



Perceptions of GHG emissions and renewable energy sources in Europe, Australia and the USA

Yaming Zhang^{1,2,3} · Majed Abbas¹ · Wasim Iqbal⁴

Received: 24 April 2021 / Accepted: 8 August 2021 / Published online: 25 August 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

People's sentiments and perceptions of greenhouse gas emission and renewable energy are important information to understand their reaction to the planned mitigation policy. Therefore, this research analyzes people's perceptions of greenhouse gas emissions and their preferences for renewable energy resources using a sample of Twitter data. We first identify themes of discussion using semantic text similarity and network analysis. Next, we measure people's interest in renewable energy resources based on the mentioned rate in Twitter and search interest in Google trends. Then, we measure people's sentiment toward these resources and compare the interest with sentiments to identify opportunities for policy improvement. The results indicate a minor influence of governmental assemblies on Twitter discourses compared to a very high influence of two renewable energy providers amounts to more than 40% of the tweeting activities related to renewable energy. The search interest analysis shows a slight shift in people's interest in favor of renewable energy. The interest in geothermal energy is decreasing while interest in biomass energy is increasing. The sentiment analysis shows that biomass energy has the highest positive sentiments while solar and wind energy have higher interest. Solar and wind energy are found to be the two most promising sources for the future energy transition. Our study implies that governments should practice a higher influence on promoting awareness of the environment and converging between people's interests and feasible energy solutions. We also advocate Twitter as a source for collecting real-time data about social preferences for environmental policy input.

Keywords Twitter · Sentiment analysis · Topic modeling · GHG emission · Renewable energy

Introduction

Greenhouse gas emissions and consequent climate change have become a major issue of concern. Therefore, social engagement in the public discourse has risen considerably through academic articles, scientific research, or informal social media platforms (Zhang et al. 2020). The success in the implementation of GHG and climate change mitigation policies depends on public involvement and willingness to adapt their energy consumption behaviors (Keramitsoglou 2016). People as significant contributors to energy consumption can influence energy policy by sharing their opinions and views regarding the unfair distribution of renewable energy development burdens and benefits (Oluoch

et al. 2020). Successful energy and GHG mitigation policies entail inclusive people input to improve their institutional and regulatory environment (Omengue et al. 2019; Yang et al. 2020). Therefore, studies on people perceptions and preferences of renewable energy help policy-makers to design sustainable renewable energy policies (Oluoch et al. 2020)

Many previous studies have analyzed preferences, acceptance and willingness to participate in energy transition covering various dimensions and determinants. Among these determinants, Lee et al. (2020) highlight the impact of change in energy policy such as electricity tariff spikes in South Korea. The study of Fischer et al. (2021) on the other hand focus on social values such as Patience, altruism, reciprocity in addition to economic incentives and environmental values. While the study of Kim et al. (2018) stress the education level of consumer. Energy sources were also identified as an important dimension; Bengart and Vogt (2021), for example, highlight the importance of disclosing renewable energy sources information in the fuel mix to improve people's willingness to consume more renewable energy. However, only a few studies have discussed people's

Responsible Editor: Baojing Gu

✉ Majed Abbas
abbasmajed@gmail.com

Extended author information available on the last page of the article.

preferences of different energy sources (Bakkensen and Schuler 2020; Yu 2021) which find out that citizens in Vietnam and China prefer coal over renewable resources. On the other hand, Oluoch et al. (2021) found that people in Kenya prefer solar energy then wind and biomass; similarly, Peng et al. (2021) claimed that the interest in wind energy is growing in the recent decades worldwide.

Researches on GHG and renewable energy policies are normally conducted using statistical data gathered by local administration councils. These statistics are organized and well structured, and utilized; nonetheless, they have some flaws like delay, inadequate coverage of the subject, and the top-down process of data creation (Pfeffermann 2015). Statistics on preferences of renewable energy are insufficient. In Europe, there are only two surveys (Eurobarometer and the Social Survey). These surveys do not have enough questions related to renewable energy preferences. In addition, they are not conducted every year; thus, data related to renewable energy preferences and GHG emission are not current. Furthermore, public administrators' activities and decisions should not be performed based on a top-down approach, their actions and decisions should develop from interactions with people (Margetts and Dunleavy 2013).

Social media has recently become an important platform for people and organizations since it became a cyber market where people perceptions, ideas, and experiences influence others' awareness (Hee et al. 2012). Thus, social media analytics is an important area of research that involves the development of information applications and frameworks to collect, model, analyze and present data to obtain useful data patterns that serve for policy input and decision-making (Fan and Gordon 2014). The sheer volume of data posted by social media users provides opportunities for measuring satisfaction and enhance services to meet people requirements (Zikopoulos et al. 2012). Among social media platforms, Twitter has become a popular medium to share information, publicize perceptions, and attitudes (Ibrahim and Wang 2019).

There are limited studies that discuss how to derive insight into social preferences, awareness, and attitudes toward GHG emissions and renewable energy consumption, and the associated climate change from Twitter analytics. The yellow vest protest in France has proven the lack of information about social perception regarding the issues at hand, and point out the need for effective data collection and handling in real time to understand people's attitudes toward energy policies and update these policies accordingly. Therefore, there is an immense need to develop an updated comprehensive and comparable social indicators when studying public issues such as GHG emissions and renewable energy policy (Zhao et al. 2020).

In this article, we prove that Twitter can be used as a source for real-time data related to renewable energy that

can serve to develop social indicators for energy policy inputs. We generate social information related to GHG emissions and renewable energy concerns from Twitter. We propose a method to analyze social perceptions and interests related to renewable energy consumption to provide valuable inputs for renewable energy policy and discover correction opportunities. Our method includes a new tweet classification algorithm based on network theory and text document semantic similarity. In addition, we also develop an indicator of people satisfaction and an indicator of energy source importance. People satisfaction is measured by measuring their sentiments toward a specific energy source while importance, on the other hand, is measured by its mention rate. Eventually, we assign each topic a score of opportunity by implementing an opportunity algorithm on the satisfaction and importance values. This allows policy-makers to lead policy development focusing on dimensions that have higher correction opportunity potential.

This research has three major contributions. Firstly, We generate up-to-date information about people's perceptions and preferences of renewable energy. This information is valuable for renewable energy policy inputs. Secondly, our methodology can estimate the likelihood of renewable energy policy improvement. This enables policy-makers to identify important development directions based on people's preferences and attitudes. Thirdly, our method can serve as a real-time tool for assessing people's attitudes and a base for developing intelligent policy planning applications. The organization of this document goes as follows. The following section presents a literature review and defines important concepts. "Data and methodology" explains the methodology and "Result and discussion" presents the findings and implications of the study, the last section draws the conclusion

Literature review

In this section, we provide a comprehensive review of the literature related to social media analysis of environmental issues. We further provide a theoretical background on the concept related to the methodology specifically, Topic modeling to extract the topics of discussion and sentiment analysis as a tool to measure people's attitude.

Twitter as a medium for environmental studies

Social media analysis for environmental studies is gaining prominence over time. Though limited works have analyzed on-line people's perceptions of greenhouse gas emissions, considerable researches have studied people's awareness of the associated climate change. The work of Kirilenko et al. (2015) proved the correlation between temperature

abnormalities and tweeting activities. This was taken one step further by Kryvasheyeu et al. (2016) who tested synchronization between activity on Twitter and environmental disasters in the USA. They show that environmental threats are directly observable through the intensity of Twitter streams. Similarly, the work of Sisco et al. (2017) used Twitter data to analyze short-term awareness of climate change in relation to extreme environmental events. They found that people's attention correlates with the financial loss of climate events and the degree of event abnormality.

The role of Twitter influencers in shaping people awareness of climate change was studied by Kirilenko and Stepchenkova (2014) who analyzed major climate discourses on Twitter and their geographies. The study of Cody et al. (2015) also pointed out the potential of social media as a channel for spreading climate change awareness by studying on-line people emotions related to environmental events. Similarly, Maynard and Bontcheva (2015) analyzed public engagement in climate change and Holmberg and Hellsten (2015) studied the gender differences in Twitter discourses related to climate change.

The dynamics of grouping and forming communities on Twitter was studied by Pearce et al. (2014) who analyzed the discussions and communities on Twitter related to an intergovernmental report on climate change. The finding suggests the tendency of users to form communities of users with similar views. The study also pointed out the geographical impact on topics of discussion and community formation. On the other hand, Jacques and Knox (2016), assessed the reasons for people denial of climate change on Twitter discourses. They found that mistrust in government, opposition to energy taxation, and rejecting climate science are the main reasons for denial of climate change.

