sources (12%).

Despite thematic coherence of the crawl, Figure 1 illustrates that the websites in the sample, represented by various nodes, are only loosely connected. Using graph theory

Table 4. Cross-tabulation for health-related hyperlinks by type of health news website expressed in % ($N = 254$).

Health-relatedness	Website type			Total
	Net-native	Mixed	Legacy	
No	22.9 _a (-5.50)	34.5 _a (-2.94)	83.9 _b (8.35)	47.6
Yes	77.1 _a (5.5)	65.5 _a (2.94)	16.1 _b (-8.35)	52.4

$$\chi^2(2, N = 254) = 72.05, p = .05.$$

Notes: Values in the same row not sharing the same subscript are significantly different at the adjusted alpha-level using the Bonferroni method. Bracketed numbers below column proportions represent the adjusted residuals. All adjusted residuals are significantly different at the adjusted alpha-level using Bonferroni method. Critical z value = 2.64.

Table 5. Cross-tabulation of (pseudo-) external hyperlinks and health-relatedness (expressed in %, $N = 254$).

Hyperlink type	Health-relatedness	
	Yes	No
Pseudo-external ($n = 85$)	24.7	75.3
Genuinely external ($n = 169$)	66.3	33.7

$\chi^2 (1, N = 254) = 39.17, p = .05.$

terminology, one could say that the mapped hyperlink network of health news websites is low in density. Conventionally, density is expressed as a number from 1 to 0. A complete graph in which each individual node is connected to every other node in the graph would have 100% density, i.e., graph density of 1 (Scott, 2017, p. 81). The density of the graph in Figure 1 is 0.009. Every node is in some way connected to the network, hence the network does not contain isolates. Nevertheless, website B, *telegraaf.nl*, is incidentally connected to the network through a single shared domain (i.e., *ec.europa.eu*) with *e-gezondheid.be*. Therefore, *telegraaf.nl* is not an isolate.

The north–south orientation of Figure 1 is of no importance. What is important, is the distance between the nodes and their centrality. A force-based algorithm (ForceAtlas2), in which linked nodes attract and non-linked nodes repel each other, was used for the

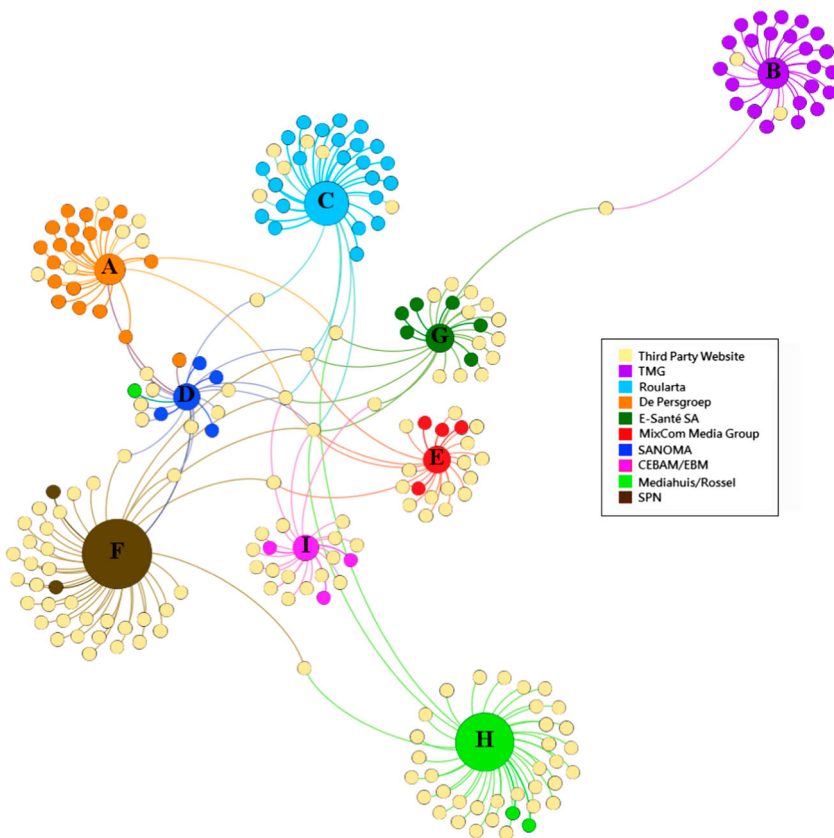
**Figure 1.** Hyperlink map indicating individual health news websites' affiliation with overarching media conglomerates.

