The creation of the climategate hype in blogs and newspapers: mixed methods approach

Iina Hellsten
Department of Organization Sciences, VU University Amsterdam, Amsterdam, The Netherlands, and
Eleftheria Vasileiadou
School of Innovation Sciences, Eindhoven University of Technology, Eindhoven, The Netherlands

Received 11 May 2014
Revised 28 May 2014
29 August 2014
24 September 2014
Accepted 28 September 2014

Abstract

Purpose – Research into the emergence of a hype requires a mixed methods approach that takes into account both the evolution over time and mutual influences across different types of media. The purpose of this paper is to present a methodological approach to detect an emerging hype in online communications.
Design/methodology/approach – The paper combines Auto Regressive Integrated Moving Average (ARIMA) time series modelling and semantic co-word networks, and this combination of methods provides a view on the emergence and development of a hype at the level of mutual influences across a heterogeneous set of newspaper and blog data. The subject scope of the paper is the climategate hype. The climategate hype was triggered by the online publication of a set of hacked e-mails belonging to climate researchers at the East Anglia University in November 2009.
Findings – The main findings show that the climategate hype was initiated in the blogs, and the newspapers were reacting to the blogs. At the level of semantics, the blogs and the newspapers framed the issue from opposite perspectives.
Research limitations/implications – The combination of methods contributes theoretical insights to how blogs interact with more traditional media on hype generation and methodological insights to internet researchers investigating emergent online hypes. The method calls for further validation.
Practical implications – Investigating the emergence and evolution of a hype, and the interaction of the two media is relevant for journalists in becoming more reflexive in their practices and the cues from the outside world.
Originality/value – The paper is novel in its combination of the two specific methods, ARIMA time series modelling and co-word networks and its attempt to identify the media origins of a hype, and especially the interaction between blogs and newspapers.

Keywords Hype, ARIMA modelling, Blog analysis, Climategate, Newspaper analysis, Semantic co-word networks
Paper type Research paper

1. Introduction

In November 2009 the e-mails of climate scientists at the University of East Anglia, Climate Change Unit were published on the internet without their authors' consent. This internet-based event caused a public hype around what was named as "climategate". This naming served as a turning point in the public debate on climate change (Nerlich, 2010), the ethics, and even governance of climate science (Grundmann, 2012) and contributed to polarization of the discussion of climate change and climate science

The first author gratefully acknowledges the funding provided by the Netherlands Organisation for Scientific Research (NWO), ORA-NWO Grant No. 464-10-077.

Internet Research
Vol. 25 No. 4, 2015
pp. 589-609

1066-2243
DOI 10.1108/IntR-05-2014-0130

(Painter, 2011). Climategate is a good case of an online event with a sudden increase in public attention that we call a hype in this paper.

Our paper introduces a mixed method approach to: trace the emergence of an online hype; and to assess the role of online and offline media (here blogs and newspapers, respectively) in the creation and maintenance of an online hype.

A hype is characterized by a sudden change in the way a specific issue, such as global warming, is framed in public, accompanied by an increase of the amount of public attention to the issue (Vasterman, 2005). Often such attention is accompanied by emotional reactions to the issue (Van Ginneken, 2003), and a dominant framing of the issue – as represented in the use of specific key-phrases. Although such hypes have wide-spread scientific, economic and political consequences, it is largely unknown how online and offline media affect each other in triggering a hype. This has partly been caused by methodological limitations in analysing the mutual influence of two types of data sources, for instance blogs and newspapers. In order to fill this gap in existing research, we combine two methods to trace and analyse the emergence of this online hype: time series modelling to detect the mutual influence of different media (Auto Regressive Integrated Moving Average (ARIMA) modelling) and semantic co-word networks to trace framing of the hype in the different media. The former analyses changes in the public attention to a topic, while the latter analyses the framing of the topic, combining in this way the two key elements of hype definition: attention and framing (Vasterman, 2005). To our knowledge, these two methods have not been combined previously. While there have been several articles with blogs and newspapers as their data sets (e.g. Cacciatore *et al.*, 2012; Meraz, 2009; Wallsten, 2007), there has been, to date, no methodological insight in how the mutual interaction of the two different media can be studied at the level of attention and framing.

Social media, such as blogs, open up a wider range of options for (interested) lay as well as expert publics to get involved in current debates, and (technically) facilitate the spread of certain discourses and frames that may enhance the emergence of a hype. The thematic content of online coverage, especially in blogs, differs from that in print media (Cacciatore *et al.*, 2012; Robinson, 2009).

The potential of internet-mediated communication, in particular social media, for the involvement of lay publics in the debate around scientific issues has been hailed as an example of democratization of science (Delfanti, 2010) that has been argued to lead to changes in the relation between the sciences and society (Heimeriks and Vasileiadou, 2008). Blogs and online forums can act as public spaces for citizens to share knowledge, and develop their own ideas about science (Birch and Weitkamp, 2010). However, it has been argued that the inclusion of non-scientists in science-related discussions can also destabilize and confuse scientific debates and make them more adversarial (Berkhout, 2010). A previous study emphasized that evaluations and comments on science blogs are often trivial, caustic and petty, resembling bar chats rather than encompassing an ideal, dialectic, public sphere (Kouper, 2010; Van Zoonen *et al.*, 2010). In this setting, it is interesting to compare newspapers and blogs in the climategate debate. If blogs and other online spaces do provide cues for newspapers, and, depending on the tone they provide, that would mean an opening up for public sphere for topics and viewpoints coming from the public, with potentially positive impact on science communication, and democratization of scientific journalism.

In this paper, we analyse the hype around climategate in a heterogeneous set of blogs and newspapers during the first three months after the introduction of the term "climategate", i.e. in the period 20 November 2009-28 February 2010. We chose to

analyse blogs as the issue is intrinsically linked to the use of new media: the hype was triggered by hacked e-mails that were posted online and widely circulated in blogs. The comparison with newspapers seems logical, since the interaction of different media, at the level of attention and framing, has mainly been studied as an influence of newspapers on other media, within intermedia agenda setting (Sweetser *et al.*, 2008). We are especially interested in the interaction of the two media for the hype creation and maintenance. Our research questions are:

RQ1. How did the term climategate spread between blogs and newspapers?

RQ2. What was the relation between the attention in the blogs and in the newspapers?

RQ3. To what extent did the semantics of the blogs and the newspapers overlap?

2. Literature review

The intermedia agenda-setting theory focuses on the extent to which media act as content sources for each other, defining each other's topics (Vliegenthart and Walgrave, 2008). Previous research has mainly focused on the relationships between traditional news media, such as national news agencies and daily newspapers or newspapers and television (Vliegenthart and Walgrave, 2008). Intermedia agenda-setting has its origins in agenda-setting theory which is interested in the transfer of salience (attention to a topic) from mass media to the public agenda, i.e. the way that the media influence not only which issues people talk about, but also how they perceive those issues.

As we have mentioned, previous studies confirm that newspapers generally set the agenda for other media, especially elite newspapers such as *New York Times* (Sweetser *et al.*, 2008). However, as Meraz (2009) and Lee *et al.* (2005) point out, empirical evidence of intermedia agenda-setting in new media has been limited. More generally, there has been little empirical attention to the interaction between online and offline media. For example, Graham and Greenhill (2013) have studied the synergy between print and online news media, but focused on the print and online format of newspapers. Meraz (2011) investigated intermedia agenda-setting across political blogs and news entities, and concludes that traditional media were unable to influence the blog agenda. In these studies, the content of the media was not taken into account, but only circulation figures (Graham and Greenhill, 2013), or mutual influence of blogs and media using Granger causality (Meraz, 2011). These studies remain among the few who explored interaction between print and online media. In an earlier study, Lee *et al.* (2005) confirmed the influence between new and traditional media by showing how Korean newspapers and internet bulletin boards shaped each other's agendas on issues during the 2000 South Korean elections. Further, Sweetser *et al.* (2008) also found that blogs and television shape each other's content during election periods.