Apart from climate change, energy consumption was explored by Holmberg and Hellsten (2015) who analyzed Twitter discussions relevant to energy to understand energy consumption behaviors over time and the influence of social media discussion over these behaviors, they found that social media can be a tool to test society awareness. Similarly, attitude and sentiments toward carbon taxation were analyzed by Zhang et al. (2021). These studies pave the way to utilize Twitter data for more researches on people's perceptions and preferences related to GHG emission and renewable energy policy. Thereby, overcoming the limitation of conventional methods of data collection. We particularly address the query of how to measure interest and satisfaction of energy sources using Twitter data and investigate opportunities for renewable energy policy improvement. Our work complements the previous studies by developing an indicator of online public perceptions and sentiments related to greenhouse gas emission and renewable energy consumption behaviors relevant to

environmental policy and introduce a method to discover corrective opportunities of this policy.

Topic modeling

Topic modeling refers to the process of discovering hidden topics of text documents. The basic method for topic modeling is the TF-IDF approach (Salton and McGill 1983), which calculates the frequencies of occurrence of each word in each document, then normalizes these frequencies and saves them into a term-document matrix. This method reduces lengthy text documents to the set of unique words; however, this approach is not efficient for a large number of documents that have a large number of unique words. To overcome this weakness Latent semantic analysis LSA approach was proposed by Deerwester et al. (1990). LSA utilizes singular value decomposition to reduce term-document matrix dimensionality while capturing the textual aspect of the text. However, this method lacks the ability to model documents into multiple topics. LSA was improved by Hofmann in 1999 when he proposed PLSA. Despite its ability to assign multiple topics, PLSA is prone to overfitting and can not be a generative model for new documents that it was not previously estimated on. Hence, latent Dirichlet allocation (LDA) was developed by Blei et al. (2003). The LDA does not need any previous training rather the topics are learned directly from original texts and it is a good generative model. However, the accuracy of LDA depends on the variability of subjects in the text document. Despite being used on a variety of text documents like email (Blei 2012) and newspaper archives (Wei and Croft 2006). LDA is not a good model when all the documents have the same topic. In addition, it is not a good choice for short text documents such as tweets. We use a novel approach based on text semantic similarity proposed by Mihalcea et al. (2006) and network theory.

Sentiment analysis

Text can be classified into objective texts such as facts, entities and events, and subjective text related to people's feelings such as opinion (Liu 2010). Sentiment analysis concerns with text expressing opinions. Nasukawa defines sentiment analysis as a big data analytics method that serves to identify the polarity of sentiments in expressions or judgments made by consumers (Thelwall 2018). This technique focuses on extracting the negative or positive opinions in a text and the intensity of the sentiment in the text. Sentiments and their intensity can be measured using three main methods: lexicon, text classification, and deep learning method. The lexicon-based methods are usually conducted by using a set of dictionaries that list all sentimental

words and map them to specific values (Baccianella et al. 2010). Based on the words' sentiments, it recognizes the sentiments of a document (Turney 2002). The text classification method recognizes document sentiments based on supervised machine learning techniques such as support vector machine, naive bias, and maximum entropy using labeled examples of text documents (Pang et al. 2002; Taboada et al. 2011). Finally, the deep learning method uses a deep neural network model (Santos and dos Gatti 2014). For our study we use a lexicon-based sentiment analysis.

Data and methodology

Our method is chosen based on the assumption that people share their beliefs, opinions, and perspectives on social media (Fig. 1). These opinions and beliefs are revealed in short expressions, comprising words, that express their interests and attitudes to concepts, businesses, and activities. In our study, the words and terms symbolize people's interests in renewable energy sources and their perception of GHG emissions. Therefore, we first collect two samples of tweets related to renewable energy and GHG emission and eight samples related to renewable energy sources as shown in Table 2. We identify topics of discussion related to renewable energy and GHG emissions. We then measure people's sentiments toward each energy source and measure the importance of energy sources based on each source mention rate. Based on importance and sentiments we measure opportunities for renewable energy policy development. This section provides a detailed explanation of our methodology.

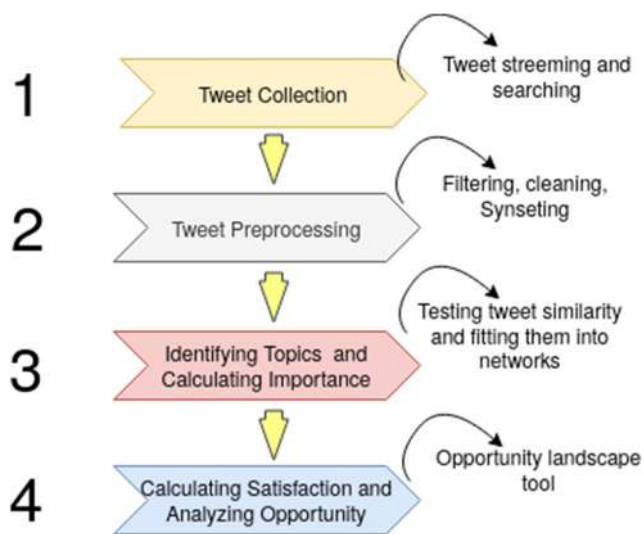


Fig. 1 Methodology

Tweet collection

Twitter provides an application programming interface to enable access to tweets posted by users through searching and streaming. Through searching, the API can obtain the tweets published at a specific time by defining the start and end search dates so that the tweets are searched within one week before the search time. Users can search for tweets based on term appearance or based on hashtags. The search criteria can be limited to capture tweets with certain attributes (such as tweet language). Searching also allows retrieving specific fields of a tweet instead of retrieving the entire tweet. Streaming on the other hand allows data analysts to receive real-time tweets posted by users. Analysts can filter streams to receive tweets that belong to specific topics. The stream has limited filtering capabilities, so users receive the complete tweet structure. We found that searching tweets is faster and more efficient because we can limit data retrieval to our needs. It is also more flexible because it can be done at any time. On the other hand, streaming media requires more storage space, more filtering and cleaning activities, and permanent connections.

We collected the samples of tweets from Twitter during a period of 3 months from September, 1st 2020 to November, 30th 2020. The period of sampling was selected to evade any severe climate circumstances that could influence people's opinions. Many earlier articles have examined the effect of climate experiences on tweeting. The work of Sisco et al. (2017) points out that people's care to climate change correlates with the strength of a weather abnormality. Similarly, Kirilenko et al. (2015) show that tweeting correlates with temperature anomalies. Though, no study has examined how the acceptance of environmental policies vary during severe environmental conditions. We expect more supportive public attitudes and perceptions during extreme weather conditions. Severe environmental conditions raise short-term people's awareness of environmental issues (Sisco et al. 2017; Kirilenko et al. 2015). It is further presumed that the pandemic of coronavirus has turned people's attention to public health. This may somewhat decrease people's tweets related to the environment.

We searched tweets for the following terms “#greenhouse gas”, “#GHG” and “#renewable energy”, “#coal”, “#natural gas”, “#solar energy”, “#wind energy”, “#biomass”, “#hydro energy”, “#geothermal energy”, “#tidal energy”. The collected data contained seven fields of the tweet structure. Table 1 explains each field. We applied search restrictions to extract tweets written in English only. Eventually, we performed our analysis on 10 distinct Twitter samples. Two samples (greenhouse gas and renewable energy) were

Table 1 Collected fields of the tweet structure

Field name	Explanation
id	Unique number that identifies each tweet.
text	Contains the text of the tweet
favorited_count	The number of Twitter users who liked this post
screen_name	Unique identification code for the user who posted the tweet
created_at	The date the tweet was posted
retweet_count	The number of users who retweeted this tweet
user.location	user location

used to analyze people's perceptions of greenhouse gas and renewable energy. The other 8 samples were used to analyze people's interest in renewable energy resources and their sentiment toward these different resources. Coal and natural gas were included to compare people's interests and sentiments in renewable and nonrenewable energy resources. Table 2 shows the number of tweets in each sample.

Tweet preprocessing

In this step, we use the popular regular expression programming tool to clean up the text. Regular expressions are a very effective method for text matching. They are a series of characters and special characters that define the search pattern, so they can be effectively used for text removal. First using the regular expression library in python, we removed all non-alphabetic characters, including tags, URLs, and punctuation marks. The order of spaces is also removed and replaced with a single space. After deleting non-alphabetic characters, we use the splitter library in python to split compound words Spell checker library to correct all spelling errors. Tweets contain up to 140 characters, so many abbreviations and acronyms are used. Tweets are also informal short essays, which contain many slang words. Therefore, the commonly used words by Twitter users are collected and added to the dictionary. The dictionary lists each abbreviation or word and its corresponding formal complete term. A Python function is written to search for these abbreviations and replace them with corresponding terms. New terms can be easily added to the dictionary without affecting functional efficiency. This provides users with greater flexibility and can add specific terms for different situations.