Table 6. Who provides the linked to health information? (expressed in %, $n = 133$).

News websites	Commercial	Associations of health professionals and hospitals	Government	Academic	Patient support and advocacy	Civil society and non-profit	Sickness funds	Personal homepage or community website	Consumer organization	Other	Social media
19.5	15	15	14.3	12	9	7.5	2.3	2.3	1.5	0.8	0.8

spatialization of the nodes (Jacomy, Venturini, Heymann, & Bastian, 2014). A clear divide can be observed between, on the one hand, the legacy news websites in the upper part of the graph, and, on the other hand, net-native and mixed tradition news websites. The spatialization thus corresponds to tendencies found in the contingency tables above whereby the mixed tradition websites sometimes lean towards the linking behavior of legacy websites and other times to that of net-native websites.

The black and grey shaded nodes, which represent organizational affiliations of the nine media conglomerates, illustrate that websites A, B and C (i.e., the legacy websites) each form a hub of websites pertaining to the respective overarching media conglomerates. There are no direct links between the seeds except for *nu.nl* (D) which links out to *hln.be* (A) and to *gezondheidsnet.nl* (F). Furthermore, *nu.nl* is also the only website in the sample that links to competing news websites; i.e., *trouw.nl* (*De Persgroep*) and *standaard.be* (*Mediahuis*). The general absence of hyperlinks to the competition combined with a relatively low number of health-related websites (cf. Table 4) contributes to the graph's low density. Hyperlinks to social media, mainly *Twitter*, *Facebook*, *Pinterest* and *Google Plus* hold this network together. It is common for news websites to include social media handles above every news article to increase traffic. In other words, despite thematic coherence between the crawled websites, it is unlikely that news users will encounter content from other online health news sites if they directly go to a specific health news website as opposed to using search engines for locating health news.

Conclusion and discussion

Drawing on an automated content analysis of hyperlinks on nine different health news websites that were identified via search engines, this study examined how hyperlinks mediate access to online health information in search engines and health news websites across three different types of news websites, i.e., legacy news websites, net-native websites with a mixed tradition and exclusively net-native websites. Viewing hyperlinking as a process of normalization shaped by professional journalistic norms, this paper adopts a technological constructivist approach to describe patterns of hypertextuality at different kinds of health news websites (Coddington, 2014; Cui & Liu, 2017; Vobič, 2014).

Most importantly, this paper demonstrates that the distinction between genuine and pseudo-external hyperlinks is meaningful for journalism because pseudo-external hyperlinks less frequently present health-related information than genuine external hyperlinks. The occurrence of pseudo-external hyperlinks, therefore, resonates well with the financial motivations underlying the occurrence of internal hyperlinks. Both pseudo-external hyperlinks and internal hyperlinks reflect protectionist linking strategies that run counter to the normative ideals of diversity and source transparency as envisioned by early online journalism scholars (Deuze, 2003). It is paramount that future quantitative analyses move beyond the traditional internal-external distinction to include pseudo-external hyperlinks that refer to other websites in the portfolio of overarching media conglomerates and compare them to internal hyperlinks.

Furthermore, the inclusion of external hyperlinks to sources, original documents and raw materials, as opposed to pseudo-external hyperlinks to thematically unrelated content, encourages independent evaluation of news content (Kovach & Rosenstiel, 2001). As a ritual of transparency hyperlinks can increase credibility and trust in journalism

(Chung et al., 2012; Karlsson, 2010). As recent surveys indicate, trust in journalism is low (European Broadcasting Union [EBU], 2017; Swift, 2017). Hyperlinks, therefore, serve a higher – and much needed – goal for the survival of journalism in general. But also for health reporting in particular, since the latter constitutes a journalistic niche that is prone to rely on ‘linkable’ documents such as peer-reviewed journal articles or statistical reports to support claims.