Cacciatore *et al.* (2012) focused on differences in the volume of coverage and thematic content of print news media and online media in the case of nanotechnology. They found that blogs provide more variety in their portrayals of nanotechnology than print news media, instead of merely amplifying the portrayals of the print media. Campbell *et al.* concluded that blogs play three different roles in influencing news media. As originators, blogs sometimes can feed new issues to media agenda; as resuscitators, blogs provide sustained attention to issues that receive low priority in news media, and, as re-framers, blogs can challenge the dominant frames provided in the news media. In our study, we are interested in studying the influence between

blogs and newspapers in the case of climategate hype. On the basis of previous work, we expect that there will be mutual influence at the level of the content (framing), and the level of attention.

3. Hype dynamics

The mutual influence of media has also been discussed in media studies in terms of a media hype. Hypes have been defined as news waves that are generated by self-reinforcing processes in the news production (Van Ginneken, 2003; Vasterman, 2005). We define a hype as initial coverage of an external event, such as the circulation of hacked e-mails online, which suddenly gains increasing attention in the mass media. Due to positive feedback loops, the news flow begins to accelerate generating ever more news stories on the topic. Media hypes have only seldom been studied in internet settings (see Hellsten and Nerlich, 2010). In this process of a positive feedback a particular framing of the issue becomes dominant, and rapidly spreads across different media. In this process, the hype itself, i.e. previous reports on the topic, seem to lower the threshold in the mass media for publishing more news items on the topic. (Van Ginneken, 2003; Vasterman, 2005).

Prominent key-terms, such as "climategate" in the global warming debate – or "Frankenstein foods" in the debate on genetically modified foods in the late 1990s often function as crystallization points, or as catalysts of meaning (Hellsten, 2000) that may play a crucial role in triggering a public hype in the interaction between several media. Labeling the debate on the hacked e-mails as "climategate", and the controversy over errors in a prominent IPCC report on climate change report as "glaciergate" served as turning points in the public debate, framing climate change sciences as "fraud" (Nerlich, 2010). The climategate event has been argued to cause a decline in public belief in human induced climate change in the USA (Leiserowitz *et al.*, 2012).

We expect that the climategate incident triggers high visibility which is sustained for a period of time, then the issue dies out (because of gradual shift of media attention to other issues), and this will be followed by a period of somewhat moderate attention. This is because issues remain on the public agenda on a higher level after an attention cycle (Downs, 1972), and can be easier activated for a new cycle (Newig 2004). Ruef and Markard (2010) found a very similar dynamic pattern of media attention of German newspapers on fuel cells.

4. Case study: climategate

The climategate hype emerged in November 2009, when e-mails by climate scientists working at the University of East Anglia (UK) were published on the internet without the scientists' consent. The hacked e-mails were first posted on the 17 November on the RealClimate web site, and were later circulated on several blogs. The e-mails suggested unethical procedures for scientific discovery and publication of results, while later investigation into the issue rejected these suggestions. The leaking of the e-mails occurred two weeks before the Conference of Parties 15, the United Nation's Copenhagen Climate Conference that took place on 6-18 December 2009. For a thorough analysis of the event see Garud *et al.* (2014).

The debate was coined climategate by climate skeptics, referring to the earlier Watergate scandal. Naming the topic a gate immediately evoked images of anthropogenic climate change as a political conspiracy and was readily used to draw associations between climate sciences and politics (Watergate) as a cultural emblem of fraud that moved from the political to the scientific fraud.

5. Approach and methods

We collected newspaper and blog data from the LexisNexis Academic database from 17 November 2009 to 28 February 2010 with the search term "climategate" occurring anywhere in the text. This period covers the emergence of the hype as well as its evolution. The first appearance of the term climategate on the blogs in our sample was on 20 November and in newspapers on 23 November 2009.

We concentrated on English language blog postings and newspapers. For the newspapers, we searched for all international, English language newspapers. This covers a heterogeneous set of newspapers, including major newspapers, such as *Washington Post* as well as local ones, such as *Yukon Times*. The climategate issue was a global debate, which took place prior to an international climate change meeting. This prompted us to look at international English-speaking newspapers[1]. Therefore we did not expect regional and national differences to be crucial for this particular debate, and focused on global news media. For the blogs we used the NewsTex blogs, archived in the LexisNexis Academic database that includes a rich set of different types of blogs covering, for example, official blogs, such as EU Referendum as well as more local ones, such as PA Pundits and Midwest Voices. Including as heterogeneous sample of blog postings and news items was considered important for testing our novel combination of methods, After cleaning up the data (removing duplicates) we obtained a data set of 1,007 newspaper articles and 1,309 blog postings (Figure 1).

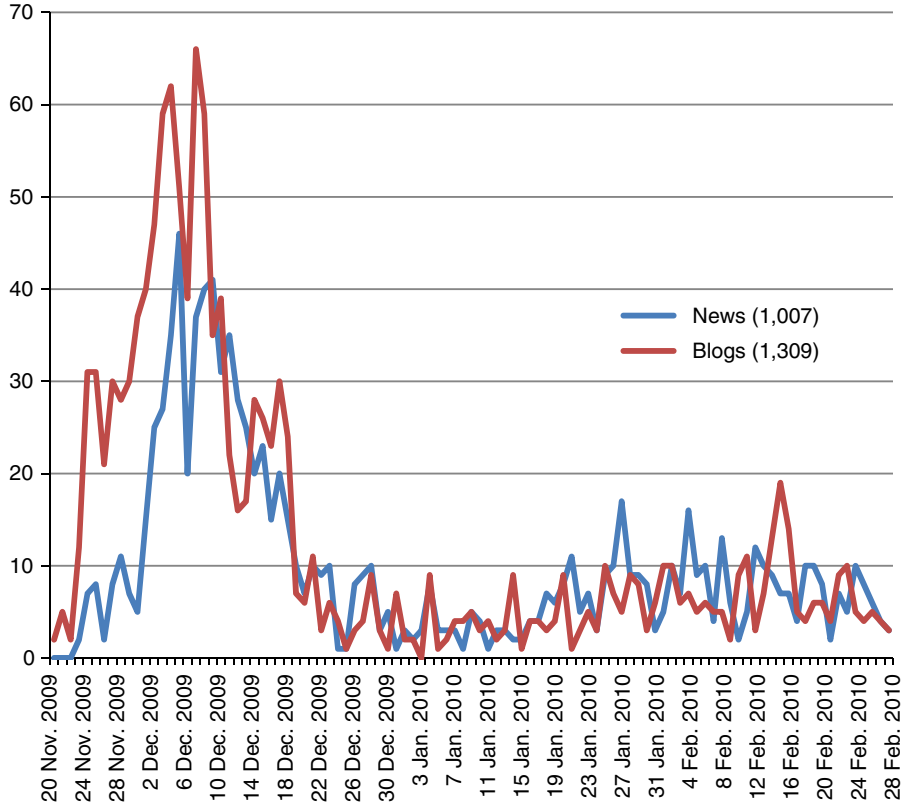


Figure 1.
Number of newspaper articles and blog postings on climategate, per day

We study the interaction of blogs and newspapers during this online hype by analysing:

(1) the attention to climategate in newspapers and in blogs, as measured by the daily number of newspaper articles or blog postings with the word "climategate"; and

(2) the semantics around climategate (framing), as networks of co-occurring headline words in newspaper articles and title words in the blogs.

We analyse: the attention to climategate with ARIMA time series modelling and cross-correlations, which can capture lead-lag relationship between time series variables; and the framing of climategate with semantic co-word networks which explore the language used in the blogs and newspapers. Below we briefly describe our methods.

5.1 ARIMA modelling

In order to study the interaction between newspapers and blogs at the level of attention, we used ARIMA modelling, in particular univariate models and cross-correlations (see description below). Time series modelling, and in particular univariate models aim to detect past regularities to predict future developments (Schoen *et al.*, 2013, p. 531) ARIMA modelling originates in econometrics, but has been applied successfully in media studies (Boomgarden and Vliegenhart, 2007; Hollanders and Vliegenhart, 2008) as well as online communication (Vasileiadou, 2009). Here we provide a brief overview of the method, based on previous work (Vasileiadou, 2009) (see Appendix 1 for the steps in the analysis and the models). For an extensive treatment of the method see (Romer, 2006; Vasileiadou and Vliegenthart, 2014).

