The use of social media to determine the most important articles in an online debate

Abstract

When someone wants to study how an issue is portrayed in the media, a common strategy is to perform a content analysis: the researcher gathers a significant sample of the texts from the corpus of all published stories on that issue and scrutinizes those texts. A significant sample might mean a sample that is large enough to be representative of the overall media landscape and therefore useful to make statements that can be generalized. However, in Communication studies, the distribution of media attention is usually uneven: a few stories grab almost all the attention, while the remaining stories are barely noticed by the public. Therefore, it is reasonable to suggest that a significant sample for Communication studies is a sample that prioritizes the articles with more impact on public opinion. This paper argues that internet data can offer a very convenient and easily available solution for the problem of determining the most significant sample. An increasing amount of people access news pieces through social networks. Therefore, we can use social network data to identify those articles that people pay attention to (and therefore are more valuable when it comes to the influence on public opinion). As a case study, this paper analyzes the corpus of stories (and associated social network data) related to the People's Climate March in New York City.

1 Introduction

The traditional approach to select a representative corpus for Communication research has relied on readership data or perceived prestige of media outlets. For instance, when Cavender & Mulcahy (1998) tried to answer the question "does the low salience in public opinion of corporate deviance come from a lack of media coverage?", they scrutinized only three newspapers: *New York Times*, *Washington Post*, and *Wall Street Journal*. Their justification for such a restrictive criterion was that "these influential daily newspapers share the status of 'newspapers of record'" and therefore are representative of the broad media picture.

An alternative approach has been to explore academic databases like LexisNexis – the most widely used news archive in the social sciences – or Web-based portals like Google News. Those digital tools aggregate content from thousands of media sources every day. However, they also come with downsides. Academic databases are usually blind to wire stories that are published in major papers (Weaver & Bimber, 2008). On the other hand, in a Web-based portal, it is not trivial to separate the substantive content from noise, e.g. ads (Arizaleta, 2009). And, in both cases, the result sets can be so big that it is virtually impossible to perform meaningful content analysis with each one of the individual articles.

In addition, both aforesaid approaches share the same serious limitation: they are blind to articles that come from independent media or from unusual and occasional players in the media game. Before the advent of the Internet, those texts could be ignored because, if they did not pass through the gateways of established media, they would hardly reach a large audience. But the world has changed.

Benkler (2006) suggests that a new and more democratic public sphere has been born and social actors who had remained silent so far can now engage in the public dialogue: "The networked information economy makes it possible to reshape both the 'who' and the 'how' of cultural production [...]. It adds to the centralized, market-oriented production system a new framework of radically decentralized individual and cooperative nonmarket production."

In a study about the online coverage of SOPA-PIPA – the Hollywood-sponsored anti-piracy legislation –, Benkler et al. (2013) present an wealth of instances to make the point that, in the *networked public sphere*, every node can have a significant impact on the media ecosystem. For example, "an often-repeated meme from the earliest stages of the [SOPA-PIPA] debate had been that the cost of piracy to the United States is $58 billion each year." That frame, sponsored by Hollywood and widespread in mainstream media, was challenged by a blog post at the Cato Institute Website. Notwithstanding its marginal position in the news environment, the blog post's argument rapidly spread across the network and echoed in prestigious media outlets. Eventually, it discredited the $58 billion frame.

In order to overcome the aforementioned limitations of traditional approaches, this paper proposes an alternative method that includes the gathering of the corpus from online tools and the use of social network data to identify the articles in the corpus that had the highest impact on public opinion. Those articles should be submitted to the content analysis (instead of the entire corpus that usually is mostly comprised by irrelevant articles).

The People's Climate March – that occurred on September 21st in New York City – was chosen as the study case to test the methods described by this paper.

2 Case Study: Contextualization

On Sunday, September 21st, New York hosted the largest climate march in history (Foderaro, 2014). With about 300,000 participants, it advocated for climate action on the verge of the United Nations Climate Summit of world leaders that took place in the city two days later, on September 23rd.

The event was proposed by 350.org in May 21st 2014 through an article published by writer and activist Bill McKibben – 350.org's founder – on the website of the Rolling Stone magazine (McKibben, 2014). In that article, he blames the fossil-fuel industry, which "by virtue of being perhaps the richest enterprise in human history, has been able to delay effective action, almost to the point where it's too late." He claims that the global climate justice movement "needs to come together and show the world how big it's gotten," and to allow for "opening up space for change." "A loud movement – one that gives our 'leaders' permission to actually lead, and then scares them into doing so – is the only hope of upending [the prophecy that it's already too late to act]."

After that first call, other 1,500 organization joined the effort, "including many international and national unions, churches, schools and community and environmental justice organizations." (Boehrer & Visser, 2014) The organizers stressed many times that the event was not devised as a protest, but as a response to the UN Climate Summit. They intended the march to be the largest single event on climate that has been organized to date, one so large and diverse that it could not be ignored. According to the official Website, other 2,646 "solidarity events" happened in 162 countries.

