Climate Change Perception in Scientific and Public Sphere

Didier Henry, Nathan Jadoul, Reynald Eugénie, Erick Stattner

Climate Change Perception in Scientific and Public Sphere

Didier Henry
LAMIA
Université des Antilles
Pointe-A-Pitre, France
didier.henry@univ-antilles.fr

Nathan Jadoul
LAMIA
Université des Antilles
Pointe-A-Pitre, France
nathan.jadoul@etu.univ-antilles.fr

Reynald Eugénie
LAMIA
Université des Antilles
Pointe-A-Pitre, France
reynald.eugenie@etu.univ-antilles.fr

Erick Stattner
LAMIA
Université des Antilles
Pointe-A-Pitre, France
erick.stattner@univ-antilles.fr

***Abstract*—Climate change is one of the major concerns of humankind during this 21st century. Indeed everywhere on the planet, effects climate change is now widely perceptible to everyone. However, while a lot of efforts have been made on the proposal of efficient models able to provide reliable climate projections for the future years, very little work has been done on the global population's feelings regarding the climate changes. Thus in this paper we focus to this social dimension of the problem and we adopt a data analytics approach that aims to study people's perception of climate change. First, we analyze the scientific papers published during the last forty years and we try to understand what are the topics addressed and their evolution. Then, we have collected data on Twitter to understand what are the feelings and the topics addressed by the individuals regarding climate change. To the best of our knowledge, this is the first data analysis approach that highlights the feelings and topics of the global population about climate change and compares them with scientific articles.**

***Index Terms*—climate change, text mining, social media, data analysis**

I. Introduction

Climate change is one of the major concerns of humankind during this 21st century. Indeed, everywhere on the planet, climate change is now widely perceptible to everyone. More particular, people feel climate changes at various levels: seasonal disruption, increasingly serious drought, increased of extreme events (storm, hurricane, tornado, heat waves, forest fire, etc.), disappearance of species, and so on. In this context, understanding environmental changes and assessing their potential effects is a major challenge for the population resilience and the adjustment of practices.

Thus, numerous works have been conducted to study the impact of climate change on various aspects of our societies: health [1], economy [2], agriculture [3], biodiversity [4], environnement [5], and even on some natural resources [6]. More specifically, one of the major challenges for understanding and adapting to climate change is the ability to provide reliable projections over the long term [7], [8]. Consequently, several studies have conducted a modeling work to provide relevant climate models that shed light on the possible changes for the next years. These models are now widely used to obtain various kinds of climate projections over the next few years according to different scenarios. For instance, in [9] M. H. Dore reviews the main observations in global precipitation patterns and details the possible evolution.

However, while a lot of efforts have been made on the proposal of efficient models able to provide reliable projections for the future years, very little work has been done on the global population's feelings regarding the climate changes. This perception of climate change at the level of individual, and more generally at the level of a population, is nevertheless very important since it allows to better understand the position of individuals in relation to the phenomenon, the perception they have of the problem, as well as the actions they are likely to accept and perform. This last point is indeed important, because in a decision-making context, the capacity that individuals can have to feel concerned by a problem is fundamental in the acceptance and the respect of the actions put in place by the decision-makers.

In this paper, unlike the large majority of works that are interested in climate projections, we focus here to the social dimension of the problem. More specifically, we adopt a data analytics approach that aims to study people's perception of climate change. The work we have conducted is twofold.
(i) First, we analyze the scientific papers published during the last forty years and we try to understand what are the topics addressed as well as their evolution. For this purpose we begin by characterizing the possible evolution on the topics addressed by the scientific community and we then propose a model able to identify them in the data. By this way, we were able to observe what are the most discussed topics around climate change, as well as that which were neglected by the community for many years.
(ii) In a second study, we sought to understand what was the feeling of the population in relation to climate change. For this purpose, we have collected data on the social media Twitter

and we have performed sentiment analysis on messages posted to map feelings about climate change. The topics addressed in Tweets have also been compared to topics addressed by the scientific community.

To the best of our knowledge, this is the first data analysis approach that highlights the feelings and topics of the global population about climate change and compares it with scientific articles. This approach, guided by field data, allows to better understand climate tendencies of recent years and the perception of individuals.

The paper is organised as follows. Section II reviews the main works involving climate change and social media. Section III describes the model proposes to follow the evolution of topics addressed in scientific papers and in Tweets. Section IV and V are respectively devoted to the results obtain on scientific papers and Twitter. Section VI compares the topics addressed by the two types of community Scientists and Twitter users. Finally, Section VII concludes the paper and presents our future works.

II. Related Works

In this era of data expansion, we can easily see the interest behind collecting data, stocking them and extracting information from them. Nevertheless, they are many ways to format the data, and even more ways to treat them. For decades, the most usual format was a table of values, but nowadays the interest in the other support of information rose in order to fit more accurately the observed elements. In this study, we will focus on two formats: the text corpus, with which we can make clusters of documents that share similarity, and the networks, which became more and more relevant due to the democratization of social networks.

In [10], Gupta and Lehal pinpoint 10 main uses for text mining, and we will focus on two of them : (1) categorization in order to classify a document for a given taxonomy, and (2) clustering in order to aggregate the documents of similar structure. We can see the interest of regrouping words of similar topics by using a taxonomy with the work of Wu and al [11], while [12] and [13] draws our attention on different text mining pre-processing methods useful for documents clustering. More specifically, we can conclude from those papers that the use of the term frequency-inverse document frequency (TF-IDF) methods in a study on a very large text corpus, especially when it is applied to social networks messages which can generate millions of messages per years.

In just a few years, online social media has transformed the way we create, share and access information. These platforms based on huge networks allow the free exchange of information between hundreds of millions of people (celebrities, organizations, individuals, unions, etc.) around the world, and this instantaneously. Every day these people broadcast hundreds of millions of messages on a wide variety of topics (politics, sport, health, news, technology, etc.). Whether related to a global event, that is to say, specific to all or part of individuals, or with a local event, that is to say, specific to an individual, these messages can influence a society and contain useful information for the detection or prediction of real-world phenomena.