As a first step, the Augmented Dickey-Fuller test was run, in order to detect whether the two variables (number of newspaper articles per day and number of blog posts per day) were stationary[2]. The ADF test (see Table AI) suggested that neither variable was stationary, so we proceeded with differencing the variables, and used the new (differenced) variable for further modelling.

Next, we performed univariate modelling, so we created a model of the daily number of newspaper articles as a function of its past. First, an Auto Correlation and Partial Auto Correlation function (ACF/PACF) was run in SPSS. The auto correlation shows the extent to which the variable depends on its past values. The shapes of the ACF/PACF indicate the type of ARIMA model suitable for the variable under investigation. Following this, different univariate ARIMA models were tried on the two variables, to identify the most suitable one. We used the following criteria (also shown in this order in Table AII): highest explanatory variable of the model (highest R^2); Ljung-Box test on the residuals not statistically significant; and statistical significance of AR or MA components[3].

After correctly modelling each variable (univariate modelling), we conducted cross-correlations function (CCF) on the residuals (the two variables without memory of their own past). Cross-correlations indicate whether the variables correlate with each other in different time lags. We used the results of CCF as an indication to create multivariate ARIMA models: a time series model which indicates that one variable, e.g. newspaper frequencies is dependent on its own past, as well as an independent variable (here blog frequencies). In the new multivariate models, the same criteria apply, with the addition of estimates for the independent variable (also shown in this order in Table AII in the Appendix 1): highest explanatory variable of the model (highest R^2); Ljung-Box test on the residuals not statistically significant; statistical significance of AR or MA components[4]; significance test of independent variable; and coefficient of independent variable.

5.2 Framing

To compare the framing of climategate in the blogs and newspapers, we draw semantic co-word maps (Leydesdorff, 1997; Leydesdorff and Hellsten, 2005; Hellsten *et al.*, 2010) of the title words of the blog postings' and newspaper headline words. The titles of the blog postings and the newspaper headlines give an indication of the content of the postings and news items. In particular we focused on the main hype in the period 23 November-17 December 2009.

Semantic co-word maps is a method for analysing implicit frames in text documents. This tool helps to compare the levels of codification in different types of texts, such as scientific articles and newspapers (Leydesdorff and Hellsten, 2005) and to detect changing implicit frames in newspapers over time (Hellsten *et al.*, 2010). The semantic maps method uses specific routines and open software (www.leydesdorff.net) to construct networks of co-occurring words using texts as data. This method has been validated with several case studies (Leydesdorff and Hellsten, 2006; Hellsten *et al.*, 2010, 2014).

We used the tool to words occurring in the headlines of the blog postings and newspaper items. The headlines provide well-codified data for the semantic maps, which is more insightful than the full-text (Leydesdorff, 2012). The method compares not only the co-occurrences of words in sentences but also the positions of the single words, co-words and triad-words in a set of documents (in our case the titles or headlines)[5].

We used the resulting cosine normalized matrices for the network visualization in Pajek, an open source network analysis and visualization tool (http://vlado.fmf.uni-lj.si/pub/networks/pajek/). For the visualization in Pajek, we used the Kamada-Kawai graph layout algorithm that results in most readable maps (Kamada and Kawai, 1989). The cosine matrices use the vector space for the visualization that represents coordinates instead of relational space. Distances in the visualization are based on similarity in the distributions of words in documents and not on the relations among words (Leydesdorff and Vaughan, 2006). The visualization of the patterns of co-occurring words in the sets of text documents (e.g. the headlines of newspaper articles on climategate for a specific period in time) shows different sub-themes and implicit frames in the debate (see Appendix 2 for detailed description of the method). We use this method to compare to what extent the frames in the blogs and in the newspapers overlap.

6. Findings

6.1 Dynamics of the hype

From our data set, we derived daily frequency figures for both newspaper items and blog postings, for the period of 20 November 2009-28 February 2010, in order to observe attention to climategate in both media (see Figure 1).

The climategate hype evolves through three different periods. First, the debate hypes in the period 23 November-17 December 2009 in both blogs and newspapers, followed by a relatively low-attention period between 23 December 2009 and 16 January 2010. From 16 January onwards, the issue gains more attention, but far less than in the first hype phase. The hype cycle develops through a peak in both media, then a decrease in attention, and then a moderate-level attention, as expected by previous studies (Ruef and Markard, 2010).

The term climategate first appears in our data set of blogs on 20 November 2009 on the web site of RealClimate, and in fast pace the term is taken up in several blogs. The newspapers pick up the topic three days later, 23 November 2009, with two articles, both in Australian newspapers. The attention in both blogs and newspapers is increasing he following day. The phase of increased attention lasts until around 17-23 December,

i.e. until the end of the Copenhagen COP 15 summit that ended on 17 December 2009. The term climategate does not die out completely, but it seems to have become a regular term in use, until the end of our research period, i.e. end of February 2010.

6.2 Mutual influence of newspapers and the blogs

We first conducted ARIMA modelling on the whole data set of three months, to study the interaction between the attention in newspaper and attention in blogs. The augmented Dickey-Fuller tests on both variables (attention in the newspapers and attention in the blogs) indicated non-stationarity (Table AI). We thus differenced the variables and proceeded with modelling. An ARIMA model (0,1,1) was the best fit for the newspaper articles (Model 1, Table AII), while for the blog postings the ARIMA model (0,1,1) with square root transformation provides the best fit (Model 2, Table AII).

Next, we run cross-correlations of the residuals of both models, in order to identify whether the attention in newspapers correlates with attention in blogs. The results show that there is a statistically significant correlation at lag 0 (Figure A1). This means that the newspapers' and blogs' attention influence each other. Since the issue emerged in blogs, and because of the time needed for newspapers to print their issue, we expected blog postings to be the independent variable, and the newspapers the dependent variable. Indeed, a multivariate ARIMA model of newspaper attention, using blog attention as a predictor was statistically significant (Model 3, Table AII). This means that the number of blog posts has a positive impact on the change6 in the amount of newspaper articles the same day. The higher the amount of blog posts one day, the bigger the change of newspaper articles the same day. Low amount of blog posts one day means little change in newspaper attention (either sustained attention, or sustained in attention). This suggests that the two media react to the same external events.

Last, we investigated whether newspaper articles of the previous day influence the blog postings one day later. This was indeed the case, with a multivariate ARIMA (0,1,1) model with the lagged newspaper articles as a predictor, explaining blog postings better than the univariate model (Model 4, Table AII). The fewer articles on newspaper articles one day, the biggest the change in the amount of blog postings (either sudden increase or sudden decrease) the following day.

The results suggest that blogs also react to newspapers (statistically significant relationship), but in the following way: if there are many newspaper articles one day, this means that the next day, the difference in blog postings is very small. The blogs were either discussing the topic already the day before, and they continued discussing it the next day, or (less likely, but possible) they did not discuss about it the previous day, and do not discuss it the next day. The newspapers react to blog postings in a positive way; the more blog postings one day, the higher the change of attention in newspapers the same day (so sudden increase of attention, or sudden decrease). In order to probe deeper into the creation of the hype, we ran ARIMA models on the data set covering the period November 23-December 2009, the initial hype period in Figure 1. Again we followed the same steps as described in the methodology.

Both variables, newspapers attention and blog attention, were not stationary according to the ADF tests and needed to be differenced (Table AI). For newspaper attention the univariate ARIMA (1,1,0) model was fit (Model 5, Table AII), while for blog attention the univariate ARIMA model (0,1,0) was fit, with log transformation (Model 6, Table AII). We tested whether newspaper attention affected blog attention, and the other way around, by applying different multivariate models. We found a statistically significant effect of blog

attention on newspaper attention, for the first hype period. A multivariate ARIMA (1,1,0) model was fit for newspaper articles, with blog attention at the same lag and blog attention three days before as predictors (Model 7, Table AII).

