# Gender differences in the climate change communication on Twitter

## Abstract

*Purpose:* We present a study about gender differences in the climate change communication on Twitter and in the use of affordances on Twitter.

*Design/methodology/approach:* Our dataset consists of about 250,000 tweets and retweets for which the authors' gender was identified. While content of tweets and hashtags used were analyzed for common topics and specific contexts, the usernames that were proportionately more frequently mentioned by either male or female tweeters were coded 1) according to the usernames' stance in the climate change debate into convinced (that climate change is caused by humans), sceptics, neutrals and unclear groups, and 2) according to the type or role of the user account (e.g. campaign, organization, private person).

*Findings:* The results indicate that overall male and female tweeters use very similar language in their tweets, but clear differences were observed in the use of hashtags and usernames, with female tweeters mentioning significantly more campaigns and organizations with a convinced attitude towards anthropogenic impact on climate change, while male tweeters mention significantly more private persons and usernames with a sceptical stance. The differences were even greater when retweets and duplicate tweets by the same author were removed from the data, indicating how retweeting can significantly influence the results.

*Practical implications:* On a theoretical level our results increase our understanding for how women and men view and engage with climate change. This has practical implications for organizations interested in developing communication strategies for reaching and engaging female and male audiences on Twitter. While female tweeters can be targeted via local campaigns and news media, male tweeters seem to follow more political and scientific information. The results from the present research also showed that more research about the meaning of retweeting is needed, as we have shown how retweets can have a significant impact on the results.

*Originality/value:* Our findings contribute towards increased understanding of both gender differences in the climate change debate and in social media use in general. Beyond that this research showed how retweeting may have a significant impact on research where tweets are used as a data source.

## Introduction

Social media in general and Twitter in particularly provide new venues and research opportunities for computational social sciences as researchers can mine tweets for public opinions and attitudes, social activities and networks, and trends in conversations and information sharing patterns. Social media may even have an important role in creating public opinions about different issues as it has been discovered to be a very efficient medium to spread ideas and news (Jansen et al., 2009; Cha et al., 2012) and to engage people in environmental activism (Cheong & Lee, 2010). Online campaigns have been found to have offline impacts, as was the case with for instance the Twitter campaign connected to Earth Hour in 2009 during which a connection between activity related to the campaign on Twitter and reduced energy consumption was detected (Cheong & Lee, 2010). On the other hand, data mined from social media may provide insights into public perceptions and opinions about various topics, such as climate change (Auer et al., 2014; Kirilenko et al., 2014). Environmental issues, climate change and global warming are hotly

debated topics on many forums, social media none the least. In this paper, we focus on Tweets about climate change and in particular gender differences (and similarities) in the climate change debate between male and female tweeters. Our results have implications for organizations interested in targeting male and female stakeholders and customers via Twitter, as well as for other researchers using tweets and retweets as their data source.

**Literature review**

Climate change is something that many of us are in some way concerned about, but in general, women have been found to be more concerned about climate change than men and possessing greater scientific knowledge about the issue, in the context of American public (McCright, 2010). This is related to a more general tendency of women being more concerned about local environmental issues than men are. In addition, women tend to be more willing to take voluntary actions to mitigate climate change, such as trying to reduce greenhouse emissions, while men "*are more willing to support government policies to impose public sacrifices in order to reduce greenhouse gas emissions*" (O'Connor et al., 1999). These results indicate that women are more concerned about local issues and campaigns related to climate change and more willing to take private action, while men feel more comfortable in the political world. However, more recently Whitmarsh (2011) discovered that although demographic factors, such as age and gender, played a role in the discovered gender differences in perceptions of climate sceptisism, the strongest correlations were discovered in political orientation and environmental values. Whitmarsh (2011) concludes that "*those with right-of-centre political views and low pro-environmental values tend to be most skeptical about the reality and severity of climate change. In other words, beliefs about climate change are fundamentally linked to existing values and worldviews*". These findings are also supported by Davidson and Haan (2012) who discovered that the gender differences were predominately associated with socioeconomic factors and political views, at least in the province of Alberta, Canada. They conclude their findings by writing that "*gender distinctions appear to be related to the lower tendency for women to ascribe to a conservative political ideology relative to men*". These studies do not reject the earlier findings that women tend to be more aware and concerned about anthropogenic impact on climate change, but they give us one explanation for this difference between men and women. But do these gender differences about the anthropogenic impact on climate change appear online and in social media use as well? Can we mine social media to study gender differences in the attitudes about the reasons for climate change?

In contrast to more traditional research methods such as surveys, data from the web in general and social media in particular can be automatically and non-intrusively mined with relative ease and analyzed statistically to discover patterns in user behavior, attitudes and emotions (Barbier & Liu, 2011). While content and hyperlinks on webpages are typically collected with a web crawler (e.g., Wilkinson, Thelwall & Li, 2003; Thelwall, 2009), many social media sites provide an Application Programming Interface (API) through which researchers can mine the user generated data on that specific site. These APIs are, however, often restricted so that only a fraction of the data can be retrieved. Twitter for instance restricts the data access to their free API to a maximum of 1% of the total volume of tweets (Bruns & Stieglitz, 2015). With about 350,000 tweets a minute (November 11, 2014, http://www.internetlivestats.com/twitter-statistics/) only about 3,500 tweets with any given search parameters could be collected through the API. For popular topics this means that not all of the tweets could be mined as the maximum limit would be reached, but for less popular topics, or with stricter search parameters, all the tweets could in fact be collected. Aiming at highest possible coverage would have a positive impact on representativeness of the data, as it has been shown that the restricted feed of

tweets accessible through Twitter's free API may not be representative of the total number of tweets, depending on the search parameters and the type of analysis (Morstatter et al., 2013: see also Bruns & Stieglitz, 2015, for a discussion about the representativeness of Twitter data). It is also important to acknowledge that not everyone uses social media, yet it has been suggested and proven that social media data can, at least in some cases, function as "real-world sensors" providing insights into people's opinions, feelings and attitudes (e.g., Bollen, Pepe & Mao, 2011; Takahashi, Abe & Igata, 2011).

