Microblogging as an extension of science reporting

Büchi, Moritz

Abstract: Mass media have long provided general publics with science news. New media like Twitter have entered this system and provide an additional platform for the dissemination of science information. Based on automated collection and analysis of more than 900 news articles and 70,000 tweets, this study explores the online communication of current science news. Topic modeling (latent Dirichlet allocation) was used to extract five broad themes of science reporting: space missions, the US government shutdown, cancer research, Nobel Prizes and climate change. Using content and network analysis, Twitter was found to extend public science communication by providing additional voices and contextualizations of science issues. It serves a recommender role by linking to web resources, by connecting users, and by directing users' attention. The paper suggests that microblogging adds a new and relevant layer to the public communication of science.

DOI: https://doi.org/10.1177/0963662516657794

Originally published at:
Büchi, Moritz (2017). Microblogging as an extension of science reporting. Public Understanding of Science, 26(8):953-968.
DOI: https://doi.org/10.1177/0963662516657794

POSTPRINT VERSION OF:

Büchi, M. (2016). Microblogging as an extension of science reporting. *Public Understanding of Science*. Advance online publication. http://dx.doi.org/10.1177/0963662516657794

Microblogging as an extension of science reporting

Moritz Büchi

m.buechi@ipmz.uzh.ch

University of Zurich, Switzerland

Abstract

Mass media have long provided general publics with science news. New media like Twitter have entered this system and provide an additional platform for the dissemination of science information. Based on automated collection and analysis of more than 900 news articles and 70,000 tweets, this study explores the online communication of current science news. Topic modeling (latent Dirichlet allocation) was used to extract five broad themes of science reporting: space missions, the US government shutdown, cancer research, Nobel Prizes and climate change. Using content and network analysis, Twitter was found to extend public science communication by providing additional voices and contextualizations of science issues. It serves a recommender role by linking to web resources, by connecting users, and by directing users' attention. The paper suggests that microblogging adds a new and relevant layer to the public communication of science.

Introduction

Mass media and the internet are the main sources of science information and research-based knowledge for a general public (Wade & Schramm, 1969; Horrigan, 2006; Brossard, 2013; National Science Board, 2014). Science news outlets thus constitute an important interface between the scientific community and a broader public. Twitter, the dominant microblogging platform (see van Dijck, 2011), functions as an 'ambient news network' (Hermida, 2014). It potentially complements traditional science communication by affording new ways to disseminate, consume and discuss scientific issues and findings. Despite initial dismissal of microblogging as just another way of spreading trivialities, Twitter does seem to provide useful functionalities that have promoted its diffusion (Arcenaux & Schmitz Weiss, 2010); currently, 23% of online American adults use Twitter (Duggan, Ellison, Lampe, Lenhart, and Madden, 2015).

The communication of scientific knowledge in a networked public sphere is part of larger structural changes in how modern complex societies produce information, knowledge, and culture (see Benkler, 2006; Weinberger, 2012; Han, 2010). Effective public engagement in science requires mediated forms of communication where differences in expert knowledge, values, and goals can be articulated and discussed (Nisbet & Scheufele, 2009)—traditional media however, have by design largely supported a unidirectional form of science communication. Recent studies that took new media into account explored the communication of single science topics such as climate change (Veltri & Atanasova, 2015), nanotechnology (Veltri, 2014), human genome research (Gerhards & Schäfer, 2010) or food contamination (Shan et al., 2014). The question of how internet-enabled services impact science communication—from web sites promoting scientific knowledgeability (Eveland & Dunwoody, 1998) to blogs and videos (Ranger & Bultitude, 2014) to Google and Wikipedia (Segev & Sharon, 2016)—deserves continued attention due to accelerated media change (see Brossard, 2013).

The flow of knowledge in mediated social systems

Producing science news

The flow of knowledge from a semi-public scientific community to a broader general public is generally mediated by dedicated science reporters (Trench, 2007). They pick up scholarly output and produce news. Turning research results into valuable information is vital for the public understanding of science as well as the democratic decision-making processes in public policy (e.g. Kennedy, 2010). Still, this normative role of mass media need not degrade popular science to a mere derivative of 'real' science as the production of knowledge is always subject to various social influences (see Hilgartner, 1990). Science reporting is a form of specialist journalism where formal training in the science fields reported on is less important than journalistic professionalism and news values (Hansen, 1994). In new media such as blogs, professionals and amateur enthusiasts are driven more by personal motivations (Ranger & Bultitude, 2014).

Given the vital societal role of science news, it is noteworthy that scientists often regard traditional science reporting as poor (Ashwell, 2014)—this opens up opportunities for new media. Or as Edward J. Robinson (1963) concluded more than 50 years ago, 'the mass media should not be expected to carry the whole burden of reporting science news' (p. 313). He suggests 'supplementary media and techniques such as the use of company publications, speeches to civic groups, open houses, etc.' (p. 313). Does Twitter serve as a supplementary medium for science news today?

The integration of new media in science communication

The technologies and platforms that enable online public communication—natively web-based or adapted from offline—also structure the production, representation and distribution of scientific knowledge (also see Schäfer, 2014). Negatively expressed, new media such as Twitter interfere with the traditional model of scientific and journalistic knowledge production and representation. The positive formulation is that new media can enhance the communication processes shown in Figure 1, and eventually transform science communication as new communication models and nontraditional actors gain relevance (Bucchi, 2013). New media as platforms constitute a technical frame for communication that is interpreted and put into practice by various users, including journalists and scientist. As Figure 1 shows, their practical uses may materialize in several forms: conversation, dialogue, collaboration and exchange (two-way communication); self-communication, status updating and monitoring (one-to-many communication); and information sharing, news sharing and marketing (many-to-many communication) (van Dijck, 2011, p. 337).

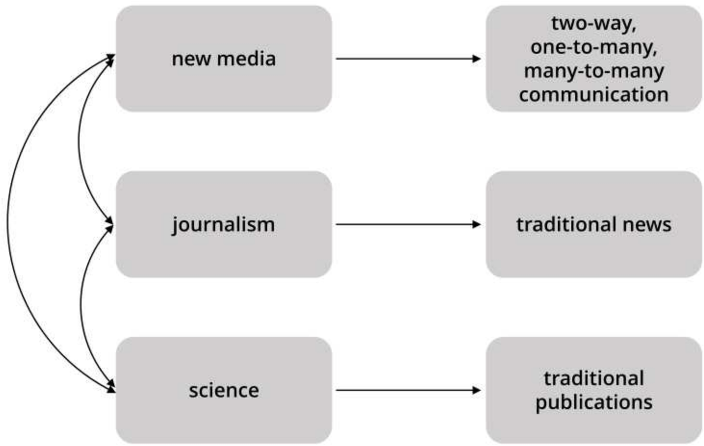

Figure 1. Science, journalism and new media as connected fields of communication.