High number of blogs means that there is high change in newspapers attention (either high increase or high decrease of attention) the same day. In addition, high number of blogs on one day means low difference of the newspapers attention three days later (either sustained interest or sustained disinterest). During the same period, the attention in blogs is not influenced by the attention in the newspapers. These results can be explained with respect to the online nature of the hype: blogs seemed to have acted as originators of the topic as well as resuscitators, since the attention in the blogs was influencing the attention in newspapers and not the other way around. We come back to this in the discussion.

In conclusion, during the whole period of 20 November 2009-28 February 2010, blog attention and newspaper attention influenced each other. However, during the initial period of hype emergence, the attention in blogs influenced positively the attention in newspapers while the attention in blogs was independent of the attention in newspapers. In this sense, the blogs were responsible not only for the event, but also, mainly, the hype around it.

We now turn our attention to the semantics around climategate to investigate the interaction of blogs and newspapers at the level of the semantic content. In other words, we analyse what the blogs and newspapers were discussing about, and how they framed the climategate issue.

6.3 Frames in semantic maps

In order to examine the interaction of the two media at the level of semantics we draw semantic co-word maps of the title words in the blogs and in the newspapers. First, we present the co-word of the period 20 November-17 December, the initial hype phase in both media, and thereafter more focused semantic maps around the main peaks in the attention in both media.

We have focused on the first intensive period from 23 November to 17 December, 2009 because this period provided us enough articles to conduct the semantic co-word maps. We conducted three subsequent sets of comparisons between semantics in the blogs and the newspapers. First, the emergence of the hype for the period 23-26 November 2009, and the period 29 November-1 December. Second, anticipation of the external event, the Copenhagen Climate Conference (4-5 December 2009) and third, fading away of the hype after the Copenhagen meeting (17 December 2009).

We first present an overview of the words in the two media, before we go into the semantic maps. During the whole hype period, the most frequently used words in blogs are: "leak", "climategate", "climate change", "Al Gore", "leaked e-mails", "denier", "scandal", "data", "EPA", "fraud", "question", "Obama", "UN", "scientists" and "Copenhagen summit scandal". These words provide the angle to the various sub-debates, such as endangered findings on greenhouse gas. Later on we focus on some of these sub-debates as they evolve. In the newspapers, the most frequently used words were partly the same as in the blogs: "climategate", "global warming", "climate change", "Copenhagen Summit", "UN", "scientist", "e-mail", but also words such as "debate", "hacked", "controversy" and "hot air" are used in framing the climategate topic – words that were not used in the blogs.

The overall words used show some interesting differences between the blogs and the newspapers: the newspapers seem to react and respond to a debate originating in the blogs by, for example, calling the event "hot air" and a "controversy". Here we ;focus on the emergence of the hype, especially the first days 23-26 November (Figures 2 and 3).

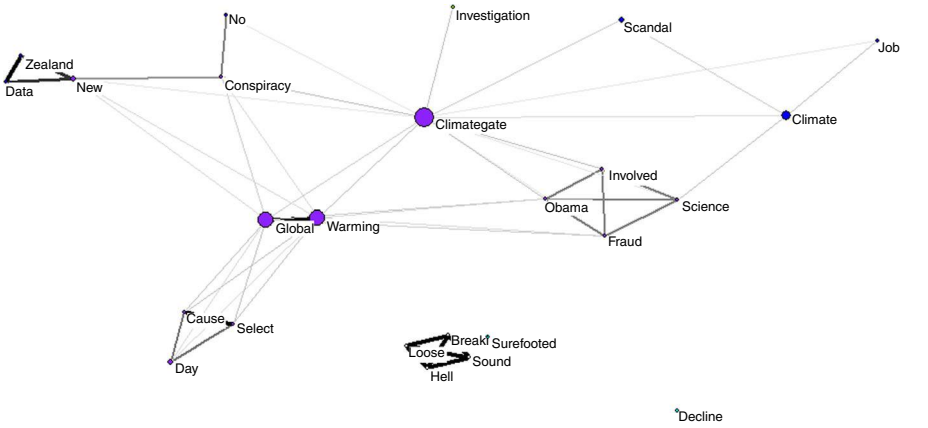


Figure 2.
Blogs on "climategate", 23-25 November 2009

Notes: 64 blog postings, 25 unique words, occurring more than twice, cosine>0>0.433, using Kamada-Kawai algorithm

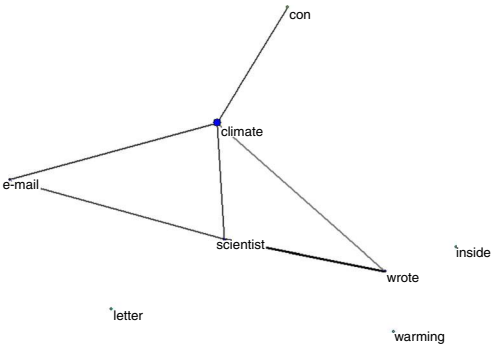


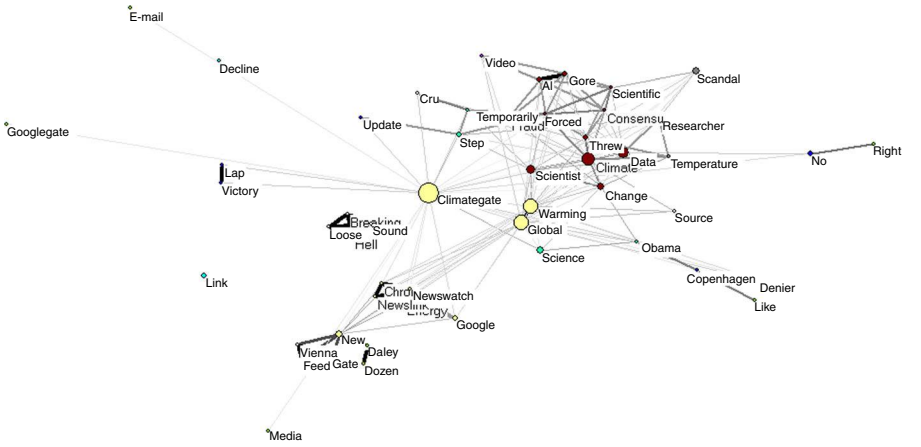
Figure 3.
Newspaper articles on "climategate", 24-26 November 2009

Notes: 16 news items, eight unique words occurring more than twice, cosine>0>0.531, using Kamada-Kawai algorithm

In the blogs, the newly coined term climategate is very central, together with the phrase "global warming", and the semantic context for climategate is about "scandal", "conspiracy" and "investigation", "Obama involved in science fraud" (Figure 2).

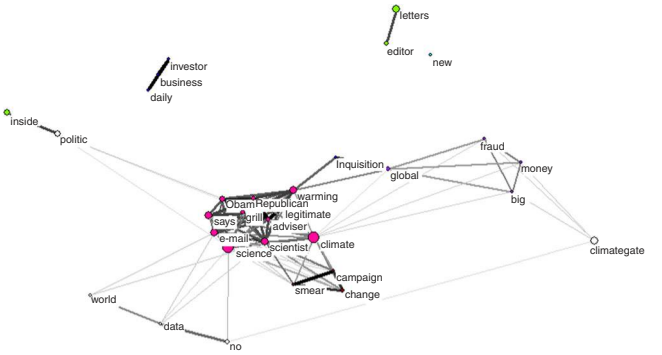
In the newspapers (Figure 3), the term climategate is not yet central in the initial phase of the hype. Instead, the newspapers discuss about climate scientists writing e-mails. This may indicate that the newspapers were at first hesitant of using the term climategate. The newspapers pick up the term climategate in the headline on 24 November 2009, in the *Financial Times* article entitled "Climategate; what the climate scientists wrote and when they wrote it". This is not visible in the semantic map because we apply here a threshold of words appearing twice or more: The term appeared only once in the newspapers in this three days set of newspaper headlines.