For example, Sakaki et al. [14] present a model using the messages and their position as sensors to inform the population in the event of an earthquake. Similarly, Gomide et al. [15] use messages to detect and locate dengue cases in Brazil. In addition, social networks are also widely used in emergencies such as floods [16], [17], forest fires [18], and hurricanes [19].

Sentiments and emotions expressed by users can be extracted through text analysis tools [20], [21] and can be useful for event prediction in the real world and to know the opinion of individuals.

In his work, Paul Ekman [22] has defined a set of six universally recognizable basic emotions: anger, disgust, fear, joy, sadness and surprise. Researchers [23], [24] have shown that emotions expressed in social media may affect the decision to share information or to be contagious [25], [26] and thus disseminate information more widely.

Regarding sentiments, Asur and al. [27] propose a model to predict the revenue generated by movies beyond the fourth week after the release of the film. Their model uses the rate of messages per week, the rate of messages per hour, the polarity and the subjectivity of the messages. Thanks to these parameters, the model is able to predict weekly film revenue in the short term with a high probability. In the same way, Bollen et al. [28] use messages to predict stock market developments. By analyzing the subjective nature of the messages, they note that the mood changes of the users are correlated with the changes in the financial markets. In addition, messages submitted on social media may be used to predict the outcome of a political election [29]–[31].

Events related to climate change spread largely in social media [32], [33]. Recently, some researchers [34], [35] have attempted to understand the human reaction on climate change thought Twitter. For instance, Cody and al. [36] have measured the happiness in messages on Twitter posted between September 2008 and July 2014. An et al. [37] have performed a sentiment analysis of messages posted on Twitter in terms of polarity and subjectivity between October 2013 and December 2013. They have noted that Twitter can be a valuable way to yield insights on climate change opinions. In their work, Holmberg and al. [38] have collected messages spread on Twitter between October 2013 and January 2014 and they have observed gender differences in the climate change communication on Twitter.

To the best of our knowledge, not any work in this field has attempted to consider climate change taxonomy, sentiments, emotions and location of individuals. In this paper, we observe evolution of climate change thematic addressed in the scientific area over the last 40 years, then we focus on the general public opinions and feelings related to climate change through Twitter between 2009 and 2018.

III. Modeling evolution of climate change perception

In order to study the evolution of the themes tackled on climate change both by scientists and by the general public, we introduce the BCDR model to analyse 4 behaviours:

- the **B**irth of a set of themes,
- the **C**ontinuity of a set of themes,
- the **D**eath of a set of themes,
- the **R**eappearance of a set of themes.

The proposed model can be applied both to the study we have conducted on scientific articles of Google Scholar and to the study of messages posted on Twitter.

More formally, the different behaviours are described below:

Let x_j a theme among the set of themes I ={Arctic, Agriculture/Forestry, Economy, Energy, Politics/Opinion, Risk/Disaster, Ocean/Water, Weather}.

Concerning the general public opinion that is to say messages diffused on Twitter, we have chosen to assign them a subset of themes of I noted X such that $X \subset I$ because of the limited number of characters of tweets. For example, X ={Arctic, Economy, Weather}. We calculate the proportion of messages where Xappears noted $Support(X)$. A subset X is said to be frequent if and only if $Support(X) \geq \theta$ where θ is a threshold defined in advance. We define $X_t = [X_{0t}, X_{1t}, ..., X_{nt}]$ the set of frequent X at time t. Regarding the scientific circle we assign to each abstract article a theme, thus $X = \{x_j\}$. We define $X_t = [X_{0t}, X_{1t}, ..., X_{nt}]$ the set of themes at time t. We define $Before_t$ the set of frequent X or X before time t.

- **Birth**: a set X is born at the moment t if and only if X does not belong to any X_k such that k belongs to $[0, 1, ..., t-1]$,
- **Continue**: a set X continues at the moment t if and only if X belong to X_{t-1},
- **Death**: a set X die at time t if and only if X does not belong to any X_k such that k belongs to $[t+1, ..., t_{max}]$,
- **Reappearance**: a set X reappears at a time t if and only if it exists $t1 \geq t-1$ such as:
  - $X \in Before_{t1}$,
  - $X \in X_t$,
  - $\forall k \in]t1, ..., t-1] X \notin X_k$.

IV. Climate change perception in scientific sphere

In a first step, we focus on the scientific community and attempt to understand how climate change is treated in scientific articles. For this purpose, we collect scientific papers related to climate change published the past forty years and we analyze them to extract the topics addressed each year. Topics are indeed a good indicator of the strong concerns around climate change over time. This section details the methodology used for collecting papers and the results obtained.

A. Collecting scientific papers

To study the evolution of topics related to climate change in scientific articles, we extract papers from Google Scholar since 1980. Papers are then labeled according to a taxonomy of climate change to identify the topics addressed each year. The whole process is described on Figure 1.

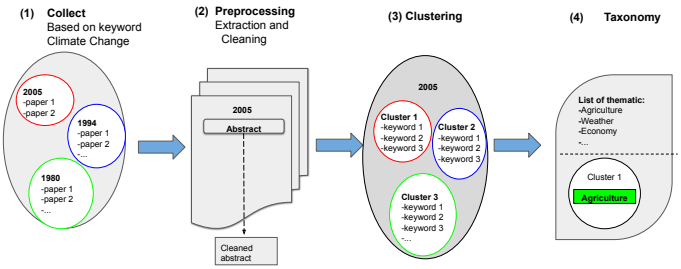


Fig. 1. Extraction of topics from scientific articles

(1) Collecting scientific papers. The corpus has been built by downloading scientific articles on Google Scholar search engine with the keyword "*climate change*". Articles have been collected on the first five pages of results and the dataset covers the period 1980-2019. The returned documents that were not scientific articles (such as books, images, presentation, flyers, etc.) were removed from the study. By this way, about 750 articles have been collected, namely about 20 articles for each year.

(2) Extracting and preprocessing textual content. In our analysis approach, we have focused on the abstracts of the articles studied. Abstracts are relevant since they condense the important points of the article. Once the abstract has been extracted, the preprocessing phase has been conducted with the R language and performs as follows. First of all, blank, punctuation and all numbers has been removed from the text. Then the text was put in lowercase and all stop words were deleted. Finally, only the root of the remaining words has been preserved to group the words with the same root (for example: *like*, *liked* or *liking* will become *lik*).