Although so far most of the research about climate change communication have focused on the role of traditional news media (e.g., Schmidt et al., 2013; Grundmann & Scott, 2014; Schäfer & Schlichting, 2014), an increasing number of research in both natural and social sciences are using online data to get insights about people's opinions and attitudes about climate change. In the context of climate change, it has been suggested that especially in developing countries where the mobile Internet is growing rapidly, social media sites, such as Twitter, may even have a better potential to reach audiences than traditional news media (Bosch, 2012). Recently the micro-blogging site Twitter has been used as the data source in some studies about climate change communication. Kirilenko et al. (2014) investigated tweeting about climate change and mapped users, topics and news sources used by the tweeters. They discovered that "*the flow of information [about climate change] is highly centralized, with few media outlets, celebrities, and prominent bloggers leading the debate*". Pearce et al. (2014) analyzed tweeting about the release of the report by Working Group 1 of the Intergovernmental Panel on Climate Change (IPCC) and discovered that birds of the feather flock together, as tweeters were most likely to converse with other users that shared their stance in the climate change debate. Both of these studies have shown that climate change is a hotly debated topic on Twitter, making it an interesting forum to investigate people's opinion and attitudes in a non-intrusive way. Gender differences in the use of social media have not gained wide attention so far, partly because the social media phenomenon is relatively new. In studies on social media use in general, Whiting and Williams (2013) studied the reasons for using social media from the point of view of uses and gratifications theory and reported that most people use social media for social interaction, information seeking and passing time (ibid. p. 368), but do not report gender differences. Similarly, Kietzmann et al. (2011) compare the use of different types of social media (Facebook, YouTube, LinkedIn and 4Square), but without taking into account gender differences. They propose a model for comparing different social media sites according to whether these are used for identity building, sharing of information or sustaining relationships, for example (2011). This is consistent with the results by Naaman et al. (2010) that Twitter is used for presenting 'self' in order to maintain relationships, and for sharing information. Similarly, Chen (2013) reports that women bloggers tend to use social media for information, engagement and recreation.

In their study comparing female and male mentions in tweets and in news stories, Armstrong and Gao (2010) found that both genders were portrayed in a similar way in news and in Twitter. They conclude that news media are not using Twitter to gather more female audience. Lasorsa (2012), on the other hand, found little differences in how male and female journalists used and appeared on Twitter, with the exception of female journalists being more transparent about their personal lives compared to men. Some of the earlier studies have focused on linguistic gender differences in online context. Bamman et al. (2012) contrasted female and male linguistic styles in Twitter to the social networks of these tweeters. Male linguistic markers include numbers, quantifiers and technology words while female markers tend to consist of pronouns, emotion terms and family terms (ibid. 15). The results of Bamman et al. (2012) show that social networks of women tend to consist mostly of other women, and those of men mostly out of other men. Cunha et al. (2014) focused on gender differences in the use of hashtags on Twitter and found that while hashtags used by women show a personal involvement (e.g. I vote for…), the hashtags used by men show a persuasive strategy, for instance by expressing a command (e.g. Vote for…). Similarly it has been discovered that when choosing political hashtags men opt for

more overt ways in persuasion in their hashtags, while women opt for more indirect ways in their hashtags (Cunha et al., 2012).

These topics, climate change debate, gender differences, and social media, have not to the best of our knowledge been combined in a research before. With this research we will fill this gap. Our aim with this research is to study gender differences in the tweeting behavior in the climate change debate. The goals of this research can be summarized into the following research questions:

1. What kind of differences (in the content of tweets, use of hashtags, mentioning of other usernames) are there between male and female tweets about the climate change?

2. How do retweets and duplicate tweets from the same author affect the results? Do retweets emphasize or skew the results?

Based on earlier results from research about gender differences in the attitudes towards climate change our hypothesis is that women are more concerned about the climate change while men show more sceptical attitudes. This hypothesis will be tested using statistical methods. By answering the above mentioned research questions we will contribute towards increased understanding of both gender differences in the climate change debate and in social media use in general. Beyond that this research will investigate the possible impact retweeting may have on research where tweets are used as a data source.


**Data and methods**

A total of 556,517 tweets containing the words "climate change" were collected between October 26, 2013, and January 10, 2014, via Twitter's API with Webometric Analyst (Thelwall, 2009). We used Mozdeh (http://mozdeh.wlv.ac.uk/) to identify the gender of the authors of these tweets. The identification of the gender of the tweeters was done by comparing their first names (as written in their Twitter profiles) to common US first names by gender. A total of 94,579 (17.0%) tweets were sent by female tweeters and a total of 145,275 (26.1%) tweets were sent by male tweeters. The gender of the remaining tweets (56.9%) could not be determined because the names were either not identified among the common US first names or the names were not provided in the profile information. As we are limited to the information that the Twitter users have entered in their profiles on Twitter, we cannot rule out any non-gender related differences between the groups, such as education level, age, political opinions, or geographical location. Only the tweets for which the tweeters gender could be identified were used for analysis.

In order to analyze the possible impact of retweeting and duplicate tweets from the same author we created another data set by removing these from the data. This was done by removing any tweets that started with RT or that otherwise was clearly forwarded or quoted. Duplicate tweets were identified by searching for identical tweets from the same author. This second data set contained a total of 43,323 (7.8%) tweets that were sent by female tweeters and a total of 79,736 (14.3%) tweets by male tweeters. We analyzed the content of the tweets using different approaches and methods in order to triangulate the semantic meaning of the content and to investigate possible differences in the positions the two groups had in the climate change debate. In the content of the tweets we focused our analysis on the differences in the hashtags used, usernames mentioned, in the textual content of the tweets and the sentiment of the tweets. A Spearman rank correlation was calculated to measure the similarities between the noun phrases, hashtags and usernames used by both groups and for both data sets.

In both data sets and in both groups about one third of the tweets contained a hashtag and roughly one in every ten tweets contained a unique hashtag (Table 1). Unique usernames were mentioned in about quarter of the tweets in both data sets and both groups. However, while about 9 out of 10 tweets contained a username in the first data set, when the retweets were excluded about half of the tweets contained a username. This is most likely due to the structure of retweets that usually include the username of the person who published the tweet before, and hence the ratio of usernames mentioned in retweets is higher than for original tweets.

**Table 1. Number of tweets, hashtags in total, unique hashtags, usernames in total, and unique usernames mentioned in the tweets by female and male tweeters when including and when excluding retweets and duplicate tweets (means are in parentheses).**

|  |  | Tweets (mean) | Total number of tweets | Hashtags | Unique hashtags | Usernames | Unique usernames |
|---|---|---|---|---|---|---|---|
| **All tweets** | **Female** | 94,579 | 34,925 (0,369) | 6,157 (0,065) | 88,501 (0,936) | 21,254 (0,225) |
|  | **Male** | 145,275 | 54,234 (0,373) | 8,024 (0,055) | 127,792 (0,880) | 28,244 (0,194) |
| **Excluding retweets and duplicates** | **Female** | 43,323 | 13,994 (0,323) | 4,347 (0,100) | 23,785 (0,549) | 9,790 (0,226) |
|  | **Male** | 79,736 | 27,025 (0,339) | 6,349 (0,080) | 44,355 (0,556) | 16,838 (0,211) |

For the hashtags and the usernames mentioned in the tweets we calculated the differences in proportions from normal distribution between the two groups of tweeters. This gave us a list of hashtags and usernames that were proportionately more frequently used by either male tweeters in comparison to female tweeters and vice versa. These are not necessarily the most frequently used hashtags or usernames, but they are significantly more used by one group of tweeters in comparison to the other. For the analysis of the usernames and hashtags we chose to analyze only those that scored high on the z value (higher than 4.5 or lower than -4.5), i.e. indicating larger proportionate difference in the use between the genders. When calculating large numbers of z values there is a chance to gain a high value by chance which means that the calculation of z values may be unreliable, however, the z values are still useful to indicate trends in the proportional differences between the two groups. Because the data was also analyzed qualitatively (in which case possible false results and anomalies in the results would have been detected) we decided not to apply Bonferroni correction on the data to counter for the possibility of false results.