Second, we focused in the period 29 November 29-2 December, comparing the blogs and the newspapers at three-day aggregation level (Figures 4 and 5). In the blogs, "climategate" and "global warming" still hold central positions in the network of co-occurring words, but



Notes: 107 blog postings, 53 unique words occurring more than twice, cosine>0>0.363, using Kamada-Kawai algorithm

Figure 4.
Blogs on "climategate", 29 November -1 December 2009



Notes: 50 news items, 33 unique words occurring more than twice, cosine>0>0.570, using Kamada-Kawai algorithm

Figure 5.
Newspaper articles on "climategate", 30 November -2 December 2009

also "climate change scientists threw data" (brown clustered nodes, central right) holds an important position in the network. The anticipation of the climate summit (which would take place on 6 December) has already started, as indicated by the words "Obama", "science" and "source" (Figure 4), as main issues to be expected to be discussed during the meeting. The blogs also discuss about "forced scientific consensus", "Al Gore" and "scandal". The word "e-mails" is peripheral to the semantic network.

In the newspapers, the term climategate is peripheral (right side), and the issue is discussed in the context of "smear campaign" (central, lower side), "global inquisition" (central, right side) and "big money fraud" (right). The words "Republican", "adviser", "Obama", "grill", "climate", "science" are centrally positioned in the newspaper discourse (Figure 5). In summary, both blogs and newspaper started anticipating the Copenhagen meeting, but they do so from opposite positions: for the blogs the issue is about climate sciences withholding temperature data; and for the newspapers it is about a smear campaign and inquisition of science.

Third, we drew semantic maps of the first peak day for both media separately (Figures 6 and 7). The first peak in the attention to climategate is on 4 December 2009 in the blogs and on 5 December 2009 in the newspapers (see Figure 1). This peak anticipates the Copenhagen Summit that was to start on the 6 December 2009.

On 4 December 2009, the word "climategate" is less centrally positioned in the blogs as "Copenhagen summit" is coming up (right side), but it is still rather central. In the blogs, the anticipation of the summit is discussed in terms of "hoax" and "liar" (lower right), and "who hold leaked e-mails" (lower left). "Obama" gains attention as well (see Figure 6).

In the newspapers, the term "climategate" has gained in prominence, and is used slightly less than "climate" itself (Figure 7). The debate is about "climategate e-mails", but also sub-debates, such as "British data exaggerate" (left side) and "UN climate panel report" (below) are present. The cluster around "Steve McIntyre" and "Ross McKitrick", who investigated the data behind the hockey-stick graph, form a separate sub-debate in the newspaper discourse during this period.

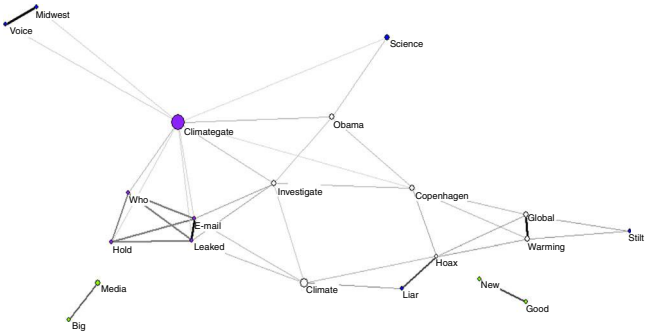


Figure 6.
Blogs, 4 December

Notes: 62 blog postings, 21 unique words occurring more than twice, cosine<0<0.420, using Kamada-Kawai algorithm

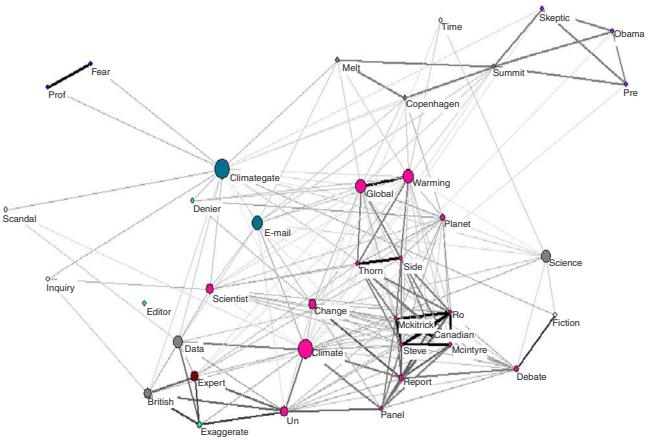


Figure 7.
Newspaper articles, 5 December

Notes: 46 news items, 40 unique words occurring more than twice, cosine<0<0.328, using Kamada-Kawai algorithm

Finally, we drew semantic maps of the last day of the Copenhagen summit, which was the moment that the hype started fading away. Because of smaller amount of blog postings and newspaper items, we were able to include all words in the visualization of the semantic maps. This final peak occurred on the same day in the blogs and in the newspapers, 17 December 2009 (Figures 8 and 9).

In the blogs, the discussion around climategate starts fragmenting into several, unconnected sub-debates. The word "climategate" connects clusters of words, such as "CRU manipulated data" (left) and "leaked UN emission report" (lower left). On the right side, the word "Copenhagen" connects clusters on "James Inhofe failed conference" (right) and "Hugo Chavez roaring video" (lower right).

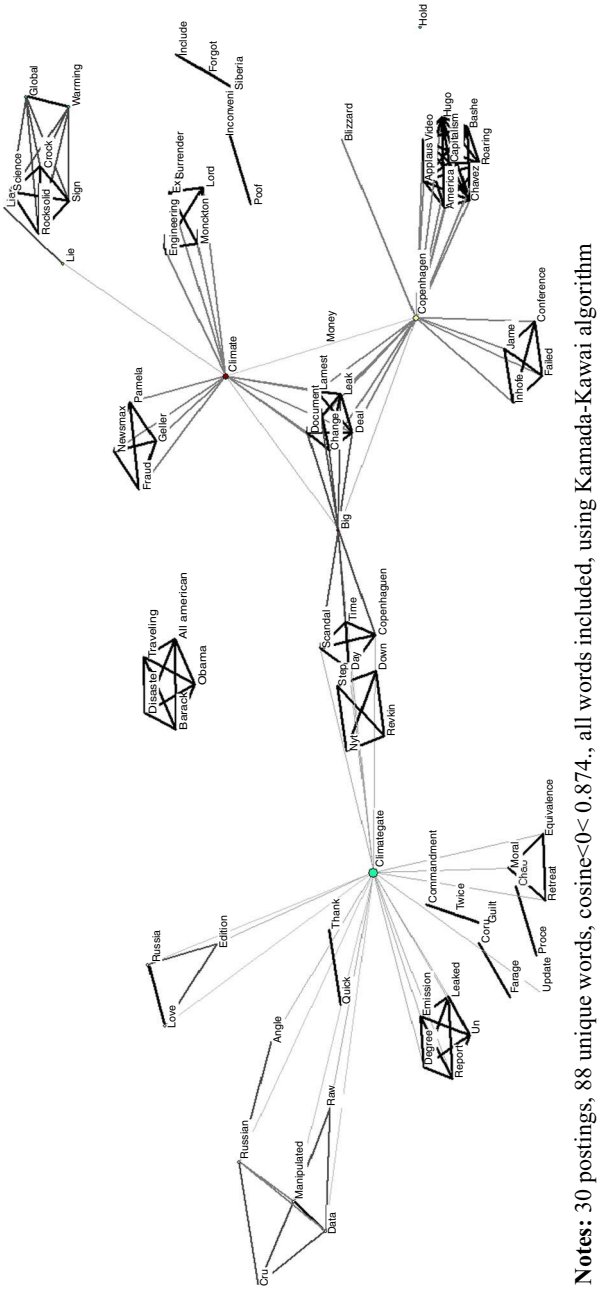
In the newspapers, the debate has evolved into separate sub-debates with little overlap in the word use. The only cluster consisting of more than four words, on the right side, represents a reflection on the Copenhagen summit with words such as "money drive" and "big ideas", "Copenhagen conference", "underestimated Russian". The debate is highly fragmented.