3 Method

The methodology described in this paper is comprised by two different and complementary processes: the effort to gather a representative and extensive corpus on the one hand and the selection of the most popular articles based on Facebook data on the other.

At the end, there is also a supplementary explanation on how to use Twitter data to interpret the audience's reaction to the most significant stories.

3.1 Gathering a corpus

The first step in any study that involves content analysis is the selection of a set of texts from which the researcher picks the pieces that will constitute her corpus. In this section, I describe three methods of gathering that corpus.

3.1.1 Media Cloud

Taking into account the aforementioned inadequacy of traditional methods for corpus gathering in the highly diverse online environment, researchers at the Berkman Center for Internet and Society at Harvard and the MIT Center for Civic Media developed Media Cloud.

Like LexisNexis and Google News, Media Cloud is a news database. Unlike them, it is not restricted to established media outlets, but encompasses a much broader picture of the media ecosystem that includes independent, governmental, institutional, and alternative media.

Media Cloud checks RSS feeds of about 40,000 websites every day and downloads the new content available. An algorithm removes the undesired content – usually ads and menus

```
import mediacloud
mc = mediacloud.api.MediaCloud('HERE_YOUR_API_KEY')
stories = mc.storyList('(climate AND march)',
                    '+publish_date:[2014-08-21T00:00:00Z TO 2014-11-21T23:59:59Z]',
                    0, 1000)
```

Figure 1 – Code snippet in Python for obtaining all stories about "climate march" between Aug 21st and Oct 21st. Complete code at https://github.com/alexgonca/ClimateMarch/blob/master/articlesaboutmarch.py.

– and stores the substantive information in the Media Cloud database. So far, it has already collected 235 million of stories, most of them in English, but with a significant and a fast-growing corpus in other languages, like Portuguese and Russian.

Researchers can use a search engine to find information in that database. The easiest and most powerful way of accessing the search engine is through the Media Cloud API[1] – a set of routines and protocols that describe how to query the information one needs. There is an interface for that API in Python[2]. It is worth noting that an API key is needed in order to send any request to the Media Cloud server. Therefore, researchers have to register in the Media Cloud website (www.mediacloud.org).

Figure 1 shows a code snippet for obtaining all the stories with the keywords "climate" and "march" during the month before and after the People's Climate March. That specific query returned 4,152 articles.

The distribution of those articles around the period of the event is presented in Figure 2. The shape of the curve is a very traditional one when compared to similar studies about news coverage. Walter Lippman famously observed that "the press is… like the beam of a searchlight that moves restlessly about, bringing one episode and then another out of darkness

[1] Available at http://mediacloud.org/api/
[2] Available at https://github.com/c4fcm/MediaCloud-API-Client

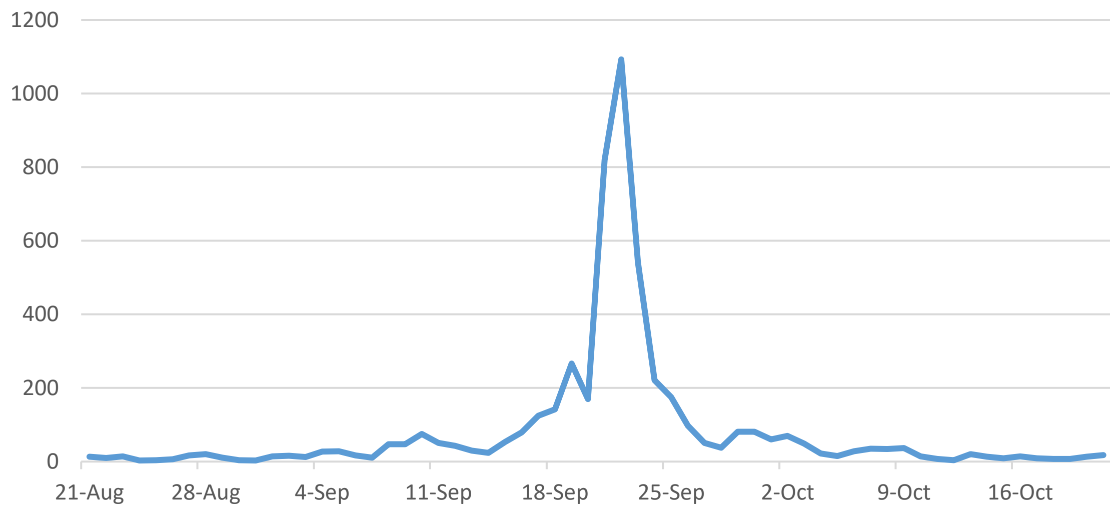

Figure 2 - Number of articles about "Climate March" extracted from Media Cloud.

into vision" (Lippmann, 1922). But, as the metaphor suggests, the attention span of a restless beam is usually short. The focus quickly migrates to the next item of the agenda. Accordingly, there was a considerable increase of media attention in the week before the Climate March, followed by an impressive surge in the number of articles around September 21st and a predictable plunge in the subsequent week. On average, in the weeks before and after the march, about 275 articles were published every day. In the other six weeks of the research period, the average was only 25 articles per day, less than a tenth of the frequency in those two weeks around the event.