The usernames mentioned were coded 1) according to the users' stance in the climate change debate (convinced that the climate change is caused by humans, neutral, sceptic, unclear) and 2) according to type of account. The categories for type of account were: campaign (connected to a campaign or online movement, usually related to environmental issues or climate change), climate scientist (a person who works with climate science or related fields of research), company (for-profit company), news sharing

(news organizations or news feeds, in some cases automated distribution of news from another online source), organization (non-profit organizations, usually related to environmental issues), private person (a private person not affiliated with any organization), technical (usually generated by the system, e.g. tweeting from YouTube), other (accounts that could not fit into the existing categories), and unclear (accounts for which the type could not be determined, usually because of lack of information). A username could only belong to one category. The coding was done based on the information provided on the respective Twitter profile page and by visiting other external webpages, when such were linked to from the Twitter profile. When determining the users' stance in the climate change debate we also looked at the content of their tweets. In order to ensure reproducibility (Krippendorff, 2004, p. 217), the coding was done independently by two researchers and Cohen's Kappa was used to measure inter-coder agreement (Krippendorff, 2004, p. 246). Statistical methods were used to test our hypothesis and to confirm detected differences between the groups.

The hashtags were also classified by two researchers according to what the hashtags represented or to what kind of general topic they were related to. The codes from both researchers were compared and any discrepancies were discussed. The hashtags were categorized to following categories: related to politics or policies, campaigns and movements, media (as in news organizations and media sites like YouTube), environment, and other for the remaining hashtags. Originally we also prepared to code hashtags related to science, geographical areas, and new technologies, but no hashtags related to these areas could be found. Many of the hashtags represent a very general level of metadata describing the content or context of the tweet, and because of that we feel that an appropriate approach to present the data is to present it descriptively rather than statistically. Hence inter-coder agreement was not measured for the coding of hashtags.

We also analyzed the semantic content and the sentiment of the tweets by both groups separately (excluding hashtags, usernames and URLs). In order to get an overview of the content of the tweets we used VOSviewer (Van Eck & Waltman, 2010) to extract the noun phrases from the tweets and analyzed the differences between the frequent noun phrases used by male and by female tweeters. Sentiment analysis was conducted using SentiStrength (Thelwall, Buckley & Paltoglou, 2011; 2012); a sentiment analysis tool especially designed for short texts such as tweets. SentiStrength gives each analyzed text a negative and a separate positive sentiment strength which can vary between -5 and -1, and between +1 and +5 respectively. From these a total sentiment score for the whole data set can be calculated.

## Results

*Content of tweets*

To focus on the most frequently published content we used VOSviewer (Van Eck & Waltman, 2010) to extract the noun phrases that appeared at least 20 times from the tweets sent or forwarded by male or female tweeters separately. This resulted in about 3400 unique noun phrases in both groups and a total of 793,456 noun phrases, of which 306,700 were in the tweets by female tweeters and 486,756 in the tweets by male tweeters. A high Spearman rank correlation (0.738) showed that the noun phrases used by the both groups and the relative frequencies of them were very similar.

In order to analyze the most popular semantic content of the tweets we focused on the about 500 most frequently used noun phrases in both groups. For the noun phrases used by female tweeters a total of 505 noun phrases, all with a frequency of or over 113 were selected, and for the noun phrases used by male tweeters a total of 502 noun phrases, all with a frequency of or over 172 were selected. Noun phrases shared by both groups and unique to both groups were analyzed. A total of 425 noun phrases were shared by both male and female tweeters, 79 noun phrases were unique to male tweeters and 76

noun phrases were unique to female tweeters. Among the most frequent noun phrases shared by both groups were noun phrases related to climate change (e.g., *climate change, climate, change, global warming*), and to the science of climate change (e.g., *scientist, science, study, report*). The words "climate change" of course appears in every tweet collected because they were the search terms when mining the tweets. Some of the noun phrases were on a more general level (e.g., *world, person, country, man)* or related to time (e.g., *time, year, today,*). Noun phrases related to two specific cases were also visible among the most frequently used noun phrases. The first case contained noun phrases from a frequently retweeted tweet with a photo of a pool in Mumbai, India, that looked like a flooded Manhattan:

> *rt @[...] a pool in mumbai that look like manhattan flooded to raise awareness of climate change*

The second case contained noun phrases from several different tweets that were all related to the typhoon Haiyan that hit the Philippines in November 2013. A typical tweet about this event looked like:

> *rt @[...] typhoon haiyan what really alarm filipino is the rich world ignoring climate change [...]*

In the set of about 500 most frequently used noun phrases the ten most frequently used noun phrases unique to male tweeters were: *bbc*, *conservative*, *source*, *new report finds climate change*, *global*, *smart person*, *marijuana*, *revolution*, *telling chart*, and *trend*. The one that seems a bit surprising in the context of climate change is *marijuana*. This is due to a frequently retweeted tweet:

> *rt @[...] these debates are settled for smart people marijuana evolution gay marriage climate change*

The ten most frequently used noun phrases by female tweeters were: *bible prove climate change*, *extreme event*, *gop reps*, *read*, *marine carbon cycle*, *oceans s10*, *second international symposium*, *advance climate change adaptation*, *alarming climate change effect*, and *Norwegian army*. The most frequent one of these was due to frequent tweeting of slight variations of the following tweet by just three female tweeters:

> *god bible prove climate change a hoax say gop rep #climatedesk #jimmydore #tyt #youngturk*

This tweet, as we will later see, also had an impact on the analyzed hashtags.

The one that does not seem to fit into the climate change debate is the phrase *Norwegian army*. This was due to frequent retweeting of:

> *rt @[...] norwegian army goes vegetarian to war against climate change*

Overall the noun phrases used by both groups were very similar and no clear patterns could be detected in the tweeting between genders.

These examples above also demonstrate how the results were clearly influenced by popular retweets and the activities of a few tweeters. When excluding the retweets and duplicate tweets from the same author from the analysis the Spearman correlation showed still a high similarity (0.677) in the used noun phrases. When removing the retweets from the data the most frequently used noun phrases that were unique to female tweeters changed significantly; *fund*, *social effect*, *young filmmaker*, *video competition*, *new news*, *climate change daily*, *plea*, *climate change video*, *nuclear wire*, and *essay*. Some of these hashtags were due to slightly different versions of the tweet:

> *@[...] new competition invite filmmaker to make a climate change documentary #action4climate*

The unique noun phrases for male tweeters included *source*, *project*, *level*, *tax*, *party*, *ice age*, *weather event*, *crisis*, *lie*, and *model*. This suggests that while female tweeters were concerned about the social

effects of climate change and sharing information about specific videos, male tweeters focused more on politics, economics and science issues. This difference in tweeting behavior across the genders has implications for organizations targeting these two groups via Twitter. Female and male audiences may be reached via different types of tweets.