In summary, the semantics in the blogs and newspapers differ in that, quite expectedly, the blogs use more informal language while the newspapers maintain mainstream language. Apart from the differences in style, the blogs frame the climategate debate in more negative terms than the newspapers. In the maps representing shorter periods of time the differences between the blog postings and newspaper headlines become visible. In the beginning of the hype, the blogs focus on the term climategate while the newspapers hesitate using this term. All in all, both blogs and newspapers pick up similar sub-topics, but they use different words to discuss the sub-topics. The blogs initiated the debate on climategate and the newspapers responded to the accusations in the blogs.

7. Conclusion and discussion

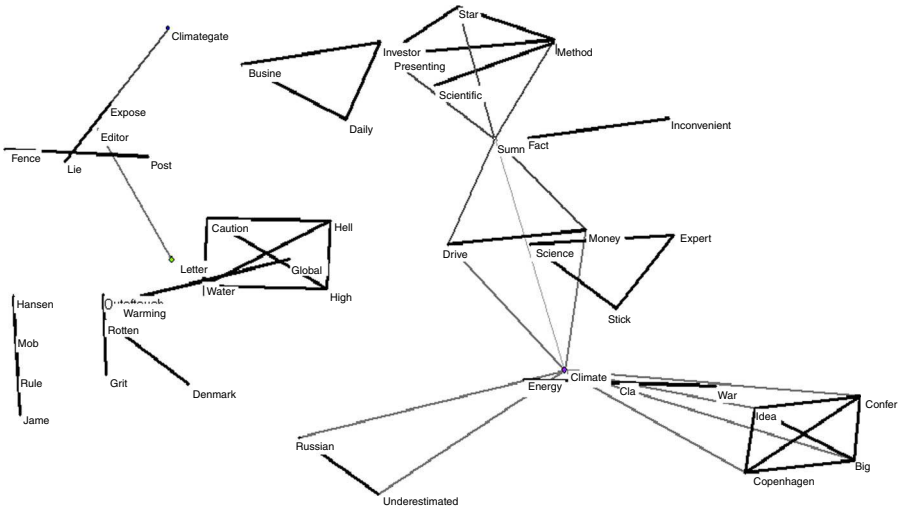
In conclusion, the combination of the ARIMA modelling and the semantic co-word maps reveals that the hype emerges not only at the level of frequencies (attention) but also at the level of semantics. Blogs seemed to heighten newspaper attention on the topic, and the sub-topics (semantics) with which climategate was discussed. Thus, the combination of the two methods to study both attention and frames is vital in order to understand the emergence of the hype. During the initial hype period, the blogs seemed to have influenced both the attention and the semantics of newspapers. The climategate term was initiated in blogs and spread quickly in both media following a pattern of: rapid increase in the attention during which blogs influenced the attention in newspapers; a period of very little attention after the Copenhagen summit and, finally a period of sustained, lower level of attention, from February 2010 onwards.

Our results show that new media (in our case blogs) and traditional media (newspapers) influenced each other during the whole three month period. This confirms the results of Meraz (2011) on that blogs can, in particular cases, be able to influence media agenda. More specifically, blog attention positively influenced the changes in newspaper attention, whereby more blog posts meant high change of newspaper articles the same day (sudden increase or decrease in attention). Newspaper attention negatively influences the changes in blog attention, whereby an increase in newspaper attention one day results in little change in blog attention (sustained interest or disinterest) the following day. This relationship of mutual influence suggests that blogs operate as forums initiating the debate, whereas newspapers are the mainstream media that pick up the issue from the blogs. During the initial hype period, this relationship between the two media became clearer; the blogs were influencing the newspapers,



Notes: 30 postings, 88 unique words, cosine<0< 0.874., all words included, using Kamada-Kawai algorithm

Figure 8.
Blog postings,
17 December 2009



Notes: 20 news items, 49 unique words, cosine<0<0.894, all words included, using Kamada-Kawai algorithm

Figure 9.
Newspapers, 17 December 2009

while they were independent of the attention in newspapers. The blogs were responsible not only for the event, but also the hype around it.

Moreover, in the beginning of the online hype, the blogs and newspapers discussed similar topics, but provided opposing perspectives on what climategate means. For the blogs it was a case of science fraud, and for the newspapers a smear campaign of science. Despite of these different terms that the blogs and newspapers used, they both discussed the same topic, climategate. When the hype was fading away, the discussions in the blogs and newspapers became fragmented to the extent that these two media did not discuss the same topics anymore. This is interesting for hype research, because it indicates that during a hype emergence, different media, temporarily, seem to react to each other while framing the shared topic from different perspectives. During the fading away of the hype different media start discussing separate sub-topics.

There are a number of practices that may be responsible for the delay in newspapers picking up blog semantics and topics. Caution and adherence to journalistic standards can provide one reason why newspapers may have been reluctant in using the online term, especially until more facts were confirmed. Delays in publication times can provide another, more mundane reason. While it is impossible to establish strict causality with this set of methods (or any other method, for that matter), it is nevertheless clear that there was mutual influence of the two media in the creation of the hype, and that, in the emergence phase, the blogs played a leading role, for whatever reasons. This is a novel contribution to intermedia agenda setting, which has not, up to date, empirically confirmed influence of online to offline media. Contributing to intermedia agenda setting, our work points to the fact that: intermedia agenda-setting also functions in topics such as scientific controversies; social media can be the agenda setters for influential newspapers, addressing a par excellence online event; and intermedia agenda-setting is one of the mechanisms responsible for the hype creation itself.

In summary, our results show that the blogs initiated the online hype of climategate, and the polarization in the framing of the climategate issue contributed to the acceleration

of the hype, a week before the Copenhagen meeting started. Blogs and newspapers kept their opposite position to the issue also during the summit itself, and in the end of the summit, their semantics became differentiated from each other. In newspapers, the debate also fragmented into separate sub-topics. A previous study comparing the attention to a political issue in the blogosphere and on newspapers, suggested that the debate in the blogosphere is indeed fragmented, as indicated in our results as well, in contrast to newspaper discussions (Xenos, 2008). These findings show that the creation of the hype seems to result from the interaction between different media, reinforcing each other, something that has not been discussed in the theory on media hypes previously (Vasterman, 2005). In practice this result can help journalists become more aware of their practices, and reflective towards the extent to which they feed into stories from different media and become each other's sources. In addition, what spreads between the blogs and newspapers is a set of word plays and metaphors, such as "dog ate homework" that seem to play a crucial role as providing common ground for the two media. An interesting further research question is to take a serious look at the circulation and spreading of such popular, amusing word plays, and their role in accelerating a hype.

Blogs acted in this case of online hype as originators, resuscitators and re-framing, suggesting that these three different roles can come together at different stages during a hype creation and maintenance. It would be interesting to validate this hypothesis, specifying the stages in which these roles operate, in more cases of online hypes.

The pivotal role of social media in the discussion of the scientific controversy is indicated not only by the leading role of blogs in the climategate hype in newspapers, but also by the wealth of sub-topics with which blog authors connected the incident itself. While one may say that such sub-topics remind more of bar chats (Van Zoonen *et al.*, 2010), for example, by the link to "dirty" politicians, it remains nevertheless a case of wider audience, more or less knowledgeable about the inside of climate change field itself, that became engaged in a topic related to scientific conduct and ultimately science governance. This remains a far cry from a democratization of science, but it provides evidence of the possibilities that social media can help relate scientific discussions to one's everyday life and expectations, as well as for politicization of science. In addition, social media can play a large role in politicization of science At the same time, the climategate hype served as a heated debate beyond mere bar chat (Van Zoonen *et al.*, 2010). The democratization of science and policy seemed to backfire as a heightened hype and controversy over climate sciences governance. Science journalists need to aware of this double-edged role of online scientific discussions.

Methodologically we used a novel combination of tools to analyse the emergence of the climategate hype and study the two main hype elements: increased attention to a topic and change in framing, with ARIMA modelling and semantic co-word maps, respectively. Future research in this direction can apply time series modelling on the semantic network data itself, which can provide robust information about the creation and the role of the sub-debates, the dynamics of centrality of specific terms, and how different sub-debates may depend on each other.