3.1.2 Google

Another way of gathering a corpus is through Google Search. In the "advanced" options of the search engine, it is possible to choose a specific date to search for webpages that mention the desired keywords. Therefore, one can go through the search results for each one of the dates in a controversy and store them in a database.

```python
from selenium import webdriver
import selenium.common.exceptions
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support import expected_conditions as EC
from jdcal import gcal2jd

# opens Firefox
fp = webdriver.FirefoxProfile()
fp.set_preference("browser.private.browsing.autostart", True)
driver = webdriver.Firefox(firefox_profile=fp)
# access Google Search
driver.get("http://www.google.com/")
# for google search, dates must be in the Julian calendar.
julian_date = int(gcal2jd(2014, 8, 21)[0] +
                  gcal2jd(2014, 8, 21)[1] + 0.5)
# perform query with "climate march"
inputElement = driver.find_element_by_name("q")
inputElement.send_keys('"climate march" daterange:' + str(julian_date) +
                       '-' + str(julian_date))
inputElement.submit()
```

Figure 3 – Code snippet in Python for obtaining the search results for "climate march" in Google. Complete code at https://github.com/alexgonca/ClimateMarch/blob/master/google.py.

That seems theoretically simple. However, it demands some computer automation. Otherwise, it can take forever to perform the data scraping of the Google results, especially for long periods of time.

Then, another challenge arises: when Google realizes that it is a bot – and not a human – who is performing the searches, it temporarily blocks the IP address of the requester. Again, it can take forever if one has to wait Google "quarantines" to end before resuming the collection of search results.

The solution is to build a software capable of mimicking human behavior. The easiest way of doing that is using a method called "browser automation". The requests are done by the traditional interface a human agent would use: a Web browser – like Google Chrome, Internet Explorer, or Mozilla Firefox. The only difference is that the browser is controlled by a bot instead of a human.

Selenium WebDriver (www.seleniumhq.org) is probably the most robust programming package for this kind of automation. Figure 3 shows part of the computer code used to automate the Google Search.

The use of this method for the search terms "climate march" provided 12,976 articles for the period between August 21st, 2014 and October 21st, 2014.

3.1.3 A hybrid approach

Each one of the aforementioned methods for gathering an initial corpus have their own advantages and limitations.

Google search usually provides meaningful results mixed with a lot of noise since the Web crawler distinguishes quality stories from internet debris only to a limited degree. On the other hand, Media Cloud works with a manually curated list of sources. Therefore, its results are usually more significant.

In addition, Google search does not give a sense of the evolution of a controversy or event. As Figure 4 shows, based on Google data, it is hard to know when the People's Climate

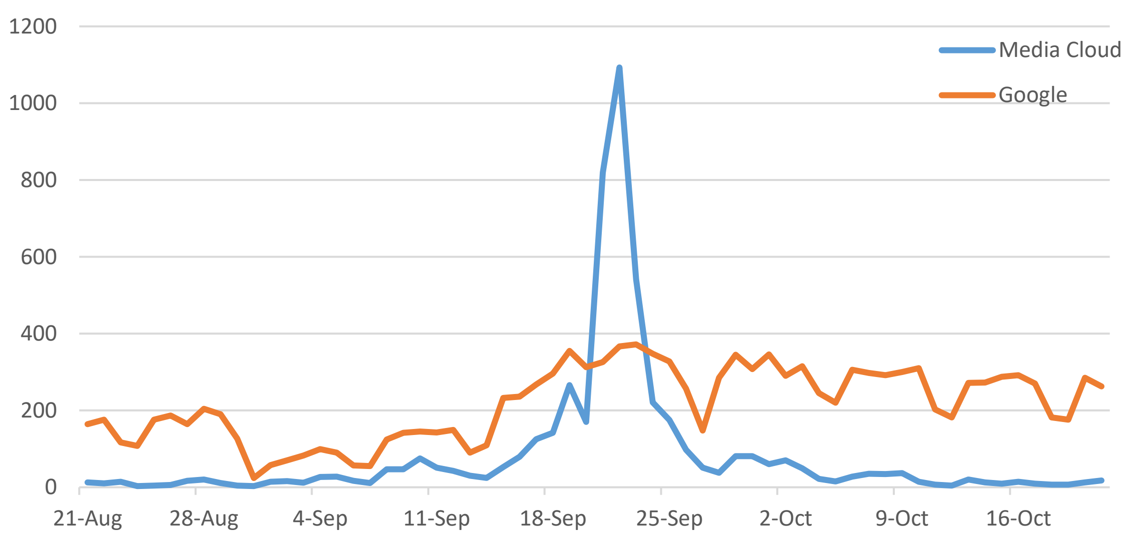

Figure 4 – Comparison between the numbers of articles extracted from Google and Media Cloud.

March occurred (September 21st), although it is apparent in the Media Cloud data. The number of articles for each day is much more constant in the Google results – a well-known behavior for everybody who is familiar with Google search engine. Therefore, the Google method does not offer insights related to temporal trends.

