Social Impact and Risk Analysis Using Twitter

Measuring Sentiment about Infrastructure Sectors on the Example of Wind Power Generation

June 2023

EDHEC Infrastructure & Private Assets Research Institute

Contents

1. Introduction

Efficient infrastructure networks bring essential public services to communities, including electricity, transport, and water. However, infrastructure projects can also create significant disruptions, like loss of amenities, increased noise, air or water pollution, or impacts on local wildlife and human health. These adverse effects may lead to negative sentiments and reduced public support, resulting in delays or even cancellations of infrastructure projects. Hence, successful infrastructure developments require identifying and addressing deteriorating public endorsement in a timely manner.

Studies have shown that opposing movements from residents or environmental groups are a significant factor for infrastructure projects to be canceled, delayed, and more expensive. For example, McCarthy et al. (2020) found that community opposition is the leading cause for the cancellation of road infrastructure projects in the United States. In Australia, developers had to cancel the East West Link road in Melbourne after contracts were agreed because a deterioration in support for the project triggered protests (Alcorn, 2014). According to a report by the Victorian Auditor-General (2015), this cancellation cost Victorian taxpayers AU$1.1 billion. Similarly, the HS2 rail project in the United Kingdom reported to parliament that protests have already added £200 million to its costs (Plimmer and Pickard, 2022).

The transport sector is not alone in facing pushback; the energy sector also often faces public opposition, especially from those living close to potential project sites. For example, the La Compagnie du Vent wind farm had to remove four turbines from an already operating wind farm in 2013 due to their impact on local residents (Dodd, 2010). Despite the growing need for greener energy generation, the public often turns against new developments (Duxbury, 2021), which can lead to additional costs for infrastructure developers, governments, and investors.

As a result, monitoring public attitudes toward infrastructure sectors during development processes is necessary to be able to intervene promptly. Traditional approaches, like public opinion surveys or long-term panel interviews in the field, require time, money, and human resources. As a consequence, infrastructure developers can find that they have detected a change in public opinion too late to react and ensure their project is developed without interruptions. We propose using sentiment analysis to measure social sentiments across various infrastructure sectors and gain insights into public opinions in an immediate and relatively cost-effective approach.

In this paper, we describe our approach, develop indices of social sentiment, apply this method to wind power generation in the United States and the United Kingdom, and validate sentiment as a proxy for public acceptance. Wind power generation is a relatively mature renewable technology. Given the current push toward cutting emissions from electricity production, focusing on the wind power sector provides an insightful analysis. The literature review discusses the primary concept of social acceptance and the latest developments regarding using social media platforms (specifically Twitter) as a source of sentiment and public opinion. Our results show that the sentiment indices for the wind power sector correlate well with other measures of public acceptance.

2. Literature Review

Climate-related disasters have increased by almost 75% in the 21st century compared with the previous 20-year period. Storms, floods, and heat, among others, not only affected 4 billion people but also led to almost US$3 trillion in economic losses (Thacker, S. et al., 2021). More than ever, climate change requires massive actions and investments in efficient infrastructure networks to address present challenges and to build a more resilient environment, economy, and society. Currently, the Global Infrastructure Hub (2023) estimates an infrastructure investment gap of US$18 trillion. The road and energy sector have one of the most significant gaps, with US$8 trillion and US$5.6 trillion, respectively, required to reach the Paris Agreement and meet the UN Sustainable Development Goals.

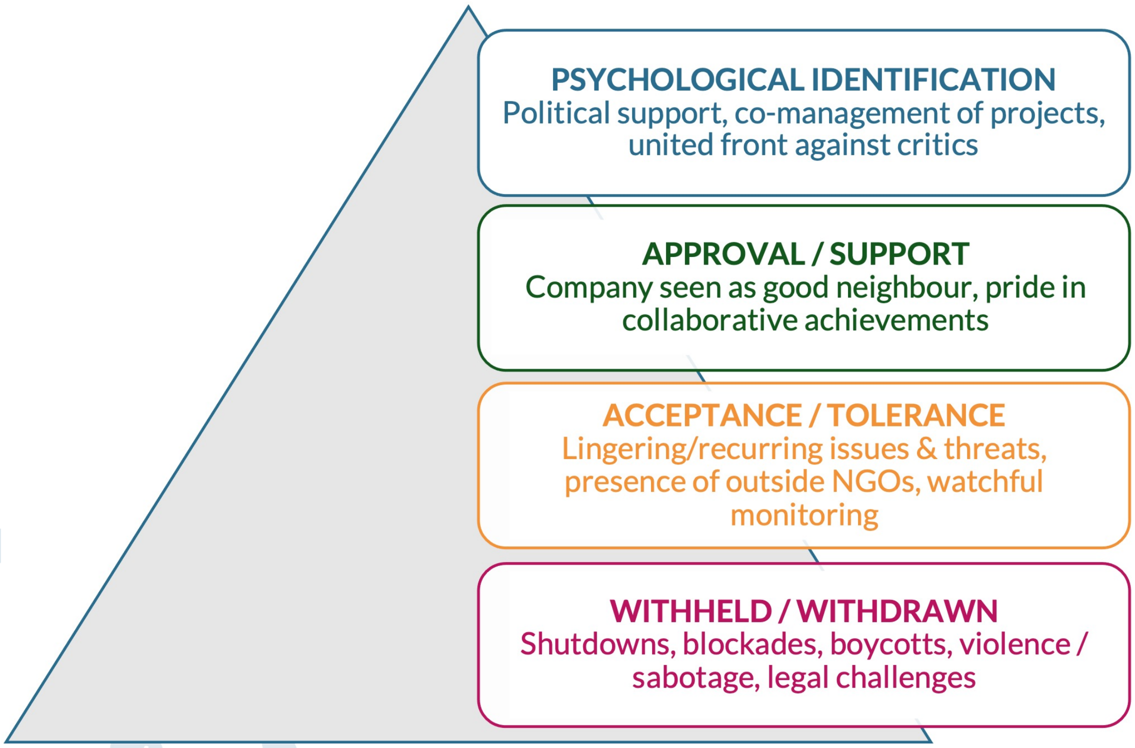
In order to build new infrastructure successfully and efficiently, infrastructure developers need not only a legal but also a social license to operate. A social license to operate is "an informal 'social contract' between all stakeholders giving project holders the consent they need to develop, deliver, and operate a project, through engaging with stakeholders and aligning interests" (Vauban Infrastructure Partners, 2022, 4). Whether the public accepts an infrastructure project can significantly influence project costs and development time. The past has shown that local residents have had significant legal wins against infrastructure developers. For example, local protests in Ireland stopped a wind farm development after four years of costly legal proceedings (The Kerryman, 2018). In Australia, the Victorian Supreme Court ordered the Bald Hills Wind Farm to compensate local residents AU$260,000 in damages for noise pollution and issued a permanent injunction to emit noise at night (Field, 2022). Accordingly, infrastructure developers need to monitor the social acceptance of infrastructure projects to prevent timely and costly legal proceedings and maintain a social license to operate.

2.1 Social acceptance

A social license to operate is a sequential, dynamic, and ongoing process in which communities and stakeholders form an opinion on whether a project and its developers are legitimate, credible, and trustworthy. Social legitimacy - based on a community's legal, social, and cultural norms - builds the basis, and its absence leads to the rejection of a project. If stakeholders and communities perceive a project and its developers as legitimate and credible, they can accept (lower level of tolerance or consensus) or - with an increased perception of credibility and trust - approve (higher level of agreement and support) and even co-own a project (see Figure 1). Additionally, a social license to operate requires a shared sentiment of all levels across the involved stakeholders and communities (Thomson and Boutilier, 2020). When analysing social support for infrastructure projects, most studies focus on the level of social acceptance and do not differentiate between acceptance and approval. Accordingly, our paper follows the dichotomous distinction between social acceptance - including all levels of tolerance of, support for, and identification with a project - and opposition.

Due to its interdisciplinary nature, the literature on social acceptance hardly provides a clear and consistent definition of the concept. Generally, the research field should differentiate between social acceptability and social acceptance (Busse and Siebert, 2018). Social acceptability describes the intersubjective and dynamic process of socio-political dimension influenced by individual attitudes, inter- and intrapersonal evaluations,

Figure 1: Levels of social license to operate and respective indicators (Thomson and Boutilier, 2020)



and perceptions of involved stakeholders, circumstances, and the broader economic and political situation. In contrast, social acceptance is the positive result and outcome of the acceptability process at a specific point in time. Due to the ongoing acceptability process and interactions between stakeholders, social acceptance can change over time (Busse and Siebert, 2018).

Wüstenhagen et al. (2007) provide one of the most cited definitions and classifications of social acceptance. They divide the concept into three categories of socio-political acceptance, community acceptance, and market acceptance. These three categories refer to different groups that represent various interests. Socio-political acceptance relates to the broad acceptance of policies and new technologies by the public, key stakeholders, and policymakers. Community acceptance relates to local acceptance by communities and authorities directly affected by siting decisions around infrastructure assets. Market acceptance relates to consumers' and investors' willingness to actively demand and invest in new technologies and renewable infrastructure assets (Wüstenhagen et al., 2007).

Often, the public's socio-political acceptance of renewable energy on a global level is high, while the community acceptance of specific infrastructure projects on a local level or the market acceptance to invest in such solutions is low. The concept of NIMBYism ("not in my backyard") explains this phenomenon. NIMBYism describes people's general acceptance of an innovation or technology but their disapproval of a specific project when it affects them directly (Carley et al., 2020; Devine-Wright, 2005; Wüstenhagen et al., 2007). Besides the controversial "NIMBY syndrome" (Busse and Siebert, 2018, 240), Wüstenhagen et al. (2007) add a time dimension that contributes to the community acceptance. Accordingly, we suggest applying not just the community acceptance but all three of Wüstenhagen et al.'s (2007) categories to the different levels of the social license to operate approach (Thomson and Boutilier, 2020): a broad socio-political acceptance might refer to a relatively neutral tolerance, but it requires community acceptance to receive the approval from local stakeholders to plan and build an infrastructure project. While communities might be sceptical toward infrastructure projects early in the development cycle, they often become more supportive when they experience a project's benefits over time. Finally, market acceptance is needed for consumers and investors to support and run the asset in the long run. Combining both theoretical frameworks supports the definition of social acceptability as a process that runs through various intersubjective levels that lead to socio-political, community, and market acceptance and, in the final step, provides a project with a social license to operate.

For the purpose of this study, our main interest lies in the outcome of the social acceptability process and the risks for investors related to the rejection of infrastructure projects. While the process and dimensions can support explaining the results of our analysis, we do not differentiate between levels and dimensions of acceptance but focus on measuring social acceptance.

