Comparing Events Coverage in Online News and Social Media: The Case of Climate Change

Alexandra Olteanu
École Polytechnique Fédérale
de Lausanne, Switzerland
alexandra.olteanu@epfl.ch

Carlos Castillo
Qatar Computing
Research Institute
chato@acm.org

Nicholas Diakopoulos
College of Journalism
University of Maryland
nad@umd.edu

Karl Aberer
École Polytechnique Fédérale
de Lausanne, Switzerland
karl.aberer@epfl.ch

Abstract

Social media is becoming more and more integrated in the distribution and consumption of news. How is news in social media different from mainstream news? This paper presents a comparative analysis covering a span of 17 months and hundreds of news events, using a method that combines automatic and manual annotations. We focus on *climate change*, a topic that is frequently present in the news through a number of arguments, from current practices and causes (e.g. fracking, CO_2 emissions) to consequences and solutions (e.g. extreme weather, electric cars). The coverage that these different aspects receive is often dependent on how they are framed—typically by mainstream media. Yet, evidence suggests an existing gap between what the news media publishes online and what the general public shares in social media. Through the analysis of a series of events, including awareness campaigns, natural disasters, governmental meetings and publications, among others, we uncover differences in terms of the triggers, actions, and news values that are prevalent in both types of media. This methodology can be extended to other important topics present in the news.

1 Introduction

The study of anthropogenic (human induced) climate change goes back more than 100 years,[1] with a scientific consensus on the topic beginning to emerge in the 1980s. By 2014 our planet had registered the warmest year since 1880, when records began to be kept, and 14 of the 15 warmest years on record have all fallen in the first 15 years of this century.[2] Climate change is an issue with myriad impacts being felt and discussed across the globe. The increased salience of the topic has lead to many publications in scientific journals and in the general press, campaigns for legal reforms, and high-profile meetings and talks including the establishment of the IPCC, the Intergovernmental Panel on Climate Change (Weart 2008). These various events and publications vie for attention around the issue of climate change—each seeking to define and frame the problems, causes, or potential solutions that are worthy of consideration.

The steady presence of climate change as a topic discussed in media, due to its huge potential consequences,[3] creates a valuable research opportunity for *an in-depth comparative study on how news are communicated through different types of online media*, in particular mainstream news media (MSM) and social media. Understanding these differences offers insights into how such a complex and multifaceted topic is comparatively covered and framed in these different media. Why might some events or actors in the climate change discourse receive more attention in the mainstream media versus on social media, or vice versa? What are the types of news events that receive more attention in both? Ultimately, agenda setting serves to define the problems that are worthy of public attention (Entman 2007), and we seek to understand and compare the agenda that emerges from traditional MSM attention as compared to the agenda that organically emerges on a social media platform.

Our contributions. Our main contribution is *a comparison between social media and mainstream news on climate change*. This comparison uncovers significant differences between triggers, actions, and news values of events covered in both types of media. For instance, mainstream news sources frequently feature extreme weather events framed as being a consequence of climate change, as well as high-profile government publications and meetings. In contrast, actions by individuals, legal actions involving governments, and original investigative journalism, feature frequently as viral events in social media.

We also introduce *a methodology for comparing news agendas online*. This methodology is based on a comparison of spikes of coverage. We analyze two large-scale datasets, both covering a period of 17 months, on news (a global database of about 30 million news articles) and social media postings (a sample of about 2 billion tweets, corresponding to 1% of all Twitter posts). We perform automatic processing to discover terms and topics related to climate change using an iterative procedure. Next, we automatically detect a set of candidate events which are curated through a crowd-sourced step of manual annotation. Along this process, we attempt to keep an uniform treatment of both media under analysis. This process offers a starting point for future com-

[1] http://en.wikipedia.org/wiki/History_of_climate_change_science accessed 01.2015.

[2] http://www.huffingtonpost.com/2015/01/16/2014-hottest-year-on-record_n_6479896.html accessed 01.2015.

[3] http://en.wikipedia.org/wiki/Media_coverage_of_climate_change accessed 01.2015.

parative studies extending to other issues of global attention such as pandemics, global terrorism, or human rights issues.

There are *challenges and limitations* in our approach, including:

– We seek a deeper understanding of the climate change discourse, and we do not attempt to test and validate our methodology across multiple domains—we outline how this can be done towards the end of this paper.

– We do not describe patterns of consumption attention, but rather patterns of coverage, or output attention. In other words we are not claiming that a certain issue is more read, but that it is more written about.

– We use a period of time of 17 months, while climate change has been discussed in the news for decades.

– We cover only one language (English), but we note that it is the language in which most reports triggering this debate are written, including the ones by IPCC.

– We cover only one social media source (Twitter), but it is a large one and it is frequently associated with news (Kwak et al. 2010).

– Both data collections have their own biases. For instance, while the global database of news we use (GDELT) is considered a reliable source of news media coverage across the world (Arva et al. 2014), it may also be biased towards US news media, which are comparatively more active media organizations (Kwak and An 2014).

The next section outlines previous work related to ours. Sections 3 and 4 present our data processing and annotation methodology. Section 5 presents the analysis of the results. The last section summarizes our findings and describes future work to extend this methodology to other domains.

2 Related Work

In this work, we compare media coverage of a broad and long lived social issue: *climate change*. We outline relevant work on climate change discourse (§2.1), and describe other comparative studies of social media and news media (§2.2).

2.1 The Discourse on Climate Change

Climate change has been singled out as one of the most urgent global challenges (Hoornweg 2011), generating a great deal of interest from communication scholars in recent years. Schmidt et al. (2013) perform a transversal study regarding news media coverage of topics related to climate change across 27 countries over a 15-year time frame. They look at the mainstream newspapers of each country and define the relevant articles as matching a specific search query. They found that events such as governmental meetings and report releases trigger increased conversations on climate change, and that such debates are more intensive in carbon-dependent countries. Across all countries, they observe that media attention about climate change fluctuates and peaks around specific events, which are usually of global interest. This pattern is typical of media reporting in general, which is often characterized by topic peaks (Rödder and Schäfer 2010; Boykoff and Rajan 2007). In contrast with these studies, that typically focus on a handful of news outlets, in this paper we analyze the news coverage of events across global

news media. Furthermore, looking at the coverage volume alone does not reveal nuances about the actors involved in the debate or how climate change is framed.