Another advantage of Media Cloud is that it downloads the texts of the articles and separates them from ads, menus, and other unrelated content. Google only provides the URLs for the stories and the researcher must download them later.

On the other hand, Media Cloud also has its drawbacks. For instance, Media Cloud can have blind spots: outlets that are important but that have been overlooked by the administrators of Media Cloud.

Moreover, as it is based in RSS, it cannot look into the past. Therefore, the articles from a specific media source will only be available after the inclusion of that source in Media Cloud's list of RSS feeds. Stories published before that date will not appear in the database.

In order to overcome those limitations, it is advisable to combine data from Google and Media Cloud. According to this approach, the same search terms should be used in both systems and with the same time period. Probably, there will be many duplicated records that will have to be discarded.

Then, the next step is to get rid of the unrelated articles both in Google (arguably many) and Media Cloud (hopefully only a few) that were included by accident because they mention the search terms in other contexts. This method can look a bit overwhelming since the number of results can be enormous and it is very difficult to automate this kind of judgment. In the case of the People's Climate March, the combined result set has more than 17,128 URLs.

However, it is not necessary to look at each one of those articles, but only at the most important ones: those that people are reading and talking about. One can, for instance, read the most popular articles that grab around 70% of the social media attention. Even for a corpus with dozens of thousands of articles, that selection would usually amount to a few hundreds. How to identify those articles is the topic of the next section.

3.2 Finding the most important articles

After gathering a corpus, the next step is to discover which articles received attention from the audience and therefore can be deemed as important in shaping public opinion. If the number of stories were smaller, that filtering would not be necessary. It is fairly simple to manually analyze a few hundred articles. However, with thousands of pieces, there are only two solutions: the use of computational methods to make sense of the complex information conveyed by each story – a technique that is still in its infancy – or the determination of the articles that are really worthwhile to read – approach preferred by this paper.

In order to determine the most relevant articles, a number of approaches has already been tested. The SOPA/PIPA (Benkler, Roberts, Faris, Solow-Niederman, & Etling, 2013) and the Trayvon Martin (Graef, Stempeck, & Zuckerman, 2014) studies, for instance, used a tool called Controversy Mapper to estimate the relative influence of each one of the articles in the corpus. Controversy Mapper represents the corpus as a network. Each edge is a link from one media outlet to another. Presumably, the relevance of a media outlet will be directly proportional to the number of inlinks it has – in fact, an analogous criterion to Google's PageRank algorithm.