2.2 Social media as a source of public opinion

Initially, most studies conducted surveys to measure social acceptance (for an overview, see Batel, 2020; Busse and Siebert, 2018). For example, Ribeiro et al. (2014) analysed the social acceptance of renewable energy technologies in Portugal. Surveying people from different regions with and without renewable energy technologies and including specific questions on community and market acceptance, the researchers concluded that the overall acceptance would not lead to local opposition. However, due to its design (conducting telephone interviews), the study faced some limitations: the exclusion of specific project developments and proximity to actual infrastructure assets might limit their conclusion. Another survey addressed the gap between the abstract perspective of socio-political acceptance and the concrete perspective of community and market acceptance (Sütterlin and Siegrist, 2017). Based on 1,211 telephone interviews in Switzerland, they found that people's social acceptance of solar power declines when confronted with drawbacks from that energy system. Important to note is that the interviews were conducted between one and five years before the actual publication of the results.

Designing and conducting surveys - especially on a representative level - is a costly and time-consuming process. Furthermore, public opinion surveys mainly measure socio-political acceptance but often ignore the communities' or markets' perspective (which studies have shown remains a more significant risk factor than the broad acceptance of the general public). Finally, public opinion surveys primarily represent a snapshot at a specific point in time but are rarely conducted as longitudinal panel studies that present the development of public opinion over time.

With the growth of social media, the clear boundaries between mediated and interpersonal communication dissolve, and people can share their opinions online with a wide audience. As a result, the latest research has moved to alternative approaches and developed novel methodologies to measure public opinion and social acceptance in written text people share online. This data includes posts on social media platforms like *Twitter*, reviews on *Google* or topic-specific websites like *Tripadvisor*, forum discussions (e.g., *Reddit*), or blogs.

The microblogging service *Twitter* is an online social network platform that combines news and information and uses hashtags to make different discourses visible. Twitter users can follow each other unilaterally and according to interest. They can disseminate information through posting *tweets* or sharing so-called *retweets* (other users' tweets). Both can contain text, emojis, pictures, videos, and links and are limited to 280 characters. [1] Users can engage in discourses

1 - In November 2017, Twitter doubled the maximum tweet length from 140 to 280 characters. In early 2023, Twitter offered a subscription service that allows "Twitter Blue" users to publish tweets with up to 10,000 characters.

by liking and retweeting a tweet or replying and mentioning a user directly to initiate a conversation.

Many Twitter users are experts who aim to share and receive the latest information in their field of interest and who use the network to discuss professional content rather than connect personally. Of the 556 million Twitter users in January 2023[2], most of them are between 25 and 49 years old (59%), male (63%), well educated (42% of users in the United States have at least a bachelor's degree), and based in the United States (95.4m) and Japan (67.5m), followed by India, Brazil, Indonesia, and the United Kingdom (Kemp, 2023; Wojcik and Hughes, 2019). Although Twitter has much fewer active users than other social media platforms, it is the leading platform for users to keep up to date with news and current affairs and the most visited website after Google, YouTube, and Facebook (based on website traffic between December 2021 and November 2022; Kemp, 2023), indicating that more people follow the platform's discourses without having an account.

Regarding social acceptance toward renewable energy, Vågerö et al. (2023) and Kim et al. (2021) analysed Twitter data to analyse social acceptance toward wind energy in Norway and solar power in the United States, respectively. Both studies applied similar machine learning techniques (NorBERT and RoBERTa). These are Natural Language Processing (NLP) approaches that measure sentiment in textual data. Vågerö et al. (2023) and Kim et al. (2021) used geospatial information Twitter users share on their profiles to understand regional differences. While the study in the United States showed a clear correlation between sentiment and political orientation, Vågerö et al. (2023) found that the sentiment in Norway varied between regions but did not provide clear patterns or explanations for the differences.

2 - Until Q3/2021, Twitter used to report the average daily active users but changed its reporting approach to the potential audience reach of advertisements. Hence, user numbers may not be comparable over time.

2.3 Sentiment analysis using Twitter data

The increasing sophistication of sentiment analysis has fuelled its popularity across many disciplines and beyond the use of Twitter data. For example, Shapiro et al. (2022) applied sentiment analysis to **economic** and financial newspaper articles and detected that positive sentiment shocks increase consumption, decrease inflation, and let to a rise in interest rates. O'Connor et al. (2010) applied a lexicon containing positive and negative words to Twitter data to estimate sentiment word frequencies and hence, the public's **political** opinion. In comparison, Loughran and McDonald (2011) found that general lexicons misclassify words in **financial** texts and thus, developed a finance-specific dictionary. They applied the dictionary to 10-K reports and found that negative word counts correlate significantly with announcement returns. In terms of social acceptance of **energy technologies**, Nuortimo and Härkönen (2018) applied machine learning techniques and compared social sentiment on various social media platforms and news outlets and of different energy technologies.

While all these studies follow different focal points, their approaches share the employment of NLP techniques to measure sentiment on an extensive collection of textual data. The aim is to understand how the public, stakeholders, or the industry think about products, services, and innovations and what influences them in decision-making processes (Liu, 2012). Algaba et al. (2020, 547) defines sentiment in this context as:

"Sentiment equals the disposition of an entity toward an entity, expressed via a certain medium. This working definition consists of (1) the expression by an entity of its disposition, in the form of verbal or non-verbal communication, (2) the expression has a polarity or a semantic orientation measurable on a discrete or a continuous scale, and (3) the expression is oriented toward (an aspect of) an entity."

Entities can be, for example, consumer confidence serving as a proxy for the state of the economy or investor sentiment to monitor cash flows and investment risks. These sentiments can be expressed subjectively in consumer or investor surveys or objectively in textual sentiment like market data. Algaba et al. (2020) hypothesise that media sentiment translates into consumer or investor sentiment. This hypothesis follows the idea of the so-called *two-step flow of information* that describes how information flows from the mass media to opinion leaders who - in a second step - shape the public opinion of the masses (Lazarsfeld et al., 1944; Katz and Lazarsfeld, 1955). While the mass media still influences the public to some extent, people use social media platforms to share their opinions and influence others.

As a result, monitoring sentiment on social media can provide necessary insights into people's opinions and acceptance of infrastructure projects and sectors. In this study, the entity of positive and negative expressions from residents in the United Kingdom and the United States serve as a proxy to measure the entity of social acceptance toward wind power generation. Textual data from the social media platform Twitter serve as a medium.

Overall, two main methodological directions to sentiment analysis apply either statistical or lexicographic models to identify the polarity[3] of a text in order to determine the sentiment expressed (Shapiro et al., 2022). The lexicographic approach involves linking dictionaries or phrases with their corresponding polarity valence, while the statistical - or machine learning - approach involves constructing statistical models that 'learn' the polarity of the text passage.

The latter sentiment analysis approach employs NLP techniques from machine learning. These techniques develop statistical models to determine the polarity of a text based on pre-annotated datasets. Researchers can either naturally label[4] or professionally curate[5] text to develop those pre-annotated datasets. Creating pre-annotated datasets - especially those manually labelled - requires employing human annotators, which is not only cost- and time-intensive but can also lead to biases. As summarised by Paullada et al. (2021), annotation work is interpretative, and failure to properly train and supervise the annotators can result in biases creeping into the dataset. The pre-annotated dataset is then used to train the machine on patterns, structures, and word combinations for each polarity direction. Finally, the machine can be applied to large datasets to identify positive, negative, and neutral sentiments in text passages. As the process shows, statistical models require large datasets as each step (pre-annotation, learning, and analysis) requires different data to analyse sentiment reliably.

Alternatively, the lexicographic approach involves curating a set of pre-defined words or phrases (a dictionary) and ranking each by their valence 1, 0, and -1 for positive, neutral, and negative sentiment. Dictionaries match and count word occurrences and summarise the expressed

3 - The polarity of a text means whether it expresses a positive, neutral, or negative view.

4 - Texts like movie or customer reviews are often naturally labelled texts that provide specific tags automatically when collecting the data. Usually, the reviewing author provides those tags; hence, it can be assumed that they accurately reflect the person's sentiment.

5 - In this case, researchers would manually label text as positive, negative, or neutral to create a dataset similar to naturally labelled data.

sentiment based on the valence in the dictionary (Shapiro et al., 2022). As previous studies have shown, dictionaries are more precise in analysing sentiment when treated as domain-specific and compiled by subject matter experts (Loughran and McDonald, 2011). For the lexicographic approach, constructing domain-specific dictionaries is the most labour-intensive step. Once the dictionary is developed, it can be applied to any type of text to count the occurrences and valences of the words and estimate the overall sentiment of the text. However, by only including words, the lexicographic approach loses information around the context of how the words are used in a sentence or longer text passage. To mitigate such problems, researchers have switched from employing words to employing phrases in dictionaries, which has the side effect of increasing the complexity of building a lexicon.

Following Shapiro et al. (2022), Shen and Whittaker (2023) combined both approaches: they applied the VADER dictionary (see more in Chapter 3.3) to news articles to measure news sentiment and identify social acceptance toward wind farms. In the second step, they manually labelled news articles. They used this pre-annotated 'ground truth' dataset to compute the pointwise mutual information (PMI) and improve the results of the sentiment analysis. Finally, they built a sentiment index and time series to present sentiment development over time. Shen and Whittaker (2023) compared their results with public opinion surveys to validate the sentiment results as a proxy for social acceptance.

In this paper, we build on previous research and aim to measure the social acceptance of wind power generation by examining the sentiment of Twitter posts from users residing in the United Kingdom and the United States. Our methodology (see Chapter 3) uses the lexicographic approach and follows studies conducted by Shapiro et al. (2022) and Shen and Whittaker (2023). In line with other studies, we validate our results (see Chapter 4) by comparing the results of the sentiment analysis with results from measuring news sentiment and public opinion surveys.

3. Methodology and Data

The methodology of our study is based on two research aims. Firstly, we aim to measure social support and risk for infrastructure sectors through the lens of sentiment shared by Twitter users. Secondly, we aim to compare and - in future studies - combine the results of our sentiment analysis with results from sentiment analysis using news articles (Shen and Whittaker, 2023) to further advance our methodology. Accordingly, our methodology is designed in line with Shen and Whittaker's (2023) study, as visualised in Figure 2 that explains the sentiment analysis process.

1. We **developed infrastructure- and ESG-specific dictionaries** that enabled us to identify relevant Twitter data.
2. We **pre-processed the imported Twitter data** to identify a final sample of ESG-related tweets.
3. We **measured sentiment scores** for each tweet in the data set.
4. We **built a sentiment index** to provide a smooth time series of sentiment development over the past ten years.
   Steps 1-4 are described in more detail in the following sections of this chapter.
5. We **validated** the sentiment results using public opinion surveys and news sentiment. Chapter 4 presents the results and the validation of the constructed indices.