After clearing the text, words in each tweet are tokenized. Word tokenization is the process of dividing a character string into multiple word tokens, which can be further processed in the generated tokens. Part of speech tagging is the process of identifying the grammatical categories of words as nouns, verbs, adjectives, or adverbs based on word definitions and context. After tweets are cleaned, all tweets are word-tokenized and tagged; then, synsets are

created. Synsets are used when testing semantic similarity in the next step instead of normal words. While testing tweets' similarities, synsets in each tweet can be restricted to specific parts of speech, such as verbs, adjectives, nouns, and adverbs. This improves the efficiency and effectiveness of similarity testing and gives analysts more flexibility. Different discussion topics have different unique words, so the optimal settings are different in different situations. Finally, stop words are deleted. Stop words are common words that often appear in the text, such as the word "the". Removing stop words improves similarity testing between two documents, this is essential for finding unique words in the text. The NLTK python library has a list of English stop words that can be extended using the built-in extensions. Depending on the environment in which the algorithm is applied, expanding the stop word list can improve similarity testing and topic modeling.

Tweets that have less than 80 characters were deleted to avoid any inaccurate similarity values that could result from very short sentences. We coded algorithm 1 in a python function and used it to test similarities between the collected tweets. The function returns a float number between 0 and 1. 1 represents an identical pair of tweets and 0 represents a completely different tweet. For each identical pair, one tweet was removed. When the similarity between two tweets is greater than five the representing nodes of these tweets are connected.

Identifying topics and measuring their importance

In this step, we measure similarities between tweets to classify them based on the similarity value. There are many methods for measuring text similarity. Generally, they can be divided into three categories. The string-based similarity uses the vector space model to store the number of occurrences of each word in a text. The similarity between the two texts is measured by the angle formed by their two representing vectors. The corpus-based similarity is based on knowledge gained from the corpus. A corpus is a collection of text or speech used for language searches. The knowledge-based similarity is a kind of semantic similarity that measures the relationship between words based on a semantic network (Gomma and Fahmy 2013). Semantic similarity is more suitable for measuring similarity between short texts because the probability of words appearing repeatedly in short texts is low. The semantic similarity methods reduce the error rate of the traditional vector-based model by 13% (Mihalcea et al. 2006). Tweets are very short collections of words that are limited to 140 characters so semantic similarity tests provide better results than vector-based similarity tests. Wordnet corpus is used to test semantic similarity. Wordnet is a very popular semantic network that contains a large database of cognitive synonyms

Table 2 Number of tweets in each sample

GHG	Renewable energy	Coal	Solar	Gas	Wind	Biomass	Hydro	Geothermal	Tidal
12000	72000	48000	21600	19200	14400	6000	4800	2400	1200

(synsets) (Mihalcea et al. 2006). Synsets are groups of words that express a specific concept and are linked by semantic and lexical relationships. Wordnet enables testing semantic similarity between two words by measuring the shortest path between them. The wordnet path similarity function in python measures how similar are two synsets then returns a value between 0 and 1 based on the shortest path between these two synsets. Mihalcea et al. proposed an algorithm for calculating semantic similarity between two short texts based on semantic similarity between words (Mihalcea et al. 2006). Equation 1 illustrates this method. For two text documents T_1 and T_2 First for each word w in T_1 find the word in T_2 that has the maximum word semantic similarity $\max sim(w, T_2)$. The same process is repeated for each word in T_2 . Then word similarities are weighted with the corresponding word specificity $idf(w)$ and then summed up and normalized to the number of words in each document. The resulting similarity scores are averaged together using simple average. Similar tweets discuss the same topic so we group tweets based on their similarity using a network. Algorithm 1 is a detailed process for document semantic similarity testing.

$$sim(T_1, T_2) = \frac{1}{2} * \left(\frac{\sum_{w \in T_1} \max sim(w, T_2) * idf(w)}{\sum idf(w)} + \frac{\sum_{w \in T_2} \max sim(w, T_1) * idf(w)}{\sum idf(w)} \right) \quad (1)$$

Algorithm 1 Testing tweet similarity.

```

function document_path_similarity(tweet1, tweet2)
  list1 = []
  #For each synset in tweet1
  for a in tweet1 do
    # find the most similar synset in tweet2 to each
    synset in tweet2 and append its similarity to list1
    list1.append(max([i.path_similarity(a) for i in
    tweet2 if i.path_similarity(a) is not None], default < - 0))
  end for
  #return the mean of list1
  return np.mean(list1)
end function

```

To form a network of tweets, we treat each tweet as a node and the similarity value between two tweets as an edge. We construct the distance matrix of the tweets based on semantic similarity. Then initialize the network between

tweets based on the tweet similarity distance matrix. If the similarity is greater than a specified value, an edge between the two nodes is created with a weight equal to the similarity value. The value of similarity ranges from 0 (completely different tweets) to 1 (identical tweets). A higher similarity edge results in more accurate detailed tweet classification however; it reduces the clustering coefficient resulting in a higher number of components which leads to a larger categorization of topics. Lower value instead reduces accuracy and improves the clustering coefficient.

A community in the network is a group of highly connected nodes. As the process continues the number and size of communities will increase. The degree of a node in the weighted network is the sum weight of each edge connecting it. When the whole process ends, a collection of communities of various sizes will be generated. Each community represents a group of connected nodes. Communities are identified using Louvain's algorithm as this algorithm is suitable for large networks of textual data. Unconnected nodes can be ignored. Firstly, there is a higher probability that these nodes are not relevant since tweets discussing major topics have some similarities and tend to aggregate. Besides, they can be substituted by increasing the number of tweets to be analyzed. This improves the accuracy of the analysis by removing irrelevant tweets.

Each community represents a group of connected tweets with the same theme, we find the theme for each community by using word collocation and frequencies and reading the tweets with the highest degree. We then calculate the importance of each topic on a scale of 0 to 10 based on its mention rate. The following equation explains how to compute importance.

$$Importance_i = \frac{ND_i - ND_{min}}{ND_{max} - ND_{min}} \quad (2)$$

where ND is the number of the processed tweets, ND_i number of tweet discussing feature i , ND_{min} is the number of tweets discussing least mentioned issues ND_{max} is the number of tweets discussed the most mentioned feature.

Measuring people sentiments and analyzing opportunities

The sentiment toward each energy source is measured and a distribution graph of these sentiments is plotted. We focus on the distribution of sentiments rather than the mean value

because the attributes of the distribution reflect the variance of sentiments. After analyzing the sentiments, we calculate the mean of each distribution and save it in a list. Then, we compute the satisfaction of each source on a scale from 0 to 10 using the following formula.

$$satisfaction_i = 10 * \frac{sn_i - \min(sn)}{\max(sn) - \min(sn)} \tag{3}$$

where satisfaction_i represents the satisfaction value of energy source i, sn is the list of sentiments mean, and sn_i is the sentiment mean for source i.

After measuring the sentiment toward each energy source, the opportunity algorithm proposed by Ulwick (2005) is used to identify potential opportunities for policy improvements. The opportunity algorithm estimates the Fulfillment of customer needs to discover development opportunities based on important yet unmet needs. Therefore, opportunities can be discovered by identifying the most dissatisfied and most important customer needs. The algorithm is a significant tool to measure customer need based on two basic dimensions, importance, and satisfaction. The algorithm has been applied to many objectives including determining resources competitive advantage (Hinterhuber 2013), designing management structure based on customer satisfaction (Chung 2007), designing the best products based on maximum opportunistic customer need (Killen et al. 2005). According to the algorithm, if the importance of an energy source increases without a corresponding increase in satisfaction, the improvement opportunities increases. Based on this stand the opportunity algorithm assigns each source an opportunity score based on the importance and satisfaction. As the importance increases and satisfaction decreases the opportunity score gets higher. The algorithm uses the following formula to calculate opportunity scores for each source.

$$Opportunity_i = Importance_i + \max(Importance_i - Satisfaction_i, 0) \tag{4}$$

The opportunity landscape tool proposed by Ulwick (2005) helps visualize improvement opportunities and monitor resource allocation. The opportunity landscape is a plane composed of two coordinates, satisfaction, and importance. We display different energy sources in the plane according to their importance and satisfaction. The plane is then divided into three parts. The no-opportunity segment that contains energies with high satisfaction compared to importance, the low-opportunities segment, and the high-opportunity segment contains features that have high importance and low satisfaction.