The sentiment analysis showed that overall the tweets by both male and female tweeters and both including and excluding retweets and duplicate tweets, were slightly negative (Table 2). In each case the positive sentiment strength was about 1.3, while the negative sentiment strength was about -1.7. In general the sentiment scores for female tweeters are slightly less negative, and excluding the retweets slightly increased the sentiment score for female tweeters, but these differences are too small to be considered as indicative of any trend or any real differences between the two groups and two data sets.

**Table 2. Sentiment strength for the tweets by male and female tweeters, both including and excluding retweets.**

|  | Positive sentiment strength | Negative sentiment strength | Total sentiment score |
|---|---|---|---|
| Female tweeters, excl. retweets | 1,276 | -1,654 | -0,378 |
| Male tweeter, excl. retweets | 1,264 | -1,686 | -0,422 |
| Female tweeters, incl. retweets | 1,267 | -1,687 | -0,420 |
| Male tweeter, incl. Retweets | 1,257 | -1,696 | -0,439 |

*Hashtags in tweets*

The hashtags used in the tweets by both groups were extracted and analyzed separately. The Spearman rank correlation between the hashtags used by both groups was a very low 0.011. Among the most frequently used hashtags by both groups were hashtags related to climate and climate change (e.g., *#climate*, *#climatechange*, *#globalwarming*), some were related to campaigns calling for action on climate change (e.g., *#climateaction*, *#thisisreal*, *#action4climate*, *#actonclimate*, *#O29*[1]), while others were related to politics (*#auspol*[2], *#tcot*[3], *#p2*[4], *#teaparty*[5], *#cdnpoli*[6], *#thisislabor*[7], *#policy*). Some of the frequently used hashtags were related to specific events (e.g., *#COP19*[8], *#AGU13*[9]). Many of the most frequently used hashtags are clearly connected to the climate change debate and reflect the interests of the tweeters participating in that debate.

We also analyzed the hashtags that had the highest proportionate differences in the frequencies between the two genders. We chose to focus this analysis on the 15 hashtags with the highest proportionate differences (Table 3). Among the hashtags by male tweeters (positive z score in Table 2) appeared some hashtags related to politics (*#policy*, *#tcot*, *#gop*), while only a few may be connected to climate change or to environmental issues on a general level (e.g. *#agw*, *#eco*, *#sustainable*). Many of the hashtags could be traced to a single very active male tweeter or to a very popular tweet that was retweeted frequently. The hashtag *#responsibility* for instance, was frequently connected with *#policy* in the tweets of a single tweeter, as were the hashtags *#occupyinfo*, *#owsinfo*, *#revolution* and *#anonymous* by another very active tweeter. Hashtags *#sustainable* and *#youtube* appear in identical tweets by two very active tweeters. The hashtag *#conradnew* was also traced back to a single user that used the hashtag to label his own tweets.

Among the hashtags that were proportionately more frequently used by female tweeters (negative z score in Table 3) some were related to campaigns and online movements connected to climate change (*#action4climate*, *#thisisreal*, *#climatenamechange*). Five of the top hashtags (*#youngturk*, *#tyt*, *#climatedesk*, *#jimmydore*, *#michaelshure*) were connected to The Young Turks (an online commentary program on YouTube) and appeared on the list because two very active female tweeters sent almost the same tweet frequently. Both *#extinction* and *#prevent* were also traced back to a single very active tweeter sending the same tweet multiple times. Some of the hashtags (*#freethearctic30*, *#freecolin*) were connected to a campaign to release the 30 Greenpeace activists arrested by Russian authorities in September 18, 2013. One of the campaign hashtags, *#freecolin*, was part of a campaign to demand the release of Australian activist Colin Russell who was the last of the activists to be released on November 29, 2013.

**Table 3. Hashtags used proportionately more by male and by female tweeters compared to the other (M. freq. = number of tweets sent by male tweeters in which the hashtag appeared, M. prop. = proportion of male tweets in which the hashtag appeared, F. freq. = number of tweets sent by female tweeters in which the hashtag appeared, F. prop. = proportion of female tweets in which the hashtag appeared, z = test statistic for comparing the two proportions).**

| Term | M. freq. | M. prop. | F. freq. | F. prop. | z |
|---|---|---|---|---|---|
| #policy | 426 | 0,002932 | 68 | 0,000719 | 11,7 |
| #sustainable | 228 | 0,001569 | 6 | 0,000063 | 11,5 |
| #youtube | 216 | 0,001487 | 4 | 0,000042 | 11,4 |
| #bbcqt | 847 | 0,00583 | 251 | 0,002654 | 11,3 |
| #conradnew | 151 | 0,001039 | 0 | 0 | 9,9 |
| #tcot | 1116 | 0,007682 | 449 | 0,004747 | 8,7 |
| #occupyinfo | 98 | 0,000675 | 0 | 0 | 8 |
| #owsinfo | 92 | 0,000633 | 0 | 0 | 7,7 |
| #agw | 145 | 0,000998 | 17 | 0,00018 | 7,5 |
| #responsibility | 105 | 0,000723 | 7 | 0,000074 | 7,2 |
| #transition | 175 | 0,001205 | 31 | 0,000328 | 7,2 |
| #eco | 194 | 0,001335 | 41 | 0,000434 | 6,9 |
| #revolution | 74 | 0,000509 | 0 | 0 | 6,9 |
| #forest | 168 | 0,001156 | 32 | 0,000338 | 6,8 |
| #gop | 195 | 0,001342 | 43 | 0,000455 | 6,7 |
| #theel | 0 | 0 | 38 | -0,000402 | -7,6 |
| #santaclaus | 0 | 0 | 39 | -0,000412 | -7,7 |
| #thisisreal | 343 | 0,002361 | 391 | -0,004134 | -7,7 |
| #freethearctic30 | 57 | 0,000392 | 128 | -0,001353 | -8,3 |
| #climatenamechange | 209 | 0,001439 | 296 | -0,00313 | -8,8 |
| #o29 | 332 | 0,002285 | 411 | -0,004346 | -8,9 |

| #freecolin | 77 | 0,00053 | 168 | -0,001776 | -9,3 |
|---|---|---|---|---|---|
| #michael shure | 0 | 0 | 75 | -0,000793 | -10,7 |
| #extinction | 14 | 0,000096 | 117 | -0,001237 | -11,7 |
| #jimmydore | 0 | 0 | 105 | -0,00111 | -12,7 |
| #prevent | 0 | 0 | 110 | -0,001163 | -13 |
| #action4climate | 29 | 0,0002 | 173 | -0,001829 | -13,4 |
| #tyt | 1 | 0,000007 | 180 | -0,001903 | -16,5 |
| #climatedesk | 0 | 0 | 180 | -0,001903 | -16,6 |
| #youngturk | 0 | 0 | 218 | -0,002305 | -18,3 |