Secondly, this study provides supporting evidence for the idea that news websites’ hyperlinking behavior varies according to the type of news website and its presumed associated socialization within institutional journalism (Coddington, 2012, 2014; Cui & Liu, 2017; Sjøvaag et al., 2012). The results suggest that media brands originating in print behave differently than those originating online, since no significant differences were found between net-native brands by publishers engaging in print and digital activities, and net-native brands by publishers who exclusively operate online.

This is also illustrated by the spatialization of the health news websites in the hyperlink map (Figure 1). The legacy websites are clearly pushed towards the periphery of the network and make scant use of genuine external hyperlinks (Table 3). Additionally, contrary to what cross-tabulations tell us, a visualization of the network shows that both legacy and net-native news media refrain from establishing connections amongst each other. Only one health news website directly links to other competing health news websites. Therefore, despite variations in hyperlinking patterns between legacy brands and net-native brands, the overall absence of direct links to competing health news websites suggests that such hyperlinks are considered as competitive threats by nearly all news websites in the sample (Vobič, 2014). Nevertheless, while this might be true in the short-term, Weber (2012) has demonstrated that, in the long run, establishing hyperlink ties with other (younger) news organizations strengthens the position of that organization in the network thus boosting traffic.

Interestingly, as illustrated in table 5, hyperlinks often refer to other news media for presenting additional health-related information. While this finding might seem to contradict the idea that news websites’ commercial incentives discourage redirecting the end-users’ attention towards competing news outlets, closer inspection reveals that the linked to news websites are international and mostly in English or French. In other words, these news websites are not considered as direct competition in the marketplace of attention (Pearson & Kosicki, 2016; Webster, 2012).

Finally, this study illustrates that using search engines as starting points for navigating online health information landscapes is problematic. The SERPs are fraught with biased, unreliable and commercial websites. Nevertheless, while commercial biases can be relatively easily identified by checking an organization’s ‘about us-webpage’, recognizing various types of misinformation or fake news such as, conspiracy websites and aggressive promotional content disguised as genuine news to boost credibility, is much harder (Tandoc, Lim, & Ling, 2017). The environmental scan also highlights the need for broadening the scope of research into misinformation, and more recently fake news, to include topics such as health alongside political topics. Given the obstacles encountered in this case study and since large-scale surveys emphasize the widespread reliance on search engines for news and health information (McDaid & Park, 2010; Newman et al., 2016), future research seems warranted. Yet, so far, investigations of SERPs for news are still limited (Ørmen, 2016).

Notes

1. For more information on the classification and on the organizational background of these websites, please contact author.
2. For all crawled seeds the *robots.txt*-file was checked to ensure the crawler was not disallowed to visit (parts of) the seed website. The robots.txt-file is publicly available and can be viewed by adding '/robots.txt' after the website URL, e.g., <http://www.bbc.com/robots.txt>. With this Robots Exclusion Protocol web designers provide explicit instructions for bots. For example, search engines use bots to crawl and index the Internet's webpages. Web designers may refuse bots from certain areas, or impose a crawl delay in order not to overload the server. Nevertheless, some bots can still ignore the robots.txt, e.g., malware bots harvesting e-mail addresses (<http://www.robotstxt.org>).

Disclosure statement

No potential conflict of interest was reported by the author.

Funding

This work was supported by the Bijzonder Onderzoeksfonds (Special Research Fund) under Grant BOFGOA 2014 000 604 '(De)constructing Health News'.

Notes on contributor

Joyce Stroobant is a PhD student at Ghent University (Department of Communication Sciences) and a member of the Center for Journalism Studies and of the interdisciplinary research group Health, Media and Society. Her research focuses on journalistic sourcing practices for health news in a rapidly changing digital environment. A recent peer-reviewed publication is 'Tracing the Sources', published in *Journalism Practice* [email: joyce.stroobant@ugent.be].