Result and discussion

Major topics of the greenhouse gas sample

We found 9 communities that resemble 9 major topics of discourses. The ratio of the number of nodes in each community to the number of nodes in the network indicates the dominance of each topic. The word collocations for each community show that the discussed topics are: emission reduction (24%), global carbon dioxide rise (13%), greenhouse gas intensity (13%), climate crises(12%), energy production(10%), Intergovernmental Panel on Climate Change report(IPCC 9%) renewable energy (7%), carbon taxes (6%), the Japanese government new year resolution to face climate change(5%). Figure 2 shows the word collocation of the greenhouse gas sample. For brevity we deleted all nodes that have a degree less than 50 from the figure. It shows similar terms and associations to the terms in the detected communities.

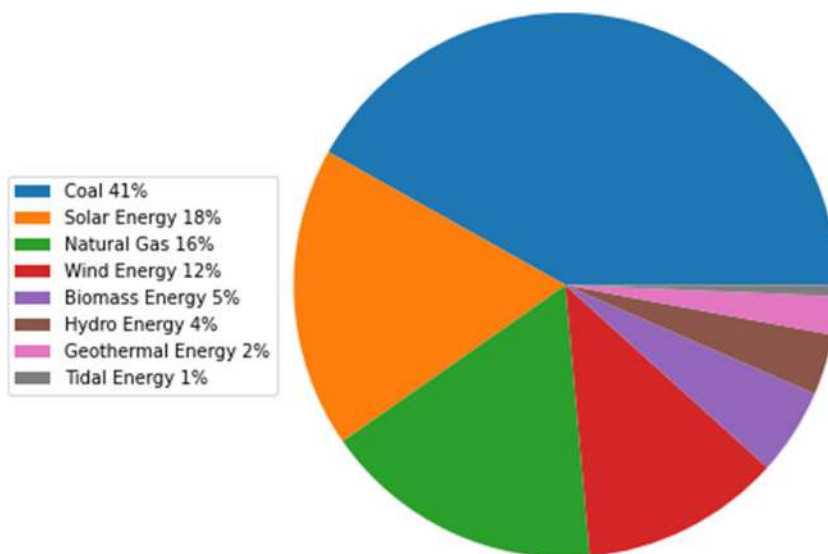
Major topics of the renewable energy sample

We have detected five communities in the renewable energy sample indicating five major themes of discourses. The percentage of the participating nodes in each community and word collocations shows that the topics discussed are Werusys energy management systems(35%), global sustainable energy transition(25%), energy consumption structure (19%), energy efficiency(13%) and solar energy(7%). It is worth noting the influence of renewable energy companies on Twitter discourses. Two major renewable energy providers Werusys and AMP are dominating more than 40 % of the discourses. Werusys dominate the first community and AMP partially appears in the discourses related to the energy consumption structure. Figure 3 shows the collocations of the top 50 frequent words in the sample and the subtopics of the five general topics.

People interest in renewable energy sources

Figure 4 shows the mention rate of each type of energy. According to the collected sample, the major two important sources of energy are coal and solar energy. Coal is the most important source mentioned in 41% of the tweet sample followed by solar energy in the second rank with an 18% mention rate. The other sources of energy have a much lower mention rate. Natural gas lies in the third rank with a 16% mention rate followed by wind energy with a 12% mention rate then, biomass and hydro energy with 5% and 4% mention rate respectively and lastly, geothermal and tidal energy with 2% and 1% mention rate.

Fig. 4 Energy sources importance based on tweet mention rates



We compare the results of our analysis with Google trends of the search for each kind of energy. Figure 5 shows the search interest of different sources of energy compared to the interest of coal between 2004 to 2018. In general, we can see in Fig. 5A a decreasing search interest for all kinds of energy sources between 2004 and 2018; however, the decrease in interest is most apparent in natural gas followed by coal. Relatively, the decreased interest of the other sources of energy is lower. This means that there is a slight shift in people’s interest from traditional energy sources to renewable sources despite the overall decreasing interest in energy over time. Similarly, Fig. 6 shows a slight decrease in the interest in Tidal, hydro and geothermal energy. However, the interest in biomass energy is stable and gaining more interest over time than the other sources. The current search trends show that biomass energy is gaining more interest than geothermal energy in 2020. This is demonstrated in Fig. 7C.

Figure 5A shows that the highest interest is in coal energy followed by natural gas with approximately two-thirds of the coal interest. Interest in solar energy comes in the third rank then wind energy in the fourth rank. Figure 6 shows the relative ranks of search interest between the rest of the energy resources. Until 2018 geothermal energy comes before biomass energy; however, this changes in 2020 as shown in Fig. 7C. So biomass energy becomes in the fifth position then geothermal energy in the sixth followed by tidal and hydro energy in the last rank with the lowest search interest.

Comparing the mention rates in Twitter with search interests in Google trends we find one difference only. The mention rate of solar energy is much higher in Twitter, it is approximately equal to the mention rate of coal and it comes in the second rank while it comes in the third rank after natural gas in Google trends. The difference can

be attributed to the sampling areas. Our Twitter sample is collected from Europe, the USA, and Australia and it is limited to tweets in English while Google trends are taken worldwide. To overcome this limitation, we have made a comparison of these interests based on the Google trends tool. This enriches our understanding of the difference in interest in renewable energy resources between different areas.

Figure 7 shows the trend of interest in renewable energy across different areas. Figure 7A shows that solar energy is the most popular source in Europe, the USA and Australia except three European countries: Norway where wind energy is preferred, France and Poland where biomass energy is more popular. Figure 7B shows that wind energy is the second most popular source of energy in the USA and Australia while biomass energy is more popular in Europe. Specifically, biomass is popular in Sweden, the UK, France, Spain, Poland and Italy while wind energy is popular in Norway and Germany. Figure 7C shows the preference for biomass energy over hydro and geothermal energy worldwide. Figure 7D shows that solar energy is preferred over natural gas in Australia, Germany, and Ireland. Since these three countries are part of the area where our study is conducted, this explains why the mention rate of solar energy is higher in the Twitter sample than natural gas while it is the opposite in Google trends. Comparing interest in biomass energy and hydro energy in Fig. 7E and 7:F hydro energy is more popular in Australia, Norway, Germany and Austria while biomass is more popular in the USA and the rest of European countries.

People sentiment toward different energy sources

Using the sentiment python library, we measured the sentiments of tweets for each topic. The sentiment is calculated

Fig. 5 Interest of different energy source compared to coal based on Google trends between 2004 and 2018. **a:** interest in natural gas, solar, wind and hydro energy compared to interest in coal **b:** interest in tidal, geothermal and biomass energy compared to interest in coal