Although the results presented above clearly demonstrate how the results are influenced partly from frequent retweeting of some very popular tweets and partly from the actions of a few very active tweeters we can still see some differences between the two groups. These differences were even clearer when the retweets were removed. When excluding the retweets and duplicate tweets from the analysis the Spearman rank correlation between the hashtags used by female tweeters and by the male tweeters decreased to -0.295, indicating an even clearer difference in how both groups used hashtags. While female tweeters mentioned proportionately more frequently hashtags that were connected to different campaigns and online movements related to climate change (e.g., #action4climate, #O29, #climatenamechange), male tweeters used proportionately more frequently hashtags related to politics (e.g., #policy, #tcot, #auspol, #gop). Male tweeters also used many hashtags that were on a more general level related to climate and the environment and that were more descriptive (e.g., #climate, #forest, #agw, #eco), while female tweeters used more specific hashtags related for instance to a specific event (e.g., #cop19, #glfcop19), campaign or person. This systematic difference in the tweets shows that female tweeters seem to be more involved in local events related to climate change and male tweeters to political and scientific debates.

*Usernames mentioned in the tweets*

We calculated the Spearman rank correlation between the usernames mentioned in both groups and it showed no correlation (-0.004). When analyzing the proportional differences in the mentioned usernames by both groups some additional differences became visible. A total of 37 usernames that were mentioned proportionately more frequently by male tweeters and 77 usernames that were mentioned proportionately more frequently by female tweeters met the threshold described in the Methods section and were chosen for a closer analysis. These usernames were coded by the authors according to a) type of user account or role of the user (i.e., private person, news, organizations, campaigns, etc, see Table 4), and b) the mentioned usernames' stance in the climate change debate (i.e., convinced of anthropogenic impact on climate change, neutral, sceptic, or unclear, see Figure 1). Any discrepancies between the results of the coding were discussed and rectified for the final tables.

The results indicate that both male and female tweeters frequently mention private persons in their tweets (Table 4), although men do so much more frequently (male: 51.4%, female: 29.9%). Usernames related to news or news sharing were also frequently mentioned by both groups (male: 18.9%, female: 15.6%). While female tweeters mention organizations frequently in their tweets, men do not so much (male: 2.7%, female: 27.3%). Female tweeters also mentioned usernames related to different campaigns

in their tweets, while none of the usernames mentioned more frequently by male tweeters were related to campaigns or online movements (male: 0%, female: 14.3%). Some male tweeters on the other hand mentioned usernames of climate scientists, while none of the usernames mentioned more frequently by female tweeters belonged to climate scientists (male: 8.1%, female: 0%). The inter-coder agreement between the two researchers was calculated using Cohen's kappa, which gave an agreement of 0.765 (81.6% agreement), which constitutes as good agreement.

**Table 4. Usernames mentioned more frequently by male and more frequently by female tweeters according to type or role of the user account.**

| Type of user account | Mentioned by male tweeters | Mentioned by female tweeters |
|---|---|---|
| campaign | 0 (0.0%) | 11 (14.3%) |
| climate scientist | 3 (8.1%) | 0 (0.0%) |
| company | 1 (2.7%) | 1 (1.3%) |
| news sharing | 7 (18.9%) | 12 (15.6%) |
| organization | 1 (2.7%) | 21 (27.3%) |
| other | 2 (5.4%) | 4 (5.2%) |
| private person | 19 (51.4%) | 23 (29.9%) |
| technical | 1 (2.7%) | 2 (2.6%) |
| unclear | 3 (8.1%) | 3 (3.9%) |
| Total | 37 (100.0%) | 77 (100.0%) |

The results from the coding of the usernames based on the usernames' stance in the climate change debate showed that female tweeters mentioned significantly more convinced users in their tweets and retweets, and that the male tweeters mentioned significantly more sceptics, while among the usernames mentioned more frequently by female tweeters there were no sceptics at all (Figure 1). These differences were also tested for statistical significance by running a chi-square test which confirmed that the proportion of female tweeters mentioning convinced tweeters and the proportion of male tweeters mentioning sceptic tweeters were higher than expected from random tweeting (chi-square=31.28, p=0.000). We also ran a Z test for the differences in proportions of convinced and sceptic usernames mentioned in both groups. The null hypothesis is that if there is no difference in the proportions from two populations the Z-score would result in 0. The Z-score for the difference in proportion of usernames coded as convinced was -3.67 (p=0.00024) and the Z-score for the difference in proportion of usernames coded as sceptic was 5.03 (p=0.000). Both results significant at $p < 0.05$. With that we can reject the null hypothesis and state that there are statistically significant differences in how male and female tweeters mention usernames that have been coded as sceptic and as convinced of anthropogenic impact on climate change. The inter-coder agreement was 0.522 on Cohen's kappa (68.4% agreement), which constitutes as moderate agreement.

**Figure 1 Usernames mentioned more frequently by male and more frequently by female tweeters according to the username's stance in the climate change debate. Retweets are included in the data.**
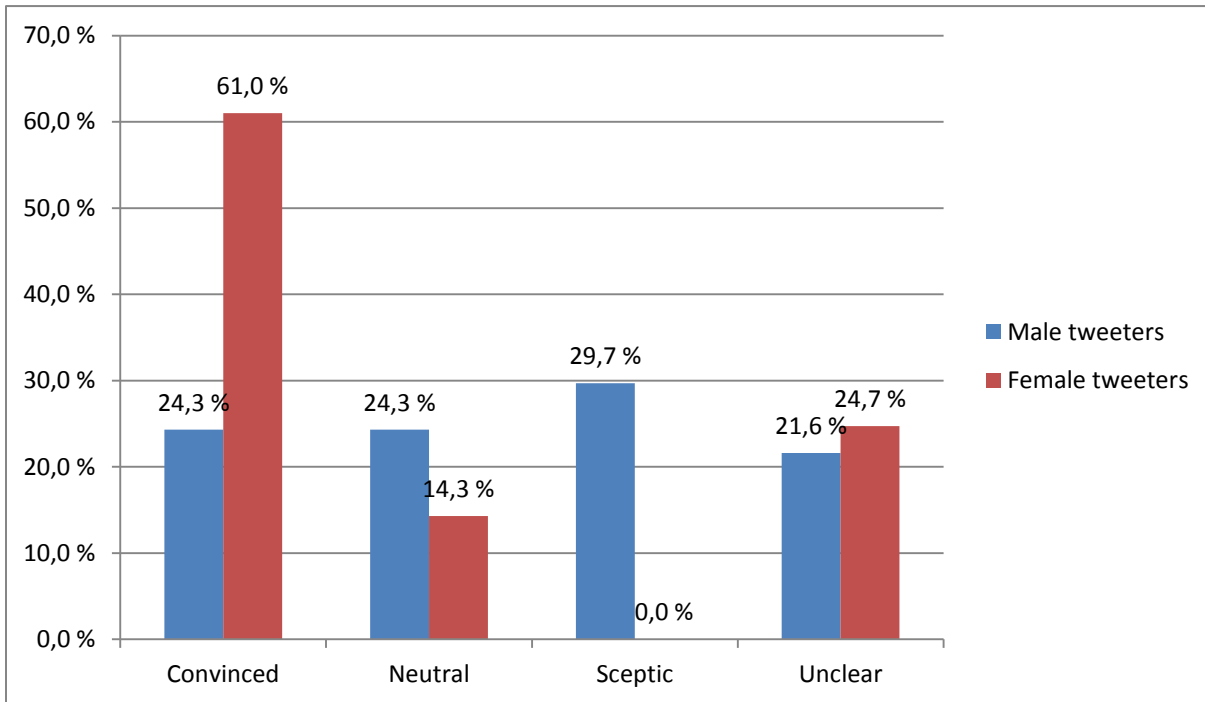