Molodtsova et al. (2013) show that the number of tweets on climate change correlates with extreme weather events, a correlation that also holds for opinion polls on climate change (Donner and McDaniels 2013). Along with weather events, Kirilenko and Stepchenkova (2014) and Segerberg and Bennett (2011) found that other major events of global or local interest ignited discussions about climate change on Twitter as well, including political elections, governmental meetings, and climate-related demonstrations. The study by Kirilenko et al. (2015) is closest to ours, as they look at both mainstream media coverage (14 news outlets) and attention patterns in Twitter. They analyze the influence of local weather anomalies on the volume of climate change publications in mainstream media and Twitter. In contrast, we juxtapose these media across a wide range of issues (not only weather) to understand the selection gap between them. Given our goal of *comparing climate change agendas*, we look at certain types of events that are often related to climate change, to seek an answer to which types of events are more prominent in one media or another.

2.2 Discourse Comparisons

There is a well-documented difference between what news journalists select to publish, and what their readers consume and share (Boczkowski 2010). Journalists have to adhere to deontological ethics and balance between "public interest and what the public is interested in" (Tandoc Jr 2014), which, in turn, might lead to different attention patterns between social and mainstream news media. Users tend to rely more on their social entourage to filter the news rather than on journalists (Hermida et al. 2012). Such research motivates our current study examining a comparison of climate change events that emerge in news media and on Twitter.

Comparative research (Esser and Hanitzsch 2012) of Twitter communications includes studies on hashtag lifecycles (Lehmann et al. 2012), usage across users of different languages (Hong, Convertino, and Chi 2011), or food consumption (Abbar, Mejova, and Weber 2015). In this work we study the prominence of different types of climate-related events as found on Twitter and in online news media.

Newspapers vs blogs. In contrast to Twitter, comparative research has been applied much more frequently to the analysis of the coverage and framing of various issues across newspapers and/or blogs, including religion (Baker 2010), surveillance (Diakopoulos et al. 2015), and immigration (Diakopoulos et al. 2013). Some studies have also examined and compared the climate change discourse between clusters of blogs corresponding to climate change acceptors and skeptics (Diakopoulos et al. 2014; Elgesem, Steskal, and Diakopoulos 2014). Instead, the focus of this study is not the debate between acceptors and skeptics, but the ways in which different news events feature in different media.

Other studies compare news with blogs, showing that there is a few hours lag between the attention peak of a meme (short sentence or phrase) in mainstream media and
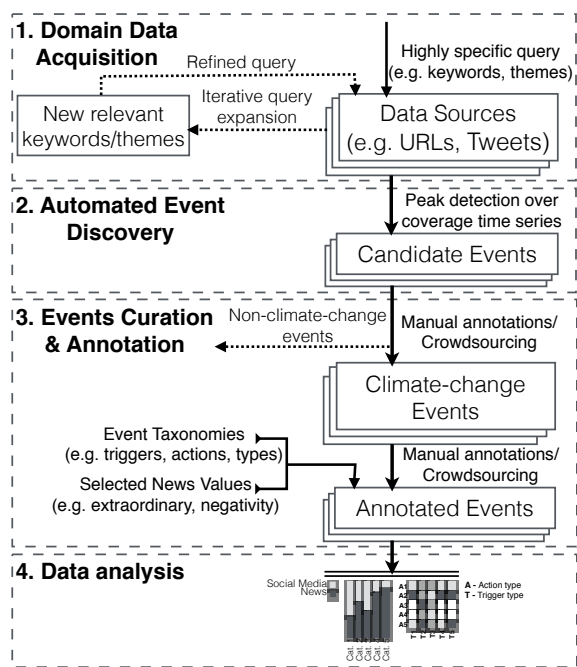
Figure 1: The main steps of our analysis framework: (a) domain data acquisition (§3.2 and §3.3), (b) automated event discovery (§3.3), (c) events curation and annotation (§4), and (d) data analysis (§5).

blogs (Leskovec et al. 2009). The media frames—the different ways of communicating about an issue—have also been studied to gain understanding into their impact on the perceptions about news (Papacharissi and de Fatima Oliveira 2008), as they are one important tool to shape public opinion (De Vreese 2005). Here we depict events to reveal nuances about the factors related to spikes in coverage of an event in mainstream news media and social media.

3 Data Collection and Candidate Events

In this section, we define the class of events we are interested in (§3.1), explain how we collected news articles (§3.2) and social media postings (§3.3) and describe the event detection framework used to generate candidate events (§3.4).

3.1 Defining "Climate Change" News

Our analysis is grounded in the current understanding of the discourse on climate change. For the purposes of this study, by the *discourse on climate change* we mean the discussion around its anthropogenic causes, adopting the definition used in the United Nations Framework for Climate Change (emphasis added): "a change of climate which is attributed directly or indirectly to *human activity* that alters the composition of the *global atmosphere* and which is in addition to natural *climate variability* observed over comparable time periods."[4] The three elements we have emphasized in

this definition delimit the scope of the news we consider: (i) the presence of human activity as causes, (ii) effects in the global atmosphere, and (iii) variations of climate.

Even with this definition, the boundary delimiting which news are related to climate change and which are not, is by no means absolute. Articles about climate change cover a large number of topics that vary from causes (e.g., CO_2 emissions, deforestation) to consequences (e.g., melting Arctic ice, extreme weather), current practices (e.g., fracking, coal use) and actions to stop it (e.g., electrical cars, recycling), just to name a few. Topics such as climate-induced migration and risks to food security, among others, are also frequently included in a long list of consequences of climate change: "we will continue to see rising oceans, longer, hotter heat waves, dangerous droughts and floods, and massive disruptions that can trigger greater migration, conflict, and hunger around the globe."[5]

We consider that a news article is about climate change if *it operates within the climate change frame*, in which framing is defined as a set of actions described by Kuypers (2009): (i) defining the problem, (ii) diagnosing its causes, (iii) making a moral judgment, or (iv) suggesting a remedy. We do not look for causation links between a news event and climate change, e.g. whether a severe storm is caused by climate change or not. Instead, we examine the way in which the storm is framed in the news, in this case, if it is described as being part of the climate change problem.

Our data sampling attempts to have a broad coverage of events related to climate change, which results in a set of *candidate events*, including several false positives. In Section 4 we describe a manual annotation process by which we remove spurious events.