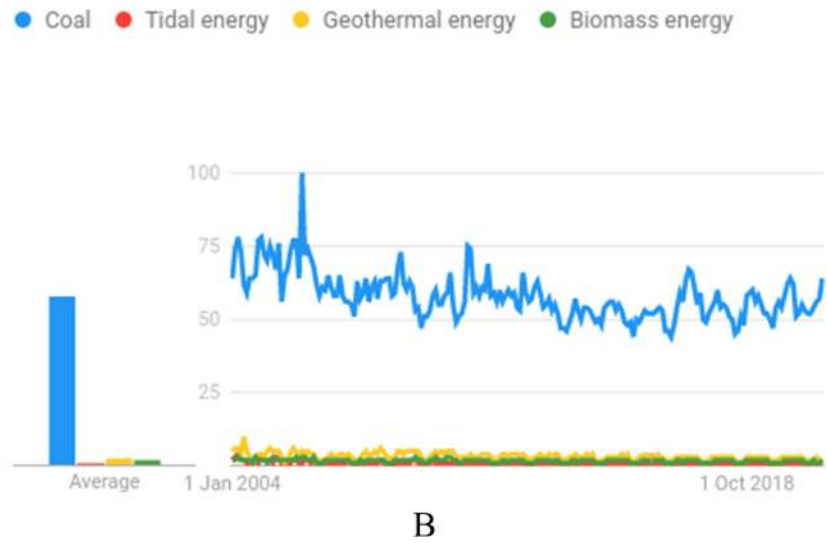
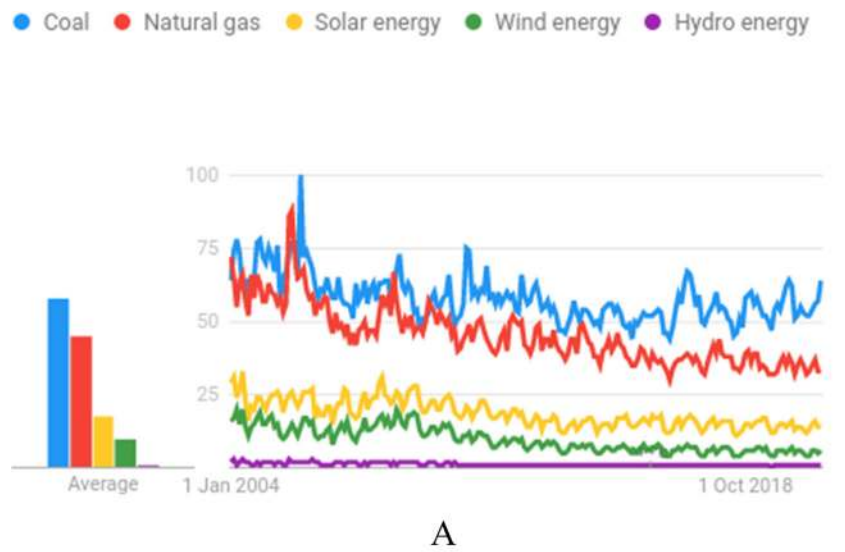
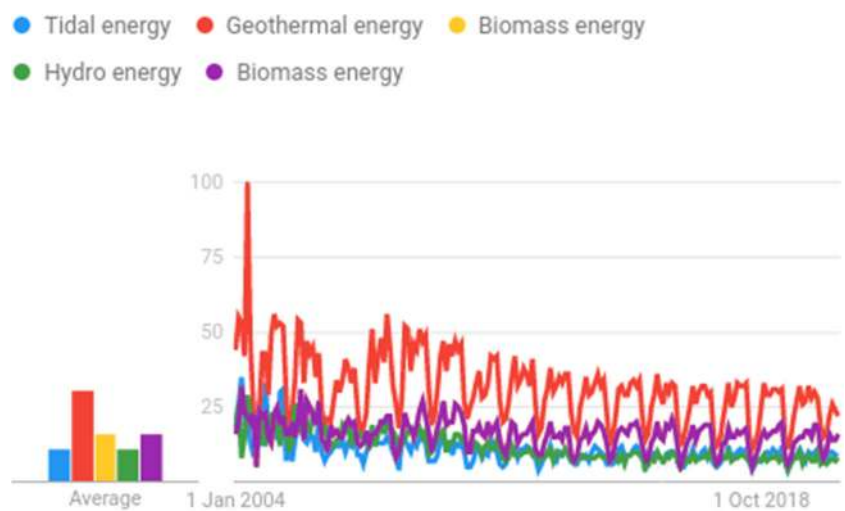


Fig. 6 Google trend of search interests in different renewable energy resources



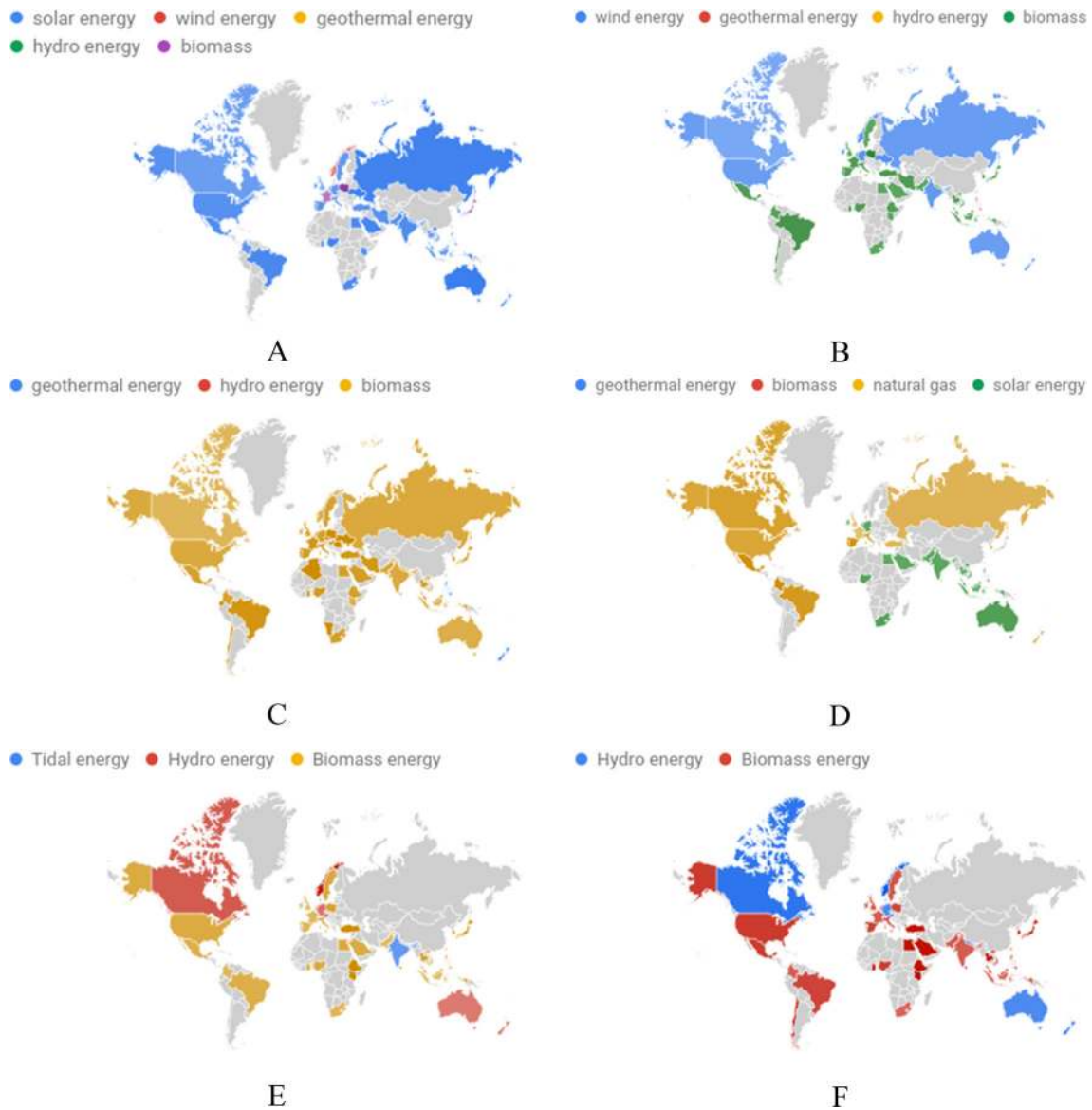


Fig. 7 Google trend of interest in renewable energy across different areas. **a:** Interest in solar, wind, geothermal, hydro and biomass energy sources **b:** Interest in wind, geothermal, hydro and biomass energy **c:**

Interest in geothermal, hydro and biomass. **d:** Interest in geothermal, biomass, natural gas and solar energy **e:** Interest in tidal, hydro and biomass energy **f:** Interest in hydro and biomass energy

in a range between -1 and 1 . Negative sentiment has a value greater than -1 while the positive one has a value of less than 1 . We adopt the moderate sentiment to refer to sentiment value less than 0.6 or greater than -0.6 and extreme sentiment to refer to sentiment value greater than 0.6 or less than -0.6 . Figure 8 shows the sentiment values distribution for different energy resources. The graph contains one plot for each source.

Figure 8A shows the sentiments distribution for solar energy, It appears that there are few negative sentiments toward solar energy. The majority have a positive to neutral sentiment, approximately 20% have moderately positive sentiments, and very few with extreme positive sentiments. The sentiment distribution toward wind energy in Fig. 8B is

similar to that of solar energy but slightly flatter with a little more negative, moderately positive sentiments.

On the other hand, hydro energy seems to be the most controversial source of renewable energy as it appears from Fig. 8C. The majority of people show moderately positive sentiments and few extreme positive sentiments; however, some people show moderate negative sentiments. Compared to all other sources of renewable energy, the distribution of hydro energy seems to have the broadest variety of sentiments.

In contrast to hydro energy, tidal energy has the narrowest distribution between all energy sources as Fig. 8D shows. Approximately all people are neutral toward tidal energy. Few people have moderately positive sentiments and few have negative sentiments.