Essentially, new media are remediatizing the 'science-media interface' (Peters et al., 2008) and the science-public interface. For instance, science is affected in that researchers have new possibilities of staying up-to-date or collaborating (e.g. Meyer & Schroeder, 2009; Bik & Goldstein, 2013). Traditional scholarly output such as journal articles are now complemented by forms of self-publishing, multi-media content and data (e.g. Rzepa, 2011). New media can link interested amateurs with science projects to foster citizen science (e.g. Silvertown, 2009). Future developments may even lead to an increased merging of the production and the publication subsystem including general publics—a development discussed under the label of 'open science' (Neuberger, 2014, p. 337). Journalists can use new media as sources for investigation and interact with scientists in perhaps more informal and immediate ways (e.g. Fahy & Nisbet, 2011). The traditional formats of news such as newspaper articles can be tied to commenting features or news organizations can distribute their content via application programming interfaces (APIs) for reuse on different platforms (see Aitamurto & Lewis, 2013).

The simplistic idea that scientists transmit their 'superior knowledge' to the public via mass media has faded (Schäfer, 2011); the notion of completed facts being unidirectionally transferred to the public needed revision even before interactive online media emerged (Hilgartner, 1990; Bucchi, 1998; Bucchi, 2004; Weinberger, 2012). New models therefore incorporate the evident mutual interrelations of science, the media and the public (Schäfer, 2011, pp. 400–401), but new media have not been systematically integrated into theories of science communication and public engagement with science. New media afford the possibility for public discussions, new voices and new contexts (see Atton & Wickenden, 2005; Batts, Anthis, & Smith, 2008; Shan et al., 2014), but research has also shown that online sources' actors and frames hardly differ from those of print media (Gerhards & Schäfer, 2010). It is thus argued that the comparison of old and new forms of science communication deserves continued attention from social science research.

Twitter, the de facto standard platform for microblogging, has shifted from being primarily a social communication tool to a global news and information following tool (Van Dijck, 2011). On the basis of graph analysis, recent in-house research from Twitter unsurprisingly concluded that the platform is both an information network and a social network (Myers, Sharma, Gupta, & Lin, 2014). Another large-scale study showed that the trending topics are mostly news in nature and points to the potential of Twitter to rapidly diffuse such content via retweets and the large implied audiences (Kwak, Lee, Park, & Moon, 2010). Traditional news outlets can drive traffic to their own media products through strategic use of Twitter (Hong, 2012). Beyond reposting prominent messages, ordinary users also extend news on Twitter by commenting on current issues (Subasic & Berendt, 2011) and may form conversational clusters around current issues (Smith, Rainie, Himelboim, & Shneiderman, 2014).

Twitter has been of interest to science communication scholars (Puschmann, 2014). For example, Shan et al.'s (2014, p. 924) content analysis of the 2008 Irish dioxin crisis showed that 'Twitter mainly functioned as a news information disseminator.' Their study investigated 175 social media documents (68 tweets) and compared them with 141 newspaper articles (Shan et al., 2014). In terms of topic contextualization, they find that social media emphasize global reaction, the government's handling and public perception more (Shan et al., 2014, p. 921). Veltri (2014) collected 24,634 tweets on nanotechnology and found that communication is not conversational but rather dominated by very few 'power users' while issue framing is congruent with newspaper reports.

The diverse features of new media have the potential to make them relevant during the entire process of information diffusion—from first exposure to potential attitude changes. The popularization of (science) issues depends on broadcasting functionalities afforded by new media like Twitter while the high trust put in interpersonal relationships maintained on such platforms supports their perpetuation (see Karnowski, 2011). This means that the multi-step flow of communication (Katz & Lazarsfeld, 1955) can be contained in a single medium. While the roles of sources, opinion leaders and audience are dynamic in the new media environment, empirical findings show that mainstream news organizations can hold considerable influence across a variety of topics (Cha, Haddadi, Benvenuto, & Gummadi, 2010).

The primary substantive interest of this work lies in assessing the role of science tweets: Has the emergence of Twitter as a new media representative extended science communication? Do topic contextualizations between traditional online news outlets and microblogs diverge? Does Twitter add anything to science news beyond disseminating the reports of established media outlets? An a priori confinement to nanotechnology or climate change, for example, is inept for such questions as the interest does not lie in the single issue but potentially generalizable platform differences. Research question 1 thus involves the identification of popular science topics and the comparison of two platforms regarding their contextualization operationalized as co-occurring terms.

> *RQ1—Content of news and tweets.* How does the contextualization of science topics differ between online news and microblogs?

Given that traditional news have received much more attention from science communication research than new media—apart from blogs perhaps—a closer look at the science information sharing practices of Twitter users is warranted. The focus here lies on the structure of science tweets (e.g. references to other users or the use of web links) as well as the topology of the network produced by users' communicative behavior.

> *RQ2—Structure of tweets.* How are science news topics communicated on Twitter?

Both research questions are used to explore the 'science news Twitterverse' and to draw conclusions about the affordances of microblogs in science coverage.

Methods and data

Linking news and tweets via automated topic extraction

In contrast to numerous empirical studies on media representations which selected scientific issues a priori (see Schäfer, 2007, pp. 63–75)—nanotechnology appears to be particularly popular (e.g. Veltri, 2013; Retzbach & Maier, 2014)—our approach was to dynamically detect current and prevalent public science topics (Figure 2).

Based on Kleinberg and Lawrence (2001), Weber and Monge (2011) identified three relatively distinct roles in the flow of online news: sources, authorities and hubs. The selection of news sites considered in our study thus aimed to reflect this spectrum. The dedicated online science sections of the following sites were used: *Reuters* and *Associated Press* (sources), *The New York Times* and *BBC* (authorities), and *The Huffington Post* and *Yahoo* (hubs). These news platforms represent traditional media (see Gerhards & Schäfer, 2010) whereas Twitter represents new media.

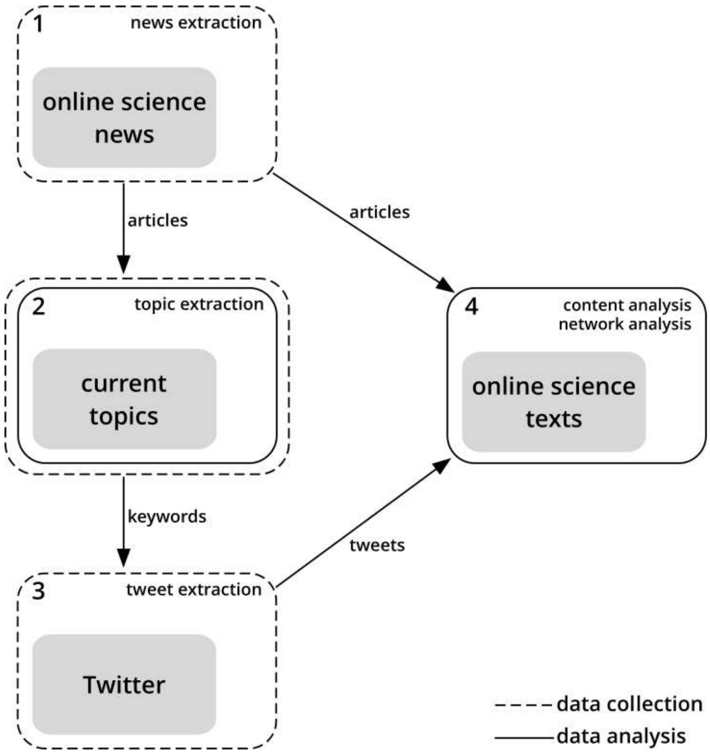

Figure 2. Study overview.