3.2 News Data Acquisition

We use news data collected by GDELT (Global Data on Events, Location, and Tone)[6] and social media data from Twitter covering the same time interval.

Mainstream Media Collection. We use GDELT, currently the largest global event catalog, to automatically discover relevant events with high MSM coverage. GDELT releases data about daily media coverage in two formats: the Event Database and the Global Knowledge Graph (GKG).[7] GDELT covers a "cross-section of all major international, national, regional, local, and hyper-local news sources, both print and broadcast, from nearly every corner of the globe"[8] including major international news sources.

We use GDELT's GKG, as it provides the number and the list of news articles covering each event from their database, to discover the list of climate-change related events that received moderate to high media coverage between 1st April 2013 to 31st September 2014, baring January 2014 for which Internet Archive missed Twitter data; this covers 17 months. However, given that we are interested in the peak in

---

[4]http://unfccc.int/key_documents/the_convention/items/2853. php accessed 03.2015.

[5]US President Obama on climate change in the State of the Union Address: http://whitehouse.gov/sotu accessed 01.2015.

[6]http://www.gdeltproject.org accessed 03.2015

[7]http://gdeltproject.org/about.html accessed 03.2015

[8]http://tm.durusau.net/?p=47505 accessed 03.2015

the coverage, rather than in the number of events, here we directly use the news articles, not the events automatically mapped by GDELT; applying a consistent methodology for detecting events.

To locate the URLs corresponding to news articles relevant to climate change, we rely on GDELT themes and taxonomies, which are topical tags that automatically annotate events. To systematically identify all the GDELT themes and taxonomies that are related to *climate change* we first built the co-occurrence graph among them. We start with a set of relevant themes/taxonomies containing only the ENV_CLIMATECHANGE theme, and iteratively add themes, respectively taxonomies, that co-occur for at least 25% of their corresponding URLs with the ones already in the set (the relevance test). We do so until no theme/taxonomy is added. This results in a set of 39 themes (full list in our data release, details at the end of the paper). Then, we extract all the unique URLs corresponding to events annotated in GDELT with one of these themes for each day. The resulting collection of 561,644 URLs contains an average of about 30,000 URLs per month, with over 80% of the tags being tagged with the theme ENV_CLIMATECHANGE.

3.3 Social Media Data Acquisition

We rely on publicly available data covering about one and a half years of Twitter's Sample API[9], which we then retrospectively sub-sample. The quality of such sub-samples is discussed in e.g. Olteanu, Vieweg, and Castillo (2015).

To locate relevant tweets we start with a set of highly-specific terms about climate change, e.g. *climate-change, global_warming* (Pearce et al. 2014; Kirilenko and Stepchenkova 2014)—see our data release for full list—which we then expand in a snowball fashion as we did for themes/taxonomies in GDELT.

Candidate Term Selection. Given a Twitter collection obtained by sampling with a set of keywords $K_{climate}$—deemed relevant for climate change—we detect new relevant keywords by (1) extracting uni-grams and bi-grams that co-occur with terms in $K_{climate}$, and (2) rejecting those infrequent (occurring in less than 25 unique tweets[10]) or that contain only verbs, adjectives or adverbs (e.g. verb: *run*, adj.: *beautiful*, adv.: *often*)—typically not specific to any domain. When both a bi-gram and the uni-grams contained on it appear in this set, we keep only the bi-gram (the more specific term) if it accounts for more than one third of the uni-grams' frequency, otherwise we keep the uni-grams. Such automated approaches tend to miss-detect less precise terms e.g. *year, park, hell, light* (Olteanu et al. 2014), which we manually filter out. We refer to the remaining set as the *candidate terms*, K_{cand}.

Then, using the remaining terms, we build the co-occurrence graph with the terms in $K_{climate}$ and K_{cand}, and select from K_{cand} the terms that co-occur in at least 25%

of unique tweets[11] with terms from $K_{climate}$ (the relevance test). The creation of a co-occurrence sub-graph in Twitter is done in a stream processing fashion, avoiding loading the entire data in memory. Thus, we extract the tweets matched by terms in K_{cand}, mimicking the way in which Twitter does keyword tracking on both tweet text and the URLs contained on it. Then, we test each term from K_{cand} for relevance to climate change as described above. We keep repeat this process 5 times, discovering a total of 230 terms (full list available in our data release). Qualitatively, terms discovered in the last passes are less obviously about climate change than the terms discovered in the initial passes. This results in a collection of 482,615 tweets, an average of about 28,000 tweets per month. Given that this is a 1% sample, our estimate is that the tweets in our sample are representative of a larger set of around 2.8M tweets per month related to climate change.

3.4 Events Discovery

We analyze attention patterns in the scale of days and roughly follow the heuristic for activity peak detection used by Lehmann et al. (2012). To identify coverage peaks we compute the time series of the aggregated daily coverage in GDELT (respectively Twitter)—where the coverage is the number of URLs (respectively, tweets), c_i for each day d_i—and use a sliding window of $2m + 1$ centered around day d_i, with $m = 15$—resulting in a month-long time window. Then, within each window we juxtapose the volume on d_i, v_i, with a baseline represented by the median volume within the window. We declare a peak if v_i deviates more than 1.5 median average deviations (MAD) from the mean;[12] and $v_i > t_r$, where $t_r = 50$ is an arbitrary value used to filter out low-frequency peaks which tend to be vague.

This resulted in 218 peaks represented as a $\langle$date, theme$\rangle$ pair for GDELT, and 428 $\langle$date, keyword$\rangle$ pairs in Twitter.

Detectable events. The attention patterns of Twitter keywords have been described as belonging to three classes (Lehmann et al. 2012): (i) continuous, i.e. having a relatively constant volume, (ii) periodic, i.e. having spikes at regular periods, and (iii) isolated, i.e. having singular peaks. Similar observations hold for news consumption (Leskovec, Backstrom, and Kleinberg 2009; Castillo et al. 2014). As detailed by Lehmann et al. (2012), the method we discussed above will miss events that do not peak when observed at a granularity of one day, e.g. events that build slowly over weeks or months, or smaller phenomena occurring at a finer granularity (i.e., at the level of hours, minutes or seconds).