(3) Clustering process. Each year, the number of word occurrences is evaluated by using the document term matrix (DTM). As a reminder, DTM is a matrix that describes the frequency of terms that occur in a collection of documents. To dismiss irrelevant words, only the fifty most frequent words in the corpus are kept. Then, we use a hierarchical clustering approach based on the *Ward* method to group scientific articles on the basis of common words. The tree thus generated by the approach is then pruned to keep only five clusters. Compared to other clustering methods, such as k-means, we made the choice of hierarchical clustering since it is inexpensive in terms of calculation.

(4) Labelling with taxonomy. Finally, each cluster is labeled with the topics associated. For this purpose, we use a taxonomy of climate change to identify the topics covered by each cluster. The taxonomy is composed of 8 topics: (i) Risk/Disaster, (ii) Politics/Opinion, (iii) Economy, (iv) Energy, (v) Ocean/Water, (vi) Weather, (vii) Agriculture/Forestry and (viii) Artic. Each of these topics has a list of keywords

that define it. We compare the most frequent words of the clusters to the lists of keywords to determine the topics of a cluster. By this way, the set of topics of all clusters provide the topics addressed around climate change for a given year.

B. Evolution of topics

In a first step, we have used the model described in Section III to study the evolution of topics since 1980. More specifically, topics around climate change are collected in 1980 and the model is used to study the evolution from 1981. By this way, 1980 is the reference year. Figure 2 shows (a) the appearance of new topics, (b) the distribution of the number of years of disappearance of topics, (c) the reappearance of topics, and (d) the death of topics.

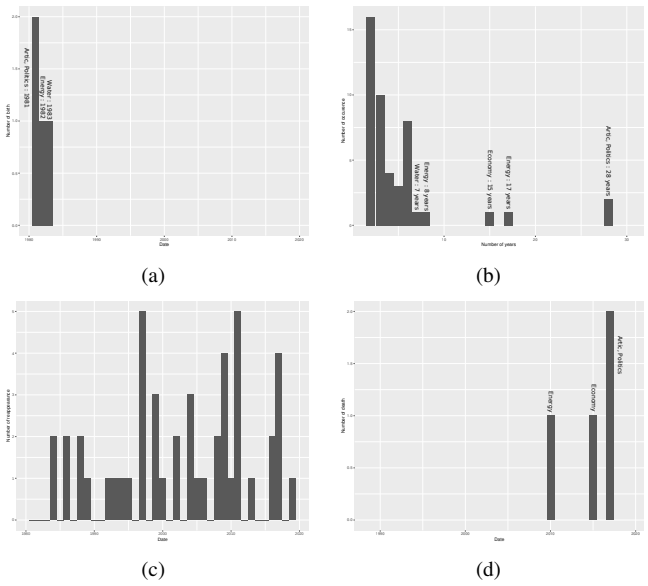


Fig. 2. Evolution of (a) births, (b) distribution of the number of years of disappearance, (c) reappearances and (d) death of topics on climate change

Regarding the appearance of topics (see Figure 2(a)), it is very interesting to observe that from 1980 to 1983 all the topics of the taxonomy have been found in scientific papers. More particularly, the vast majority of topics are found in scientific papers from 1980, the themes of the *Artic* and *Politics* appeared in 1981, *Energy* appeared in 1982 and topic of *Water* emerges in 1983. This reflects the fact that from 1980 to 1983, the scientific community seemed to be concerned by all the thematic around climate change.

The results obtained for the disappearance of topics are very surprising (see Figure 2(b)). Indeed, we can observe that some topics are no longer treated by the scientific community during several years. More precisely, the vast majority of the topics that are no longer treated disappearance between 2 and 6 years before emerging again. Moreover, it is also very interesting to observe that some topics are no longer found in scientific papers for more than 10 years. For instance, it is the case of *Water* that disappears 7 years or *Economy* that disappears 15 years. The longest period of disappearance is for *Arctic* and *Politics* that are not found in paper for 28 years.

Thus, if we focus on the number of reappearances (see Figure 2(c)) the results confirm our previous observations. Indeed, we can observe that all topics are not treated each year since throughout the study period numerous topics that were not addressed at the previous year may be reappear at year t.

Finally, the first deaths of topics occur after 2010 (see Figure 2(d)). For instance, *Energy* is no longer found in scientific papers after 2010. In the same way, the topics of *Economy* and *Arctic* are considered as death, namely they are no longer found in scientific publications, respectively in 2015 and 2017. It is important to note that as 2019 marks the end of our study, these topics are considered dead because they have not been discussed. Obviously, it is not impossible that in the coming years these topics reappear.

Our previous results have shown that topics of the taxonomy had very different covers in scientific articles over time. To go further, we wanted to represent on a single diagram the frequency of each topic over time. Figure 3 presents these results.

Fig. 3. Topics covered in scientific papers on climate change since 1980

Figure 3 allows to observe the individual trajectories of each topic and in consequence the trends that emerge. For instance, for the topics *Arctic* and *Politic*, we can observe that they were found in scientific articles in 1981 and disappear until they reappear in 2010. On the contrary, some topics as *Energy* are mainly found in scientific papers published in the 80s. Finally, we can also observe that some topics are regularly covered by the scientific articles such as *Weather* or *Risk*.

Thus, this approach made it possible to highlight the topics covered by the scientific articles around the climate change problematic. Although scientific articles have a duty of neutrality in terms of feeling, these topics nevertheless reflect the concerns of the scientific community on the climate issue. Thus, to go further we conduct in the next section a study on the population feelings.

TABLE I
DATASETS CHARACTERISTICS.