The research questions of this study posed a methodological challenge regarding the collection and analysis of web content (see Weare & Lin, 2000): How can science texts be sampled on different platforms? To further complicate the issue, the theoretical interest required that the data were not restricted to a single theme (such as climate change). This study provides an innovative solution to building a corpus of science news articles and *corresponding* tweets by dynamically linking the two platforms with an integrated data collection and data analysis step (Figure 2, step 2).

The first step (news extraction) in Figure 2 used rich site summary (RSS) feeds of the science sections of the news sites to gather the URLs of the all newly published articles. A script then followed these links and extracted the text body of the articles. In step two (topic extraction), three topics per day were automatically detected. These current news topics were represented by two keywords each (e.g. "space station") that were then used in step three (tweet extraction) as search terms (e.g. "space *and* station") in the Twitter API. This generated corresponding Twitter data that included the publication time, user name, retweet status and tweet content. The automated data collection process (steps 1 to 3) and the content analysis (step 4) were performed in the programming language and software environment R (see Jackman, 2006). Network analysis (step 4) was conducted in the software package Gephi.

The method by which the topics of science texts were detected was latent Dirichlet allocation (LDA; Blei, Ng, & Jordan, 2003). This topic modeling approach differs substantially from communication research's traditional content analyses in that it is automated and unsupervised (Grimmer & Stewart, 2013): There is no training data, no human coding and no codebook. In essence, an algorithm detects words that are likely to appear together based on distributional assumptions. The only inputs for this generative approach are the number of topics to detect and a (large) text corpus (Mohr & Bogdanov, 2013; Weng, Lim, Jiang, & He, 2011). This method conceptualizes documents as a probabilistic mixture of latent topics, i.e. a collection of related words —each topic contributes manifest words to the document (Blei, 2012). Each document exhibits multiple topics and LDA retrospectively determines out of which pool of words, i.e., topic space, a term was most likely selected (Blei et al., 2003). This automated coding results in meaningful and plausible readings of texts with high levels of 'substantive interpretability' (DiMaggio, Nag, & Blei, 2013, p. 578). The method is key to linking online news articles with tweets (Figure 2): Even if corpus size were not an issue, near-simultaneous data collection of corresponding tweets would not have been feasible using manual topic extraction.

Comparing topic contexts across traditional and new media

As a precondition for RQ1, mallet LDA topic models (Mimno, 2015) were run on both the news article corpus and the dependent Twitter corpus. The number of topics, which necessarily needed to be specified, was set to 35 and the five terms with the highest weights were used as indicators of broad themes (see *Five broad science topics in the news*). This strategy of using automatic LDA and then interpreting the high-level classes was shown to yield very good results, comparable to inter-human agreement in manual coding (see Razavi, Inkpen, Brusilovsky, & Bogouslavski, 2013). These analyses used a random subsample of 10,000 tweets because system memory limitations did not allow for the complete corpus to be processed; all news articles were used. To then analyze the contextualizations of these topics (RQ1), a bag-of-words approach (see Grimmer & Stewart, 2014) was employed with lower case conversion and whitespace, number, punctuation and stopword removal (see Feinerer, Hornik, & Meyer, 2008). The corpora were transformed to term–document matrices for co-occurrence analysis (Scharkow, 2012). The co-occurrences between representative terms of each broad theme and terms from the rest of the corpus revealed how Twitter and traditional news outlets contextualized current science topics. The resulting terms and coefficients were qualitatively interpreted using domain knowledge from web searches (see Mohr & Bogdanov, 2013).

Tweet structure and mention network analysis

RQ2 was addressed by analyzing the structure of tweets as well as the mention network. We compared the prevalence of internal (mentions and retweets) and external (use of URLs) references in the science tweet sample to a baseline. A mention of another Twitter user in a tweet could occur due to conversational interactions, (modified) sharing of others' tweets or referencing someone as a source. Every mention thus produced a connection (edge) between two Twitter user accounts (nodes). To further investigate the communication of science topics on Twitter, these user relations were extracted and visualized as a network. Only those users in the data that had at least one mention tie were used. The mention network extracted from the Twitter data comprised 37,190 edges between 41,228 nodes. Network metrics and visual interpretation were used to analyze components (subsets of the total network) and the indegree distribution (the number of received mentions; see Easley & Kleinberg, 2010).

Five weeks of science news content

Data were collected during a period of 35 consecutive days (five weeks), from 23 September, 2013 to 27 October, 2013. A total of 55,697 unique Twitter users created 72,469 tweets matching the news topic search terms. 965 articles were obtained from the science news RSS feeds of the six online news platforms (AP, Reuters, The New York Times, BBC, The Huffington Post, Yahoo).

Despite having analyzed about 75 times more Twitter documents than news documents, the tweet corpus was only 1.45 times larger with respect to characters (8.3 and 5.7 million, respectively). The average day yielded 28 articles. The average number of tweets returned per day was 2,070, the theoretical maximum being 4,500 due to API limitations. Using all six news platforms, a clear weekly reoccurring pattern was evident with a peak around Thursdays and lows on Sundays and Mondays. The news platforms The Huffington Post and Yahoo (type hub) were most active. The news wires AP and Reuters (type source) each had several days without any articles—as did BBC (type authority). The other authority, The New York Times, matched the overall mean of about 4.5 science articles per day. The mean document length was about 18 words for tweets (skewed towards the platform's limit of 140 characters) and 980 words for news articles. Because automatic tweet extraction could fail due to unknown character encodings, the script had to be restarted on nine days.

Results

Five broad science topics in the news: space missions, US government shutdown, cancer, Nobel Prizes and climate change

Before analyzing any differences between news and tweets it is necessary to know which science topics were covered during the time of data collection. Using the news article corpus, an LDA model was run and the top words were interpreted to find general topics (see Razavi et al., 2013; Weng et al., 2011). The same was done for the Twitter corpus; since by design the sample of tweets depended entirely on the news, this yielded very similar results. Five main terms were discovered in the combined online science texts. These were *space*, *government*, *cancer*, *nobel* and *climate*. Other high-weight terms clearly assignable to one of the main five terms were also included in the co-occurrence analysis to achieve a more precise outline of each theme (Table 1 first column).

In sum, science reporting from major English-language news sites during the fall of 2013 predominantly featured NASA space missions, the US government shutdown, cancer, Nobel Prizes and climate change. These foci of science reporting reflect organizational selection processes based on news values such as relevance, reference to elite people or negativity (Galtung & Ruge, 1965). The first criterion, perceived relevance to the reader, has specifically been shown to guide science news selection

(Hansen, 1994). The five themes also show that news organizations collapse science, medicine and technology into their science sections (see Lewenstein 1995, p. 344).

Twitter has been found to reproduce the diversity of news (Bastos & Zago, 2013), therefore it is consequential that there was simultaneous tweeting on all five broad science news topics, indicating high ‘transmissibility’ of these issues (Milkman & Berger, 2014). The next step then was to evaluate the topic contexts and differences between traditional news outlets and Twitter.