Events identification. We annotated each detected peak with the most likely event that triggered it. This annotation often takes the form of a news headline. To assist the event identification, we computed the frequency of uni-grams, bi-grams and tri-grams based on the text of the corresponding URLs (respectively, tweets). Then, we manually checked the

---

[9]These tweets are collected via Twitter's Sample API and can be found in the Internet Archive: https://archive.org/details/twitterstream accessed on 01.2015.

[10]We correct term frequency to account for cases when their prominence is caused by frequent bi-grams in $K_{climate}$.

[11]Here we compute statistics over the set of unique tweets to avoid biases due to viral tweets.

[12]We chose MAD for its' robustness (Leys et al. 2013), but also experimented with standard deviation, and the deviation function used by Lehmann et al. (2012), obtaining similar results.

items containing the most frequent n-grams, based on which we annotate the event. When two different sets of frequent n-grams referred to different events (e.g., the peak was due to two concomitant events) we add both of them; otherwise, if there were not clear sets of frequent n-grams referring to a single event, we mark the peak as *ambiguous*. When two different pairs ⟨date, theme/keyword⟩ referring to the same event co-occured within a half of month time window we map them to a single entry in our event list (e.g. typically a meeting or a natural hazard that lasted for several days).

This resulted in 195 candidate events in GDELT, out of which we marked 14 as *ambiguous* (possibly related to more than one news event); and 202 candidate events in Twitter, with 22 marked as *ambiguous*. Further, we note that many of the candidate events in Twitter were duplicates. For instance, a cartoon of a polar bear mending an iceberg with duct tape[13] peaked on 4 non-consecutive days in June and July 2014. We mark 12 such cases as duplicates.

4   Events Filtering and Annotation

As noted in the previous section, some of the automatically-identified events are not related to climate change. Two authors of this paper reviewed each event to remove false positives (§4.1) and to classify each event according to a taxonomy we present in this section (§4.2). Finally, we annotate events according to how they are perceived in terms of news values (§4.3). This section describes the annotation process, with the analysis deferred to the next section.

We use a mixture of annotation done by the authors and by crowdsource workers through the Crowdflower platform,[14] selecting workers in countries having a majority of native English speakers, collecting 5 independent annotations for every element (3 for the easier task of false positives removal), resolving disagreements by majority voting, and using a set of unambiguous test questions provided by the authors to catch inattentive workers, following standard recommendations from this platform.

4.1   False Positives Removal

The automatic data collection described above was designed to be inclusive, which has the disadvantage that some non-climate-change events get included in both the mainstream news and the social media collection.

Two of the authors of this paper review each one of these candidate events to remove false positives, i.e. events that do not match the definition given in Section 3.1. Two URLs were sampled from each event, including a Wikipedia entry or official activity/publication page when available, or the URL of a tweet, when no URL was available. Some cases are trivial to label, for instance when "climate change" or "global warming" are mentioned in the headline of a news article linked from the event. In many cases, however, the reference to climate change is indirect, e.g. a protest by Greenpeace against Procter and Gamble which is presented

as an action against deforestation, a cause of climate change according to the manifesto inviting to this demonstration.

Out of the 181 non-ambiguous news candidates, 122 (67%) were accepted, 43 (24%) rejected and 16 (9%) marked as borderline.[15] From the 168 non-ambiguous and non-duplicate[16] Twitter candidates, 119 (71%) were accepted, 46 (27%) rejected and 3 (2%) marked as borderline.

Next, we contrasted our labels with annotations provided by crowdsourced annotators on the same events. The options given to them were: coding (A) related to climate change, (B) weakly related to climate change, (C) not related to climate change, (D) cannot judge (e.g. broken links, not in English, or other issues).

Mapping them to our assessment of the same events (A and B correspond to accept and borderline, C and D correspond to reject), we observe a 77% agreement with the annotations from crowdsource workers. In general, crowdsource annotators applied a more narrow definition of climate change events, which often overlooked some elements of the news being analyzed. For instance, news about the development of a "Stem cell hamburger" were accompanied by statements from the scientists, in which they indicated that the development of this synthetic meat is motivated by reducing the number of farm animals and hence the methane released to atmosphere that causes climate change. This was missed by annotators who instead indicated this news was not related to climate change.

Disagreements in which crowdsource annotators labeled an event that we accepted as "not related to climate change" were further reviewed by a third author of this paper. This annotator rejected a further 30 events from that set. The final list contains 211 events, out of which only 25 events (about 25% from News, and 22% from Twitter) appear in both lists.

4.2   Event Annotations

We annotate each event according to a series of types and sub-types from previous work, as summarized in Table 1. According to the literature we cite in the table, climate change coverage in the news is often triggered by either a disaster, or by statements or actions of a group of people, or in some cases an individual. In the case of disasters, we further classify them as natural or human-induced (Fischer 1998). In the case of statements and actions of people, we divide them into the following categories of actors, following observations from previous work cited in Table 1:

– *Governmental organization:* Any institution belonging to any government branch (executive, legislative, judicial), or any inter-governmental agency, or any government employee acting in official capacity.
– *Non-governmental organization:* Any non-profit, non-governmental group, formally established or not. We include in this category educational and research institutions, which are all universities in our dataset.

---

[13]https://twitter.com/thereaibanksy/status/526438158742081537 among many others.

[14]http://www.crowdflower.com/

[15]When the event is only marginally associated with climate chance; e.g. while Greenpeace is often involved in climate change campaigns, the "Court Hearing: Greenpeace Activists to stay in jail" story in our event list rather focuses on the trial outcomes.

[16]Duplicate processing is described in §3.4.