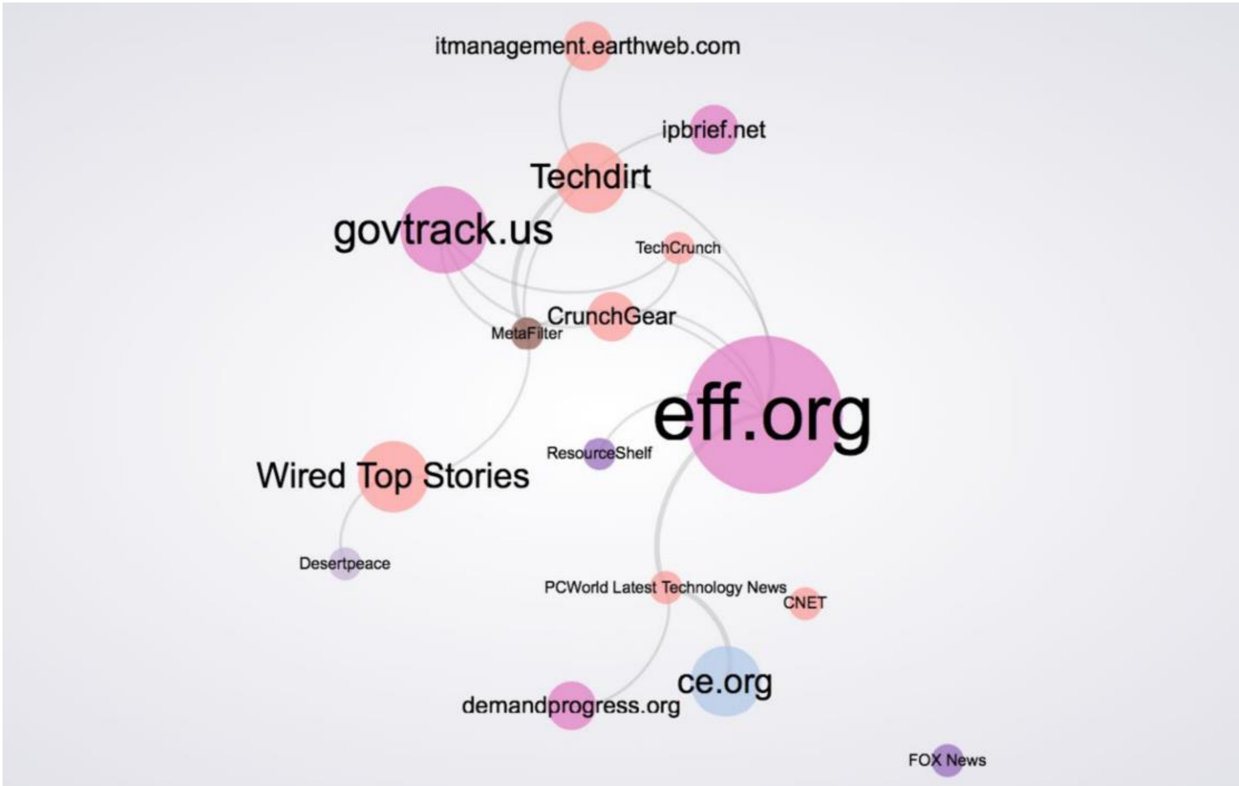

Figure 5 – Network of media outlets in the SOPA/PIPA debate generated by Controversy Mapper. (Benkler, Roberts, Faris, Solow-Niederman, & Etling, 2013)

Figure 5 shows a graphical representation of the network built with Controversy Mapper. The size of the nodes is proportional to the number of inlinks each media outlet receives. Thus, the most prominent media players become apparent in the network.

However, some corpora show a rather low number of mutual connections which precludes the use of Controversy Mapper. For this reason, it is necessary to find another estimation for the relative importance of each article. Trayvon Martin, for instance, in addition to the Controversy Mapper strategy, also had recourse to bitly data, the well-known URL shortener service. For each link in the corpus, bitly was able to provide the click count analytics related to those links that were disseminated using the shortener. As in bitly's dashboard, the click counts were broken down by where the link was shared: Twitter, Facebook, or other.

```
import requests
queryFacebook = "SELECT url, normalized_url, " \
         "share_count, like_count, comment_count, " \
         "total_count, commentsbox_count, comments_fbid, click_count " \
         "FROM link_stat " \
         "WHERE url = \"" + url.replace('\"', "%22") + "\""
facebook_url = "https://graph.facebook.com/fql"
data = requests.get(facebook_url, params={'q': queryFacebook}).json()
```

Figure 6 – Code snippet in Python for obtaining the Facebook total count for each article in the corpus. Complete code at https://github.com/alexgonca/ClimateMarch/blob/master/facebookmediacloud.py.

With the bitly data, Graef, Stempeck, & Zuckerman were able to build a ranking of the most influential media outlets in each phase of the Trayvon Martin controversy and had interesting insights about the role of race-specific media in the debate. Nowadays, it would be even easier to obtain the same information. There is a bitly API that provide access to the link statistics data[3] with minimum effort.

This paper, however, advocates for the use of Facebook data. The sum of comments, likes and shares that a link received on Facebook is a readily available information via the Facebook Graph API and seems a good representation of influence due to the widespread usage of social networks. Facebook claims to have 1.35 billion monthly active users, around 19% of the world population (Facebook, 2014). Only in the United States and Canada, about 152 million people access the social network daily.

For convenience, in the context of this research, the term **total count** will be used as a shorthand for the sum of shares, likes, and comments for a link posted on Facebook and is interpreted as a consolidated measure of the attention received by that URL on the social network. With the total count number, it is possible to create a ranking with the articles that

[3] Available at http://dev.bitly.com/link_metrics.html.

received the highest share of attention on Facebook. Figure 6 shows the Python code to obtain the total count from the Facebook API.

Then, it is important to specify some criteria to define a threshold for the articles that will be included in the analysis. A simple example is to focus on the 300 top-rated articles in the corpus. Another is to look at the articles with a total count greater than 5,000.

The latter criteria was applied to the Climate March corpus. As a result, 113 articles (only 0.65% of the 17,128-article corpus) were selected for content analysis. They represent at least 56% of the attention received by the controversy on Facebook[4]. About 57% of those 113 articles have a Media Cloud origin and 43% come from Google searches.

The most popular story in the whole corpus is a Huffington Post piece on Leonardo DiCaprio's address at the United Nations Climate Summit on September 23rd (Visser, 2014). He had been designated UN Messenger of Peace in order to be a "new voice for climate advocacy", in the words of UN Secretary General Ban Ki-moon. In the previous Sunday, he had been one of celebrities at the People's Climate March.

That Huffington Post article was commented, liked, and shared by almost 180,000 people on Facebook. In fact, that was not the only article that decided to focus on the American actor. Other six news pieces in the corpus followed its footsteps in both traditional and independent media (Stampler, 2014). Some of them described his closeness to the cause of the indigenous people in Canada (Wood & Uechi, 2014; Indian Country Staff, 2014). Others mention the fact that he received the Clinton Global Citizen Award (Lewis, 2014), his commitment with ocean conservation efforts (Valentine, 2014), or simply the last steps in his career as an

[4] A spreadsheet with those 113 articles is available at http://bit.ly/climatemarchdata.

artist (Johnson, 2014). Together, they have a total count of almost 285,000 or 16% of the overall attention received by the top 113 articles on Facebook.

On the other hand, around 23% of the top articles were critical to the Climate March. They were able to grab about 21% of Facebook attention. The most successful piece in this category was The Gothamist's "People's Climate March Leaves Trail of Trash" with a total count of 87,647 (Evans, 2014). As the title suggests, the article used Twitter pictures to prove the alleged incoherence of environmentalists who littered the city with trash.