3.1 Sector and ESG dictionary development

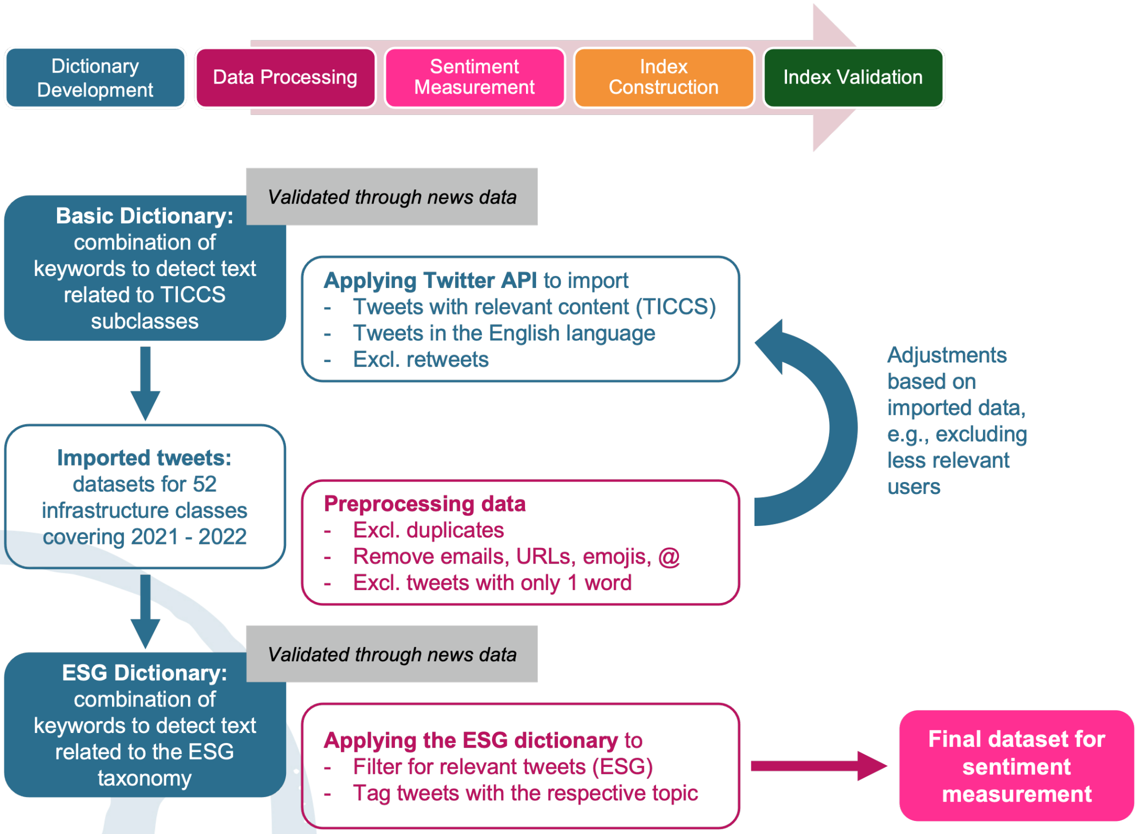
To assure quality control of our textual data and to identify specific Twitter discourses on infrastructure sectors as well as filter for ESG-related tweets, we applied a mixed methods approach to develop the **EDHEC*infra* Sector Dictionary** and the **EDHEC*infra* ESG Dictionary**. The dictionaries follow TICCS' (EDHECinfra, 2022) definition of asset classes and the topics presented in the ESG taxonomy (Manocha et al., 2022). Both dictionaries are based on a combination of keywords and terms that aim to collect and filter data in the most precise and inclusive way to assure quality control during the sample selection process. We developed the Sector Dictionary and the ESG Dictionary in three steps: 1) research and development of keywords, 2) data collection and revision process, and 3) validation against news articles.

To develop the **EDHEC*infra* Sector Dictionary**, we manually listed keywords for each asset class based on the TICCS definition, EHDEC's ESG Exposure Profiles[6], online dictionaries and thesauruses, expert interviews, and the respective Twitter discourse. Specifically the dictionaries, thesauruses, and tweets helped to identify terms with several meanings that needed to be excluded. For example, for the asset class IC605050 ("Stand-Alone Bridges"), the dictionary needs to be able to exclude text on dental bridges, the cards game Bridge, the stadium "Stamford Bridge," or the expression "a bridge too far," among others.

Second, we used this first version of the dictionary to collect tweets from Twitter (n = 973,298 tweets for 68 asset classes). We used the asset-specific samples to revise the dictionary based on two insights: the number of tweets posted daily and topic models identifying related topics and keywords for each asset class. Accordingly, if a set of keywords led to a small sample of tweets per

6 - The Exposure Profiles (EPs) aim to identify the main material factors explaining the direct and indirect exposures to ESG risks (risks related to environmental, social and governance issues) of different types of infrastructure assets. The EPs follow the definition of ESG risks and impacts developed by EDHEC*infra's* ESG Taxonomy (Manocha et al., 2022), and the definition of infrastructure assets introduced by the TICCS classification standard (EDHECinfra, 2022). The EPs set a minimum standard for ESG risk exposure assessment and provide asset owners, asset managers, investors and other stakeholders with a parsimonious view of the ESG profile of infrastructure asset types.

Figure 2: Sentiment analysis process and data processing



day, we added further terms or made the combination of keywords less restrictive to widen the selection and potentially be able to collect more relevant tweets. Conversely, if the set of keywords provided a large sample of tweets per day, we used the results of the topic model analysis to identify topics and keywords unrelated to the respective asset class to exclude those terms in future data collection and restrict the results.

To validate the revised dictionary and to test its applicability to other forms of text, we applied the dictionary to a sample of newspaper articles. The articles were provided by *Factiva*, a global news database, which tags each article by sector similarly to the TICCS classification. In total, 556,619 articles covering six sectors built our ground truth (GT) database. We applied the dictionary to the articles, compared our results with the Factiva tags, and calculated measures of performance for each sector group. Despite the F1 score (the harmonic mean between precision and recall), we also focused on the specificity score representing the ratio of correctly labelled negatives (articles not covering the respective sector). Compared to medical studies, we want to avoid having tweets in our data set that cover other topics (false positives) but would not be affected by excluding tweets that cover wind power generation (false negatives). Table 1 shows good to very good F1 scores for most asset classes and very good specificity scores for all asset classes. Accordingly, it can be assumed that the dictionary selects relevant text precisely and extensively among tweets, news articles, and potentially other forms of text. The low precision scores for solar and conventional power relate to a higher number of false positives (tweets and articles that cover sectors other than wind power generation) and result from articles covering multiple sectors when, for example, comparing renewable and conventional forms of power generation. Both dictionaries identify sectors and ESG topics, respectively, based on a simple keyword search but do not provide information on specific foci. Accordingly, it can be assumed that the dictionaries provide better results when applied to Twitter, as tweets are

Table 1: Accuracy, specificity, precision, recall, and F1 scores after revising the EDHEC*infra* Sector Dictionary and applying it to selected news articles

	Sector Dictionary				
Sector groups	Accuracy	Specificity	Precision	Recall	F1
Wind	0.983	0.984	0.864	0.973	0.915
Roads	0.905	0.898	0.817	0.919	0.865
Airports	0.873	0.975	0.921	0.647	0.760
Ports	0.949	0.991	0.894	0.609	0.725
Conv. Power	0.968	0.978	0.482	0.652	0.554
Solar	0.980	0.980	0.264	0.961	0.414

limited to a maximum of 280 characters and hence, do not provide space to discuss several sectors.

To develop the **EDHEC*infra* ESG Dictionary**, we cover topics of social risks regarding four actor groups as defined by the ESG Taxonomy (Manocha et al., 2022): the public, customers, employees, and regulators. To identify sets of keywords that represent interest groups and types of social risks, we ran anchored topic models using the previously collected sector-related Twitter data. Unlike unspecified topic models, our studies apply anchored topic models when expecting specific terms and topics to appear in the data set (Gallagher et al., 2017). We anchored topics presented in the ESG Taxonomy and EHDEC's Exposure Profiles to focus the analyses on those terms.

In addition, we added three specific topics related to EDHEC*infra's* work on carbon emissions. With the additional topics, we aim to explain social risks and the results of sentiment development over time. In detail, the topics cover:

- **public impact**: a sector's impact on the environment, wildlife, human health, and communities; the public's general acceptance; socio-economic factors; privatisation perception
- **customer service**: a sector's quality, affordability, and accessibility of service (for direct customers)
- **working conditions**: a sector's reputation regarding payment, working hours, work safety, human and labour rights, and discrimination; workforce availability
- **regulatory risks**: a sector's risks coming from (ESG) regulations, reportings, climate change, and transition goals; subsidies; corruption
- **negative reputation**: selective topics regarding the four actor groups with a focus on harmful and negatively impacted issues
- **transition risks**: a sector's transition risks; carbon emissions; carbon lock-in; stranded assets
- **carbon offset**: a sector's stand on carbon offsets and carbon certificates

Based on the results of the anchored topic models, the ESG Taxonomy, EHDEC's Exposure Profiles, and expert interviews, we created sets of keywords representing different aspects of social risks related to the four actor groups and the three additional topics. Depending on the asset class, some sets of keywords apply to all or more asset classes (e.g., to identify text on working hours and payment, a sector's negative reputation, or environmental impacts), whereas others relate to one specific asset class (e.g., the quality perception of airport areas and services or the condition of roads). Internal discussions led to several rounds of revisions.

Again, we tested the validity and applicability of the ESG Dictionary on a sample of news articles. At least four independent researchers manually

Table 2: Accuracy, specificity, precision, recall, and F1 scores after revising the EDHEC*infra* ESG Dictionary and applying it to selected news articles

	ESG Dictionary				
Sector groups	Accuracy	Specificity	Precision	Recall	F1
Wind	0.782	0.429	0.964	0.800	0.874
+non-ESG BOW	0.630	0.459	0.597	0.800	0.684
Roads	0.898	0.0	0.941	0.951	0.946
+non-ESG BOW	0.608	0.265	0.564	0.951	0.708
Airports	0.730	0.444	0.901	0.785	0.839
+non-ESG BOW	0.602	0.419	0.575	0.785	0.664
Solar	0.750	0.400	0.961	0.768	0.854
+non-ESG BOW	0.537	0.305	0.525	0.768	0.624

Note: "+non-ESG BOW" means we included additional bag-of-words from non-ESG-related articles about the respective sector that had initially been excluded from the GT dataset before the labeling exercise. We added them here to balance the samples of ESG- and non-ESG-related articles. Due to limited data availability, we added bag-of-words from airport articles to validate the ESG dictionary for the roads sector.

analysed the news sample to create GT data that defines whether each of those articles a) relates to a specific asset class, b) is ESG-related or not, and c) contains any positive or negative sentiment. In total, our GT data set contained 461 articles, out of which more than 92% were ESG-related. We used the ESG Dictionary to filter relevant articles, followed by comparing our results with the GT data set to calculate the same performance measures to determine the dictionary's quality. As the GT data was highly unbalanced, we added non-ESG-related articles to balance ESG- and non-ESG-related articles in the GT data set and recalculated the measures of performance.