Old and new media contextualizations of popular science (RQ1)

The term co-occurrence analysis to compare the topic contextualizations between news and Twitter used 12 keywords that cover the five broad topics (Table 1). The five upper-case terms in Table 1 best represent each topic in a single word, the lower-case terms frequently co-occurred with them and further specify the topic. Table 1 thus reveals which words frequently appeared together in a document (news article or tweet, respectively) and this illustrates which issue contexts or subtopics were dominant. For example, in the news corpus, the term ‘mars’ was strongly correlated with ‘curiosity’ (.73); this was the third highest correlation for ‘mars.’ In the Twitter corpus, ‘mars’ was also correlated with ‘curiosity’ (.42), albeit less strongly. Still, it was the second highest correlation for this topic word on Twitter, meaning the comparative interpretations below are based on the type and order of the terms rather than on direct comparisons of coefficients across platforms. In this case, news articles and tweets covered NASA’s Mars rover called Curiosity—a research article released during the study time period reported findings of water in Martian soil detected by Curiosity (Leshin et al., 2013).

News	w1	r1	w2	r2	w3	r3	w4	r4	w5	r5
SPACE	nasa	0.69	nasas	0.66	spacecraft	0.64	terminator	0.63	visible	0.63
nasa	cassiopeia	0.78	star-forming	0.78	cfa	0.76	doradus	0.76	riess	0.76
mars	martian	0.85	rover	0.75	curiosity	0.73	rovers	0.73	sol	0.7
GOVERNMENT	shutdown	0.93	mcconnell	0.89	partial	0.89	barrasso	0.88	barricade	0.88
shutdown	government	0.93	federal	0.9	mcconnell	0.86	partial	0.86	barrasso	0.85
CANCER	tumors	0.77	breast	0.72	nucleotide	0.66	polymorphisms	0.66	aggressive	0.65
NOBEL	prize	0.93	awarded	0.79	won	0.7	medicine	0.68	chemistry	0.67
prize	nobel	0.93	awarded	0.88	physiology	0.81	antoine	0.79	barre-sinoussi	0.79
higgs	englert	0.96	boson	0.93	particle	0.87	cern	0.83	collider	0.83
CLIMATE	change	0.89	temperature	0.68	warmer	0.67	melting	0.66	greenhouse	0.64
change	climate	0.89	arabica	0.74	blitz	0.74	bumpier	0.74	color-changing	0.74
warming	intergovernmental	0.81	global	0.71	ipcc	0.7	stocker	0.67	greenhouse	0.65
Twitter	w1	r1	w2	r2	w3	r3	w4	r4	w5	r5
SPACE	station	0.42	ship	0.32	international	0.3	carpenter	0.27	image	0.26
nasa	erupting	0.24	origin	0.23	space	0.23	volcano	0.23	exotic	0.22
mars	rover	0.74	curiosity	0.42	water	0.35	opportunity	0.31	prototype	0.29
GOVERNMENT	shutdown	0.72	federal	0.25	shut	0.25	shuts	0.25	affecting	0.21
shutdown	government	0.72	affecting	0.25	science	0.22	woods	0.22	partial	0.21
CANCER	breast	0.52	study	0.48	society	0.37	significantly	0.34	women	0.29
NOBEL	prize	0.77	higgs	0.59	peace	0.34	englert	0.32	medicine	0.31
prize	nobel	0.77	peace	0.41	medicine	0.35	wins	0.35	higgs	0.34
higgs	nobel	0.59	peter	0.56	boson	0.52	englert	0.51	francois	0.38
CLIMATE	change	0.53	catastrophic	0.36	breakdown	0.35	scientists	0.34	season	0.25
change	climate	0.53	asking	0.31	breakdown	0.24	catastrophic	0.24	small-scale	0.21
warming	global	0.71	cause	0.39	mankind	0.39	convinced	0.37	main	0.35

Table 1.

The topic of NASA space explorations, in the news articles, was associated with rather specific terms such as Cassiopeia, a supernova remnant of which new images were made by NASA's X-ray observatory. A star-forming region near the Milky Way named Doradus was also reported on. Tweets, on the other hand, seemed more concerned with NASA satellites improving volcano eruption forecasts. The government shutdown topic in the news focused on prominent individuals: John Barrasso and Mitch McConnell are both U.S. senators. The contextualization of the shutdown seemed broader on Twitter. Interestingly, the term 'science' in tweets was strongly associated with the government shutdown indicating that the payment default for science programs was discussed (see Maron, 2013). Cancer as a recurring science topic generally deals with breast cancer on both platforms. In the news, 'nucleotide' and 'polymorphism' are often mentioned which refer to a DNA sequence variation. Nobel Prizes were again associated with similar terms in news and tweets: the Higgs boson, its place of discovery CERN, medicine and peace. The Nobel Prize announcements traditionally begin in October, which was during data collection.

Climate, a major topic in science news, was clearly discussed in terms of global warming and climate change. The IPCC (Intergovernmental Panel on Climate Change) had set an upper limit for greenhouse gas emissions in late September, 2013. This United Nations panel's proceedings constituted the context of climate discussions in the news in the fall of 2013 (see Pearce, Holmberg, Hellsten, & Nerlich, 2014). Simultaneously, Twitter users' communication was more emotional: associated terms were 'catastrophic' and 'breakdown.' A very interesting finding is that tweets seemed concerned with the anthropogenic causes of global warming (see Veltri & Atanasova, 2015).

The recommender role of Twitter (RQ2)

The above results illustrated the contextualizations of science topics in Twitter and news. In the following findings on the structure of tweets and the mention network describe how science news topics are communicated on Twitter. Of the tweets collected for this study, 35.7% were retweets, 71.4% included a URL and 51.5% contained a mention (Figure 3). These structural attributes clearly show that science tweets are rarely intended to be self-contained.

These values for tweets about science topics are most informative when compared to a baseline, i.e. corresponding percentages for random non-topic-restricted tweets (Figure 3). Partly owing to the very large sample size, all deviations from the baseline were statistically highly significant. The proportion of retweets was slightly above the baseline, so disseminating existing science messages to one's followers is comparably prevalent. Even though retweets are not original content, the practice of retweeting pushes messages to new networks of followers. Roughly half of the science tweets contained a mention to another user in some form (by retweet, reply or genuine mention), which was just slightly below the general rate. The most remarkable and substantively significant deviation from the baseline was in references to external resources via URLs. This certainly makes sense when thinking about communicating science in microblogs that are limited to 140 characters. Twitter users referenced web resources in 71% of all science tweets compared to a vastly smaller 12% in general tweets.