Type	Sub-Type	Examples	Related categories in previous work
Disaster	Natural Hazards	Typhoon, Drought	(extreme) weather events (Schmidt, Ivanova, and Schäfer 2013; Molodtsova, Kirilenko, and Stepchenkova 2013; Donner and McDaniels 2013; Kirilenko and Stepchenkova 2014)
	Human-Induced Hazards	Deforestation, Oil Spill	deforestation (Boykoff and Rajan 2007)
Government (all branches) and inter-governmental agencies	Legal actions	New legislation	legislation, policies, international agreements (Schmidt, Ivanova, and Schäfer 2013; Hestres 2013), executive actions, police arrests (Hestres 2013)
	Publication/studies/research	Government-sponsored study	(inter-)government reports (Schmidt, Ivanova, and Schäfer 2013; Pearce et al. 2014), public surveys (Kirilenko and Stepchenkova 2014)
	Meetings/Conferences	IPCC meeting	(inter-)government meetings (Schmidt, Ivanova, and Schäfer 2013; Kirilenko and Stepchenkova 2014; Segerberg and Bennett 2011)
	Other (e.g. campaigns, statements)	City installs recycling bins	political elections, climate change adaptation database launch (Kirilenko and Stepchenkova 2014; Pearce et al. 2014), Keystone project (Hestres 2013)
Groups, NGOs, and universities	Legal actions by NGOs	Lawsuit initiated by NGO	petitions (Kirilenko and Stepchenkova 2014; Hestres 2013)
	Publication/studies/research	Academic research	scientific studies (Kirilenko and Stepchenkova 2014)
	Other (e.g. campaigns, statements)	Direct action e.g. cleaning a beach	labor unions, environmental groups statements (Schmidt, Ivanova, and Schäfer 2013), awareness campaign (Kirilenko and Stepchenkova 2014; Hestres 2013), march/protest organization (Segerberg and Bennett 2011)
For-profit (excl. media, universities)	Legal actions by for-profit entity	Lawsuit by for-profit group	—
	Publication/studies/research	Reports by for-profit group	scientists funded by carbon-based industries, memos (Boykoff and Rajan 2007)
	Other (e.g. campaigns, statements)	Google invests in solar energy	energy industry activities (Schmidt, Ivanova, and Schäfer 2013); paid media campaigns (Boykoff and Rajan 2007)
Media	Publication/studies/research	Newspaper investigation	reports/investigation by news media (Schmidt, Ivanova, and Schäfer 2013; Kirilenko and Stepchenkova 2014; Pearce et al. 2014)
	Other (e.g. campaigns, statements)	Campaign by newspaper	media activism (Segerberg and Bennett 2011), campaigns (Pearce et al. 2014),
Individuals	Legal actions by individuals	Lawsuit by individuals	—
	Publication/studies/research	New book	opinion/editorial (Kirilenko and Stepchenkova 2014)
	Other (e.g. campaigns, statements)	Bill Gates funds climate research	elite person campaign (Schmidt, Ivanova, and Schäfer 2013), statements/changes in opinion by individuals (Kirilenko and Stepchenkova 2014)

Table 1: Typology of events covered in media, in relation with categories described in previous work.

– *For-profit organization:* Any for-profit organization, including business and corporations but excluding media and universities, which appear in the other categories.
– *Media organization:* Any media organization.
– *Individual:* Any individual that is not acting as a representative of any of the organization types listed above.

We further categorize the actions of organizations or individuals as follows:

– *Legal actions:* Any action that is legally binding, including new executive orders and new laws, plus any action brought to a court of law, such as lawsuits.
– *Publications:* Any release of a document to the public, including reports, studies, memoranda, infographics and cartoons.
– *Meetings:* Any meeting, conference, convention, etc.
– *Other:* Other types of actions not belonging to the categories above, in our data this corresponded mostly to campaigns and brief public statements.

The annotation was done by two authors of this paper. We noted that an event can have more than one trigger, and we took this into account in our annotation, associating a second trigger to some events when deemed necessary.

Then, we again contrasted our labels with annotations provided by crowdworkers on this set of events. Workers were provided the same categories detailed above and were asked to choose the most likely one for each event (i.e., only

one type and sub-type). Mapping our assessment to theirs—we consider agreement if they choose one of our labels (either the first or the second type/sub-type)—we observe a 80.1% agreement for sub-types.

4.3 News values

Finally, to understand *why* a certain event is covered prominently, we annotate the events according to *news values*. News values are factors that determine the prominence with which an event is covered in the news. There are many news values, see e.g. the lists by Harcup and O'Neill (2001) and Stovall (2004). For the purposes of this analysis, after inspecting the list of events labeled as related to climate change, we decide to study the following six:

– *Extraordinary:* it is out of the ordinary or rare.
– *Unpredictable:* could not have been anticipated.
– *High magnitude:* has large global consequences.
– *Negative:* is bad news.
– *Conflictive:* involves two persons/groups in antagonism.
– *Related to elite persons:* rich, powerful or famous.

We do not claim these are all the news values that matter in this case, but given limited resources for annotation, bounding the number of them is necessary. This annotation is done through crowdsourcing by using instructions that echo the list above (full text of instructions, plus examples given to annotators for each class, are included in our data release).

	Disast.	Gov.	Non gov.	For-profit	Media	Indiv.
News	20.2%	62.6%	32.3%	6.1%	1.0%	4.0%
Twitter	7.1%	52.7%	29.5%	5.4%	8.9%	14.3%

| | Disaster | | | | | |
	Hum.	Nat.	Legal	Publ.	Meet	Other
News	3.0%	17.2%	19.2%	46.5%	13.1%	27.3%
Twitter	0.9%	6.3%	22.3%	37.5%	3.6%	47.3%

Table 2: Types and sub-types of events found in our dataset. Numbers add up to more than 100% because one event may have more than one type. Distributions are significantly different at $p < 0.01$.

We note that some of these tasks are more subjective than others, and hence elicit a lower level of agreement, as measured by the distribution of the agreement of annotators on each task (a value reported as the *confidence* on each annotation by the crowdsourcing provider). For instance, while references to elite persons and conflictive news are labeled with higher confidence (*median*=1.0), whether a news item is of high magnitude is a judgment in which there is less agreement among annotators (*median*=0.6). Other news values have in general a high level of agreement (*median*=0.8).

4.4 Examples

The full annotated dataset is available for research purposes. Some examples are the following:

– *"Climate refugee fighting stay in New Zealand,"* covered by news media and discussing the legal actions taken by a man from Kiribati Islands and the New Zealand government regarding an asylum request, was annotated as *neutral news* of *low magnitude*, yet *extraordinary* and *unpredictable*, and depicting a *conflict* between two entities.
– *"Climate change expert pleads guilty for fraud,"* debated on Twitter and discussing the fraud committed by a climate expert and former employee of the US Environmental Protection Agency, was annotated as *bad news* of *low magnitude*, yet *extraordinary* and *unpredictable*.
– *"Typhoon Haiyan,"* covered by news media and debated on Twitter as an event related to climate change leading to significant human and material loss, was annotated as *bad news* of *high magnitude*, *extraordinary* and *unpredictable*.