Years	Number of tweets	Number of distinct users
2009	595 293	272 380
2010	692 354	161 620
2011	969 333	236 636
2012	1 409 789	336 330
2013	1 872 662	426 944
2014	2 904 011	566 760
2015	4 328 695	778 440
2016	3 542 370	805 713
2017	4 442 785	1 109 038
2018	3 905 524	977 019

V. CLIMATE CHANGE PERCEPTION IN PUBLIC SPHERE

A. Data collection and extraction

Our aim is to observe global opinions and feelings related to climate change in social media. The main challenge we have encountered was having access to old messages related this topic on Twitter. Firstly, to collect old messages/tweets, we have used the Twitter search engine[1] by specifying the keywords "climate change", the year and the English language in the query. By following this method, we have got a large number of tweets (see TableI). In addition, we have collected account information for each user such as his current number of tweets posted on Twitter, his current number of followers, followings, his account date creation and his current location (if specified). The characteristics of datasets obtain are summary in the Table I.

Secondly, we have used TextBlob [39] an API in Python language in order to extract the sentiments of tweets in terms of polarity and subjectivity. TextBlob [2] takes an English text input and returns its polarity score and subjectivity score. The polarity score is a number in the range [-1.0, 1.0] where -1.0 is the fully negative score and 1.0 is the fully positive score. The score of the subjectivity is a real in the interval [0.0, 1.0] where 0.0 is entirely objective and 1.0 is entirely subjective. We have considered three classes for the message polarity (negative, neutral and positive) and three classes for the message subjectivity (subjective, neutral and objective).

Thirdly, we have used an emotion recognition tool proposed by Colnerič and al. [21] to extract the emotions expressed in tweets. This tool[3] contains a function that takes a text in English as input and returns the relative emotion (joy, fear, sadness, anger, surprise or disgust).

Then, we have used taxonomy proposed by a United Nations work group[4] to assign a thematic to tweets such as Risk/Disaster, Politics/Opinion, Economy, Energy, Ocean/Water, Weather, Agriculture/Foresty, Arctic.

Finally, we have used GeoPy[5] a Python library to obtain the current country of users who have indicated a location in their profile.

[1]https://twitter.com/search-advanced
[2]http://textblob.readthedocs.io/en/dev/
[3]https://github.com/nikicc/twitter-emotion-recognition
[4]http://unglobalpulse.net/climate/taxonomy/
[5]https://geopy.readthedocs.io/en/stable/

In a first step, we have focused on the evolution of the number of people posting messages related to climate change on Twitter from 2009 to 2018. We have divided this number by the number of active users on Twitter. By observing this ratio, we have noted that the number of people posting climate change messages on Twitter over the years tends to be stable (see Figure 4).

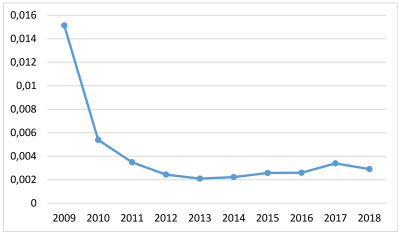


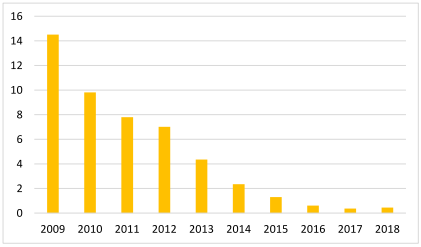
Fig. 4. Evolution of the ratio of the number of Twitter users posting messages related to climate change on the number of active users

Then, we were interested in the messages and their contents. By observing the percentage of retweets (namely the sharing an already posted message), we have noted that individuals tend to post their own messages. Indeed, we have remarked that the percentage of retweets is close to 14% in 2009 against less than 1% (see Figure 5 (a))

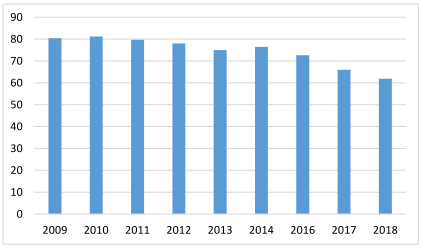
In addition, we have observed that the percentage of messages containing a leat a URL tends to decrease from 2009 to 2018 from 80% to about 60% (see Figure 5 (b)). Thus, individuals seem to become more and more personal about climate change.

B. Data analysis

In a second step, we were interested in sentiments and emotions expressed in messages. We have noted that there are

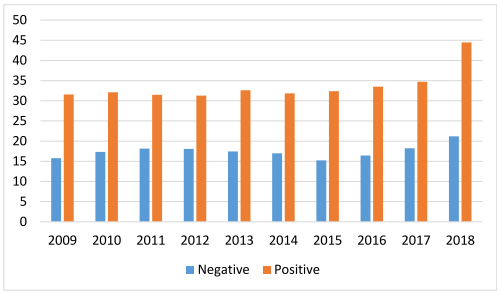


(a)

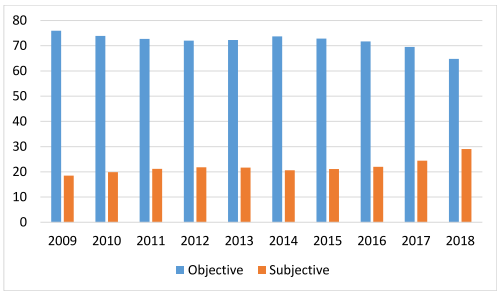


(b)

Fig. 5. Evolution of the percentage of retweets (a) and percentage of messages containing a least an URL (b) over time.

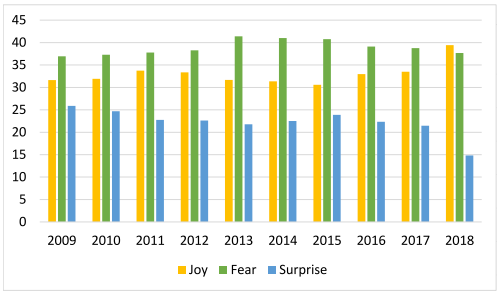


(a)

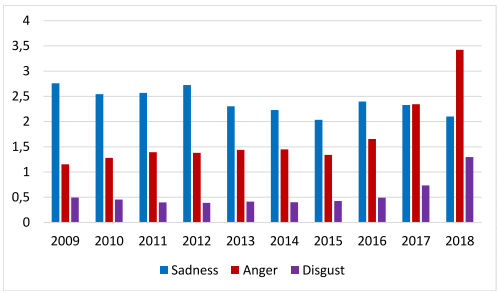


(b)

Fig. 6. Evolution of the percentage of positive and negative messages (a) and objective and subjective messages (b) over time.