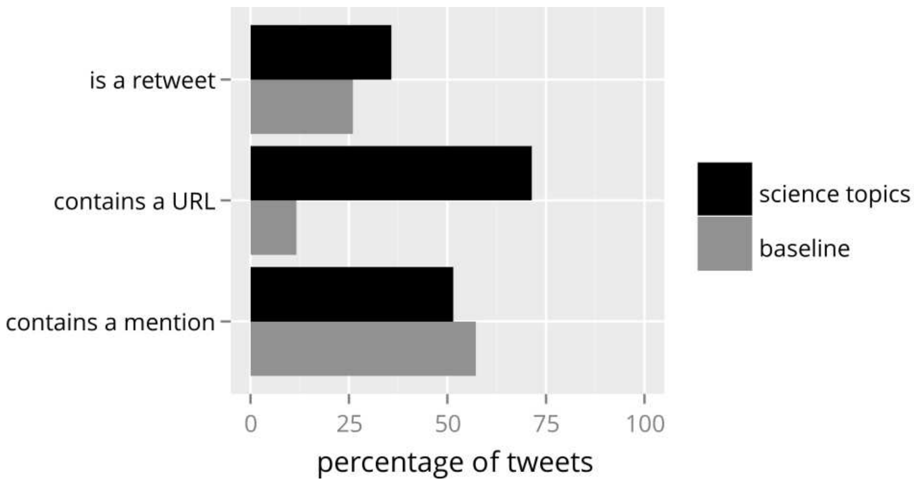

Figure 3. The structure of science tweets.
Note: The baseline is taken from recent studies using large random samples of tweets (Liu, Kliman-Silver, & Mislove, 2014 for 26% retweets; Gerlitz & Rieder, 2013 for 11.7% for URLs and 57.2% mentions).

The topology of the reference network extracted from the tweets featured several distinct characteristics. The distribution of received mentions (indegree) was L-shaped, meaning the vast majority of users received very few mentions whereas a small number of users was mentioned very frequently. A generalized linear model estimation showed a close fit to a power law distribution with an exponent of 2.28 (see Easley & Kleinberg, 2010, pp. 543–555). The largest connected component, i.e. the subset of Twitter users that are directly or indirectly connected through mentions, comprises 38.7% of all network nodes (Figure 4). The size of this component was exponentially larger than any of the remaining 8,135 components—the second largest contained only 1.5% of the nodes. The Twitter accounts of the six news platforms were all part of the giant component, where e.g. The Huffington Post employed multiple specialized accounts. Simply put, those users in the giant component were part of the science news conversation. Due to the imperfections of topic modeling, the many smaller components mostly revolved around other topics in the collected tweets such as 'cancer' as an astrological sign rather than the disease.

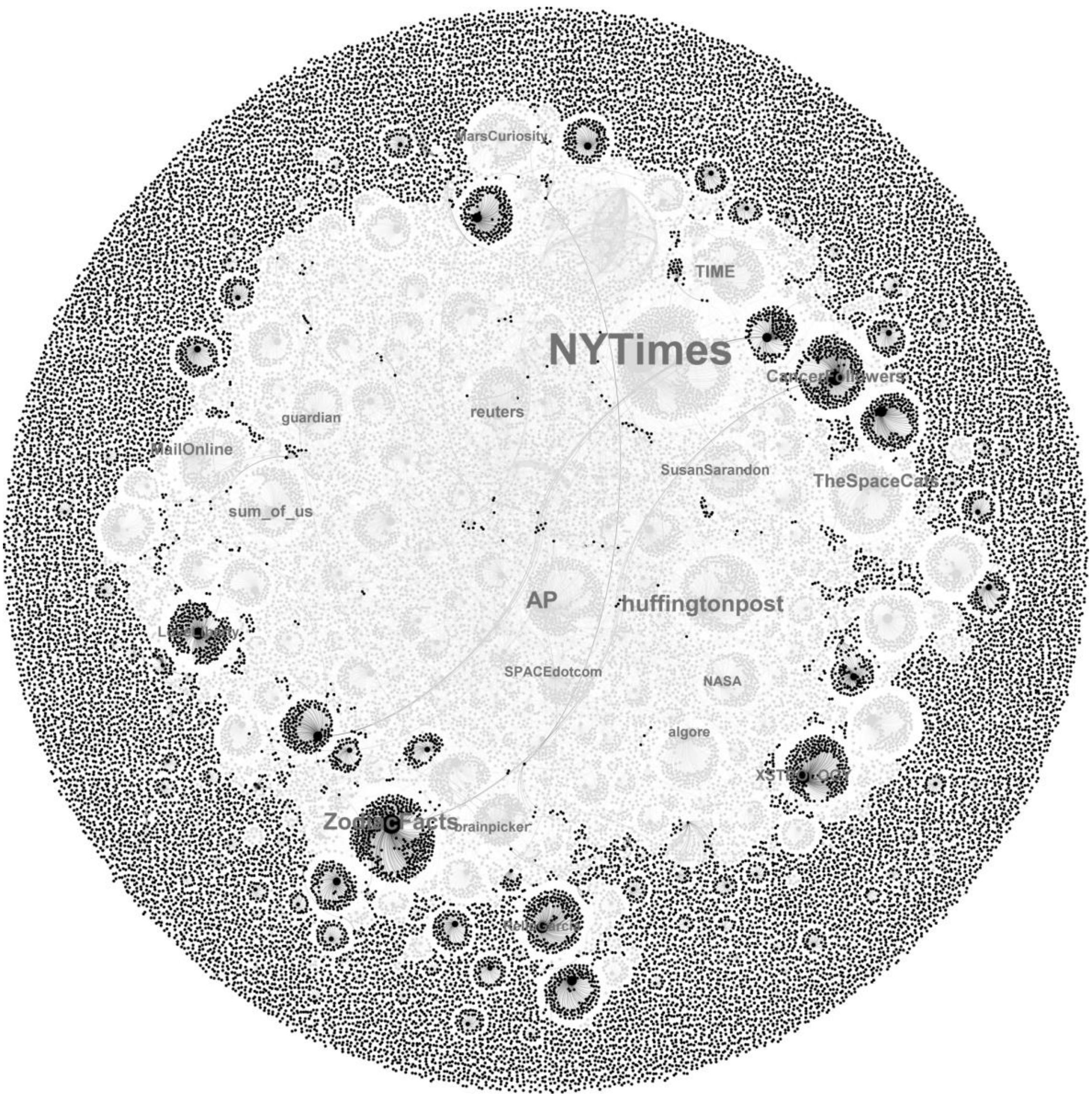

Figure 4. Twitter mention network.
Note: Nodes represent Twitter user accounts, each edge (connection between nodes) represents a mention, i.e. the addressing or referencing of another user in a tweet. Node size was scaled according to received mentions (indegree) and the 20 nodes with the highest indegree were labeled. The grey nodes represent the core network—each node in this giant component is (indirectly) connected to every other node in the component. Besides the clear hub-and-spoke structure around prominent accounts (e.g. NYTimes or NASA) there are also regions with smaller hubs and less concentrated mentioning.

Most mentioned overall was The New York Times Twitter account (NYTimes). Looking further at the giant component (Figure 4), many other news outlets besides the six used for sampling news articles were also present (e.g. TIME, guardian). Celebrity activists (Al Gore, Susan Sarandon), a scholarly blogger (brainpicker) and specialist accounts (MarsCuriosity, SPACEdotcom) complemented the high-indegree list. The rest of the giant component was made up of thousands of users, both organizational and individual, referencing and retweeting each other in tweets that mostly contained an external web link. The most mentioned accounts provided a kind of anchor for the practice of tweeting about a science topic with references to them as information sources.