3.3 A look at media effects: the use of Twitter

Although Facebook (and, to some extent, bitly) data gives a very good idea of the most popular articles in the corpus, they are only numbers. Therefore, they cannot give any glimpse of the reaction to the stories – the so-called media effects.

"Likes" on Facebook, for instance, can be ambiguous. They can refer to the story itself that was shared on someone's timeline, but they can also refer to the gloss that accompanied the story. If that gloss was critical to the article, the "like" would mean all but support to the story's argument.

However, the Facebook Graph API does not grant unrestricted access to the texts that accompany posts. So we must look for this information elsewhere if we want to guess how positive (or negative) is the attention given to a story, an event, or a controversy.

A good alternative is to use Twitter data. When a user tweets a link, she usually adds some comments. Sometimes it is just the article's title and sometimes it's her own impressions about the link. That text can convey important information, especially if the researcher has enough tweets in order to make a representative assessment of the public reaction to the link.

There are two ways of extracting data from Twitter for free. Those two methods fulfill two different needs.

If the event or controversy is occurring now or has occurred in the past week, the easiest way to get the tweets is through the Twitter API[5]. There is a maximum number of requests per hour but the sample of tweets will be arguably bigger than through the Twitter webpage. The best approach is to create a bot that downloads the tweets in real time while the event is taking place or the controversy is unfolding.

However, if the event or controversy occurred sometime in the past, there is no alternative other than scraping the Twitter webpage (with a method analogous to the one previously described for Google search results). Figure 7 presents a code snippet for extracting the tweets. The idea is to use the Twitter search function to find tweets that mention a specific URL. This approach has some limitations though. The number of tweets that the search function returns is very uneven. For some URLs, the result set includes thousands of tweets (some of them mentioning shortened links for the original webpage). For others, the result set is empty. Tweaks in the URL (for instance, omitting the http:// part) can make a difference too.

```
from selenium import webdriver
import selenium.common.exceptions
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support import expected_conditions as EC

# opens Firefox
fp = webdriver.FirefoxProfile()
fp.set_preference("browser.private.browsing.autostart", True)
driver = webdriver.Firefox(firefox_profile=fp)
# access Google Search
driver.get("https://twitter.com/search?f=realtime&q=" + url)
```

Figure 7 – Code snippet for searching a URL on Twitter.
Complete code at https://github.com/alexgonca/ClimateMarch/blob/master/twitter.py

[5] Available at https://dev.twitter.com/.

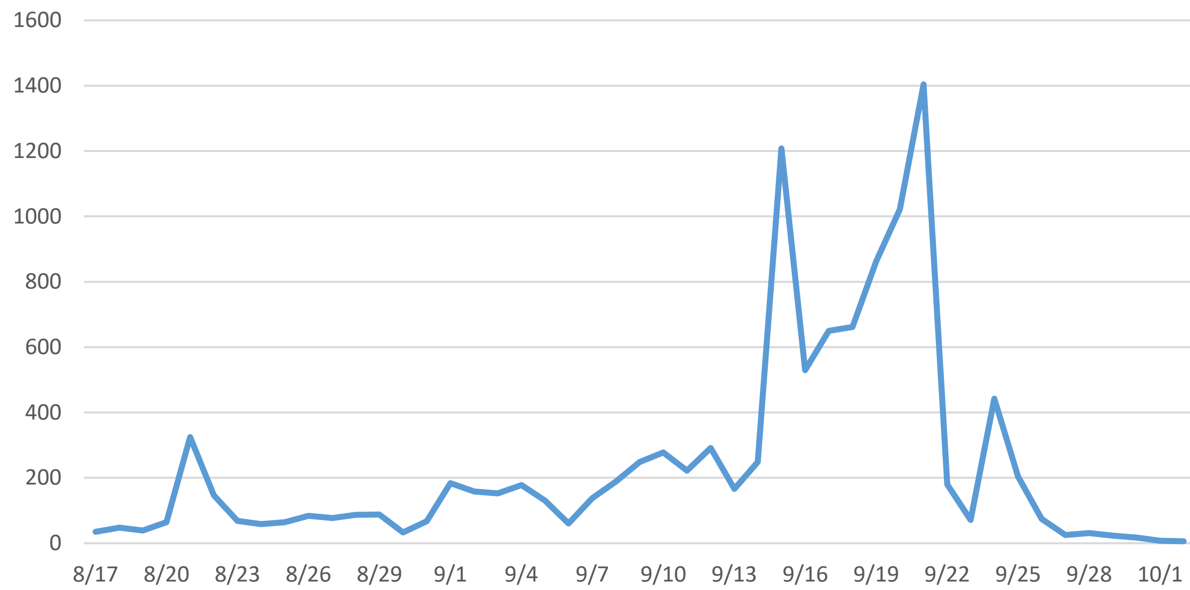

Figure 8 – Number of tweets that mention peoplesclimate.org.

Since the Twitter data for this research was collected in November and the Climate March occurred in September, the second method was used. More than 38,000 tweets that mentioned the 113 stories in the top-articles corpus were downloaded and used to infer the effects of the articles in their online readers.