The results do however not mean that female tweeters would not mention sceptic usernames at all in their tweets. The results show that among those usernames that women mentioned proportionately more frequently than men there were no sceptics and that among the usernames mentioned proportionately more frequently by male tweeters there were significantly more sceptics. If we look at the three usernames that were coded as sceptics and that were mentioned significantly more frequently by male tweeters compared to female tweeters we see that even female tweeters mentioned them, but to a significantly lesser degree (Table 5). From Table 5 we can read that *user1* was mentioned in 1,297 tweets sent by male tweeters (or mentioned in 0.008928 of all male tweets) and by 405 female tweeters (or mentioned in 0.004282 of all female tweets). This gives a proportional difference of 0.004646 between the proportions and a z value of 13.2, indicating that *user1* was mentioned significantly more by male tweeters than by female tweeters.

**Table 5. Example values of three usernames that were coded as sceptics and that were proportionately more frequently mentioned by male tweeters (descriptions of columns the same as in Table 2).**

|  | M. freq. | M. prop. | F. freq. | F. prop. | z |
|---|---|---|---|---|---|
| @[user1] | 1297 | 0.008928 | 405 | 0.004282 | 13.2 |
| @[user2] | 179 | 0.001232 | 23 | 0.000243 | 8.2 |
| @[user3] | 126 | 0.000867 | 12 | 0.000127 | 7.4 |

A closer look at the tweets revealed however that some of the usernames mentioned proportionately more frequently by either group were due to frequent retweeting. When excluding the retweets and the duplicate tweets sent by the same tweeter the Spearman correlation between the groups decreased to -0.353. Again the usernames that were mentioned proportionately more frequently by either group were coded based on their stance in the climate change debate (Figure 2). A total of 54 usernames mentioned by male tweeters and 53 usernames mentioned by the female tweeters met the threshold and were included in the analysis. Many of the usernames analyzed were found in both data sets. The results showed that the trend remains: female tweeters mention significantly more convinced usernames while male tweeters mention significantly more sceptic usernames. The numbers of neutral and unclear usernames were roughly the same in both groups. The differences were confirmed to be higher than would be expected from random tweeting by a chi-square test (chi-square=18.404, p=0.000). To confirm our hypothesis and to test whether the retweets in the data had some impact we ran the Z test for differences in the proportions of how usernames coded as convinced and sceptic were mentioned by the two groups. The Z-score for differences in the proportions for how convinced usernames were mentioned was 3.019 (p=0.00252) and the Z-score for differences in the proportions for how sceptic usernames were mentioned was -3.755 (p=0.00018). Both results were significant at $p < 0.05$. The results confirm our hypothesis and show that there are statistically significant differences in how male and female tweeters mention usernames coded as sceptic and usernames coded as convinced. The inter-coder agreement was measured with Cohen's Kappa to be a high 0.860.

**Figure 2. Usernames mentioned more frequently by male and more frequently by female tweeters according to the username's stance in the climate change debate. Retweets are excluded from the data.**
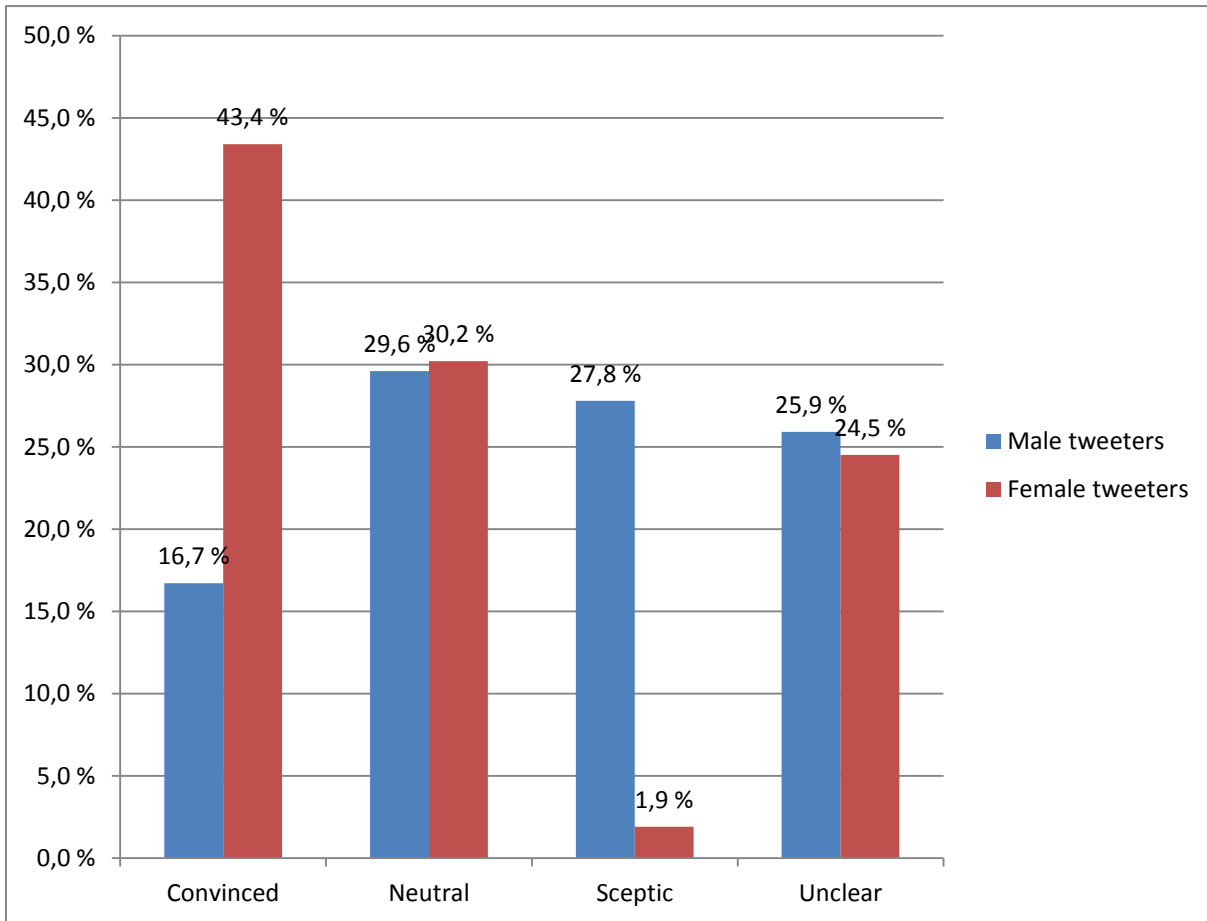