5 Data Analysis

This section presents our observations regarding event types (§5.1), news values (§5.2) and their interaction (§5.3).

5.1 Event types

Table 2 presents differences in coverage between mainstream media (MSM) and Twitter as present in our dataset. We observe significant differences in terms of coverage of disasters, which MSM favors much more than Twitter (20% vs. 7%); in the presence of media-triggered events—such as the publication of an investigation by a newspaper, which is an infrequent event in terms of global news coverage but
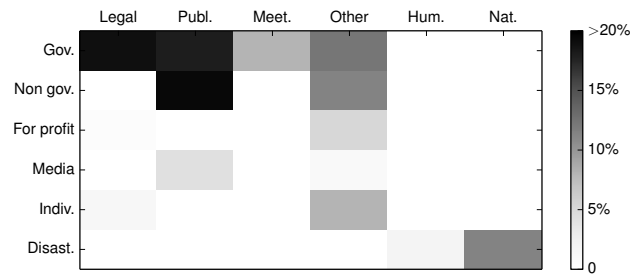
(a) Distribution of events into types and sub-types. Darker cells contain more events.
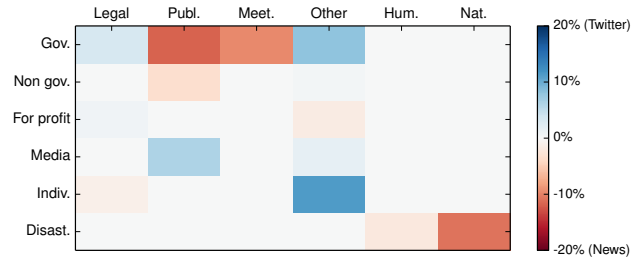
(b) Comparison of the distributions of events in main stream news and Twitter. Red indicates more events in mainstream news, blue indicates more events in Twitter (best seen in color).

Figure 2: Distribution of types and sub-types.

does trigger significant reactions in Twitter (1% vs. 9%); and in the coverage of individual actions, which are given less prominence in news compared to Twitter (4% vs. 14%).

There are interesting similarities and differences between types of actors and actions covered in both types of media, as depicted in Figure 2:

– Government/inter-governmental agencies, which receive the largest amount of coverage in both (top row of Figure 2(a)), are discussed in relation to a broad range of action types. The main difference seems to be a larger coverage of publications and meetings in MSM, contrary to coverage of legal and other types of actions which are covered more often in social media (top row of Figure 2(b)).
– Non-governmental groups (and universities), are covered in both cases mostly due to publications, and also through other actions (second row of Figure 2(a)) such as campaigns and public statements.
– For-profit organizations are covered mostly due to other actions (third row of Figure 2(b)), which are usually advertising and announcements of projects.
– Media organizations become protagonists with respect to climate change through their original investigative reporting (fourth row of Figure 2(a)), yet, the number of events they create in news is lower than in Twitter (fourth row of Figure 2(b)). This is because an original investigation by one news source will rarely be quoted by many other news sources, but it can have a significant impact in Twitter.
– Individuals are covered in both media occasionally with respect to actions (fifth row of Figure 2(a)), which are usually public statements. Individuals receive much more

	Extraordinary		Unpredictable		High Magnitude			Negative			Conflictive		Ref. Elite Persons	
	Extraord.	Ordinary	Unpred.	Pred.	High	Moderate	Low	Negative	Neutral	Positive	Conflict.	No	Elite	No
News	83.8%	16.2%	82.8%	17.2%	34.3%	54.5%	11.1%	43.4%	40.4%	16.2%	10.1%	89.9%	22.2%	77.8%
Twitter	75.9%	24.1%	80.4%	19.6%	25.0%	55.4%	19.6%	34.8%	46.4%	18.9%	18.8%	81.3%	21.4%	78.6%
Both	79.6%	20.4%	81.5%	18.5%	29.4%	55.0%	15.6%	38.9%	43.6%	17.5%	14.7%	85.3%	21.8%	78.2%
	**				**						**			

Table 3: Analysis in terms of news values of events covered in our mainstream news and Twitter datasets. Asterisks in the last row highlight statistically significant differences at $p < 0.01$ (***), $p < 0.05$ (**), $p < 0.10$ (*).
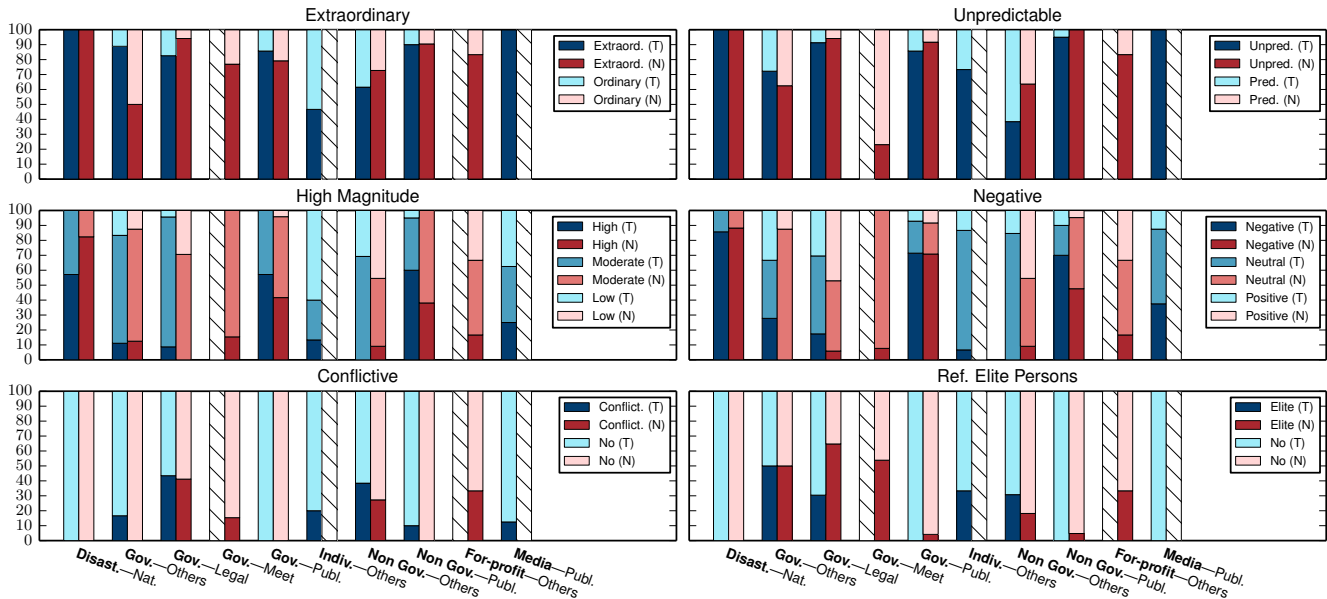

Figure 3: Distribution of news values for types/sub-types of events in Twitter (T, in blue) and mainstream news (N, in red). Hatched bars indicate insufficient data (less than 5 events). (Best seen in color.)

attention in Twitter than in traditional news media (fifth row of Figure 2(b)).