This network analysis supports the idea that Twitter enables different types of users to become part of a greater conversation (i.e. the giant component) by spreading or commenting on news (see Murthy, 2012, p. 1064; Pearce et al., 2014). The highly skewed distribution of mentions on the other hand demonstrates that established organizations, most notably The New York Times, are able to retain their dominant position as trusted news sources.

Discussion

Main findings and implications

This research was designed to explore the 'science news Twitterverse.' The study makes an important step past hashtag-coordinated or expert-constricted Twitter activity by dynamically linking tweets and news articles with automated topic extraction. Based on content analysis of science tweets, instances were highlighted where the discussion of a current science topic on Twitter (most notably, climate change) differed considerably from the dominant contextualization provided by news platforms during the period of study. Other topics exhibited virtually no differences—or at least they were not detected with the methods employed—which highlights the need to differentiate between different science issues (see Schäfer, 2009). The structure of tweets as well as the network topology of mentions showed that Twitter directs attention to science news and comments, thereby fulfilling a *recommender role*. Mentions are also a way of crediting sources and signaling credibility by referencing authorities.

Network visualization further revealed that frequently mentioned users are not only established institutions such as the news platforms themselves but also additional actors, from activist collectives (e.g YourAnonNews) to influential individuals (e.g.

brainpicker) who contribute new voices. Figure 4 reveals that these new actors produce local hub-and-spoke networks that are very similar to those around traditional news outlets. Some conversational patterns emerge in the network but a pronounced and unidirectional focus on big traditional players such as The New York Times is evident. This finding is in line with research that has found Twitter discussions to rely heavily on professional sources of information (Veltri & Atanasova, 2015). Overall, the network extracted from the science tweets features properties of two ideal type structures: 'community clusters' (Smith et al., 2014, p. 35) and 'broadcast network' (Smith et al., 2014, p. 41). This means that multiple hubs—primarily news outlets—provide different sources and perspectives on a current issue while many users retweet the central actors' messages. The majority of these users is connected only to one of the central hubs but there are also smaller groups of interconnected users who comment and discuss the science topics. In summary, these findings demonstrate that microblogging on Twitter extends public science communication by providing additional voices and contexts as well as recommending content and directing attention.

This extension of science communication may allow publics to share and find information essential for the interpretation of scientific developments and how they relate to their social realities. Vital to the understanding and judgement of new knowledge are the context and method of its creation (Kua, Reder, & Grossel, 2004)—yet science reporting frequently omits such information in an attempt to minimize complexity (Pellechia, 1997). While a 140-character tweet cannot fill this void, microblogging does generate an additional layer of science communication with new sources, voices and interpretations. This is in line with others who have observed that news reporting and tweeting have become intertwined means of (science) communication (Bastos & Zago, 2013, Subasic & Berendt, 2011). Twitter may be seen as a platform where both the public's understanding of science and scientists' understanding of the public are made visible.

Gerhards and Schäfer (2010, p. 143) demonstrated that 'internet communication does not differ significantly from the offline debate in the print media.' The present study, however, does not draw the analytical line between offline and online, but rather differentiates between traditional and new forms of communication that both happen online. Twitter did differ considerably in terms of topic contextualization, particularly for the climate change theme, and let new voices surface. Robinson's 'supplementary media' (1963, p. 313) have found their feet alongside traditional formats.

Limitations and future research

The public agenda as reflected in new media is not identical to that of traditional news, with influences in either direction (Neuman, Guggenheim, Jang, & Bae, 2014). In this paper the media agenda was taken as a pragmatic starting point for data collection (Figure 2). This implies an important limitation of this study: Topics emerging first or exclusively on Twitter could not be detected (see Rogstad, 2016).

Quantitative and automated content analyses were crucial for the analysis, yet this introduced the problem of ambiguity. Human reading of random tweets containing the search term 'cancer' revealed that several users were talking about astrology rather than the disease. In spite of this, tweets containing the word cancer dominantly discuss medical issues (Table 1). The topic models were not able to generate perfect keywords, therefore part of the collected tweets deal with non-science related issues. Furthermore, in qualitatively interpreting the output of the co-occurrence analysis, domain knowledge is crucial. For example, the term 'curiosity' is highly correlated with 'mars'—one needs to know that Curiosity is NASA's robotic rover. If unknown, most terms can be quickly researched, e.g. IPCC, which was correlated to 'warming.' Regarding the network analysis, it must be acknowledged that the topology of the mention network cannot definitively show conversational interactions—future research should therefore combine network analysis with more in-depth content analysis.

The broadening of science communication research attempted in this study was achieved by allowing several topics to emerge rather than defining one at the outset. Another path to achieving higher levels of generalizability is to include more new media sources. As only Twitter was considered here, future studies may look to simultaneously collect content from social networking sites, blogs and other web sources on several topics—an agenda that will need to be accompanied by advances in digital social research methods. In order to further characterize the layer of communication that Twitter has added to public engagement with science, future research could apply methods to automatically trace the external links included in the majority of tweets. What types of resources are referenced (news media, social media, academic literature, etc.) and what kinds of information do they offer? In general, research should continue to analyze the affordances of the web as a knowledge creation and sharing infrastructure.

References

Aitamurto, T. & Lewis, S. C. (2013). Open innovation in digital journalism: examining the impact of Open APIs at four news organizations. *New Media & Society, 15*, 314–331. doi:10.1177/1461444812450682

Arcenaux, N. & Schmitz Weiss, A. (2010). Seems stupid until you try it: press coverage of Twitter, 2006–09. *New Media & Society, 12*, 1262–1279. doi:10.1177/1461444809360773

Ashwell, J. D. (2014). The challenges of science journalism: the perspectives of scientists, science communication advisors and journalists from New Zealand. *Public Understanding of Science*. Advance online publication. doi:10.1177/0963662514556144

Atton, C. & Wickenden, E. (2005). Sourcing routines and representations in alternative journalism: a case study approach. *Journalism Studies, 6*, 347–359. doi:10.1080/14616700500132008

Bastos, M. T. & Zago, G. (2013). Tweeting news articles: readership and news sections in Europe and the Americas. *Sage Open, 3*(3). doi:10.1177/2158244013502496

Batts, S. A., Anthis, N. J., & Smith, T.C. (2008). Advancing science through conversations: bridging the gap between blogs and the academy. *PLOS Biology, 6*(9). doi:10.1371/journal.pbio.0060240

Benkler, Yochai. (2006). *The wealth of networks. How social production transforms markets and freedom*. New Haven, CT: Yale University Press.

Bik, H. M. & Goldstein, M. C. (2013). An introduction to social media for scientists. *PLOS Biology, 11*(4). doi:10.1371/journal.pbio.1001535

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM, 55*(4), 77–84. doi:10.1145/2133806.2133826

Blei, D. M., Ng, A. J., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research, 3*, 993–1022.