(a)



(b)

Fig. 7. Evolution of the percentage of expressing joy, fear, surprise (a) and sadness, anger, disgust (b) over time.

more messages marked positively than negatively (see Figure 6 (a)) and there are more objective messages than subjective messages (see Figure 6 (b)). Moreover, we have noticed that the percentage of negative messages and subjective messages follows an uptrend from 2016.

By observing the messages distribution according to emotions, we have remarked that joy, fear and surprise are the most expressed emotions about climate change. In addition, we have noted that there are more messages expressing fear from 2009 to 2017 (see Figure 7 (a)). Furthermore, while the percentage of messages expressing surprise tends to decrease the percentage of messages expressing anger tends to increase over the years (see Figure 7 (a) and (b)).

In a third step, we were interested in the theme of the messages as well as the sentiments and emotions associated with them. We have noted that the theme *Risk/Disaster* is the most talked about in the messages posted on Twitter about climate change and the theme *Arctic* is the less talked about (see Figure 8).

Then, we have focused on emotions expressed for each theme. For instance, we have remarked the part of messages expressing fear is larger in the case of politics or opinion (see Figure 9 (a)). In addition, we have noted that the part of messages expressing joy is smaller when it comes to politics, opinions, risks or disasters (see Figure 9 (b)).

Next, we were interested in sentiments expressed for each theme. We have noted that the part of subjective messages is larger in the case of weather and the part of negative messages is also greater in the case of weather since 2011 (see Figure 10 (a) and (c)). These results seem to suggest that people express their own opinion about weather and this opinion is

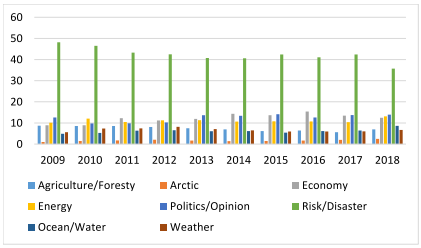
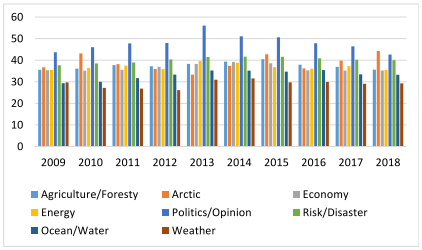


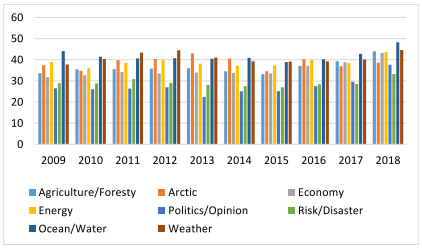
Fig. 8. Evolution of themes distribution over time.

more and more negative. Moreover, we have remarked that the part of positive messages is globally similar to the set of themes. These results may mean that there are in all fields positive ideas to fight against global warming.

Finally, we have focused on the location of individuals to observe people's feelings about climate change more finely. We have noticed that expressed feelings about climate change are different depending on the location of the individuals. However, we have observed common trends. For instance, we have remarked that the part of positive messages related to politics tends to increase at several places over time (see Figure 11 (a)). In addition, we have noted that the part of messages expressing the fear about risk also tends to rise at several places (see Figure 11 (b)). Moreover, we have observed that the part of negative messages related to weather tends to increase at several positions on earth (see Figure 11 (c)). These results seem to indicate that more and more people around the world are feeling the negative effects of climate change but solutions are being proposed, particularly at the political level.

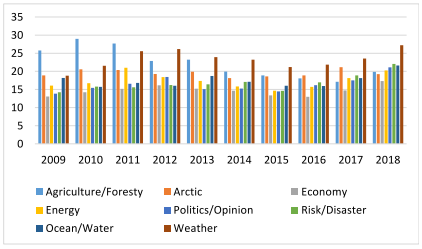


(a)

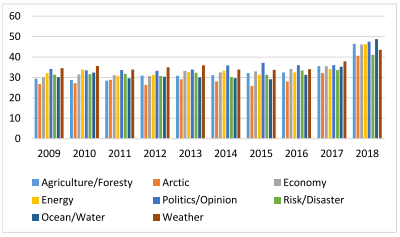


(b)

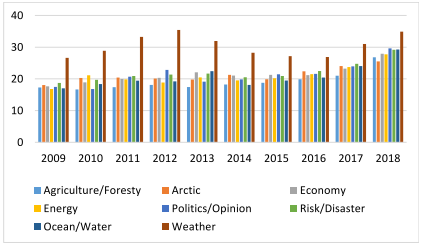
Fig. 9. Evolution of the part of messages expressing fear (a) and joy (b) by theme over time



(a)

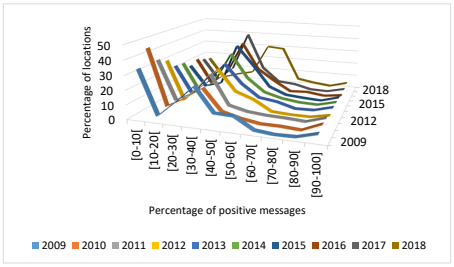


(b)

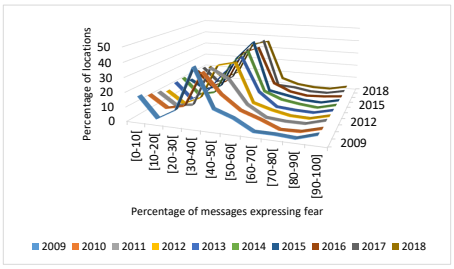


(c)

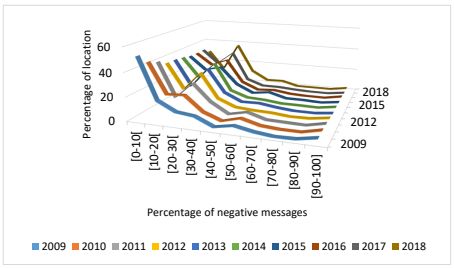
Fig. 10. Evolution of the part of negative messages (a), positive messages (b) and subjective messages (c) by theme over time.