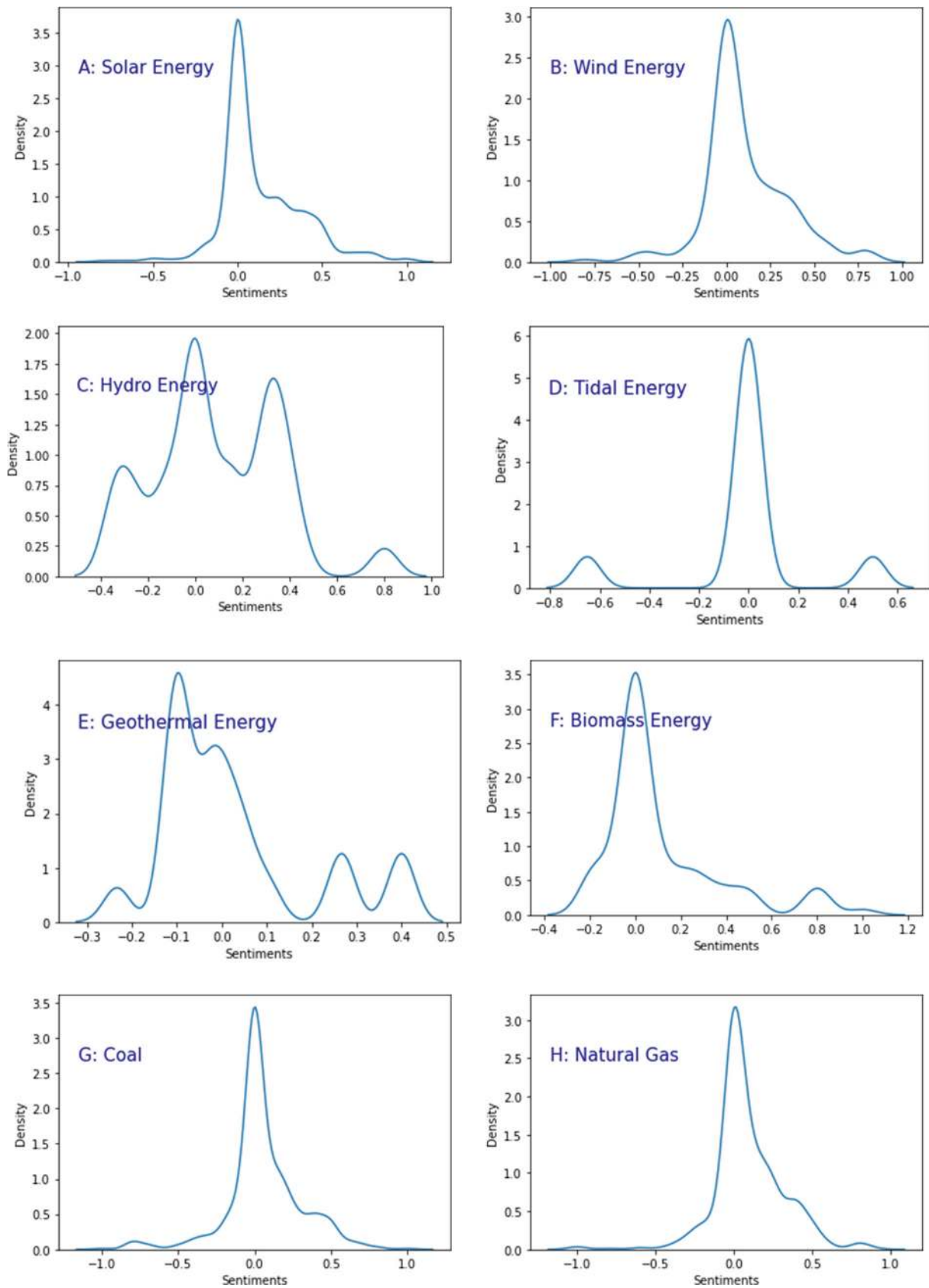


Fig. 8 Distribution of sentiment polarity toward energy. **A** Solar energy. **B** Wind energy. **C** Hydro energy. **D** Tidal energy. **E** Geothermal energy. **F** Biomass energy. **G** Coal. **H** Natural gas

The majority of people have slightly negative sentiments toward geothermal energy as Fig. 8E shows, nonetheless, some people have moderate positive sentiments. On the other hand, Fig. 8F shows that the majority are neutral toward biomass energy. Some have moderately positive sentiments and few have extreme positive sentiments. Finally, gas and coal have similar distribution as it appears in Fig. 8G and H; however, the distribution of natural gas is slightly shifted to the positive side so people are more positively neutral toward natural gas. People are mostly neutral toward these two sources with few positive sentiments. Overall, people are neutral toward tidal energy, While they show slightly positive sentiments toward solar, wind, and biomass energy. However, they show negative sentiment toward geothermal energy and mixed sentiments toward hydro energy.

The difference in people's preferences for energy resources in different countries was mentioned by Navratil et al. (2019) who documented people's origin as a determinant of their preference for a specific energy source. However, a deeper investigation suggests that these differences may originate from accumulated experiences and ideas that people were projected to in different geographical areas. We can summarize these factors based on previous research as follows. Firstly, satisfaction with the existing energy systems (Sovacool et al. 2021). This suggests that people lack information and experience of comfort that new energy technologies may provide (Schweiker et al. 2019). Secondly, resistance to changing energy technology. People concern more about their experience of comfort, cost and convenience than about what energy source brings this comfort (Mallaband and Lipson 2020; Zuo et al. 2020). So familiarity with a specific energy source increases people's preferences for this source. This explains the strong desire for natural gas in Europe since it is widely used there. Thirdly, the implementation of energy technology. A lower satisfaction or preference in a specific energy source may be due to a bad previous experience as a result of bad implementation rather than the technology itself (Magnusson 2016).

It is interesting to investigate how the differences in sentiments relate to the interests in energy resources studied in the previous section. In the following, we visualize interest (importance) and sentiments (satisfaction) as two dimensions in the opportunity landscape.

Table 3 shows sentiment mean, sentiment standard deviation, importance, satisfaction level, and opportunity score for each energy source. The sources are ordered according to their importance value. We use importance and satisfaction values to draw the opportunity landscape as shown in Fig. 9. Each point in Fig. 9 represents an energy source. The green dots represent the feature located in the low opportunity area which is the area that represents a reasonable satisfaction value compared to the importance. The red dots

indicate features located in the no-opportunity area that represent higher satisfaction compared to importance. Policy planners should shift some of the resources allocated to these features to the features represented by the blue dots. The blue dots feature are located in the high-opportunity area and require more attention from policy planner because they have greater opportunities to speed energy transition.

Practical implications

Our research implies rich policy implications. Firstly, the number of tweets collected every day during the sampling period reflected low tweeting activities related to GHG emissions (130 tweets) and renewable energy (900). This indicates that people have a low interest in environmental issues. Our results also indicate a very low influence of the governments on Twitter discourses compared to the influence of renewable energy providers (Werusys and AMP). This implies the need for higher participation of governmental bodies on Twitter for spreading awareness of GHG emissions and promoting the consumption of renewable energy resources.

Comparing different energies' opportunities, we found that solar energy is the most promising renewable energy source for energy transition since people have the highest interest in this source and lower satisfaction compared with other sources. In the second rank comes wind energy with high opportunities for improvements, while other resources have lower opportunities for improvement biomass comes in the third rank followed by hydro energy, geothermal energy, and lastly tidal energy. People's interests should be taken into consideration when planing energy transitions. Therefore, resources should be directed toward improvements of services and solutions of solar and wind energy. The technical and economic feasibility of renewable energy resources development should be always assessed in light of people's interests and preferences and a suitable marketing campaign should be designed to promote convergence between people's interests and energy production feasibility.

As our findings suggest that people's experiences and state of knowledge regarding specific energy sources or energy technology is a determinant of their interest in this source and as a consequence a determinant of the acceptance and higher participation in the energy transition. Therefore, public administrators should carefully plan and observe the implementation process of an energy system to maintain pleasant experiences for citizens. We have further noted that familiarity with energy sources is a prerequisite for raising interest in this source; thus, people should be familiarized with new promising renewable energy solutions.

Finally, in addition to familiarity, resistance to change and environmental literacy, people tend to value comfort

Table 3 Energy sources statistics

Energy source	Sentiment mean	Sentiments Std	Importance	Satisfaction	Opportunity
Solar energy	0.09	0.22	10.00	5.90	14.27
Wind energy	0.10	0.23	8.55	6.82	10.27
Biomass	0.11	0.25	5.72	7.22	5.72
Hydro energy	0.08	0.26	2.71	5.69	2.71
Geothermal energy	0.02	0.16	1.48	2.20	1.48
Tidal energy	-0.01	0.25	0.43	0	0.86

and convenience provided by nonrenewable energy sources. This raises the challenge to innovate renewable energy solutions that meet consumer demand for thermal luxury. Rather than fulfilling minimum needs and resting on changing people's consumption behavior, policy-makers should consider people's satisfaction when they plan for a sustainable energy transition. Mandating an energy transition toward non-fossil energy systems may lead to less satisfaction if the available energy solutions do not provide a similar level of thermal comfort to that of the fossil energy solutions. Even when this transition is fair and governed perfectly, people satisfaction should be highly accounted for to achieve sustainable energy transition

Study limitation

We acknowledge two limitations of the study. The first is related to the methodology, while the second is related to the treatment of data. Concerning the methodology limitation, we admit that the analysis does not account for context-specific language. Furthermore, we only focus on the text, excluding symbols that may provide extra

related information. Future studies should use deep learning techniques so that these shortcomings are addressed.