Brossard, D. (2013). New media landscapes and the science information consumer. *Proceedings of the National Academy of Sciences, USA, 110*, 14096–14101. doi:10.1073/pnas.1212744110

Bucchi, M. (1998). *Science and the media: alternative routes in scientific communication*. London, England: Routledge.

Bucchi, M. (2004). Can genetics help us rethink communication? Public communication of science as a 'double helix.' *New Genetics & Society, 23*, 269–283.

Bucchi, M. (2013). Style in science communication. *Public Understanding of Science, 22*, 904–915. doi:10.1177/0963662513498202

Cha, M., Haddadi, H., Benvenuto, F., & Gummadi, K. P. (2010). Measuring user influence in Twitter: the million follower fallacy. *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, 10–17.

DiMaggio, P., Nag, M., & Blei, D. (2013). Exploiting affinities between topic modeling and the sociological perspective on culture: application to newspaper coverage of U.S. government arts funding. *Poetics, 41*, 570–606. doi:10.1016/j.poetic.2013.08.004

Duggan, M., Ellison, N. B., Lampe, C., Lenhart, A., & Madden, M. (2015). *Social Media Update 2014*. Washington, DC: Pew Research Center. Retrieved from http://www.pewinternet.org/2015/01/09/social-media-update-2014

Duggan, M., Ellison, N. B., Lampe, C., Lenhart, A., & Madden, M. (2015). *Social Media Update 2014*. Washington, DC: Pew Research Center. Retrieved from http://www.pewinternet.org/2015/01/09/social-media-update-2014/

Easley, D. & Kleinberg, J. (2010). *Networks, crowds, and markets: reasoning about a highly connected world*. Cambridge, England: Cambridge University Press. Retrieved from http://www.cs.cornell.edu/home/kleinber/networks-book/networks-book.pdf

Eveland, W. P., Jr. & Dunwoody, S. (1998). Users and navigation patterns of a science World Wide Web site for the public. *Public Understanding of Science, 7*, 285–311. doi:10.1088/0963-6625/7/4/003

Fahy, D. & Nisbet, M. C. (2011). The science journalist online: shifting roles and emerging practices. *Journalism, 12*, 778–793. doi:10.1177/1464884911412697

Feinerer, I., Hornik, K., & Meyer, D. (2008). Text Mining Infrastructure in R. *Journal of Statistical Software, 25*(5).Retrieved from http://www.jstatsoft.org/v25/i05

Galtung, J. & Ruge, M. H. (1965). The structure of foreign news. *Journal of Peace Research, 2*, 64–90.

Gerhards, J. & Schäfer, M. S. (2010). Is the internet a better public sphere? Comparing old and new media in the USA and Germany. *New Media & Society, 12*, 143–160. doi:10.1177/1461444809341444

Gerlitz, C. & Rieder, B. (2013). Mining one percent of Twitter: collections, baselines, samples. *M/C Journal, 16*(2). Retrieved from http://journal.media-culture.org.au/index.php/mcjournal/article/viewArticle/620

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

Han, S. (2010). Theorizing new media: reflexivity, knowledge, and the web 2.0. *Sociological Inquiry, 80*, 200–213. doi:10.1111/j.1475-682X.2010.00327.x

Hansen, A. (1994). Journalistic practices and science reporting in the British press. *Public Understanding of Science, 3*, 111–134. doi:10.1088/0963-6625/3/2/001

Hermida, A. (2014). Twitter as an ambient news network. In K. Weller, A. Bruns, J. Burgess, M. Mahrt, & C. Puschmann (Eds.), *Twitter and society* (pp. 359–372). New York, NY: Peter Lang.

Hilgartner, S. (1990). The dominant view of popularization: conceptual problems, political uses. *Social Studies of Science, 20*, 519–539. doi:10.1177/030631290020003006

Hong, S. (2012). Online news on Twitter: Newspapers' social media adoption and their online readership. *Information Economics and Policy*, *24*(1), 69–74. doi:10.1016/j.infoecopol.2012.01.004

Horrigan, J. B. (2006). *The internet as a resource for news and information about science*. Washington, DC: Pew Internet & American Life Project. Retrieved from http://www.pewinternet.org/2006/11/20/the-internet-as-a-resource-for-news-and-information-about-science/

Jackman, S. (2006). Data from the web into R. *The Political Methodologist, 14*(2), 11–15.

Karnowski, V. (2011). *Diffusionstheorien [Theories of diffusion]*. Baden-Baden, Germany: Nomos.

Katz, E. & Lazarsfeld, P. (1955). *Personal Influence: the part played by people in the flow of mass communication*. New York, NY: The Free Press.

Kennedy, D. (2010). The future of science news. *Daedalus, 139*(2), 57–65.

Kleinberg, J. & Lawrence, S. (2001). The structure of the web. *Science, 294*, 1849–1850. doi:10.1126/science.1067014

Kua, E., Reder, M., & Grossel, M. J. (2004). Science in the news: a study of reporting genomics. *Public Understanding of Science, 13*, 309–322. doi:10.1177/0963662504045539

Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media? *Proceedings of the 19th international conference on World Wide Web*, 591–600. doi:10.1145/1772690.1772751

Leshin, L. A., Mahaffy, P. R., Webster, C. R., Cabane, M., Coll, P., Conrad, P. G., … Grotzinger, J. (2013). Volatile, isotope, and organic analysis of Martian fines with the Mars Curiosity rover. *Science, 341*(6153). doi:10.1126/science.1238937

Lewenstein, B. V. (1995). Science and the media. In S. Jasanoff, G. E. Markle, J. C. Petersen, & T. Pinch (Eds.), *Handbook of science and technology studies* (pp. 343–360). Thousand Oaks, CA: Sage.

Maron, D. (2013, October). The government shutdown is (almost) over, but the damage to science will last. *Scientific American*. Retrieved from http://www.scientificamerican.com/article/the-government-shutdown

Meyer, E. T. & Schroeder, R. (2009). The world wide web of research and access to knowledge. *Knowledge Management Research and Practice, 7*, 218–233.

Milkman, K. L. & Berger, J. (2014). The science of sharing and the sharing of science. *Proceedings of the National Academy of Sciences, USA, 111*, 13642–13649. doi:10.1073/pnas.1317511111

Mimno, D. (2015). *Package ‚mallet‘: a wrapper around the Java machine learning tool mallet.* Retrieved from http://cran.r-project.org/web/packages/mallet/mallet.pdf

Mohr, J. W. & Bogdanov, P. (2013). Introduction—topic models: what they are and why they matter. *Poetics, 41*, 545–569. doi:10.1016/j.poetic.2013.10.001

Murthy, D. (2012). Towards a sociological understanding of social media: theorizing Twitter. *Sociology, 44*, 1059–1073. doi:10.1177/0038038511422553

Myers, S. A., Sharma, A., Gupta, P., & Lin, J. (2014). Information network or social network? The structure of the Twitter follow graph. *Proceedings of the companion publication of the 23rd international conference on World Wide Web*, 493–498. doi:10.1145/2567948.2576939

National Science Board. (2014). *Science and engineering indicators 2014*. Arlington, VA: National Science Foundation.