Notes

1. Because the scandal included e-mails from the UK scientists, we expected UK newspapers to play a prominent role as well as newspapers in the USA, where the climate debate is highly polarized. Our choice for English-speaking newspapers covers these countries. Since the focus is on the global debate, any national differences are not discussed here.

2. Stationarity is a prerequisite for applying ARIMA models. If a variable is not stationary, it needs to be treated, for instance through differencing, or a log transformation (Vasileiadou and Vliegenthart, 2014).
3. We also used as an indicator the lowest Schwarz' Bayesian information criterion but it is not shown in the appendix table, since the comparison would have been impossible to indicate.
4. In the newspapers, the section "letters to the editor" lacks headlines for each individual item, i.e. the headline of these articles is "letters-to-the-editor". Similarly, some blogs lack posting headings, and are titled by the blog name, most notably "Chemically Green via Twitter" and "Gates of Vienna News Feed", as well as "Midwest Voices". We have kept these postings in our analysis of title words and newspaper headlines.
5. Because we differenced the variable, the new variable represents the change of attention, from one day to the next.
6. The blogs also mention "Googlegate", referring to Google search engine neglecting any searches conducted with the search term climategate. This latter topic was not discussed in the newspapers.

References

Berkhout, F. (2010), "Reconstructing boundaries and reason in the climate debate", *Global Environmental Change*, Vol. 20 No. 4, pp. 565-569.

Birch, H. and Weitkamp, E. (2010), "Podologues: conversations created by science podcasts", *New Media & Society*, Vol. 12 No. 6, pp. 889-909.

Boomgarden, H.J. and Vliegenhart, R. (2007), "Explaining the rise of anti-immigrant parties: the role of news media content", *Electoral Studies*, Vol. 26 No. 2, pp. 404-417.

Cacciatore, M.A., Anderson, A.A., Choi, D.-H., Brossard, D., Scheufele, D.A., Liang, X., Ladwig, P.J. and Xanos, M. (2012), "Coverage of emerging technologies: a comparison between print and online media", *New Media & Society*, Vol. 14 No. 6, pp. 1039-1059.

Delfanti, A. (2010), "Users and peers. from citizen science to P2P science", *Journal of Science communication*, Vol. 9 No. 1, available at: http://jcom.sissa.it/archive/09/01/Jcom0901%282010%29E/Jcom0901%282010%29E.pdf

Downs, A. (1972), "Up and down with ecology. The 'issue-attention cycle'", *Public Interest*, Vol. 28, Summer, pp. 38-50.

Egghe, L. and Leydesdorff, L. (2009), "The relation between Pearson's correlation coefficient r and Salton's cosine measure", *Journal of the American Society for Information Science and Technology*, Vol. 60 No. 5, pp. 1027-1036.

Garud, R., Gehman, J. and Karunakaran, A. (2014), "Boundaries, breaches and bridges: the case of climategate", *Research Policy*, Vol. 43 No. 1, pp. 60-73.

Graham, G. and Greenhill, A. (2013), "Exploring interaction: print and online news media synergies", *Internet Research*, Vol. 23 No. 1, pp. 89-108.

Grundmann, R. (2012), "The legacy of climategate: revitalizing or undermining climate science and policy?", *WIRES Clim Change*, Vol. 3 No. 3, pp. 281-288.

Heimeriks, G. and Vasileiadou, E. (2008), "Changes or transition: analyzing the use of ICTs in sciences", *Social Science Information*, Vol. 47 No. 1, pp. 5-29.

Hellsten, I. (2000), "Dolly: scientific breakthrough or frankenstein's monster? Journalistic and scientific metaphors of cloning", *Metaphor and Symbol*, Vol. 15 No. 4, pp. 213-221.

Hellsten, I. and Nerlich, B. (2010), "The bird flu hype: the spread of a disease outbreak through the media and internet discussion groups", *Journal of Language and Politics*, Vol. 9 No. 3, pp. 393-408.

Hellsten, I., Dawson, J. and Leydesdorff, L. (2010), "Implicit media frames: automated analysis of public debate on artificial sweeteners", *Public Understanding of Science*, Vol. 19 No. 5, pp. 590-608.

Hellsten, I., Porter, A. and Nerlich, B. (2014), "Imagining the future at the global and national scale: a comparative study of British and Dutch press coverage of Rio 1992 and Rio 2012," *Environmental Communication*, Vol. 8 No. 4, pp. 468-488.

Hollanders, D. and Vliegenhart, R. (2008), "Telling what yesterday's news might be tomorrow: Modeling media dynamics", *Communications*, Vol. 33 No. 1, pp. 47-68.

Kamada, T. and Kawai, S. (1989), "An algorithm for drawing general undirected graphs", *Information Processing Letters*, Vol. 31 No. 1, pp. 7-15.

Kouper, I. (2010), "Science blogs and public engagement with science: practices, challenges and opportunities", *Journal of Science Communication*, Vol. 9 No. 1, pp. 280-282.

Lee, B., Lancendorfer, K.M. and Lee, K.J. (2005), "Agenda-setting and the internet: the intermedia influence of internet bulletin boards on newspaper coverage of the 2000 general election in South Korea", *Asian Journal of Communication*, Vol. 15 No. 1, pp. 57-71.

Leiserowitz, A.A., Maibach, E.W., Roser-Renouf, C., Smith, N. and Dawson, E. (2012), "Climategate, public opinion, and the lost of trust", *American Behavioral Scientist*, Vol. 57 No. 6, pp. 818-837.

Leydesdorff, L. (1997), "Why words and co-words cannot map the development of the sciences", *Journal of the American Society for Information Science*, Vol. 48 No. 5, pp. 418-427.

Leydesdorff, L. (2012), "Radical constructivism and radical constructedness: Luhmann's sociology of semantics, organizations, and self-organization", *Constructivist Foundations*, Vol. 8 No. 1, pp. 85-92.

Leydesdorff, L. and Hellsten, I. (2005), "Metaphors and diaphors in science communication: mapping the case of stem-cell research", *Science Communication*, Vol. 27 No. 1, pp. 64-99.

Leydesdorff, L. and Hellsten, I. (2006), "Measuring the meanings of words in contexts: automated analysis of 'Monarch butterflies', 'Frankenfoods' and 'stem cells'", *Scientometrics*, Vol. 67 No. 2, pp. 231-258.

Leydesdorff, L. and Vaughan, L. (2006), "Co-occurrence matrices and their applications in 700 information science: extending ACA to the web environment", *Journal of the American Society for Information Science & Technology*, Vol. 57 No. 12, pp. 1616-1628.

Meraz, S. (2009), "Is there an elite hold? Traditional media to social media agenda setting influence in blog networks", *Journal of Computer-Mediated Communication*, Vol. 14 No. 3, pp. 682-707.

Meraz, S. (2011), "Using time series analysis to measure intermedia agenda-setting influence in traditional media and political blog networks", *Journalism and Mass Communication Quarterly*, Vol. 88 No. 1, pp. 176-194.

Nerlich, B. (2010), "Climategate: paradoxical metaphors and political paralysis", *Environmental Values*, Vol. 19 No. 4, pp. 419-442.

Newig J. (2004), "Public attention, political action – the example of environmental regulation", *Rationality and Society*, Vol. 16 No. 2, pp. 149-190.

Painter, J. (2011), *Poles Apart: The International Reporting of Climate Change Skepticism*, Oxford University, RSIJ, Oxford.

Robinson, S. (2009), "'If you had been with us': mainstream press and citizen journalists jockey for authority over the collective memory of Hurricane Katrina", *New Media & Society*, Vol. 11 No. 5, pp. 795-814.

Romer, D. (2006), "Time series models", in Romer, D. and Kenski, K. (Eds), *Capturing Campaign Dynamics, 2000 and 2004: The National Annenberg Election Survey*, University of Pennsylvania Press, Philadelphia, pp. 165-243.

Ruef, A. and Markard, J. (2010), "What happens after a hype? How changing expectations affected innovation activities in the case of stationary fuel cells", *Technology Analysis & Strategic Management*, Vol. 22 No. 3, pp. 317-338.