(a)



(b)



(c)

Fig. 11. Evolution of the distribution of locations related to the part of positive messages about politic (a) the part of messages expressing the fear about risk (b) and the part of negative messages about weather (c) over time.

C. Data clustering

In order to extend the scope of the work, we tried to apply the classic text mining algorithms on tweets to extract clusters. We used a python library named sklearn for the processing and also for a part of the pre-processing, which was made through 5 steps : (1) lower the case of the letters, (2) suppress the URLs, hashtags and reference to other users, (3) delete all non-alphabetic characters, (4) suppress the stop words and the words which are not in the English dictionary, and lastly (5) stemming the remaining words. The main idea was to pinpoint the composition of words clusters generated for each attributes previously extracted.

We used the TF-IDF in order to minimize the importance of the words that are too frequently used, as it is a method which applies a pertinent weighting for our dataset. The TF-IDF do not only use the frequency of the word in the document, but

also the frequency of the word in the whole corpus. Formally, for a term x and a document y in the corpus D, it is defined by:

$$TF \cdot IDF(x, y, D) = tf_{x,y} \times log\left(\frac{\mid D \mid}{df_{x,D}}\right)$$

where $tf_{x,y}$ is the number of occurrences of x in y and $df_{x,D}$ is the number of occurrences of x in the corpus D

By using this method, we were able to use a more pertinent distance between the tweets, which was used for a K-Means clustering algorithm. The optimal number of clusters was determined clusters to have a maximum of significant clusters, namely the ones which contain a certain threshold β of the data set. We observed that they are always a main cluster which usually contains 70 percents of the data, leaving only few percent for each one of the other clusters. Beta was put 4%, leading to a total of 5 clusters. To mitigate the inequality in each cluster, we had to normalize the results by dividing the number of a specific attribute a in a cluster C_i by the amount of value of the very same cluster C_i. For example, we can see in Figure 12 that the subjectivity which values go between 0 and 1 was partitioned in 3 equal parts, and after the normalization the cumulative value of all cluster set to 1.

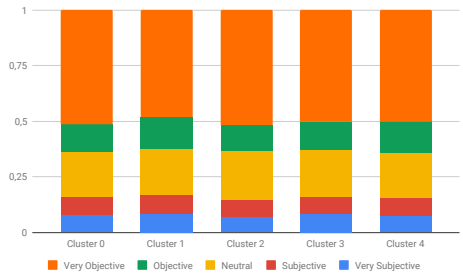


Fig. 12. Composition of the extracted cluster in regard of the subjectivity separated in 5 portions

The polarity and the subjectivity were both discretized in 3 parts (respectively negative - neutral - positive and objective - neutral - subjective) and in 5 parts (very negative - negative - neutral - positive - very positive and very objective - objective - neutral - subjective - very subjective) . With the emotions and taxonomy, we extracted 60 graphs, 6 for each year. The first observation was the similarity in the distribution of the different attributes in each cluster. For the polarity and the subjectivity, it can be explained by the fact that most of the word used for determining them should have been sorted out when eliminating the stop words, nevertheless we can conclude that except for those particular words, the structure of the messages only has very small change no matter if the message is positive, negative, objective or subjective.

However, for the emotions and the taxonomy, the results are quite different. Even throws the structure of each cluster still presents some similarity, the variation of some of them can be particularly noticed. For example, we can see in Figure 13 that despite the fact that the "risk" is usually more represented, in

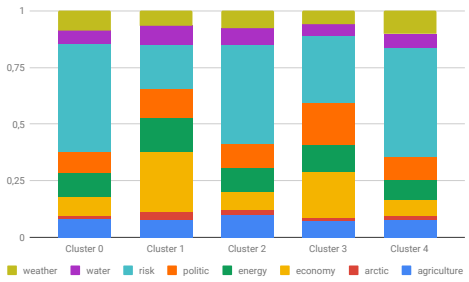


Fig. 13. Composition of the extracted clusters in regard of the taxonomy

Cluster 1 a big decrease is observed, leading to an increase the economy. In fact, for all the observed years, there is always a cluster in which the balance between the risk and the economies is shaken and since 2012, those clusters contain more tweet classified as the later than the former.

In order to see the evolution of the composition of the clusters, we used the standard deviation, then we compared it for each year.

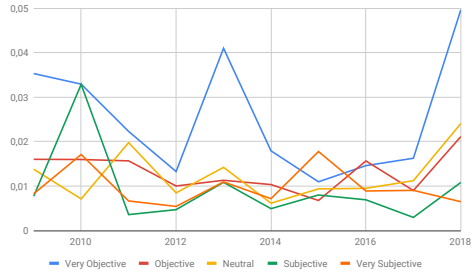


Fig. 14. Evolution of the standard deviation of the composition of the clusters in regards of the subjectivity

The Figure 14 allow to see that although the versatility in the very objective messages, the values tend to be stable and very low, showing us that the composition of the cluster does not change a lot for this attribute, a trend that is also observed for the polarity of the messages.

On the contrary, the Standard derivation for the Emotion and the Taxonomy seems overall higher, particularly the fear, the joy and the surprise for the emotions, and the risk and politic for the taxonomy (see Figure 15). In fact, it is quite logical to expect variations for the taxonomy, leaving the fact that most of the clusters still having a similar structure the most remarkable fact.

It may be because of the use of the TF-IDF which mitigate the impact of the words which are present in too many tweets, or the use of the KMeans as a method to make the clusters, but more studies have to be made in order to explain the lack of sparsity in the clusters of tweets in regards of their compositions.