Neuberger, C. (2014). Social Media in der Wissenschaftsöffentlichkeit [Social media in the science public]. In P. Weingart & P. Schulz (Eds.), *Wissen – Nachricht – Sensation* (pp. 315–368). Weilerswist, Germany: Velbrück Wissenschaft.

Neuman, W. R., Guggenheim, L., Jang, S. M., & Bae, S. Y. (2014). The dynamics of public attention: agenda-setting theory meets big data. *Journal of Communication, 64*, 193–214. doi:10.1111/jcom.12088

Nisbet, M. C. & Scheufele D. A. (2009). What's next for science communication? Promising directions and lingering distractions. *American Journal of Botany, 96*, 1767–1778. doi:10.3732/ajb.0900041

Pearce, W., Holmberg, K., Hellsten, I., & Nerlich, B. (2014). Climate change on Twitter: Topics, communities and conversations about the 2013 IPCC Working Group 1 Report. *PLoS ONE, 9*(4), e94785. doi:10.1371/journal.pone.0094785

Pellechia, M. G. (1997). Trends in science coverage: a content analysis of three US newspapers. *Public Understanding of Science, 6*, 49–68. doi:10.1088/0963-6625/6/1/004

Peters, H. P., Brossard, D., de Cheveigné, S., Dunwoody, S., Kallfass, M., Miller, S., & Tsuchida, S. (2008). Science-media interface: It's time to reconsider. *Science Communication, 30*, 266–276. doi:10.1177/1075547008324809

Puschmann, C. (2014). (Micro)blogging science? Notes on potentials and constraints of new forms of scholarly communication. In S. Friesike & S. Bartling (Eds.), *Opening Science* (pp. 89–106). New York, NY: Springer. doi:10.1007/978-3-319-00026-8_6

Ranger, M. & Bultitude, K. (2014). ‘The kind of mildly curious sort of science interested person like me’: science bloggers’ practices relating to audience recruitment. *Public Understanding of Science*. Advance online publication. doi:10.1177/0963662514555054

Razavi, A. H., Inkpen, D., Brusilovsky, D., & Bogouslavski, L. (2013). General topic annotation in social networks: a latent Dirichlet allocation approach. In O. R. Zaiane & S. Zilles (Eds.), *Advances in Artificial Intelligence* (pp. 293–300). Berlin, Germany: Springer. doi:10.1007/978-3-642-38457-8_29

Retzbach, A. & Maier, M. (2014). Communicating scientific uncertainty. Media effects on public engagement with science. *Communication Research, 42*, 429–456. doi: 10.1177/0093650214534967

Robinson, E. J. (1963). Analyzing the impact of science reporting. *Journalism Quarterly, 40*, 306–314.

Rogstad, I. (2016). Is Twitter just rehashing? Intermedia agenda setting between Twitter and mainstream media. *Journal of Information Technology & Politics*. Advance online publication. doi: 10.1080/19331681.2016.1160263

Rzepa, H. S. (2011). The past, present and future of scientific discourse. *Journal of Cheminformatics, 3*(46), 1–10. doi:10.1186/1758-2946-3-46

Schäfer, M. S. (2007). *Wissenschaft in den Medien [Science in the media]*. Wiesbaden, Germany: VS Verlag für Sozialwissenschaften.

Schäfer, M. S. (2009). From public understanding to public engagement: an empirical assessment of changes in science coverage. *Science Communication, 30*, 475–505. doi:10.1177/1075547008326943

Schäfer, M. S. (2011). Sources, characteristics and effects of mass media communication on science: a review of the literature, current trends and areas for future research. *Sociology Compass, 5*, 399–412. doi:10.1111/j.1751-9020.2011.00373.x

Schäfer, M. S. (2014). The media in the labs, and the labs in the media: what we know about the mediatization of science. In K. Lundby (Ed.), *Mediatization of communication* (pp. 570–593). Berlin, Germany: De Gruyter Mouton.

Scharkow, M. (2012). *Automatische Inhaltsanalyse und maschinelles Lernen [Automated content analysis and machine learning]*. Berlin, Germany: Epubli.

Segev, E. & Sharon, A. J. (2016). Temporal patterns of scientific information-seeking on Google and Wikipedia. *Public Understanding of Science*. Advance online publication. doi: 10.1177/0963662516648565

Shan, L., Regan,Á., De Brún, A., Barnett, J., van der Sanden, M. C. A., Wall, P., & McConnon, Á. (2014) Food crisis coverage by social and traditional media: a case study of the 2008 Irish dioxin crisis. *Public Understanding of Science, 23*, 911–928. doi:10.1177/0963662512472315

Silvertown, J. (2009). A new dawn for citizen science. *Trends in Ecology and Evolution, 24*, 467–471. doi:10.1016/j.tree.2009.03.017

Smith, M.A., Rainie, L., Himelboim, I., & Shneiderman, B. (2014). *Mapping Twitter topic networks: From polarized crowds to community clusters*. Washington, DC: Pew Research Center. Retrieved from http://www.pewinternet.org/files/2014/02/PIP_Mapping-Twitter-networks_022014.pdf

Subasic, I. & Berendt, B. (2011). Peddling or creating? Investigating the role of Twitter in news reporting. In P. Clough, C. Foley, C. Gurrin, G. Jones, W. Kraaij, H. Lee, & V. Murdoch (Eds.), *Advances in Information Retrieval* (pp. 207–213). Berlin, Germany: Springer.

Trench, B. (2007). How the internet changed science journalism. In M. W. Bauer, & M. Bucchi (Eds.), *Journalism, science and society: science communication between news and public relations* (pp. 133–141). New York, NY: Routledge.

Van Dijck, J. (2011). Tracing Twitter: the rise of a microblogging platform. *International Journal of Media and Cultural Politics, 7*, 333–348. doi:10.1386/macp.7.3.333_1

Veltri, G. A. (2013). Microblogging and nanotweets: nanotechnology on Twitter. *Public Understanding of Science, 22*, 832–849. doi:10.1177/0963662512463510

Veltri, G. A., & Atanasova, D. (2015). Climate change on Twitter : Content , media ecology and information sharing behaviour. *Public Understanding of Science*. Advance online publication. doi:10.1177/0963662515613702

Wade, S. & Schramm, W. (1969). The mass media as sources of public affairs, science, and health knowledge. *The Public Opinion Quarterly, 33*, 197–209. doi:10.1086/267691

Weare, C. & Lin, W.-Y. (2000). Content analysis of the World Wide Web: opportunities and challenges. *Social Science Computer Review, 18*, 272–292. doi:10.1177/089443930001800304

Weber, P. & Monge, M. S. (2011). The flow of digital news in a network of sources, authorities, and hubs. *Journal of Communication, 61*, 1062–1081. doi:10.1111/j.1460-2466.2011.01596.x

Weinberger, D. (2012). *Too big to know*. New York, NY: Basic Books.

Weng, J., Lim, E.-P., Jiang, J., & He, Q. (2010). Twitterrank: Finding topic-sensitive influential Twitterers. In *Proceedings of the Third ACM International Conference on Web Search and Data Mining* (pp. 261–270). New York, NY. doi:10.1145/1718487.1718520