Schoen, H., Gayo-Avello, D., Metaxas, P.T., Mustrafaj, E., Strohmaier, M. and Gloor, P. (2013), "The power of prediction with social media", *Internet Research*, Vol. 23 No. 5, pp. 528-543.

Sweetser, E., Golan, G. and Wanta, W. (2008), "Intermedia agenda-setting in television, advertising, and blogs during the 2004 Election", *Mass Communication and Society*, Vol. 11 No. 2, pp. 197-216.

Van Ginneken, J. (2003), *Collective Behavior and Public Opinion*, Lawrence Erlbaum, Mahwah, NJ.

Van Zoonen, L., Vis, F. and Mihlej, S.(2010), "Performing citizenship on youtube: activism, satire and online debate around the anti-Islam video Fitna", *Critical Discourse Studies*, Vol. 7 No. 4, pp. 249-262.

Vasileiadou, E. (2009), "Stabilisation operationalized; using time series analysis to understand the dynamics of research collaboration", *Journal of Informetrics*, Vol. 3 No. 1, pp. 36-48.

Vasileiadou, E. and Vliegenthart, R. (2014), "Studying dynamic social processes with ARIMA modelling", *International Journal for Social Research Methodology*, Vol. 17 No. 6, pp. 693-708.

Vasterman, P. (2005), "Media-hype self-reinforcing news waves, journalistic standards and the construction of social problems", *European Journal of Communication*, Vol. 20 No. 4, pp. 508-530.

Vliegenthart, R. and Walgrave, S. (2008), "The contingency of intermedia agenda-setting. A longitudinal study in Belgium", *Journalism and Mass Communication Quarterly*, Vol. 85 No.4, pp. 860-877.

Vlieger, E. and Leydesdorff, L. (2011), "Content analysis and the measurement of meaning: the visualization of frames of collections of messages", *The Public Journal of Semiotics*, Vol. 3 No.1, pp. 28-50.

Wallsten, K. (2007), "Agenda setting and the blogosphere: an analysis of the relationship between mainstream media and political blogs", *Review of Policy Research*, Vol. 24 No 6, pp. 567-587.

Xenos, M. (2008), "New mediated deliberation: blog and press coverage of the alito nomination", *Journal of Computer-Mediated Communication*, Vol. 13 No. 2, pp. 485-503.

Further reading

Cooper, C.B., Dickinson, J., Phillips, T. and Bonney, R. (2007), "Citizen science as a tool for conservation in residential ecosystems", *Ecology and Society*, Vol. 12 No. 22, p. 11, available at: www.ecologyandsociety.org/vol12/iss2/art11/

Gerhards, J. and Schäfer, M.S. (2010), "Is the internet a better public sphere? Comparing old and new media in the USA and Germany", *New Media & Society*, Vol. 12 No. 1, pp. 143-160.

Lim, J. (2006), "A cross-lagged analysis of agenda setting among online news media", *Journalism and Mass Communication Quarterly*, Vol. 83 No. 3, pp. 298-312.

McCombs, M. and Shaw, D. (1976), "Structuring the 'unseen environment'", *Journal of Communication*, Vol. 2, pp. 18-22.

Salton, G. and McGill, M.J. (1987), *Introduction to Modern Information Retrieval*, McGraw-Hill, New York, NY.

Appendix 1. ARIMA models

	ADF	Number of lags
Newspapers	−2.52	4
Blogs	−2.12	4
Newspapers1	−0.844	3
Blogs1	−2.11	3

Table AI. ADF tests for the amount of newspaper articles and blog posts

Notes: ADF tests for the amount of newspaper articles per day for the three-month period (newspapers), the amount of blog posts per day for the three-month period (blogs), the amount of newspaper articles per day the first period (newspapers1) and the amount of blog posts per day the first period. (blogs1)

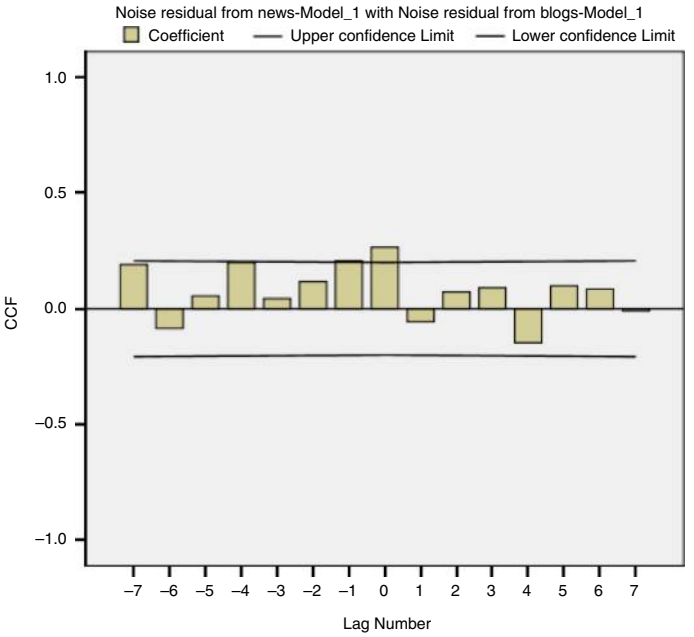
	R^2	Ljung-Box Q sig.	Constant sig.	AR/MA lag 1 sig.	Predictor sig. (multivariate)	Predictor estimate (multivariate)
Model 1 difference	0.724	0.102	0.914	0.000 MA	–	–
Model 2 square root difference	0.794	0.125	0.865	0.000 MA	–	–
Model 3 difference	0.742	0.129	0.096	0.000 MA	0.010 blogs lag 0	0.055
Model 4 difference	0.815	0.163	0.061	0.000 MA	0.019 newspapers lag −1	−0.015
Model 5 difference	0.696	0.242	0.760	0.072 AR	–	–
Model 6 difference natural log	0.357	0.637	0.754	–	–	–
Model 7 difference	0.795	0.133	0.225	0.008 AR	0.001 blogs lag 0 0.012 blogs lag −3	0.224 −0.149

Table AII. ARIMA models and specifications

Note: The sixth and seventh columns apply only for the multivariate models, and indicate the significance level and the coefficient of the independent variable, respectively

Appendix 2. Semantic co-word maps procedure

The procedure for drawing the semantic co-word maps is as follows: the headlines of the news items were saved as a text.txt document, and a word frequency list, words.txt, was created using TextStat open software tool (www.niederlandistik.fu-berlin.de/textstat/). A stopword list of common words with little semantic relevance, such as "the," "he" and "with" were compiled to the file stopword.txt to be removed from the analysis, and saved in the same folder with the text.txt and words.txt. Plural s was automatically removed (e.g. the words "car" and "cars" were stemmed into the single word, "car"). The routine calculates the co-occurrences of the headline words in the set of all the headlines, and automatically constructs matrices of words vs documents that result both in relational co-occurrence matrices and positional cosine normalized matrices (Egghe and Leydesdorff, 2009), using Salton's Index (Salton and McGill, 1972). While the analysis was conducted using ti.exe programme, Pajek was used for the visualization of the cosine.dat file (for more information on the ti-exe routine, see Vlieger and Leydesdorff, 2011).



Notes: The CCF coefficient exceeds the upper confidence limit, which indicates correlation between the two variables (newspaper attention and blog attention)

Figure A1. Cross-correlations between blog and newspaper attention, period 23 November-28 February

About the authors

Dr Iina Hellsten is an Associate Professor in the Department of Organization Sciences at the VU University Amsterdam, affiliated to the Network Institute at the VU. Her research interest focuses on tracking communication networks. Dr Iina Hellsten is the corresponding author and can be contacted at: i.r.hellsten@vu.nl

Assistant Professor Eleftheria Vasileiadou holds a Doctorate in Social and Behavioural Sciences from the University of Amsterdam. She is an Assistant Professor in the School for Innovation Sciences, at the group Technology, Innovation and Society. She has been working on issues related to the role of climate science in society, as well as the use of ICTs in science.