Concerning data collection, the low diffusion rate of the Internet in some areas causes a notable selection bias. This is higher among low-income, old people and low-educated segments. This point is important especially in Southern Europe compared to Northern Europe where the Internet use rate is higher in aged segments. Second, a bias may be present in the data collected from social media because only users inclined to share their views contribute to tweeting activity, whereas people whose views are doubtful may not be active. Further, important details associated with the social and demographic attributes of participants are missing. This leads to decreased opportunities to create profiles of responses. Last, we may consider other ethical issues that can appear in different forms.

Conclusion

The objective of this article is to investigate Twitter discourses related to GHG emission and renewable energy. We base our analysis on a sample of tweets collected from Europe, the USA, and Australia. Specifically, we concentrate on evaluating people's perceptions of GHG emissions and preferences for renewable energy resources. Firstly we employed a new method for tweet topic modeling based on text semantic similarity and network theory then, we measure people's interest in renewable energy resources based on the mentioned rate in Twitter and search interest in Google trends. Then we measure people's sentiment toward these resources and compare the interest with sentiments. Then, we used the opportunity algorithm to discover policy improvement opportunities.

The results indicate that the most discussed topic of greenhouse gas emissions is the rise of global GHG and carbon dioxide emissions, the climate crisis, and emission reduction policies. Besides we find an insignificant influence of governmental assemblies compared to a high influence of two renewable energy providers amounts to more than 40% of the tweeting activities related to renewable energy. The topics related to renewable energy include

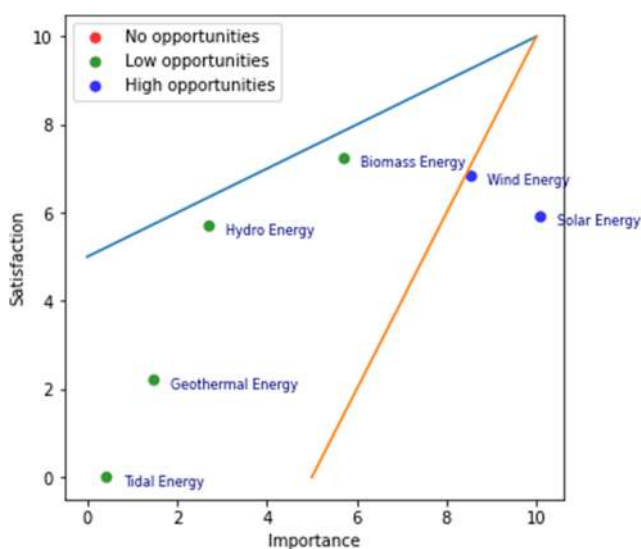


Fig. 9 Visualization of renewable energy resources in the opportunity landscape

global energy transition, energy structure, and efficiency. We found an overall decrease in interest in energy between 2004 and 2018 and a slight shift in people's interest from non-renewable energy to renewable sources of energy. We also found a shift in interest from geothermal energy to biomass energy. The opportunity algorithm shows that biomass energy has the highest positive sentiments while solar and wind energy have higher interest. Solar and wind energy are found to be the two most promising sources for the future energy transition.

Our results are analogous to those of previous European surveys. Thus, the messages related to GHG emissions and preferences for renewable energies as essential factors to decrease emissions and promote the energy transition still predominates over other sources of data. We further show that interests obtained via Twitter correspond with the behaviors of Internet searches related to renewable energies. We think that policy dialogue should fully consider the perceptions of civilians to establish sustainable and acceptable solutions.

We believe that this study will contribute to addressing the limitation of the practical application of social media analysis. We advocate the use of social media to evaluate sentiment and emotion in real-time. We think that those sources of data can afford valuable, updated data, which may be beneficial during the lack of formal statistics. Various statistical organizations are running a road-map to integrate social media data sources into official statistics. We claim that social media platforms are good resources of information to complement other official surveys.

Author contribution Yaming Zhang: supervision. Majed Abbas: conceptualization, writing original draft, data curation and methodology. Wasim Iqbal: editing and review.

Funding Project funded by National Social Science Foundation of China (Grant No. 18BXW118), National Natural Science Foundation of China (Grant No. G72101227), Natural Science Foundation of Hebei (Grants Nos. G2020203008, G2020203003), S&T Program of Hebei (Grant No. 19456232D) Science and Technology Research Project of Hebei Provincial Department of Education (Grant Nos. ZD2020199, QN2020250), Research Project of Social Science Development in Hebei Province (Grant No. 20200202012), Special Project of New Coronary Pneumonia Epidemic Research of Yanshan University (Grant No. 20FYA01), Hebei Key Research Institut of Humanities and Social Sciences at Universities (No.JJ2010).

Code availability Source codes are available on request.

Declarations

Conflict of interest The authors declare that there is no conflict of interests regarding the publication of this paper.

References

- Baccianella S., Esuli A., Sebastiani F. (2010) SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In: Proceedings of the seventh international conference on language resources and evaluation (LREC'10). Valletta, Malta, European Language Resources Association (ELRA). <http://www.lrec-conf.org/proceedings/lrec2010/pdf/769.Paper.pdf>
- Bakkensen L., Schuler P. (2020) A preference for power: Willingness to pay for energy reliability versus fuel type in Vietnam. *Energy Policy* 144:111696
- Bengart P., Vogt B. (2021) Fuel mix disclosure in Germany The effect of more transparent information on consumer preferences for renewable energy. *Energy Policy* 150:112120
- Blei D. M. (2012) Probabilistic Topic Models. *Commun. ACM* 55(4):77–84. <https://doi.org/10.1145/2133806.2133826>
- Blei D. M., Ng A. Y., Jordan M. I. (2003) Latent Dirichlet Allocation. *J. Mach. Learn. Res.* 3(null):993–1022
- Chung Y. C. (2007) Business Opportunity Algorithm for ISO 9001: 2000 customer satisfaction management structure. *Res J Bus Manag*: 1–10. <https://doi.org/10.3923/rjbm.2007.1.10>
- Cody EM, Reagan AJ, Mitchell L, Dodds PS, Danforth CM (2015) Climate change sentiment on Twitter: An unsolicited public opinion poll. *PLOS One* 10(8):e0136092. <https://doi.org/10.1371/journal.pone.0136092>
- Deerwester S., Dumais S. T., Furnas G. W., Landauer T. K., Harshman R. (1990) Indexing by latent semantic analysis. *J. Am. Soc. Inf. Sci.* 41(6):391–407. [https://doi.org/10.1002/\(sici\)1097-4571\(199009\)41:6<391::aid-asi1>3.0.co;2-9](https://doi.org/10.1002/(sici)1097-4571(199009)41:6<391::aid-asi1>3.0.co;2-9)
- Fan W, Gordon MD (2014) The Power of Social Media Analytics How to use, and influence, consumer social communications to improve business performance, reputation, and profit. *Commun ACM* 57(6). www.alexacom
- Fischer B., Gutsche G., Wetzels H. (2021) Who wants to get involved? Determining citizen willingness to participate in German renewable energy cooperatives. *Energy Research and Social Science* 76:102013
- Gomma W. H., Fahmy A. A. (2013) A Survey of Text Similarity Approaches. *International Journal of Computer Applications* 68(13):13–18
- Hee D. L., Lee A., Law R. (2012) Examining hotel managers' acceptance of web 2.0 in website development: A case study of hotels in Hong Kong. *Social media in travel, tourism and hospitality: Theory, practice and cases*, pp 53–65
- Hinterhuber A. (2013) Can competitive advantage be predicted? *Manag. Decis.* 51(4):795–812. <https://doi.org/10.1108/00251741311326572>
- Holmberg K., Hellsten I. (2015) Gender differences in the climate change communication on Twitter. *Internet Research* 25(5):811–828. <https://doi.org/10.1108/intr-07-2014-0179>
- Ibrahim N. F., Wang X. (2019) A text analytics approach for online retailing service improvement: Evidence from Twitter. *Decis. Support. Syst.* 121:37–50. <https://doi.org/10.1016/j.dss.2019.03.002>
- Jacques P. J., Knox C. C. (2016) Hurricanes and hegemony: A qualitative analysis of micro-level climate change denial discourses. *Environ Pol* 25(5):831–852. <https://doi.org/10.1080/09644016.2016.1189233>
- Keramitsoglou K. M. (2016) Exploring adolescents' knowledge, perceptions and attitudes towards renewable energy sources: A colour choice approach. *Renew. Sustain. Energy Rev.* 59:1159–1169
- Killen C. P., Walker M., Hunt R. A. (2005) Strategic planning using QFD. *Int J Quality Reliab Management* 22(1):17–29. <https://doi.org/10.1108/02656710510572968>