Table 2 presents the results of good to very good F1 scores for all sector groups. Focusing on the balanced data that includes non-ESG articles, the specificity scores remained stable or improved in some cases. As expected, most scores across all sectors decreased when we added non-ESG data to the sample. However, it needs to be considered that the additional articles were neither GT data nor tested. Accordingly, none of the data sets – neither the balanced but untested nor the GT but unbalanced data – offered an ideal but rather an exploratory database to test the ESG Dictionary.

The appendix provides a complete list of all included and excluded keywords needed to identify text on wind power generation. Further, we introduce selected subtopics and respective keyword combinations covered by the ESG Dictionary.

3.2 Data selection and pre-processing

While the Sector Dictionary can identify relevant text for 52 infrastructure sector groups, this study focuses on wind power generation. Accordingly, the following data analysis presents the respective data and results for on- and offshore wind farms but can be applied to any other TICCS infrastructure sector.

To identify tweets discussing wind power generation, we used the EDHEC*infra* Sector Dictionary to create a query for Twitter's Academic API. The query specified all included and excluded keywords (see Appendix A), the exclusion of retweets[7] and tweets in languages other than English, and the time frame from January 2013 to March 2023.

After importing the available tweets from Twitter, we pre-processed the raw data to clean it from outliers[8], duplicated tweets, tweets that start

7 - Retweets are shared tweets of an original post by another user. As retweets would lead to extensive duplications in the data set, we only included original tweets, replies, and quotes.

8 - Here, outliers are users with an extreme amount of tweets with less value to the Twitter discourse on infrastructure, for example, users who report on the amount of wind power generation in a specific region daily or even hourly.

with "RT,"[9] and tweets that contain only one word. Further, we created a different version of the original tweet in which we deleted email addresses and links, removed punctuation, numbers, whitespaces, and emojis, and saved mentioned users separately from the tweet. We used this standardised version of the tweet to apply the ESG Dictionary to and filter relevant tweets but kept the original version of the tweet for the following sentiment analysis.

Furthermore, we applied the ESG dictionary and identified tweets that related to ESG topics. Our final data set contained N = 241,938 global ESG-related tweets on wind power generation covering 10 years between January 2013 and March 2023. The identified ESG-related tweets represent slightly above 5% of all tweets on wind power generation. Overall, the intensity of the wind power discourse on ESG topics varies between 11.5% in 2023 and about 2% yearly between 2015 and 2017.

Despite the filtering for ESG-related tweets, another crucial pre-processing step was the identification of users' locations. Our primary interest lies in analysing potentially different sentiment levels from people living in and referring to infrastructure assets in the United States, Canada, the United Kingdom, Australia, and New Zealand. Assuming that users primarily share their opinions on events and issues related to their area of residence, we filtered for tweets from users who publicly share their profile location. To avoid those cases where users would discuss wind power generation in another country, we compared the location information from the profile with locations shared in the tweet. We dismissed those tweets if users mentioned opposing locations (e.g., a user living in the United Kingdom referring to wind power projects in Australia). Overall, 45% of the tweets came from users in the United States, Canada, United Kingdom, Australia, and New Zealand. By contrast, 27% of the tweets came from users in other countries, and 29% had no location information.

For the use case of this paper and to validate the results of the sentiment analysis with previous results from news sentiment and public survey data on wind power generation, we focus our sentiment analysis on tweets from Twitter users in the two most prominent locations - the United States (n = 57,651 - 23.8%) and the United Kingdom (n = 31,048 - 12.8%).

Until 2015, more than 90% of the global tweets covered the sector's impact on the public. After, other topics joined the agenda, but the overall discourse remained relatively low until 2018.[10] Figure 3 presents the development of the topics discussed on Twitter. For better visibility and due to its disproportionately high share, we excluded the topic of public impact. In both countries, the overall number of tweets jumped in 2018 and 2019, and the discourse increased again in 2022 toward the end of the global COVID pandemic. In the United Kingdom, transition risks dominated the discourse between 2019 and 2021, with tweets about the sector's negative reputation rising in 2022. Similarly, tweets around transition risks and negative reputation are the most dominant topics in the United States, with the latter being slightly more discussed.

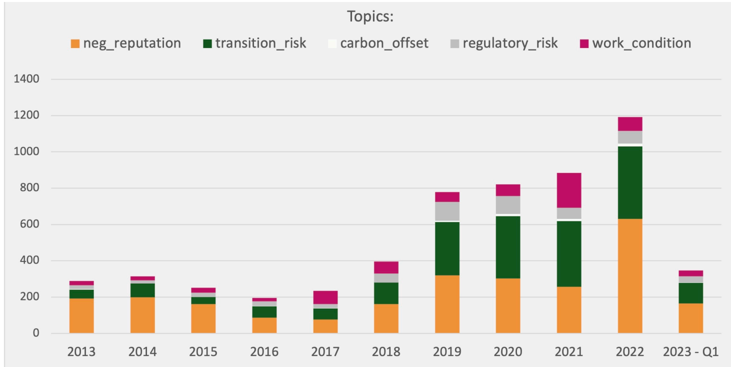
3.3 Sentiment analysis using VADER

In order to measure the "social acceptance" of infrastructure industries by examining the public sentiment around Twitter discourses on infrastructure assets, we follow the lexicographic approach (Shapiro et al., 2022) and use a similar methodology as in our previous study on news sentiment (Shen and Whittaker, 2023) to determine the sentiment of each tweet. For this,

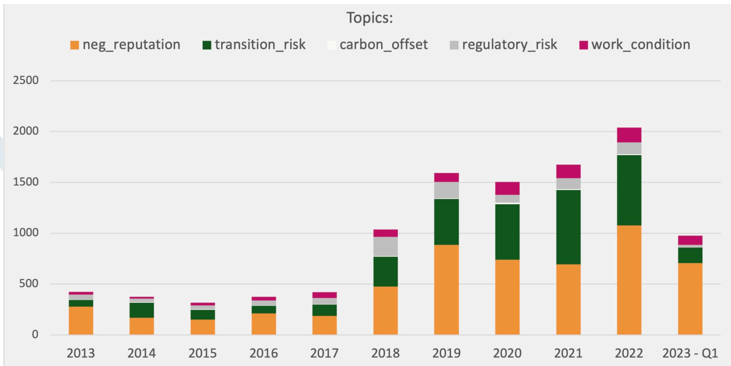
9 - Tweets that start with "RT" are retweets that users post as original tweets, and hence, the Twitter API does not identify those tweets as retweets when importing the data.

10 - Despite the assumption that users might have discussed ESG-related topics less on social media, changes on the Twitter platform could be another explanation for the increase in ESG-related tweets. In November 2017, Twitter doubled the maximum tweet length from 140 to 280 characters, changing users' ability to share sentiments on complex topics.

Figure 3: Number of tweets covering ESG-related topics within the Twitter discourse on wind power generation in the UK and the US



(a) United Kingdom



(b) United States

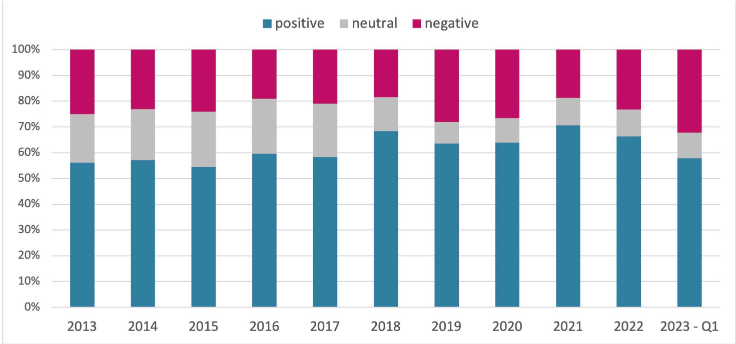
we employed the VADER dictionary that Hutto and Gilbert (2014) developed on and for short social media text.

The **VADER dictionary** provides a sentiment value (between -4 and +4) to each word in a given text. Based on the sentiment value, VADER computes a sentiment compound score between -1 and +1, indicating the sentiment's *polarity* (positive or negative) and *intensity*. The compound score is a normalised score of the sum of the sentiment values for all words in a given text (Hutto and Gilbert, 2014). Here, we calculated the compound score for each tweet in our final dataset. A tweet is classified as *positive* if the compound score is above 0.05 and *negative* if the compound score is below -0.05. All other tweets are classified as *neutral*.

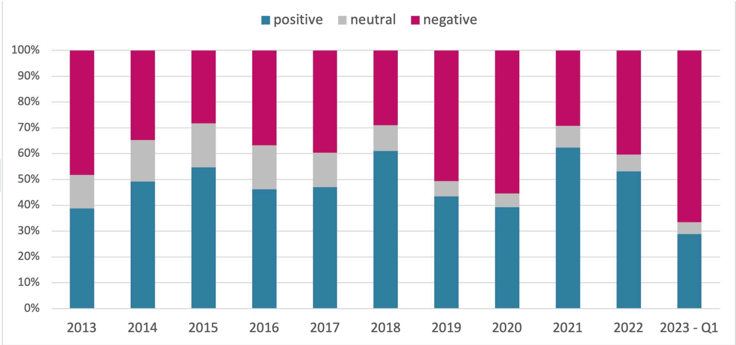
In comparison to other lexicographic approaches, the VADER dictionary includes not only words but more than 7,500 lexical features that also include Western-style emojis, sentiment-related acronyms (e.g., LOL, WTF), and slang used on social media (e.g., "meh," "giggly"). Overall, 44.5% of the features are classified as positive while 55.5% are classified as negative, which is a similar proportion to other sentiment lexicons. For each feature, the lexicon provides the polarity (+/-) and sentiment intensity (on a scale from 0 to 4). Furthermore, the method includes additional heuristics and grammatical rules that incorporate word-order sensitive relationships and influence the polarity and intensity of the sentiment (Hutto and Gilbert, 2014). These heuristics include:

- punctuation - to increase the sentiment intensity for text with one or more exclamation marks
- capitalisation - to increase the sentiment intensity for words written in capital letters
- degree modifiers - to increase or decrease the sentiment intensity for words that follow a modifier (e.g., very bad, extremely delicious, hardly affordable)

Figure 4: Distribution of ESG-related tweets on wind power generation in the UK and the US with positive, neutral, and negative VADER sentiment scores (based on the original threshold of +/-0.05)



(a) United Kingdom



(b) United States

- negations - to change the polarity of a text that follows a negation (e.g., "the food is not good")
- the conjunction "but" - to shift the polarity and the focus of the sentiment toward the text following the conjunction (e.g., "the service is great, but the food is bad")

We applied the VADER dictionary to all tweets in the dataset to compute the compound score for each tweet. We also calculated each month's average sentiment compound score for each of the five English-speaking countries to analyse how the public's sentiment for wind power generation develops over time in the different regions. Figure 4 captures the distribution of positive, negative, and neutral tweets in the United Kingdom and the United States for each year between 2013 and the first quarter of 2023. While most tweets toward wind power generation in the United Kingdom have a positive sentiment, tweets from the United States show an apparent increase in negative sentiment since 2019 (except in 2021, when the overall discourse on wind power generation decreased). The number of tweets with negative sentiment in the first quarter of 2023 (n = 4,019) already exceeded the number of negative tweets in 2022 (n = 3,717).