VI. Discussion

In the section IV, we saw the comportment of the scientific community in regards to climate change is quite heterogeneous

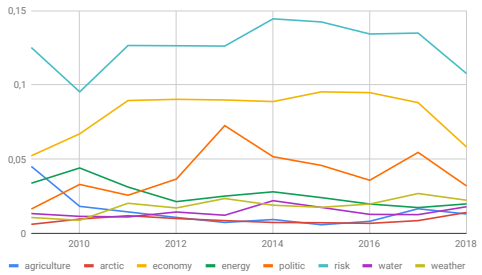


Fig. 15. Evolution of the standard deviation of the composition of the clusters in regards of the taxonomy

with respect to time but also to the used taxonomy. Between 2009 and 2018, we can observe in the Figure 3 at least one death of two years or more for each topic of the taxonomy, and for the topic of Energy, there is almost no article at all. Conversely, the twittos topics of interest seems relatively constant between 2009 and 2018.

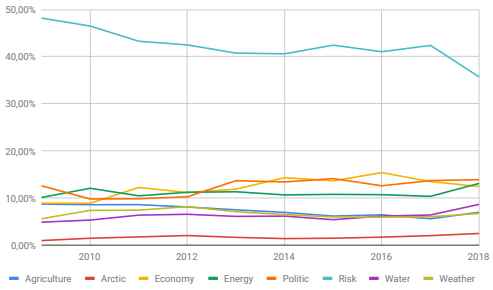


Fig. 16. Evolution of the composition of the tweets per year in regard of the taxonomy

In the Figure 16, we can effectively see that except for the Risks which has a decrease from 48% to 36%, we can hardly consider that the composition of the tweet change for the other topics

This information can be quite representative of a determining difference between tweets and scientific papers. In fact, the tweets don't really have any restrictions about the duplication of information. It is even quite frequent to spread an existing and popular topic. This particularity can explain why in the case of twitter, the topic of the taxonomy never dies.

On the contrary, the scientific article is to be different, the authors have to contribute with new information in order to have a pertinent content. This restriction can lead to an assumption of why some subjects may be more often discussed when they are more diversified (Risk, Weather, ...) while other which seems more specific are less present (Arctic, Politic, ...).

Another interesting point is precisely about the topics which are more discussed. In the 10 years from 2009 to 2018, the scientific community seemed more focused on the Risk and the Agriculture and less on other topics such as Energy and Politic but for the tweets, even if the Risk is still quite present, the Agriculture is among the less popular topics for Twitter users while Energy and Politics, the only ones that appeared in more than 10 % of the tweets for almost each year, are in the 4 more discussed topics.

This opposition in the popular topics is also very fascinating. Considering that twitter can be a great representation of the society, it shows that the scientific communities and the global population look differently the problem of climate change, but unanimously the two community consider it as a potential problem, a risk.

We can theorize about the reasons of the topics discussed, such as the fact that Twitter users have an easier access on subjects like politics, energy or economy which are more "captivating" nowadays than other subjects like agriculture, even if it is also an important aspect of the problem which is kept in the mind of the scientific community.

VII. Conclusion

Numerous works have been conducted to study the impact of climate change on our societies. While many works focused on climate projections, in this work we have adopted a new point of view by addressing the climate change issue with the study of the perception of individuals facing the problem. Thus, we have compared the topics treated about the climate issue by two categories of users, both scientific community and users of Twitter. The contributions of the paper can be summarized as follows.

- We have first defined evolutions that may occur on topics over time, through four possible evolution trajectories. Then we have proposed a model that identifies them in data collected each year.
- We have applied the model to conduct an analysis work of the scientific papers published during the last forty years and we have highlighted how the topics addressed by the scientific community have evolved over time. The model proposed has also allowed to observe interesting trajectories, since we observed that some topics may emerge after several years without being addressed by the scientific community.
- Finally, we have collected messages posted on Twitter on climate change and we have conducted a sentiment analysis to highlight the evolution of sentiments on the climate issue. We have also compared the topics addressed on Twitter on those address by the scientific community.

As perspectives, we plan to extend the work conducted on scientific papers. Indeed, we have limited the analysis to the abstracts of the scientific papers published in conferences or journals In our future works, we would like to extend the analysis to all kinds of documents and to consider the whole document in the analysis task.

Regarding the work conducted on the perception of the population, we chose to collect the messages published on Twitter. The challenge is that Twitter exists only since 2006, which makes the comparison with scientific articles difficult for which we have data since 1980. Thus, in our future directions we plan to extend the data collected on the population

feeling by targeting web sites, comments on articles, forums, etc.

REFERENCES

[1] A. J. McMichael, R. E. Woodruff, and S. Hales, "Climate change and human health: present and future risks," *The Lancet*, vol. 367, no. 9513, pp. 859–869, 2006.

[2] R. Mendelsohn and J. E. Neumann, *The impact of climate change on the United States economy*. Cambridge University Press, 2004.

[3] J. Ramirez, A. Jarvis, I. Van den Bergh, C. Staver, and D. Turner, "Changing climates: effects on growing conditions for banana and plantain (musa spp.) and possible responses," *Crop adaptation to climate change*, vol. 19, pp. 426–438, 2011.

[4] C. Bellard, C. Bertelsmeier, P. Leadley, W. Thuiller, and F. Courchamp, "Impacts of climate change on the future of biodiversity," *Ecology letters*, vol. 15, no. 4, pp. 365–377, 2012.

[5] A. Venäläinen, H. Tuomenvirta, M. Heikinheimo, S. Kellomäki, H. Peltola, H. Strandman, and H. Väisänen, "Impact of climate change on soil frost under snow cover in a forested landscape," *Climate research*, vol. 17, no. 1, pp. 63–72, 2001.

[6] N. W. Arnell, "Climate change and global water resources," *Global environmental change*, vol. 9, pp. S31–S49, 1999.

[7] G. A. Meehl, F. Zwiers, J. Evans, T. Knutson, L. Mearns, and P. Whetton, "Trends in extreme weather and climate events: issues related to modeling extremes in projections of future climate change," *Bulletin of the American Meteorological Society*, vol. 81, no. 3, pp. 427–436, 2000.

[8] D. W. Gamble and S. Curtis, "Caribbean precipitation: review, model and prospect," *Progress in Physical Geography*, vol. 32, no. 3, pp. 265–276, 2008.