ORCID

Joyce Stroobant  <http://orcid.org/0000-0003-1460-748X>

References

- Ackland, R. (2011). Chapter 12: WWW hyperlink networks. In D. Hansen, B. Shneiderman, & M. Smith (Eds.), *Analyzing social media networks with NodeXL: Insights from a connected world* (pp. 181–200). Burlington, MA: Morgan-Kaufmann.
- Astell, K. J., Mathai, M. L., & Su, X. Q. (2013). Plant extracts with appetite suppressing properties for body weight control: A systematic review of double blind randomized controlled clinical trials. *Complementary Therapies in Medicine*, 21, 407–416.
- Bakker, P. (2014). Mr. Gates returns: Curation, community management and other new roles for journalists. *Journalism Studies*, 15(5), 596–606.
- Bardoel, J., & Deuze, M. (2001). Network journalism: Converging competences of media professionals and professionalism. *Australian Journalism Review*, 23(2), 91–103.
- Barnhurst, K. G. (2002). News geography & monopoly: The form of reports on US newspaper internet sites. *Journalism Studies*, 3(4), 477–489.
- Beasley, T. M., & Schumacker, R. E. (1995). Multiple regression approach to analyzing contingency tables: Post hoc and planned comparison procedures. *The Journal of Experimental Education*, 64(1), 79–93.

- Borah, P. (2014). The hyperlinked world: A look at how the interactions of news frames and hyperlinks influence news credibility and willingness to seek further information. *Journal of Computer-mediated Communication*, 19, 576–590.
- Bowler, L., Hong, W.-Y., & He, D. (2011). The visibility of health web portals for teens: A hyperlink analysis. *Online Information Review*, 35(3), 443–470.
- Bruns, A. (2005). *Gatewatching: Collaborative online news production*. New York, NY: Peter Lang.
- Chang, T.-K., Southwell, B. G., Lee H.-M., & Hong, Y. (2012). Jurisdictional protectionism in online news: American journalists and their perceptions of hyperlinks. *New Media & Society*, 14(4), 684–700.
- Chung, C. J., Nam, Y., & Stefanone, M. A. (2012). Exploring online news credibility: The relative influence of traditional and technological factors. *Journal of Computer-mediated Communication*, 17, 171–186.
- Clarke, A. E., Shim, J. K., Mamo L., Fosket, J. R., & Fishman, J. R. (2010). Biomedicalization: Technoscientific transformation of health illness, and U.S. Biomedicine. In A. E. Clarke, L. Mamo, J. R. Fosket, J. R. Fishman, & J. K. Shim (Eds.), *Biomedicalization: Technoscience, health and illness in the U.S* (pp. 47–87). Durham, NC: Duke University Press.
- Coddington, M. (2012). Building frames link by link: The linking practices of blogs and news sites. *International Journal of Communication*, 6, 2007–2026.
- Coddington, M. (2014). Normalizing the hyperlink: How bloggers, professional journalists, and institutions shape linking values. *Digital Journalism*, 2(2), 140–155.
- Cui, X., & Liu, Y. (2017). How does online news curate linked sources? A content analysis of three online news media. *Journalism: Theory, Practice & Criticism*, 18(7), 852–870.
- De Maeyer, J. (2012). The journalistic hyperlink: Prescriptive discourses about linking in online news. *Journalism Practice*, 6(5–6), 692–701.
- De Maeyer, J. (2013). *L'usage journalistique des liens hypertextes. Étude des représentations, contenus et pratiques à partir des sites d'information de la presse belge francophone* [Journalistic use of hyperlinks: A study of representation, contents and practices on news websites from the French-speaking Belgian press] (PhD thesis). Université Libre de Bruxelles, Belgium.
- De Maeyer, J. (2017). Journalists' uses of hypertext. In B. Franklin & S. ElderidgeII (Eds.), *The Routledge companion to digital journalism studies* (pp. 302–310). New York: Routledge.
- Deuze, M. (2003). The web and its journalisms: Considering the consequences of different types of newsmedia online. *New Media & Society*, 5(2), 203–230.
- Dimitrova, D. V., Connolly-Ahren, C., Williams, A. P., Kaid, L. L., & Reid, A. (2003). Hyperlinking as gatekeeping: Online newspaper coverage of the execution of an American terrorist. *Journalism Studies*, 4(3), 401–414.
- Doherty, S. (2014). Hypertext and journalism: Paths for future research. *Digital Journalism*, 2(2), 124–139.
- EBU. (2017, May). Trust gap between traditional and new media widening across Europe. *Market insights: Trust in media 2017*. Retrieved from <https://www.ebu.ch/news/2017/05/trust-gap-between-traditional-and-new-media-widening-across-europe>
- Engebretsen, M. (2006). Shallow and static or deep and dynamic? Studying the state of online journalism in Scandinavia. *Nordicom Review*, 27(1), 3–16.
- Gans, H. J. (2011). Multiperspectival news revisited: Journalism and representative democracy. *Journalism: Theory, Practice & Criticism*, 12(1), 3–13.
- Giomelakis, D., & Veglis, A. (2016). Investigating search engine optimization factors in media websites. *Digital Journalism*, 4(3), 379–400.
- Groeselj, D. (2014). A webometric analysis of online health information: Sponsorship, platform type and link structures. *Online Information Review*, 38(2), 209–231.
- Higgins, O., Sixsmith, J., Barry M. M., & Domegan, C. (2011). *A literature review on health information-seeking behaviour on the web: A health consumer and health professional perspective*. Stockholm: ECDC.
- Hu, Y., & Shyam Sundar, S. (2010). Effects of online health sources on credibility and behavioral intentions. *Communication Research*, 37(1), 105–132.

- Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS ONE*, 9(6), e98679. doi:10.1371/journal.pone.0098679
- Karlsson, M. (2010). Rituals of transparency: Evaluating online news outlets' uses of transparency rituals in the United States, United Kingdom and Sweden. *Journalism Studies*, 11(4), 535–545.
- Karlsson, M., Clerwall, C., & Örnebring, H. (2015). Hyperlinking practices in Swedish online news 2007–2013: The rise, fall, and stagnation of hyperlinking as a journalistic tool. *Information, Communication & Society*, 18(7), 847–863.
- Karpf, D. (2012). Social science research methods in internet time. *Information, Communication & Society*, 15(5), 639–661.
- Kenney, K., Gorelik, A., & Mwangi, S. (2000). Interactive features of online newspapers. *First Monday*, 5(1–3). doi:10.5210/fm.v5i1.720
- Kovach, B., & Rosenstiel, T. (2001). *The elements of journalism: What newspeople should know and the public should expect*. New York, NY: Three Rivers Press.
- Lee, K., Hoti, K., Hughes, J. D., & Emmerton, L. (2014). Dr Google and the consumer: A qualitative study exploring the navigational needs and online health-information seeking behaviors of consumers with chronic health conditions. *Journal of Medical Internet Research*, 16(12), e262. doi:10.2196/jmir.3706
- Leman, P. (2007). The born conspiracy. *New Scientist*, 195(2612), 35–37.
- Macias, W., Lee, M., & Cunningham, N. (2017). Inside the mind of the online health information searcher using think-aloud protocol. *Health Communication*. Advance online publication. doi:10.1080/10410236.2017.1372040
- McDaid, D., & Park, A. L. (2010). Online health: Untangling the web. Evidence from the Bupa health pulse 2010 international healthcare survey. Retrieved from https://www.bupa.com.au/staticfiles/Bupa/HealthAndWellness/MediaFiles/PDF/LSE_Report_Online_Health.pdf
- Napoli, P. M. (2008). Hyperlinking and the forces of massification. In J. Turow & L. Tsui (Eds.), *The hyperlinked society: Questioning connections in the digital Age* (pp. 57–70). Ann Arbor: University of Michigan Press.
- Neuendorf, K. (2002). *The content analysis guidebook*. Thousand Oaks, CA: Sage.
- Newman, N., Fletcher, R., Levy, D. A. L., & Nielsen, R. K. (2016). Reuters institute digital news report 2016. Retrieved from <http://reutersinstitute.politics.ox.ac.uk/sites/default/files/research/files/Digital%2520News%2520Report%25202016.pdf>
- Oblak, T. (2005). The lack of interactivity and hypertextuality in online media. *International Communication Gazette*, 67(1), 87–106.
- Ørmen, J. (2016). Googling the news: Opportunities and challenges in studying news events through Google search. *Digital Journalism*, 4(1), 107–124.
- O'Sullivan, J. (2005). Delivering Ireland: Journalism's search for a role online. *International Communication Gazette*, 67(1), 45–68.
- Paulussen, S. (2004). Online news production in Flanders: How Flemish online journalists perceive and explore the internet's potential. *Journal of Computer-mediated Communication*, 9(4), doi:10.1111/j.1083-6101.2004.tb00300.x
- Paulussen, S., & D'heer, E. (2013). Using citizens for community journalism. *Journalism Practice*, 7(5), 588–603.
- Pearson, G. D. H., & Kosicki, G. M. (2016). How way-finding is challenging traditional gatekeeping in the digital age. *Journalism Studies*, doi:10.1080/1461670X.2015.1123112
- Powell, J., Inglis, N., Ronnie, J., & Large, S. (2011). The characteristics and motivations of online health information seekers: Cross-sectional survey and qualitative interview study. *Journal of Medical Internet Research*, 13(1), e20. doi:10.2196/jmir.1600
- Ryfe, D., Mensing, D., & Kelley, R. (2016). What is the meaning of a news link? *Digital Journalism*, 4(1), 41–54.
- Scott, J. (2017). *Social network analysis* (4th ed.). London: Sage.
- Sjøvaag, H., Moe, H., & Stavelin, E. (2012). Public service news on the web. *Journalism Studies*, 13(1), 90–106.