Based on the sentiment compound scores, we then created a sentiment index for wind power generation that enables us to observe the change in sentiment over time, compare sentiment development in different countries, and validate the results by comparing it with sentiment development from news articles and public surveys' results.

3.4 Sentiment index building

We followed the same approach employed in the study on newspaper articles (Shen and Whittaker, 2023) to construct sentiment indices for on- and offshore wind power generation. The **Social Support Index** measures the public's social acceptance of the wind power sector. Additionally, the **Social Risk Index** measures the level of disagreement within the public and hence, represents risk factors for investors, regulators, and the sector stemming from polarising oppositions.

To compute the **Social Support Index**, we followed Shapiro et al.'s (2022) work. However, to constrain any sudden sharp changes in public sentiment and to avoid misleading impressions of a volatile sentiment index, we improved the original approach by introducing the stochastic movement of the public sentiment between time steps. Further, we assumed that the tweets' sentiment score at time T can be split into systematic and idiosyncratic parts. The following equation expresses our approach:

$$\tilde{s}(T) = f_s(T) + \sum_k \beta_k x_k(T) + \varepsilon(T) \qquad (3.1)$$

$$f_s(T) = f_s(T-1) + \eta(T) \qquad (3.2)$$

where:

- $\tilde{s}(T)$ is the averaged tweets' sentiment score at time T,
- $f_s(T)$ is the systematic effect at time T and has the random movement $\eta(T)$,
- β_k and $x_k(T)$ are the k-th idiosyncratic effect and averaged feature of the tweet at time T, and
- $\varepsilon(T)$ is the observed noise at time T.
- Both signal $\eta(T)$ and noise $\varepsilon(T)$ follow a random walk of normal distribution $N(0, \sigma_\eta)$ and $N(0, \sigma_\varepsilon)$, respectively.

Shapiro et al. (2022) view the systematic effect $f_s(T)$ as the direct sentiment index. The regression takes place at each time step T. The idiosyncratic factors $x_k(T)$ include the type of Twitter user (news media organisation = 1; no news media organisation = 0) and the infrastructure subtypes (on- or offshore wind power generation = 1; no mentioning of a specific type = 0). We use the Kalman Filter in the State-Space model to calibrate the equation. As a result, the smoothed values of the systematic effect $f_s(T)$ (denoted as $\tilde{f}_s(T)$) build the raw index values that capture the sentiment trend.

However, since the calibration is against the averaged tweets' VADER score $\tilde{s}(T)$, the $\tilde{f}_s(T)$ have the magnitude of VADER scores (i.e., a continuous numeric number from -1 to +1) whose meaning is not direct. Therefore, to help the audience better understand $\tilde{f}_s(T)$, we applied a scaling function to transform $\tilde{f}_s(T)$ to a range from 0 to 100 to represent the overall acceptance or support of the public. Specifically, we collected a dataset of government surveys[11] stating the overall public satisfaction or support toward various infrastructure sectors. Then, we mapped the survey results and $\tilde{s}(T)$ at the corresponding time steps to create a dataset for our scaling purpose. We manually selected seven anchor points[12] from the scaling dataset to calibrate the following scaling function:

$$index(T) = \frac{A}{1 + \exp\left(k(\tilde{f}_s(T) - b)\right)} - F \qquad (3.3)$$

where A, k, b and F are the parameters to be calibrated. The proposed function is a nonlinear transformation to state that the marginal increase of social acceptance is more difficult when the acceptance rate is already at a high level. $index(T)$ is the final index value of the public sentiment at time T.

We used Twitter data about the wind power sector from the United Kingdom and the United States covering a period from January 2013 to March 2023 to construct the two sentiment indices and describe the social acceptance of on- and offshore wind power generation.

- The **Social Support Index** is computed by $index(T)$ directly and measures the overall public's social acceptance of the wind power sector.

11 - We used surveys on infrastructure sectors conducted by the British government because they maintain consistent methodologies and provide results for multiple years. The surveys cover public opinion on sectors like wind power, solar power, airports, and roads.

12 - We chose anchor points to calibrate the parameters in the scaling function to bring $\tilde{f}_s(T)$ to a proper level to compare the survey results.

- The **Social Risk Index** measures the level of disagreement within the public and represents risk stemming from polarising oppositions.
  To compute the Social Risk Index, we split the Twitter data at the median sentiment score at each time step into two groups to differentiate between the upper half with higher sentiment scores and the lower half with lower sentiment scores. We calculated $index(T)$ for both groups (denoted as $index_{up}(T)$ and $index_{low}(T)$), respectively. The difference between $index_{up}(T)$ and $index_{low}(T)$ builds the Social Risk Index and represents the level of disagreement within the public.

Section 4 presents the results and discusses the relationship between the Social Support and the Social Risk Indices.

4. Results

This section presents the Social Support and Social Risk Indices[13] for wind power generation in the United States and the United Kingdom. The results compare the social acceptance development between both countries and describe the relationship between the support and risk indices. An upward (or downward) support index trend implies an increasingly positive (or negative) sentiment toward wind power generation. An upward (or downward) risk index trend represents an increasing (or decreasing) gap between opposing opinions and hence, a higher (or lower) risk for the wind power sector. Furthermore, we compared the Social Support Index to sentiment from news articles and public opinion surveys to validate the applied methodology and the results.

4.1 Social Acceptance in the United Kingdom

Figure 5 presents the time series of the Social Support and Social Risk Indices of wind power generation between January 2013 and March 2023. In the United Kingdom, the Social Support Index (blue line) increases until it peaks in early 2018, implying a positive sentiment toward wind power generation. Smaller upward and downward cycles followed the peak. Since April 2021, the sentiment trend has decreased again, suggesting an increasingly negative sentiment.

Unlike the Social Support Index, the Social Risk Index (grey line) decreased until 2018, followed by an upward trend before it hit its lowest value in early 2021. After, the risk trend increases again. Thus, the support and the risk trends change in line with each other: the lower the risk index, the more positive the support index.[14]

The increasingly positive sentiment until mid-2018 is in line with several policy changes and initiatives that aimed to promote renewable energy sources in the United Kingdom, which may have contributed to the increasingly positive sentiment trend (UK Department of Energy & Climate Change, 2015). Incidents, such as protests by local residents and environmental groups against the Viking Wind Farm in Shetland (Williams, 2020) or against two offshore projects that require onshore infrastructure in Suffolk (Thomas, 2022), are examples that may have contributed to the ongoing declining sentiment trend since mid-2021.

4.2 Social Acceptance in the United States

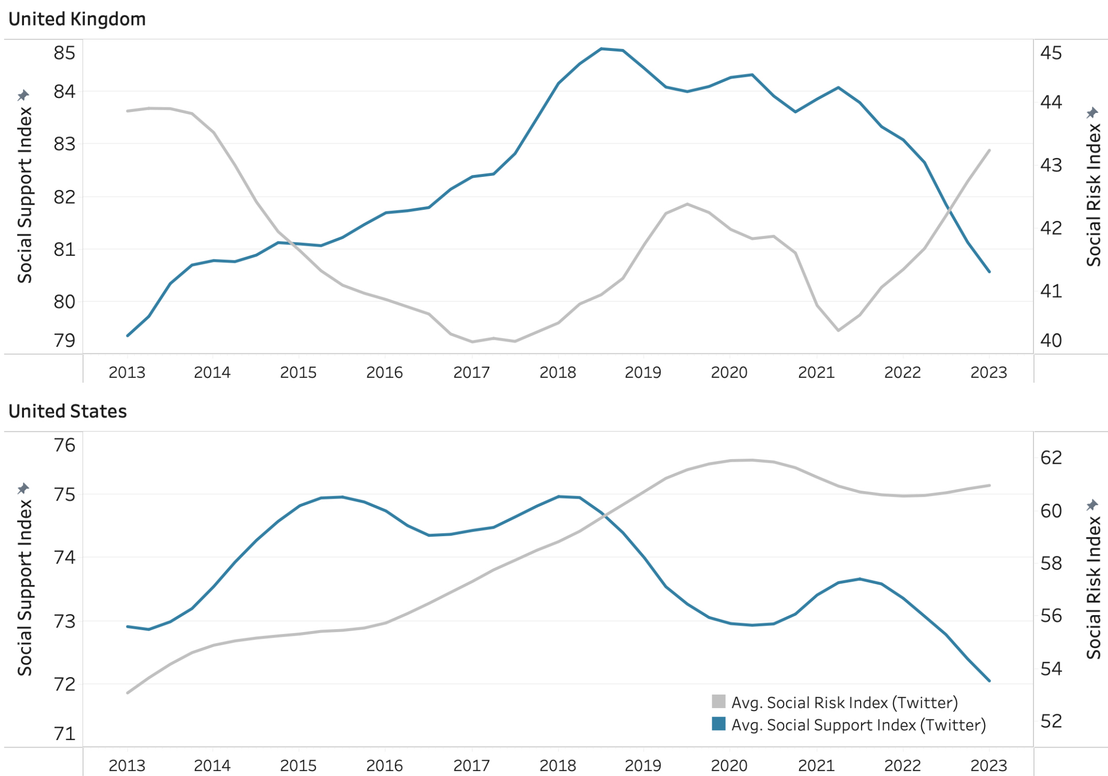
Below the United Kingdom, Figure 5 presents the time series of the Social Support and Social Risk Indices of wind power generation in the United States between January 2013 and March 2023. Much like in the United Kingdom, the support index increases first, but this positive trend only lasts until 2015. After this point, the sentiment remained stable until 2018, followed by a negative sentiment trend that still holds.