[9] M. H. Dore, "Climate change and changes in global precipitation patterns: what do we know?" *Environment international*, vol. 31, no. 8, pp. 1167–1181, 2005.

[10] V. Gupta, G. S. Lehal *et al.*, "A survey of text mining techniques and applications," *Journal of emerging technologies in web intelligence*, vol. 1, no. 1, pp. 60–76, 2009.

[11] S.-T. Wu, Y. Li, Y. Xu, B. Pham, and P. Chen, "Automatic pattern-taxonomy extraction for web mining," in *Proceedings of the 2004 IEEE/WIC/ACM International Conference on Web Intelligence*. IEEE Computer Society, 2004, pp. 242–248.

[12] W. Zhang, T. Yoshida, and X. Tang, "A comparative study of tf* idf, lsi and multi-words for text classification," *Expert Systems with Applications*, vol. 38, no. 3, pp. 2758–2765, 2011.

[13] L. Hong and B. D. Davison, "Empirical study of topic modeling in twitter," in *Proceedings of the first workshop on social media analytics*. acm, 2010, pp. 80–88.

[14] T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake shakes twitter users: real-time event detection by social sensors," in *Proceedings of the 19th international conference on World wide web*. ACM, 2010, pp. 851–860.

[15] J. Gomide, A. Veloso, W. Meira Jr, V. Almeida, F. Benevenuto, F. Ferraz, and M. Teixeira, "Dengue surveillance based on a computational model of spatio-temporal locality of twitter," in *Proceedings of the 3rd International Web Science Conference*. ACM, 2011, p. 3.

[16] S. Vieweg, "Microblogged contributions to the emergency arena: Discovery, interpretation and implications," *Computer Supported Collaborative Work*, pp. 515–516, 2010.

[17] S. Vieweg, A. L. Hughes, K. Starbird, and L. Palen, "Microblogging during two natural hazards events: what twitter may contribute to situational awareness," in *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 2010, pp. 1079–1088.

[18] B. De Longueville, R. S. Smith, and G. Luraschi, "Omg, from here, i can see the flames!: a use case of mining location based social networks to acquire spatio-temporal data on forest fires," in *Proceedings of the 2009 international workshop on location based social networks*. ACM, 2009, pp. 73–80.

[19] A. L. Hughes and L. Palen, "Twitter adoption and use in mass convergence and emergency events," *International Journal of Emergency Management*, vol. 6, no. 3-4, pp. 248–260, 2009.

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

[21] N. Colneriĉ and J. Demsar, "Emotion recognition on twitter: Comparative study and training a unison model," *IEEE Transactions on Affective Computing*, 2018.

[22] P. Ekman, "An argument for basic emotions," *Cognition & emotion*, vol. 6, no. 3-4, pp. 169–200, 1992.

[23] A. Gruzd, "Emotions in the twitterverse and implications for user interface design," *AIS Transactions on Human-Computer Interaction*, vol. 5, no. 1, pp. 42–56, 2013.

[24] S. Stieglitz and L. Dang-Xuan, "Emotions and information diffusion in social mediasentiment of microblogs and sharing behavior," *Journal of management information systems*, vol. 29, no. 4, pp. 217–248, 2013.

[25] A. Gruzd, S. Doiron, and P. Mai, "Is happiness contagious online? a case of twitter and the 2010 winter olympics," in *System Sciences (HICSS), 2011 44th Hawaii International Conference on*. IEEE, 2011, pp. 1–9.

[26] A. D. Kramer, J. E. Guillory, and J. T. Hancock, "Experimental evidence of massive-scale emotional contagion through social networks," *Proceedings of the National Academy of Sciences*, p. 201320040, 2014.

[27] S. Asur and B. A. Huberman, "Predicting the future with social media," in *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on*, vol. 1, no. 2. IEEE, 2010, pp. 492–499.

[28] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *Journal of Computational Science*, vol. 2, no. 1, pp. 1–8, 2011.

[29] A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welpe, "Predicting elections with twitter: What 140 characters reveal about political sentiment." *ICWSM*, vol. 10, pp. 178–185, 2010.

[30] F. Pla and L.-F. Hurtado, "Political tendency identification in twitter using sentiment analysis techniques," in *Proceedings of COLING 2014, the 25th international conference on computational linguistics: Technical Papers*, 2014, pp. 183–192.

[31] T. Elghazaly, A. Mahmoud, and H. A. Hefny, "Political sentiment analysis using twitter data," in *Proceedings of the International Conference on Internet of things and Cloud Computing*. ACM, 2016, p. 11.

[32] J. R. Fownes, C. Yu, and D. B. Margolin, "Twitter and climate change," *Sociology Compass*, vol. 12, no. 6, p. e12587, 2018.

[33] A. P. Kirilenko, T. Molodtsova, and S. O. Stepchenkova, "People as sensors: Mass media and local temperature influence climate change discussion on twitter," *Global Environmental Change*, vol. 30, pp. 92–100, 2015.

[34] S. Abbar, T. Zanouda, L. Berti-Equille, and J. Borge-Holthoefer, "Using twitter to understand public interest in climate change: The case of qatar," in *Tenth International AAAI Conference on Web and Social Media*, 2016.

[35] A. P. Kirilenko and S. O. Stepchenkova, "Public microblogging on climate change: One year of twitter worldwide," *Global environmental change*, vol. 26, pp. 171–182, 2014.

[36] E. M. Cody, A. J. Reagan, L. Mitchell, P. S. Dodds, and C. M. Danforth, "Climate change sentiment on twitter: An unsolicited public opinion poll," *PloS one*, vol. 10, no. 8, p. e0136092, 2015.

[37] X. An, A. R. Ganguly, Y. Fang, S. B. Scyphers, A. M. Hunter, and J. G. Dy, "Tracking climate change opinions from twitter data," in *Workshop on Data Science for Social Good*, 2014.

[38] K. Holmberg and I. Hellsten, "Gender differences in the climate change communication on twitter," *Internet Research*, vol. 25, no. 5, pp. 811–828, 2015.

[39] L. Steven, "Textblob: simplified text processing," *Secondary TextBlob: Simplified Text Processing*, 2014.