- StatCounter Global Stats. (2017). Search engine market share in Europe: Sept 2016–Sept 2017. Retrieved from <http://gs.statcounter.com/search-engine-market-share/all/europe>
- Swift, A. (2017). In U.S., confidence in newspapers still low but rising. *Gallup Poll*. Retrieved from <http://news.gallup.com/poll/212852/confidence-newspapers-low-rising.aspx>
- Tandoc, E. C., Lim, Z. W., & Ling, R. (2017). Defining ‘fake news’. *Digital Journalism*, doi:10.1080/21670811.2017.1360143
- Tang, E., & Lee, W. (2006). Singapore internet user’s health information search: Motivation, perception of information sources and self-efficacy. In M. Murero & R. E. Rice (Eds.), *The internet and health care: Theory, research, and practice* (pp. 107–126). Mahwah, NJ: Lawrence Erlbaum Associates.
- Tankard, J. W., & Ban, H. (1998, August 5–8). *Online newspapers: Living up to the potential?* In *AEJMC conference*, Baltimore, MD.
- Tremayne, M. (2005). News websites as gated cybercommunities. *Convergence: The International Journal of Research Into New Media Technologies*, 11(3), 28–39.
- Villi, M., & Hayashi, K. (2015). The mission is to keep this industry intact. *Journalism Studies*, doi:10.1080/1461670X.2015.1110499
- Vobič, I. (2014). Practice of hypertext: Insights from the online departments of two Slovenian newspapers. *Journalism Practice*, 8(4), 357–372.
- Weber, M. S. (2012). Newspapers and the long-term implications of hyperlinking. *Journal of Computer-Mediated Communication*, 17(2), 187–201.
- Webster, J. (2012). Structuring a marketplace of attention. In J. Turow & L. Tsui (Eds.), *The hyper-linked society* (pp. 23–38). Ann Arbor: University of Michigan Press.