Concurrently, the Social Risk Index remained stable until early 2016. Thereafter, the risk trend increased until early 2020 and has stabilised over the past few years. Similarly to the United Kingdom, we can observe an opposite trend direction in the Social Support and the Social Risk Indices, indicating a lower sentiment trend when the risk is going up.

13 - It is essential to note that the acceptance indices measure systematic sentiment. Hence, the sentiment trend, not the actual sentiment scores, represents the most crucial factor. The raw score of the indices do not provide any information on the "point estimate" of support or risk.

14 - While we can observe a correlation between the Social Support and the Social Risk Indices, our data does not allow for conclusion on the causation of the trend development.

Figure 5: The Social Support and Social Risk Indices on wind power generation in the United Kingdom and the United States



Explaining the sentiment development in the United States remains a challenge due to the often polarising opinions, opposing political camps, and diverse locations for on- and offshore wind projects in the country. For example, the New York State passed the Climate Leadership and Community Protection Act in 2019, which aims to transition the state to 70% renewable energy by 2030 and 100% carbon-free electricity by 2040. This includes the development of 9,000 MW of offshore wind energy by 2035 (Howe and Greene, 2019). Concurrently, the state of Ohio passed a law that placed new restrictions on wind turbine development. While the law protects properties in close proximity to wind farms, it also ended the development of wind energy in the state (Tomich, 2021).

These opposing developments are also reflected in the public's opinion: While the majority (72%) finds that power companies should use more energy from renewable sources, this support diverge heavily between Democrats (90%) with Republicans (49%; Kennedy et al., 2022). Accordingly, the sentiment trend remains relatively stable and supporting for wind power generation, while the Social Risk Index remains high (also in comparison to the risk index in the United Kingdom) and reflects the polarisation in the country. It can be assumed that the risk index indicates a challenge for the overall public support for wind power in the United States.

Overall, wind power generation seems to be more supported by citizens in the United Kingdom than in the United States. In line with our understanding of social risks for the wind power sector, the Social Risk Index in the United States is higher than in the United Kingdom. The greater support among the public might result from the greater support for renewable energy by the British government. In comparison, the Trump administration between 2017 and 2021 was especially hostile to renewable energy and more supportive of the fossil fuel industry.

4.3 Validation through Public Opinion Surveys

The Social Support Index represents the sentiment of Twitter users sharing their opinions on wind power generation on social media. However, Twitter users do not represent all citizens in the United States and the United Kingdom, and not all users share their opinions publicly (O'Connor et al., 2010). Accordingly, additional analyses are needed to validate that the sentiment on Twitter represents public opinion. We applied two additional sources of information to compare our results with those of public surveys and sentiment in news articles.

For the United Kingdom, we employed the *BEIS Public Attitudes Tracker* (UK Department for Business, Energy & Industrial Strategy et al., 2023), which includes the public's perceptions on multiple topics, including carbon emission reductions, renewable energy sources, and, specifically, the support for on- and offshore wind power generation. Considering the questions' granularity and the study's frequency, the BEIS Public Attitudes Tracker is ideal for validating the sentiment results. For the United States, we employed several surveys from the Pew Research Center that measured the public acceptance of various energy sources, including oil and gas, coal, solar, wind, and nuclear energy (Rosenstiel, 2010; Tyson et al., 2021). These surveys focus on comparing different energy sources and relate public opinions to political affiliations.[15]

Figure 6 shows that the Social Support Index for wind power generation in the United Kingdom (blue line) is closely related to the public support for on- and offshore wind farms (green lines). The performance of the sentiment index closely follows the results of the offshore wind farm support. In contrast, the support for onshore wind farms presents a similar trend but begins at a lower level of acceptance. Considering that onshore wind farms affect people more directly than offshore wind farms, the lower acceptance of onshore wind farms should not come as a surprise. Simultaneously, the higher educated and more liberal Twitter users (compared to the general population) may explain the positive sentiment toward wind power generation on Twitter (Mellon and Prosser, 2017). Overall, we can conclude that the Social Support Index closely tracks public sentiment toward wind power generation, which provides substantial evidence and validation for the methodology presented in this paper.

Due to the less frequent conduction and availability of public opinion surveys in the United States, Figure 7 presents a less clear picture of the relationship between the sentiment index and survey results on wind farm support. However, public opinion follows the same but time-delayed trend as the Social Support Index. While the sentiment shows an apparent decline from 2018 onward, most members of the population remain supportive until the end of 2019. This delayed trend may be due to the time-consuming procedures of conducting representative surveys, while Twitter sentiment can be measured without any time constraints. Accordingly, we can assume that the Social Support Index measures public opinions toward wind power generation, presumably in a more timely manner than public opinion surveys.

4.4 Validation through News Sentiment

While the comparison between results from sentiment analysis and public opinion surveys is a common approach to validate new methodologies measuring social acceptance (Vågerö et al., 2023), it needs to be taken into account that nationwide public opinion surveys represent the broader public and hence, only one dimension of social acceptance - the socio-political acceptance. Thus, other measures are needed that can

15 - All public surveys cited in this study were conducted concisely. However, differences in methodologies may impact the comparability of the results. Furthermore, the only available panel study that measures public opinion consistently since 2012 comes from the United Kingdom. For the United States, the surveys represent snapshots of public opinion at specific times.

Figure 6: Comparing the Social Support Index (blue) with public opinion survey results on wind power generation in the United Kingdom

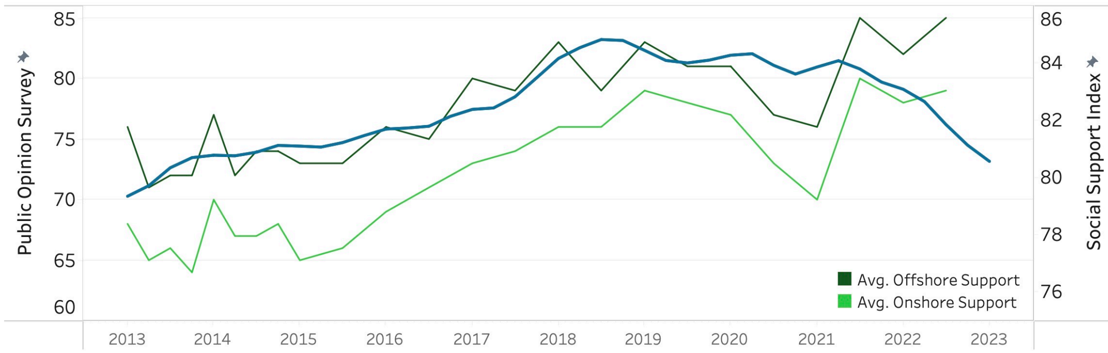
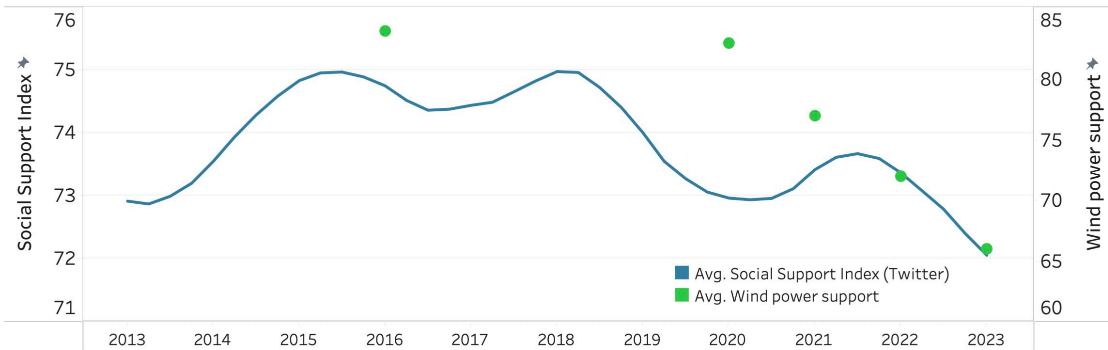


Figure 7: Comparing the Social Support Index with public survey results on wind power generation in the United States



represent community and market acceptance. In a previous study (Shen and Whittaker, 2023), we analysed news sentiment covering all dimensions of social acceptance as reported in national and local news on wind power generation.

Accordingly, we validated the Twitter sentiment results with the sentiment from news articles on wind power generation in the United Kingdom and the United States. The study on news sentiment followed a similar methodology, making the Twitter and news sentiment results comparable. For both countries, the United Kingdom and the United States, the Twitter and news sentiment indices follow a similar trend. However, in the case of the United Kingdom (see Figure 8), the Twitter sentiment is more positive than the news sentiment. In contrast, in the United States (see Figure 9), the Twitter sentiment is more negative than the news sentiment for most periods.

Considering that the news' primary objective is to inform the public and to report from a neutral perspective (Wahl-Jorgensen and Hanitzsch, 2009), it is not surprising that the Twitter sentiment trend is more extreme (in either direction) than the news sentiment trend. Interestingly, there is a difference between the United Kingdom and the United States: the Twitter sentiment trend in the United Kingdom correlates well with the news sentiment but is significantly more positive toward wind power. By contrast, the Twitter sentiment trend in the United States correlates less significantly with the news sentiment and is more negative during most periods.