## Discussion

We set out to study differences in the climate change debate between female and male tweeters. In order to investigate both the comprehensive tweeting behavior and the original tweeting content, we decided to analyze the data both including retweets and excluding them. A closer analysis of the tweets where the frequent noun phrases, hashtags and usernames appeared revealed that the retweets had a clear impact on the results, as some of the detected differences appeared to be the result of frequent retweeting of some very popular tweets or the actions of a few very active tweeters. However, when removing the retweets the trends detected remained, in fact in some cases they emerged even stronger.

The impact of including the retweets in our analyses is clear, but the question remains what a retweet actually means and whether retweets skew the results or emphasize existing trends? While this question was beyond the scope of this research we cannot ignore it completely and a discussion about the meaning of retweeting is necessary. Retweeting is different from tweeting in the sense that retweeting is the action of forwarding a message that a Twitter user has received and want to share with his or her followers. This seems to be related to one of the most common social media uses in general, namely information sharing (Kietzmann et al., 2011; Naaman et al., 2010; Chen, 2013), and in particular

forwarding information to one's followers on Twitter. Original tweeting on the other hand requires that the tweeters formulate their thoughts into a maximum of 140 characters and submit the message. Original tweeting, in turn, may be more related to the use of Twitter for engagement (Kietzmann et al., 2011; Naaman et al., 2010; Chen, 2013). Retweeting requires much less effort from the tweeter, but that does not necessarily mean that retweets would not reflect the opinions of the tweeter in the same way that original tweets do. More research is however needed on this topic.

The retweets were included in our analyses with the assumption that even retweets can reveal something about the public opinion or attitudes of the tweeters, although some tweeters (not necessarily any of the tweeters in this study) write in their profiles that retweets do not in their case equal to endorsements. Whether a retweet is meant as an endorsement or denouncement cannot be determined from tweet content alone, but a retweet is still likely to be a signal of interest or awareness. The results of this study showed that even when removing the retweets from the analysis the detected trends remained: female tweeters mentioned usernames that belonged to organizations and campaigns with a convinced stance in the climate change debate and they used hashtags that supported this trend, in contrast to male tweeters who mentioned more political and general hashtags and mentioned usernames of a sceptic stance. We can conclude that the results in our research indicate that while female tweeters tend to show more interest and belief in the anthropogenic impact on climate change and towards campaigns and organizations involved in the debate, male tweeters are more concerned with politics related to climate change and connect more (for one reason or the other) with those that have a more sceptic stance in the climate change debate. The differences in the proportions were statistically tested and the results showed that there are significant differences between the two groups. It is however unclear whether these differences are due to more fundamental differences in the way men and women use social media in general and Twitter in particular, or whether the differences reflect male and female opinions about anthropogenic impact on climate change. The latter is however supported by some earlier findings (e.g., McCright, 2010; O'Connor et al., 1999; Whitmarsh, 2011).

The present study is not without its limitations, of which the most significant concerns the data sample. It is somewhat unclear how representative the data sample is of the general population on one hand and on the other hand, the tweeting population in particularly. As discussed earlier in this paper, Twitter limits the data collection of any subtopic to 1% of the total volume of tweets. This means that for popular topics it may not be possible to collect all the tweets. As climate change is a very specific topic we can perhaps assume that we have been able to collect most of the tweets, if not all of them. Another limitation related to the data that we have to acknowledge is the fact that the tweets represent only those tweeters whose gender could be determined from their name, which could be done for about 43% of all the tweets collected. This however still means that almost 250,000 tweets were included in the study, which should constitute for large enough sample that reliable trends could be detected. Another concern that one might raise is that we calculated large numbers of proportional differences in the frequencies with which hashtags and usernames were used by the two genders and there is a chance that some of the results were gained by chance. If that was the case these outliers and anomalies should have been detected when the data was analyzed qualitatively and coded into different categories, which was not the case. Finally we need to acknowledge that although our results are supported by earlier findings about gender differences in the attitudes towards climate change science (e.g., McCright, 2010; O'Connor et al., 1999; Whitmarsh, 2011), there may have been some non-gender related differences that could have influenced the results (e.g., political views, age, location). This, however, we are unable to check due to the limited nature of the data available from Twitter profiles. But thanks to some recent developments in for instance identifying tweeters' geographic location based on the content of their tweets or based on their social networks (Compton, Jurgens & Allen, 2014; Jalal, Nichols & Drews, 2014), other aspects may be included in similar analyses in the future.

On a theoretical level our results increase our understanding of climate change communication online and especially about how women and men view and engage with climate change. This has practical implications for organizations interested in developing communication strategies for reaching and engaging female and male audiences on Twitter in general, but especially in the context of climate change communication. While female tweeters can be targeted via local campaigns and news media, male tweeters seem to follow more political and scientific information. The results from the present research also showed that more research about the meaning of retweeting is needed, as we have shown how retweets can have a significant impact on the results. Future research should aim at increasing our understanding of the meaning of retweeting and of the impact the retweets may have on Twitter research.

## Footnotes

[1] October 29th Day of Action

[2] Hashtag for political discussions in Australia

[3] Acronym for Top Conservatives On Twitter

[4] Hashtag for progressive discussions

[5] Hashtag for an American political movement

[6] Hashtag for Canadian political discussions

[7] Pro-labor hashtag in Australia

[8] UN climate change conference in Warsaw 2013

[9] The 2013 meeting of the American Geophysical Union

## References

Armstrong, C.L. & Gao, F. (2010). "Gender, Twitter and news content", *Journalism Studies*, vol. 12, no. 4, pp. 490-505.

Auer, M.R., Zhang, Y.M. & Lee, P. (2014). "The potential of microblogs for the study of public perceptions of climate change", *Wiley Interdisciplinary Reviews – Climate Change*, vol. 5, no. 3, pp. 291-296.

Bamman, D., Eisenstein, J. & Scnoebelen, T. (2012). "Gender in Twitter: Styles, stances, and social networks", *arXiv:1210.4567 [cs.CL]*.

Barbier, G. & Liu, H. (2011). "Data mining in social media", in Aggarwal, C.C. (Ed.), *Social Network Data Analysis*, Springer Science + Business Media, LLC. DOI 10.1007/978-1-4419-8462-3_12.