An analysis of Twitter mentions to the official website of the People's Climate March (http://www.peoplesclimate.org), for instance, allows to reverse engineer the social media strategy for the march. Figure 8 shows that there was a push for spreading the word about the event exactly one month before the march, on August 21st.

The second push, more successful, was one week before the march, on September 15th, with the motto "It's time to change the trajectory of this planet." It achieved almost the same amount of tweets as the day of the event itself.

Figure 8 also reveals that the Website continued to be updated and to garner attention for a short period of time after the march. The last uptick occurred around September 24th and shows the impact of a short video that was offered as a wrap-up of the event. The last image of that video – Obama's remark at the United Nations that global leaders "cannot pretend" that they "do not hear" the voice of the streets – conveys the idea of "mission accomplished" for all the participants in the march.

Obviously, after obtaining the tweets, another problem arises: its interpretation. There have been many attempts to use natural language processing to extract the general meaning or the feelings conveyed by tweets (Golder & Macy, 2011). However, there are simpler approaches to make sense of this kind of data.

It is possible, for instance, to investigate which words are used to talk about a link. It is advisable to filter out the words that appear in the title of the article (since they would appear anyway in a list of most common words for that URL and therefore do not bring any new insight) and the so-called stop words – terms with less semantical relevance in a given context (e.g. most prepositions).

An example might illustrate this strategy. The second most popular piece in the corpus was Jon Stewart's "Burn Noticed" at *The Daily Show* on September 22nd (Stewart, 2014) with a Facebook total count of 122,812. After making fun of the Climate March coverage on the TV networks, he castigated the scientific illiteracy of the House Committee on Science, Space, and Technology.

An analysis of the word frequencies on 1,891 tweets about that URL (Figure 9) offers a glimpse of the most common reactions to that piece, namely a very negative impression

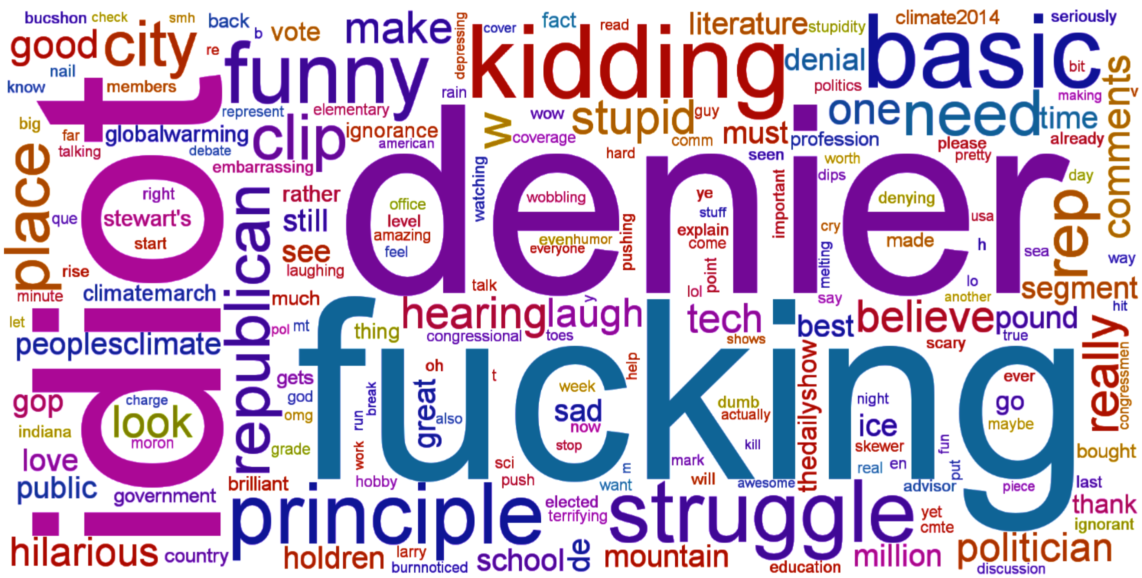

Figure 9 – Word cloud based in tweets that mention the “Burn Noticed”.

about the role of Republicans in climate discussions. Expressions like “f**king kidding me”, “denier”, “idiot”, and “republican” were among the most common terms used to spread the word about the video.

4 Critical Considerations

In this paper, there is a premise that social media attention equates impact in public opinion. Thus, stories that received a lot of likes, shares, and comments on Facebook would have a more significant influence in shaping people’s perceptions about an issue, event, or controversy. It is certainly possible to call this claim into question.

One can argue that social media data has always an embedded bias. Facebook, for instance, still skews toward women and young. Other criticism can focus on the definition of impact. An online article can arguably create a lot of buzz on social networks and be largely ignored by political or cultural elites that have a prominent role in promoting social change.

A counter-argument could be that the demographics of social networks are becoming more and more diverse. But, more importantly, the buzz generated by an article on a social network is not only a symptom of its popularity, but also its cause. When many users share, like, or comment a story, that piece will certainly appear on more people's timelines. As we say in internet lingo, the chances of "going viral" are higher. And, undoubtedly, a text that is read by a million people will have a more significant impact on public opinion than one that has an audience of a few dozens.