Despite the public opinion surveys confirming the results of the Twitter sentiment, studies also suggest that Twitter sentiment in the United States is generally more negative (Roach, na) and more polarised (Urman, 2020) than in the United Kingdom. Furthermore, differences in news coverage can also explain the converse

Figure 8: Comparing the Social Support Index with news sentiment on wind power generation in the United Kingdom

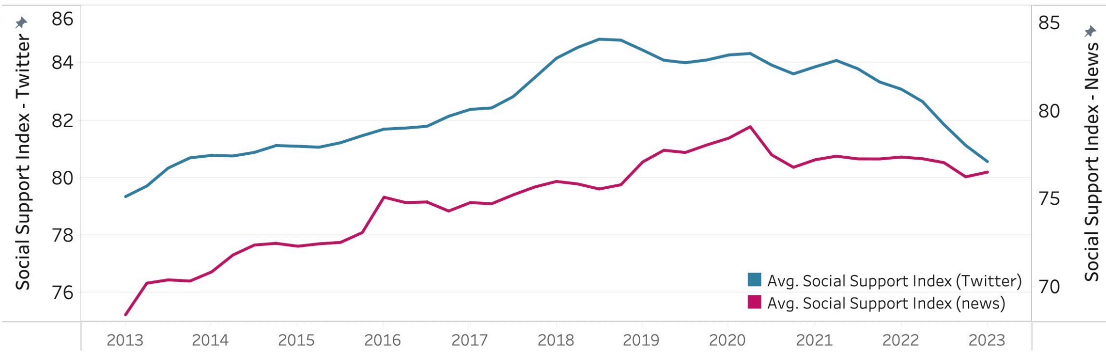
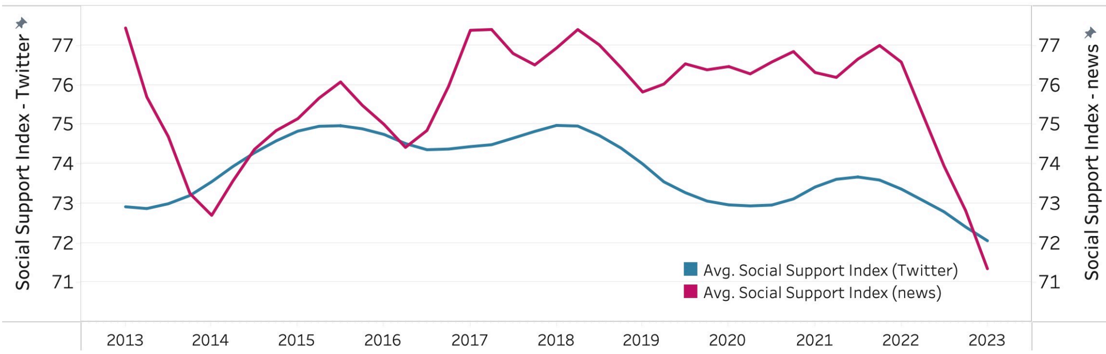


Figure 9: Comparing the Social Support Index with news sentiment on wind power generation in the United States



sentiment results in the United States and the United Kingdom. While both countries have a similar political system in which the media functions as the fourth pillar of democracy and gatekeeper between governments and the public, journalists in the United States and the United Kingdom follow different news values that can influence what and how they report (Harcup and O'Neill, 2017; Wahl-Jorgensen and Hanitzsch, 2009). For example, when reporting on the same issue, journalists in the United States report more from perspectives of patriotic values and governmental support. In contrast, journalists in the United Kingdom often have a more critical perspective emphasising consequences for the people (Wahl-Jorgensen and Hanitzsch, 2009).

Overall, the Twitter sentiment trend correlates well with the public opinion surveys and the news sentiment trend. Accordingly, we provide further evidence that our methodology is reliable and appropriate to measure social acceptance toward wind power generation.

5. Conclusion

This paper examines the potential for measuring social acceptance toward infrastructure sectors by applying sentiment analysis to social media posts that people have shared on Twitter. Infrastructure projects, including those related to energy and transport, are essential for delivering public services but can face opposition from the public, leading to delays or cancellations of projects. Inefficient approaches to monitoring public opinion toward infrastructure projects may lead to delayed reactions, resulting in additional costs for developers, investors, and governments. Therefore, this study examines the potential for measuring social acceptance toward infrastructure sectors promptly by applying the relatively cost-effective sentiment analysis approach to social media posts people share on Twitter.

By applying a sentiment index construction methodology first introduced in economics (Shapiro et al., 2022), we were able to measure social acceptance and public sentiment toward wind power generation in the United States and the United Kingdom. The indices presented a rising sentiment trend between 2013 and 2018 before the sentiment turned more negative in recent years. Overall, the public in the United Kingdom discusses wind power generation more positively on Twitter than citizens in the United States.

The results show a correlation between the Social Support Index and other measures of public acceptance, concluding that our methodology can be used to measure social acceptance effectively and efficiently. However, this conclusion varies slightly between the United Kingdom and the United States and the two validation approaches. It needs to be considered that the public opinion surveys are conducted at discrete intervals and follow different study designs. Hence, the survey results can only provide a guideline to validate rather than to generalise the findings. Furthermore, results from the news sentiment represent social acceptance as provided in the (presumably balanced) news reporting that may not reflect public opinion as polarised and extreme as it might occur in those countries. With those caveats on the validation, it is encouraging that the results of the BEIS Public Attitudes Tracker, the public opinion survey conducted over the most prolonged period and with the most consistent questionnaire design, correlate well with the sentiment index constructed for the United Kingdom and hence, validates that the index reliably measures social acceptance.

To further validate the methodology proposed in this paper, future studies should broaden the analysis and include different infrastructure sectors and other countries. Furthermore, the Twittersphere consists of various user types and includes not only individual users that represent the general public but also individuals that tweet in a specific role (politicians, journalists, researchers) and organisational accounts representing industries, political institutions, media outlets, and other interest groups. Accordingly, future studies should include the source of the sentiment to understand the interests of different actors within the discourse. Here, despite the user type, the precise location should also be considered. As the examples from New York (Howe and Greene, 2019) and Ohio (Tomich, 2021) have shown, regulations and local circumstances can differ in various regions and sentiment varies within a country (Vågerö et al., 2023; Kim et al., 2021). Additionally, scholars have found increased accuracy in studies that combine different sources of sentiment (Algaba et al., 2020). Accordingly, analysing so-called multimodal sentiment by combining news and social media sentiment could improve future results. Finally, the analysis revealed the focus on different ESG-related topics within the discourse on wind power generation. Combining the attention of the discourse with the sentiment results could explain declining social acceptance and assist in its management.

In summary, the study emphasises the importance of understanding public sentiment to ensure timely intervention and successful project completion. As a result, policy-makers, infrastructure developers, and investors can use the Social Support and Social Risk Indices as tools to identify declining social acceptance early, develop effective communication strategies and engage with the public to manage risks, and adjust and diversify investments across various sectors, regions, and social risk levels.

A. Appendix

A.1 EDHEC*infra* Sector Dictionary

The **EDHEC*infra* Sector Dictionary** offers specifications for 52 infrastructure sector groups. Each entry of the dictionary includes a list of included and excluded keywords that help specify the respective sector group and define its boundaries.[1] To detect text about on- and offshore wind power generation, we included and excluded the following keywords.

Included keywords:
(windfarm OR windfarms OR windpark OR windparks OR windstation OR windstations OR windplant OR windplants OR windpower OR windmill OR windmills OR "wind farm" OR "wind park" OR "wind plant" OR "wind power" OR "wind mill" OR "wind company" OR "wind energy" OR (wind (energy OR electricity OR power) (infra OR infrastructure OR megawatt OR turbine OR blade OR grid OR tower OR nacelle OR rotor OR onshore OR "on shore" OR offshore OR "off shore")))

Excluded keywords:
-"weather station" -concert -album -kimjaehwan -"kim jaehwan" -from:blinded_username

A.2 EDHEC*infra* ESG Dictionary

For the **EDHEC*infra* ESG Dictionary** we followed the same logic and operators. Each of the seven main topics include several sub-topics that are defined by a specific combination of keywords. Below, we define the main topics and provide a selection of keyword combinations for some of the sub-topics.

1 - The keyword combination follows Twitter's operators:
- a blank space = "and"
- OR = "or"
- quotes = exact phrases
- parentheses = to form groups that override the order of operators
- -keyword / -operator: = excluding selected keywords or other operators
- from: = operator to refer to a specific user

Impact on the public - This topic focuses on a sector's impact on the environment, wildlife, human health, and communities. It also covers sub-topics like the public's general acceptance for a sector, socio-economic factors, and the public's perception on privatisation matters. The following keyword combinations define the sub-topic of human health:
(public OR community OR citizens OR people OR locals OR human OR kids OR children OR youths OR elderly OR adults) (impact OR benefit OR risk OR harm OR threat OR protect OR destroy OR save OR damage) (health OR death OR sick OR ill OR hurt OR injured OR cancer OR asthma OR allergy OR allergies OR diseases OR "blood pressure" OR hormones OR mental OR stress OR anxiety OR wellbeing OR resilience)

Working conditions - This topic focuses on the public's perception about the working conditions in a specific sector. It covers sub-topics like payment, working hours, safety issues, human and labour rights, and discrimination but also workforce availability. The following keyword combinations define the sub-topic of discrimination at work:
(workplace OR "work place" OR worksite OR "work site" OR employer OR "work environment") (assault OR discrimination OR discriminate OR racism)

Regulatory risks - This topic focuses on regulations especially in regard to climate change and ESG reportings. It also covers sub-topics like transition goals, subsidies, and corruption. The following keyword combinations define the sub-topic of ESG-reporting:
(esg OR sustainability OR emissions) (reporting OR framework OR regulation)

Negative reputation - This topic combines selected sub-topics that focus on the four actor groups (the public, customers, employees, and regulators) but with an entirely negative perspective. The following keywords provide some examples to define a sector's negative reputation: *nimby OR "not in my backyard" OR corrupt OR corruption OR "displacement of indigenous" OR "child labour" OR "forced labour" OR "noise pollution" OR "light pollution" OR "job cuts"*

Transition risks - This topic focuses on a sector's transition risks to reduce carbon emissions. It also covers sub-topics like carbon lock-in and stranded assets. The following keyword combinations define the sub-topic of transitioning to a carbon-free sector:
(transition ("renewable energy" OR renewables OR "clean energy" OR (carbon "lock in"))) OR ("transition risk" (stranded OR climate OR carbon))

Carbon offsets: This topic focuses on carbon offsets and carbon certificates. The following keyword combinations define this topic:
"carbon offset" OR "carbon certificates" OR (carbon ("emission offset" OR "emission off set"))

Authors

Jeanette Orminski is a Senior ESG Researcher focused on public opinion and social impact factors in the infrastructure sector. Her experience spans the fields of public opinion, sustainability, and social network analyses. She received her PhD degree in Communication Science from the Nanyang Technological University, Singapore. Focusing on the sustainability discourse on Twitter, Jeanette applied network analysis and machine learning approaches to identify opinion leaders on Twitter. Prior to joining EDHECinfra, she worked as a Data Scientist at the Munich Science Communication Lab analysing the topic of Planetary Health.

Jianyong Shen is a Senior Research Engineer focusing on quantitative analysis and Machine Learning solutions in infrastructure assets' pricing/risk estimation and ESG impact. He has 15 years experience in the quantitative analysis in financial industries. He holds a Master in theoretical physics and a PhD in cosmology from Fudan University.