- Kim J., Park S. Y., Lee J. (2018) Do people really want renewable energy? Who wants renewable energy?: Discrete choice model of reference-dependent preference in South Korea. *Energy Policy* 120:761–770
- Kirilenko A. P., Molodtsova T., Stepchenkova S. O. (2015) People as sensors: Mass media and local temperature influence climate change discussion on Twitter. *Glob. Environ. Chang.* 30:92–100. <https://doi.org/10.1016/j.gloenvcha.2014.11.003>
- Kirilenko A. P., Stepchenkova S. O. (2014) Public microblogging on climate change: One year of Twitter worldwide. *Glob. Environ. Chang.* 26:171–182. <https://doi.org/10.1016/j.gloenvcha.2014.02.008>
- Kryvasheyev Y., Chen H., Obradovich N., Moro E., Hentenryck P. V., Fowler J. (2016) Rapid assessment of disaster damage using social media activity. *Science Advances* 2(3):e1500779. <https://doi.org/10.1126/sciadv.1500779>
- Lee J., Moon H., Lee J. (2020) Consumers' heterogeneous preferences toward the renewable portfolio standard policy: An evaluation of Korea's energy transition policy. *Energy and Environment*, 0958305X2094946
- Liu B. (2010) Sentiment analysis and subjectivity. In: *Handbook of Natural Language Processing*. 2nd edn. Taylor and Francis Group, Boca
- Magnusson D. (2016) Who brings the heat? – from municipal to diversified ownership in the swedish district heating market post-liberalization. *Energy Research and Social Science* 22:198–209
- Mallaband B., Lipson M. (2020) From health to harmony: Uncovering the range of heating needs in british households. *Energy Research and Social Science* 69:101590
- Margetts H., Dunleavy P. (2013) The second wave of digital-era governance: a quasi-paradigm for government on the web. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 371(1987):20120382
- Maynard D., Bontcheva K. (2015) Understanding climate change tweets: an open source toolkit for social media analysis. In: *Proceedings of EnviroInfo and ICT for Sustainability*. Atlantis Press. <https://doi.org/10.2991/ict4s-env-15.2015.28>
- Mihalcea R., Corley C., Strapparava C. (2006) Corpus-based and knowledge-based measures of text semantic similarity. *Proceedings of the national conference on artificial intelligence* 1:775–780
- Navratil J., Picha K., Buchecker M., Martinat S., Svec R., Brezinova M. (2019) Visitors' preferences of renewable energy options in "green" hotels. *Renew. Energy* 138:1065–1077
- Oluoch S., Lal P., Susaeta A., Vedwan N. (2020) Assessment of public awareness, acceptance and attitudes towards renewable energy in Kenya. *Scientific African* 9:e00512
- Oluoch S., Lal P., Susaeta A., Wolde B. (2021) Public preferences for renewable energy options: A choice experiment in Kenya. *Energy Economics* 98:105256
- Omeng P. M., Eshiamwata G. W., Makindi S. M., Obwoyere G. O. (2019) Public participation in environmental impact assessment and its substantive contribution to environmental risk management insights from EIA practitioners and other stakeholders in Kenya's renewable energy sub sector, *Energy and Sustainability VIII*. WIT Press
- Pang B., Lee L., Vaithyanathan S. (2002) Thumbs up? In: *Proceedings of the ACL-02 conference on Empirical methods in natural language processing - EMNLP 02*. Association for Computational Linguistics. <https://doi.org/10.3115/1118693.1118704>
- 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* 9(4):e94785. <https://doi.org/10.1371/journal.pone.0094785>
- Peng X., Liu Z., Jiang D. (2021) A review of multiphase energy conversion in wind power generation. *Renew. Sustain. Energy Rev.* 147:111172
- Pfeffermann D. (2015) Methodological issues and challenges in the production of official statistics: 24th annual Morris Hansen lecture. *J Survey Stat Methodology* 3(4):425–483. <https://academic.oup.com/jssam/article-pdf/3/4/425/2791023/smv035.pdf>
- Salton G., McGill MJ (1983) *Introduction to modern information retrieval*. McGraw-Hill, New York. Google Scholar Google Scholar Digital Library Digital...
- Santos C, dos Gatti M (2014) Deep convolutional neural networks for sentiment analysis of short texts. In: *Proceedings of COLING 2014, the 25th International conference on computational linguistics: technical papers*, pp 69–78. Dublin, Ireland, Dublin City University and Association for Computational Linguistics. <https://www.aclweb.org/anthology/C14-1008>
- Schweiker M., Abdul-Zahra A., André M, Al-Atrash F, Al-Khatiri H, Alprianti RR (2019) The scales project, a cross-national dataset on the interpretation of thermal perception scales. *Sci Data* 6(1)
- Sisco M. R., Bosetti V., Weber E. U. (2017) When do extreme weather events generate attention to climate change? *Clim. Change* 143(1-2):227–241. <https://doi.org/10.1007/s10584-017-1984-2>
- Sovacool B. K., Demski C., Noel L. (2021) Beyond climate, culture and comfort in european preferences for low-carbon heat. *Glob. Environ. Chang.* 66:102200
- Taboada M., Brooke J., Tofiloski M., Voll K., Stede M. (2011) Lexicon-based methods for sentiment analysis. *Comput Linguistics* 37(2):267–307. https://doi.org/10.1162/coli_a.00049
- Thelwall M. (2018) Gender bias in sentiment analysis. *Online Inf. Rev.* 42(1):45–57
- Turney P. D. (2002) Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In: *Proceedings of the 40th annual meeting on association for computational linguistics*. USA: Association for Computational Linguistics, pp 417–424. <https://doi.org/10.3115/1073083.1073153>
- Ulwick AW (2005) In: *What customers want: Using outcome-driven innovation to create breakthrough products and services*. McGraw-Hill, p 71408673
- Wei X., Croft W. B. (2006) Lda-based document models for ad-hoc retrieval. In: *Proceedings of the 29th annual international ACM SIGIR conference on research and development in information retrieval*. Association for Computing Machinery, New York, pp 178–185. <https://doi.org/10.1145/1148170.1148204>
- Yang C., Gao F., Dong M. (2020) Energy efficiency modeling of integrated energy system in coastal areas. *J. Coast. Res.* 103(sp1):995
- Yu B. (2021) Urban spatial structure and total-factor energy efficiency in chinese provinces. *Ecol. Indic.* 126:107662
- Zhang Y., Abbas M., Iqbal W. (2021) Analyzing sentiments and attitudes toward carbon taxation in Europe, USA, South Africa, Canada and Australia. *Sustainable Production and Consumption* 28:241–253
- Zhang Y., Abbas M., Koura Y. H., Su Y., Iqbal W. (2020) The impact trilemma of energy prices, taxation, and population on industrial and residential greenhouse gas emissions in Europe. *Environ. Sci. Pollut. Res* 28(6):6913–6928

Zhao X., Gu B., Gao F., Chen S. (2020) Matching model of energy supply and demand of the integrated energy system in coastal areas. *J. Coast. Res.* 103(sp1):983

Zikopoulos P., Parasuraman K., Deutsch T., Giles J, Corrigan D. (2012) *Harness the power of big data the IBM big data platform.* McGraw Hill Professional

Zuo X., Dong M., Gao F., Tian S. (2020) The modeling of the electric heating and cooling system of the integrated energy system in the coastal area. *J. Coast. Res.* 103(sp1):1022

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations

Affiliations

Yaming Zhang^{1,2,3} · Majed Abbas¹ · Wasim Iqbal⁴

Yaming Zhang
yaming99@ysu.edu.cn

Wasim Iqbal
wasimiqbal01@yahoo.com

¹ School of Economics and Management, Yanshan University, Qinhuangdao 066004, China

² Research Center of Regional Economic Development, Yanshan University, Qinhuangdao 066004, China

³ Center for Internet Plus and Industry Development, Yanshan University, Qinhuangdao 066004, China

⁴ Department of Management Science, College of Management, Shenzhen University, Shenzhen, China