– Disasters in general are covered more by mainstream news than by Twitter, mostly due to their coverage of natural disasters (last row of Figure 2(b)). Disasters have been observed to be a prominent subject in international news articles collected by GDELT (Kwak and An 2014).

As regards the 25 events that were prominently covered by both media, 60% were primarily triggered by government/inter-governmental agencies (e.g. *"UN Climate Summit 2014"*), 24% were campaigns or publications by non-governmental groups (e.g. *"2013 Earth Day/Week"*) and 16% natural disasters (e.g. *"Typhoon Haiyan"*).

5.2 News values

The analysis of news values in our data, shown in Table 3, indicates that both media tends to cover events that are (i) extraordinary, (ii) unpredictable, (iii) of moderate and high magnitude, and (iv) negative or neutral. Contrary to what one could assume given the literature on news values, events involving conflict or referencing elite persons are not the majority of news about climate change.

There are significant differences between MSM and Twitter, with relatively more coverage given in Twitter to ordinary events in comparison to MSM. Twitter also has relatively more coverage of events that are considered of relatively low magnitude and that involve two groups or individuals in antagonism.

5.3 Event types and news values

The differences in terms of news values that we observe between MSM and social media are largely correlated with the selection of events they cover, as can be observed by comparing both media in the same category. This is depicted in Figure 3, which only includes cases when there are at least 5 events. News values for the same type of event are often similar between these two media, save for small differences.

While most of the reported events on climate change are extraordinary (rare), there is one exception in which more ordinary news events are the majority, which are individual actions, featured significantly (5 events or more) on Twitter, but not in MSM. While most events are also unpredictable, there are events announced well in advance, such as govern-

mental and inter-governmental meetings. These news events only feature significantly on MSM, but not on Twitter.

As regards magnitude, disasters and publications are often the ones linked to the largest effects. In the case of disasters, it is by their consequences: most of the disasters that are associated by the press or Twitter to climate change are severe weather phenomena affecting large areas. In the case of publications, this has more to do with the content of the publications, sometimes describing existential threats to humans as a whole. The overall lower magnitude of events covered in comparison to news may be explained by the confluence of two observations: (1) Twitter focuses more on events with individual triggers which tend to have lower magnitude ratings, and (2) MSM focuses more on disaster events which tend to have higher magnitude ratings.

In terms of negativity, most news are neutral or bad. Proportionally, the most negative news are those related to disasters and to publications.

While most news events do not involve people/groups in conflict, the cases in which they have a more conflictive content are legal actions by governments (which usually are targeted at a specific group, such as a mining corporation), and statements and actions by non-governmental organizations (e.g. statements by an NGO against a certain industry).

Finally, references to elite persons (famous, rich, or powerful) are almost never included in publications (governmental or non-governmental sources), but are present in some minority amount in the remaining categories.

6 Conclusions

This section outlines our main conclusions. We include a summary of the observations about the discourse on climate change in media (§6.1), and discuss how to extend the methodology we propose to other domains (§6.2).

6.1 Climate Change in Mainstream News and Social Media

There are interesting similarities between mainstream news and social media, both in terms of the types of events they cover and the news values of those events. However, there are also striking differences, and they do tend to peak in activity around different news events, with an overlap in their peaks of attentions of about 22%-25% of the events.

Disasters. A key trigger of news coverage on climate change are disasters, both natural and human-induced. Disasters covered with respect to climate change tend to be severe atmospheric events affecting large parts of the globe. There is an important difference between MSM and Twitter, with MSM covering these events much more than Twitter.

Publications, meetings, and legal actions. News events on climate change are usually triggered by publications describing negative, global-scale consequences of climate change. News coverage of climate change is also triggered by legal actions initiated by governments, like passing new laws and bringing lawsuits against corporations. The coverage of these events differs in MSM and Twitter. In MSM, government/inter-governmental meetings and publications receive comparatively more attention than in Twitter, where legal actions and official statements have a greater impact.

Individual actions. Actions by individuals appear prominently on Twitter. In about half of the cases, these individuals do not belong to the elite: they are neither rich, nor powerful, nor famous. Twitter indeed allows those individuals, in many cases, to generate peaks of attention as large as the ones that are obtained by large organizations or governments.

Recommendations. For *activists and advocates*, publications highlighting high-impact negative effects of climate change feature prominently across both types of media, and seem to be picked up by social media even when they do not include endorsements by elite persons or references to them. For *public relations or for-profit corporations*, discussions about lawsuits involving corporations, while not appearing so prominently in mainstream media, circulate in social media. For *media organizations*, the alignment between mainstream news and social media news on this topic is significant, but there are many gaps. It would not be unreasonable to look at what are the news events in which there is the larger gap in favor of social media, particularly actions and public statements by individuals, as opportunities to disseminate information that may appeal to social media users.

6.2 Towards a General Method for Comparing Online Media

The method we have presented here can be extended to a variety of topics in the news. GDELT associates hundreds of themes to news articles, enabling analysts to perform the same procedure we have described for other themes (e.g. HEALTH_PANDEMIC, IMMIGRATION). In the case of Twitter, any topic from which a sub-set of initial hashtags can be identified is amenable to the same event discovery process.

Two important elements require adaptation. First, the triggers and actions should be specific to the topic, although some overlap with the ones we have used here is expected (i.e. government, non-government, for profit, etc.). Second, the selection of the relevant news values also requires some familiarity with the topic, and as in this paper, it is hard to claim it is in any sense an optimal selection.