References

- Alcorn, G. (2014, October). East West Link: The case for and against Melbourne's $6.8bn road. *The Guardian*.
- Algaba, A., D. Ardia, K. Bluteau, S. Borms, and K. Boudt (2020). Econometrics meets sentiment: An overview of methodology and applications. *Journal of Economic Surveys 34*, 512–547.
- Batel, S. (2020, October). Research on the social acceptance of renewable energy technologies: Past, present and future. *Energy Research & Social Science 68*, 1–5.
- Busse, M. and R. Siebert (2018). Acceptance studies in the field of land use — A critical and systematic review to advance the conceptualization of acceptance and acceptability. *Land Use Policy 76*, 235–245.
- Carley, S., D. M. Konisky, Z. Atiq, and N. Land (2020). Energy infrastructure, NIMBYism, and public opinion: A systematic literature review of three decades of empirical survey literature. *Environ-mental Research Letters 15*, 1–16.
- Devine-Wright, P. (2005). Beyond NIMBYism: Towards an integrated framework for understanding public perceptions of wind energy. *Wind Energy 8*, 125–139.
- Dodd, J. (2010, April). La Compagnie du Vent ordered to remove turbines by French court. *Windpower Monthly*.
- Duxbury, C. (2021, February). Sweden's fading enthusiasm for onshore wind farms. *POLITICO*.
- EDHECinfra (2022, May). The Infrastructure Company Classification Standard (TICCS®) 2022 Edition – includes NACE, EU Taxonomy and CPRS Mappings.
- Field, E. (2022, March). Bald Hills Wind Farm ordered to stop emitting night-time noise, pay neighbours damages in landmark ruling. *ABC News*.
- Gallagher, R. J., K. Reing, D. Kale, and G. Ver Steeg (2017). Anchored correlation explanation: Topic modeling with minimal domain knowledge. *Transactions of the Association for Computational Linguistics 5*, 529–542.
- Global Infrastructure Hub (2023). Global Infrastructure Outlook - Forecasting infrastructure investment needs and gaps. https://outlook.gihub.org/.
- Harcup, T. and D. O'Neill (2017). What is news? News values revisited (again). *Journalism Studies 18*, 1470–1488.
- Howe, C. and N. Greene (2019, July). NY takes two major leaps forward in climate change fight. *NRDC*.
- Hutto, C. and E. Gilbert (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media 8*(1), 216–225.
- Katz, E. and P. F. Lazarsfeld (1955). *Personal influence: The part played by people in the flow of mass communications.* New York, NY, US: Free Press.

- Kemp, S. (2023, January). Digital 2023: Global Overview Report. https://datareportal.com/reports/digital-2023-global-overview-report.
- Kennedy, B., A. Tyson, and C. Funk (2022, July). Americans divided over direction of Biden's climate change policies. *Pew Research Center*.
- Kim, S. Y., K. Ganesan, P. Dickens, and S. Panda (2021). Public sentiment toward solar energy -Opinion mining of Twitter using a transformer-based language model. *Sustainability 13*, 1–19.
- Lazarsfeld, P. F., B. Berelson, and H. Gaudet (1944). *The people's choice.* Oxford, England: Duell, Sloan & Pearce.
- Liu, B. (2012). *Sentiment analysis and opinion mining.* Synthesis Lectures on Human Language Technologies. San Rafael, CA: Morgan & Claypool Publishers.
- Loughran, T. and B. McDonald (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance 66*, 35–65.
- Manocha, N., D. Marcelo, and F. Blanc-Brude (2022, July). ESG Taxonomy of Impacts and Risks for Infrastructure Companies. Version 2.0. Technical report, EDHECinfra.
- McCarthy, L. N., L. Bolaños, J. Y. Kweun, and J. Gifford (2020). Understanding project cancellation risks in U.S. P3 surface transportation infrastructure. *Transport Policy 98*, 197–207.
- Mellon, J. and C. Prosser (2017). Twitter and Facebook are not representative of the general population: Political attitudes and demographics of British social media users. *Research & Politics*, 1–9.
- Nuortimo, K. and J. Härkönen (2018). Opinion mining approach to study media-image of energy production. Implications to public acceptance and market deployment. *Renewable and Sustainable Energy Reviews 96*, 210–217.
- O'Connor, B., R. Balasubramanyan, B. Routledge, and N. Smith (2010). From tweets to polls: Linking text sentiment to public opinion time series. *Proceedings of the International AAAI Conference on Web and Social Media 4*, 122–129.
- Paullada, A., I. D. Raji, E. M. Bender, E. Denton, and A. Hanna (2021). Data and its (dis)contents: A survey of dataset development and use in machine learning research. *Patterns 2*, 1–14.
- Plimmer, G. and J. Pickard (2022, May). HS2 protests set to cost high-speed UK rail line up to £200mn. *Financial Times.*
- Ribeiro, F., P. Ferreira, M. Araújo, and A. C. Braga (2014). Public opinion on renewable energy technologies in Portugal. *Energy 69*, 39–50.
- Roach, A. (n.a.). Differences in social media use between the UK and USA. https://blog.miappi.com/social-media-use-uk-versus-usa.
- Rosenstiel, T. (2010, March). Support for alternative energy and offshore drilling. Pew Research Center.
- Shapiro, A. H., M. Sudhof, and D. J. Wilson (2022). Measuring news sentiment. Journal of Econometrics 228, 221–243.

- Shen, J. and T. Whittaker (2023, February). Measuring the social acceptability of infrastructure investments. A natural language processing approach. Technical report, EDHECinfra.
- Sütterlin, B. and M. Siegrist (2017). Public acceptance of renewable energy technologies from an abstract versus concrete perspective and the positive imagery of solar power. *Energy Policy 106*, 356–366.
- Thacker, S., Adshead, D., Fantini, C., Palmer, R., Ghosal, R., Adeoti, T., Morgan, G., and Stratton-Short, S. (2021). Infrastructure for climate action. Technical report, UNOPS, Copenhagen, Denmark.
- The Kerryman (2018, March). Locals win Supreme Court battle against wind farm. *Independent.ie*.
- Thomas, N. (2022, April). UK government threatened with legal action over offshore wind farms. *Financial Times*.
- Thomson, I. and R. Boutilier (2020). What is social license? http://socialicense.com/definition.html.
- Tomich, J. (2021, August). 'Volatile place.' New laws thwart Ohio renewables. *Energywire*.
- Tyson, A., B. Kennedy, and C. Funk (2021, May). Gen Z, Millennials stand out for climate change activism, social media engagement with issue. *Pew Research Center*.
- UK Department for Business, Energy & Industrial Strategy, UK Department for Energy Security and Net Zero, and UK Department for Business and Trade (2023, March). BEIS Public Attitudes Tracker. https://www.gov.uk/government/collections/public-attitudes-tracking-survey.
- UK Department of Energy & Climate Change (2015, May). 2010 to 2015 government policy: low carbon technologies. Policy paper.
- Urman, A. (2020). Context matters: Political polarization on Twitter from a comparative perspective. *Media, Culture & Society 42*, 857–879.
- Vauban Infrastructure Partners (2022). Social license to invest, social license to operate. How to secure and leverage stakeholder engagement for infrastructure in 2022. Technical report.
- Victorian Auditor-General (2015, December). East West Link Project. Technical Report 108.
- Vågerö, O., A. Bråte, A. Wittemann, J. Y. Robinson, N. Sirotko-Sibirskaya, and M. Zeyringer (2023). Machine learning of public sentiments toward wind energy in Norway. arXiv:2304.02388 (preprint).
- Wahl-Jorgensen, K. and T. Hanitzsch (Eds.) (2009). *The handbook of journalism studies*. International Communication Association (ICA) handbook series. New York, NY: Routledge.
- Williams, M. (2020, June). 'We don't want your money' - Shetland protests as SSE presses ahead with the UK's biggest wind farm. *The Herald*.
- Wojcik, S. and A. Hughes (2019, April). Sizing up Twitter users. *Pew Research Center*.
- Wüstenhagen, R., M. Wolsink, and M. J. Bürer (2007). Social acceptance of renewable energy innovation: An introduction to the concept. Energy Policy 35, 2683–2691.

Recent Publications (2020-2023)

- Shen, J. & T. Whittaker, "Measuring the Social Acceptability of Infrastructure Investments - The case of Wind Power" (January 2023)
- Monnier, L. & Blanc-Brude, F., "Does the rise of renewable energy create new risks for investors - Insights from 20 years of energy transition in the UK" (December 2022)
- Blanc-Brude, F., D. Marcelo & F. Nugier, "Carbon Footprint and Financial Performance of Transpoert Infarstructure - Transition Risk assessement using traffic and geospatial data" (December 2022)
- Blanc-Brude, F., D. Marcelo & T. Whittaker, "Physical Risk and the Cost of Capital of Infarstructure Investments - Flood damage factor estimation and bond yields in U.S. airports" (December 2022)
- Blanc-Brude, F., A. Gupta., J.Lee, F. Nugier & T. Whittaker, "The cost of international sanctions to investors in Russia's airports" (March 2022)
- Amenc, N., F. Blanc-Brude, R. Tan, & T. Whittaker, "The cost of capital of brown gas" (February 2022)
- Blanc-Brude, F. & A. Gupta. "Robust Benchmarks for investors in private infrastructure funds" (November 2021)
- Blanc-Brude, F., A. Gupta, L. Lu & A. Wee "The Volatility of Unlisted Infrastructure Investments" (May 2021)
- Blanc-Brude, F., A. Gupta, "The Fair Value of Investments in Unlisted Infrastructure equity" (April 2021)
- N. Manocha, N. & F. Blanc-Brude "Towards a Scientific Approach to ESG for Infrastructure Investors" (March 2021)
- Amenc, N., F. Blanc-Brude & A. Gupta. "Strategic Asset Allocation with Unlisted Infrastructure" (February 2021)
- Blanc-Brude, F., L. Lum, R. Tan & T. Whittaker, "The choice of performance test benchmark of Superannuation funds' investments in infrastructure" (December 2020)
- Amenc, N. & F. Blanc-Brude. "The Cost of Capital of Motorway Concessions in France - A Modern Approach to Toll Regulation" (September 2020)
- F. Blanc-Brude & A. Gupta. "Unlisted Infrastructure Performance Contribution, Attribution & Benchmarking" (July 2020)
- Whittaker, T. & R. Tan. "Anatomy of a Cash Cow: An In-depth Look at the Financial Characteristics of Infrastructure Companies." (July 2020)

These publications are available on the infraMetrics® platform: metrics.edhecinfra.com

For more information, please contact:
Nataliia Gaidarenko
e-mail: nataliia.gaidarenko@edhec.edu

EDHEC Infrastructure & Private Assets Research Institute
EDHEC Asia-Pacific
One George Street - #15-02
Singapore 049145
Tel.: +65 6653 8575

EDHEC Europe
10 Fleet Place
London EC4M 7RB
Tel.: +44 20 7332 5601

edhec.infrastructure.institute