On the other hand, there is certainly a risk of using proprietary data. In the future, Facebook, bitly, or Twitter can restrict access to data that they have been offering for free. That would make it impossible to perform new studies using the same method.

Finally, the methods described here for accessing Twitter data does not always provide a representative picture of the overall Twitter activity (Morstatter, Pfeffer, Liu, & Carley, 2013). For a comprehensive picture, the researcher must use the paid service known as Twitter's Firehose.

5 Conclusion

The critical considerations of the previous section are certainly pertinent, but they should not be overstated. Traditional studies on print and online media have usually relied on small samples – due to limitation of human resources for performing content analysis – and many times they did not have any complementary information about readership preferences. Both limitations are greatly mitigated by social network data.

Some possible improvements in the methods describe here could include, for instance, the concomitant use of bitly and Facebook data. Bitly can offer an interesting insight about the relationship between the spread of an article on Twitter and on Facebook. Some URLs are arguably more successful in one network than in the other. Another improvement would be the use of methods of sentiment analysis to infer the media effects of specific links based on Twitter data.

6 Bibliography

Arizaleta, M. (2009). *Structured data extraction: separating content from noise on news.* Master's Thesis: Norwegian University of Science and Technology.

Benkler, Y. (2006). *The Wealth of Networks: How Social Production Transforms Markets.* Yale University Press.

Benkler, Y., Roberts, H., Faris, R., Solow-Niederman, A., & Etling, B. (2013). *Social mobilization and the networked public sphere: Mapping the SOPA-PIPA debate.* Cambridge, MA: Berkman Center Research Publication.

Boehrer, K., & Visser, N. (2014, September 21). *Why 400,000 People Marched For Climate Change Action.* Retrieved from Huffington Post: http://huff.to/1uas4gC

Cavender, G., & Mulcahy, A. (1998). Trial by Fire: Media Constructions of Corporate Deviance. *Justice Quarterly*, 697-717.

Evans, L. (2014, September 22). *People's Climate March Leaves Trail of Trash.* Retrieved from The Gothamist: http://bit.ly/1wiiEWX

Facebook. (2014, December 15). *Company Info*. Retrieved from Facebook: http://newsroom.fb.com/company-info/

Foderaro, L. W. (2014, September 21). *Taking a Call for Climate Change to the Streets*. Retrieved from The New York Times: http://nyti.ms/1rcpHKQ

Golder, S. A., & Macy, M. W. (2011). Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Diverse Cultures. *Science*, 1878-1881.

Graef, E., Stempeck, M., & Zuckerman, E. (2014). The battle for Trayvon Martin: Mapping a media controversy online and off-line. *First Monday*.

Indian Country Staff. (2014, September 23). *Star Power: Leonardo DiCaprio Climate Marches With Natives, and 9 Other Celeb Sightings*. Retrieved from Indian Country Today Media Network: http://bit.ly/1A4R10E

Johnson, Z. (2014, September 22). *Leonardo DiCaprio Raps "Scenario"—Watch Now!* Retrieved from E!: http://eonli.ne/12Q2GX1

Lewis, H. (2014, September 21). *Leonardo DiCaprio Urges World Leaders to Prioritize Environmental Issues*. Retrieved from The Hollywood Reporter: http://bit.ly/1u6ceCN

Lippmann, W. (1922). *Public Opinion*. Harcourt, Brace.

McKibben, B. (2014, May 21). *A Call to Arms: An Invitation to Demand Action on Climate Change*. Retrieved from Rolling Stone: http://rol.st/16014Lx

Morstatter, F., Pfeffer, J., Liu, H., & Carley, K. M. (2013). Is the Sample Good Enough? Comparing Data from Twitter's Streaming API with Twitter's Firehose. *ICWSM*.

Stampler, L. (2014, September 23). *Watch Leonardo DiCaprio Demand the UN Take Climate Change Seriously*. Retrieved from Time: http://ti.me/1u13YFd

Stewart, J. (2014, September 22). *Burn Noticed*. Retrieved from The Daily Show: http://on.cc.com/1zaQPyF

Valentine, K. (2014, October 17). *Leonardo DiCaprio Donates $2 Million To Ocean Conservation Efforts*. Retrieved from Think Progress: http://bit.ly/1G47y98

Visser, N. (2014, September 23). *Leonardo DiCaprio Addresses UN Climate Summit: 'You Can Make History... Or Be Vilified By It'*. Retrieved from The Huffington Post: http://huff.to/1yXp9Mt

Weaver, D. A., & Bimber, B. (2008). Finding News Stories: a Comparison of Searches Using LexisNexis and Google News. *Journalism and Mass Communication Quarterly*, 515-530.

Wood, L. S., & Uechi, J. (2014, September 21). *Leonardo DiCaprio rallies with Canadian Indigenous leaders at People's Climate March in New York*. Retrieved from Vancouver Observer: http://bit.ly/1G1E6yA