Applying this event-driven methodology to the discovery of differences between mainstream media and social media in other domains may lead to findings as interesting as the ones we uncover here. These findings can be contrasted with those from qualitative analysis, particularly of events that generate peaks of attentions in both media simultaneously.

References

Abbar, S.; Mejova, Y.; and Weber, I. 2015. You tweet what you eat: Studying food consumption through Twitter. In *Proc. of CHI*.

Arva, B.; Beieler, J.; Fisher, B.; Lara, G.; Schrodt, P. A.; Song, W.; Sowell, M.; and Stehle, S. 2014. Improving forecasts of international events of interest. In *EPSA Annual General Conf. Paper*.

Baker, P. 2010. Representations of islam in british broadsheet and tabloid newspapers 1999–2005. *J. of Language and Politics*.

Boczkowski, P. J. 2010. The divergent online news preferences of journalists and readers. *Communications of the ACM*.

Boykoff, M. T., and Rajan, S. R. 2007. Signals and noise. *EMBO reports*.

Castillo, C.; El-Haddad, M.; Pfeffer, J.; and Stempeck, M. 2014. Characterizing the life cycle of online news stories using social media reactions. In *Proc. of CSCW*.

De Vreese, C. H. 2005. News framing: Theory and typology. *Information design journal+ document design*.

Diakopoulos, N.; Zhang, A. X.; Elgesem, D.; and Salway, A. 2014. Identifying and analyzing moral evaluation frames in climate change blog discourse. In *Proc. of ICWSM*.

Diakopoulos, N.; Elgesem, D.; Salway, A.; Zhang, A.; and Hofland, K. 2015. Compare clouds: Visualizing text corpora to compare media frames. In *Proc. of IUI Workshop on Visual Text Analytics*.

Diakopoulos, N.; Zhang, A. X.; and Salway, A. 2013. Visual analytics of media frames in online news and blogs. In *Proc. IEEE InfoVis Workshop on Text Visualization*.

Donner, S. D., and McDaniels, J. 2013. The influence of national temperature fluctuations on opinions about climate change in the US since 1990. *Climatic change*.

Elgesem, D.; Steskal, L.; and Diakopoulos, N. 2014. Structure and content of the discourse on climate change in the blogosphere: The big picture. *Environ. Communication*.

Entman, R. M. 2007. Framing bias: Media in the distribution of power. *J. of communication*.

Esser, F., and Hanitzsch, T., eds. 2012. *The Handbook of Comparative Communication Research*. ICA Handbooks. Routledge.

Fischer, H. W. 1998. *Response to disaster: fact versus fiction & its perpetuation—the sociology of disaster*. Univ. Press of America.

Harcup, T., and O'Neill, D. 2001. What is news? Galtung and Ruge revisited. *Journalism studies*.

Hermida, A.; Fletcher, F.; Korell, D.; and Logan, D. 2012. Share, like, recommend: Decoding the social media news consumer. *Journalism Studies*.

Hestres, L. E. 2013. Preaching to the choir: Internet-mediated advocacy, issue public mobilization, and climate change. *New media and society*.

Hong, L.; Convertino, G.; and Chi, E. H. 2011. Language matters in Twitter: A large scale study. In *Proc. of ICWSM*.

Hoornweg, D. 2011. *Cities and climate change: responding to an urgent agenda*. World Bank Publications.

Kirilenko, A. P., and Stepchenkova, S. O. 2014. Public microblogging on climate change: One year of Twitter worldwide. *Global Environ. Change*.

Kirilenko, A. P.; Molodtsova, T.; and Stepchenkova, S. O. 2015. People as sensors: Mass media and local temperature influence climate change discussion on Twitter. *Global Environ. Change*.

Kuypers, J. A. 2009. *Framing Analysis*. Lexington Press. 181+.

Kwak, H., and An, J. 2014. A first look at global news coverage of disasters by using the gdelt dataset. In *Social Informatics*.

Kwak, H.; Lee, C.; Park, H.; and Moon, S. 2010. What is Twitter, a social network or a news media? In *Proc. WWW*.

Lehmann, J.; Gonçalves, B.; Ramasco, J. J.; and Cattuto, C. 2012. Dynamical classes of collective attention in Twitter. In *Proc. of WWW*.

Leskovec, J.; Backstrom, L.; and Kleinberg, J. 2009. Meme-tracking and the dynamics of the news cycle. In *Proc. of KDD*.

Leys, C.; Ley, C.; Klein, O.; Bernard, P.; and Licata, L. 2013. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *J. of Experimental Social Psychology*.

Molodtsova, T.; Kirilenko, A.; and Stepchenkova, S. 2013. Utilizing the social media data to validate'climate change'indices. In *AGU Fall Meeting Abstracts*.

Olteanu, A.; Castillo, C.; Diaz, F.; and Vieweg, S. 2014. CrisisLex: A lexicon for collecting and filtering microblogged communications in crises. In *Proc. of ICWSM*.

Olteanu, A.; Vieweg, S.; and Castillo, C. 2015. What to expect when the unexpected happens: Social media communications across crises. In *Proc. of CSCW*.

Papacharissi, Z., and de Fatima Oliveira, M. 2008. News frames terrorism: A comparative analysis of frames employed in terrorism coverage in us and uk newspapers. *The Int. J. of Press/Politics*.

Pearce, W.; Holmberg, K.; Hellsten, I.; and Nerlich, B. 2014. Climate change on Twitter: Topics, communities and conversations about the 2013 IPCC Working Group 1 Report. *PLOS ONE*.

Rödder, S., and Schäfer, M. S. 2010. Repercussion and resistance. an empirical study on the interrelation between science and mass media. *Communications*.

Schmidt, A.; Ivanova, A.; and Schäfer, M. S. 2013. Media attention for climate change around the world. *Global Environ. Change*.

Segerberg, A., and Bennett, W. L. 2011. Social media and the organization of collective action: Using Twitter to explore the ecologies of two climate change protests. *The Communication Review*.

Stovall, J. G. 2004. *Journalism: who, what, when, where, why and how*. Pearson.

Tandoc Jr, E. C. 2014. Why web analytics click: Factors affecting the ways journalists use audience metrics. *Journalism Studies*.

Weart, S. R. 2008. *The discovery of global warming: revised and expanded edition*. Harvard Univ. Press.