Bosch, T. (2012). "Blogging and tweeting climate change in South Africa", *Ecquid Novi – African Journalism Studies*, vol. 33, no. 1, pp. 44-53.

Bollen, J., Pepe, A. & Mao. H. (2011). "Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena", in Ademic, L., Baeza-Yates, R. & Counts, S. (Eds.), *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media*. Palo Alto, CA: AAAI Press, pp. 450-453, 2011.

Bruns, A. & Stieglitz, S. "Twitter Data: What Do They Represent?" *it - Information Technology*, vol. 56, no. 5, pp. 240-245.

Cha, M. et al. (2012). "The world of connections and information flow in Twitter", *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, vol. 42, no. 4, pp. 991-998. Retrieved on July 7, 2012, from http://www.mpi-sws.org/~gummadi/papers/cybernetics-final.pdf.

Chen, G. M. (2013). "Why do women bloggers use social media? Recreation and information motivations outweigh engagement motivations", New Media & Society. doi: 10.1177/1461444813504269

Cheong, M. & Lee, V. (2010). "Twittering for earth: A study on the impact of microblogging activism on Earth Hour 2009 in Australia", *Lecture Notes in Computer Science*, vol. 5991, pp. 114-123.

Compton, R., Jurgens, D. & Allen, D. (2014). "Geotagging one hundred million Twitter accounts with total variation minimization", *arXiv*:1404.7152 [cs.SI]. Retrieved from http://arxiv.org/abs/1404.7152 on November 5, 2014.

Cunha, E., Magno, G., Almeida, V., Goncalves, M. A. & Benevenuto, F. (2012). "A gender based study of tagging behavior in Twitter". In the *Proceedings of the 23rd ACM Conference on Hypertext and Social Media*, pp. 323-324. Retrieved on February 15, 2014, from http://www.decom.ufop.br/fabricio/download/ht-evandro.pdf.

Cunha, E., Magno, G., Goncalves, M.A., Cambraia, C. & Almeida, V. (2014). "He votes or she votes? Female and male discursive strategies in Twitter political hashtags", *PLoS ONE*, vol. 9, no. 1: e87041. doi:10.1371/journal.pone.0087041.

Davidson, D.J. & Haan, M. (2012). "Gender, political ideology, and climate change beliefs in an extractive industry community", *Population and Environment*, vol. 34, pp. 217-234.

Grundmann, R. & Scott, M. (2014). "Disputed climate science in the media: Do countries matter?", *Public Understanding of Science*, vol. 23, no. 2, pp. 220-235.

Jansen, B.J. et al. (2009). "Twitter power: tweets as electronic word of mouth", *Journal of the American Society for Information Science and Technology*, vol. 60, no. 11, pp. 2169-2188.

Kietzmann, J. H., Hermkens, K., McCarthy, I. P. & Silvestre, B. S. (2011) "Social media? Get serious! Understanding the functional building blocks of social media", *Business Horizons*, vol. 54, pp. 241-251.

Kirilenko, A. P. & Stepchenkova, S. O. (2014). "Public microblogging on climate change: One year of Twitter worldwide", *Global Environmental Change- Human and Policy Dimensions*, vol. 26, pp. 171-182.

Krippendorff, K. (2004). "*Content Analysis. An Introduction to Its methodology*". 2nd edition. Sage Publications, Inc.

Lasorsa, D. (2012). "Transparency and other journalistic norms on Twitter", *Journalism Studies*, vol. 13, no. 3, pp. 402-417.

Mahmuc, J., Nichols, J. & Drews, C. (2014). "Home location identification of Twitter users", arXiv, arXiv:1403.2345 [cs.SI]. Retrieved from http://arxiv.org/abs/1403.2345 on November 5, 2014.

McCright, A. M. (2010). "The effects of gender on climate change knowledge and concern in the American public", *Population and Environment*, vol. 32, pp. 66-87.

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 2013*. Retrieved from http://www.public.asu.edu/~fmorstat/paperpdfs/icwsm2013.pdf on November 5, 2014.

Naaman, M., Boase, J. & Lai, C-H (2010). "Is it really about me? Message content in social awareness streams", *CSCW 2010*, February 6-10, Savannah, Geogia, US: 189-192. Retrieved on July 16, 2014, from http://infolab.stanford.edu/~mor/research/naamanCSCW10.pdf.

O'Connor, R. E., Bord, R. J & Fisher, A. (1999). "Risk perceptions, general environmental beliefs, and willingness to address climate change", *Risk Analysis*, vol. 19, no. 3, pp. 461-471.

Pearce, W., Holmberg, K., Hellsten, I. & Nerlich, B. (2014). "Climate Change on Twitter: Topics, Communities and Conversations about the 2013 IPCC Working Group 1 Report", *PLoS ONE*, vol. 9, no. 4: e94785. doi:10.1371/journal.pone.0094785.

Schmidt, A., Ivanova, A. & Schäfer, M.S. (2013). "Media attention for climate change around the world: A comparative analysis of newspaper coverage in 27 countries", *Global Environmental Change – Human and Policy Dimensions*, vol. 23, no. 5, pp. 1233-1248.

Schäfer, M.S. & Schlichting, I. (2014). "Media representations of climate change: A meta-analysis of the research field", *Environmental Communication – A Journal of Nature and Culture*, vol. 8, no. 2, pp. 142-160.

Takahashi, T., Abe, S. & Igata, N. (2011). "Can Twitter be an alternative of real-world sensors?" In J.A. Jacko (Ed.): *Human-Computer Interaction*, Part III, HCII 2011, LNCS 6763, pp. 240–249, 2011.

Thelwall, M. (2009). "*Introduction to webometrics: Quantitative web research for the social sciences*", Synthesis Lectures on Information Concepts, Retrieval, and Services, 2009, 1, 1. San Rafael, CA: Morgan & Claypool.Thelwall, M., Buckley, K., & Paltoglou, G. (2011). "Sentiment in Twitter events", *Journal of the American Society for Information Science and Technology*, vol. 62, no. 2, pp. 406-418.

Thelwall, M., Buckley, K., & Paltoglou, G. (2012). "Sentiment strength detection for the social web", *Journal of the American Society for Information Science and Technology*, vol. 63, no. 1, pp. 163-173.

Van Eck, N.J., & Waltman, L. (2010). "Software survey: VOSviewer, a computer program for bibliometric mapping", *Scientometrics*, vol. 84, no. 2, pp. 523-538.

Whitmarsh, L. (2011). "Scepticism and uncertainty about climate change: dimensions, determinants and change over time", *Global Environmental Change*, vol. 21, no. 2, pp. 690-700.

Whiting, A. & Willimas, D. (2013). "Why people use social media: a uses and gratifications approach", *Qualitative Market Research: An International Journal*, vol. 16, no. 4, pp. 362-369.

Wilkinson, D., Thelwall, M. & Li, X. (2003). "Exploiting hyperlinks to study academic Web use", *Social Science Computer Review*, vol. 21, no. 3, pp. 340-351.