ROYAL SOCIETY OPEN SCIENCE

rsos.royalsocietypublishing.org

Research

Check for updates

Article submitted to journal

Subject Areas:
Interdisciplinarity, Computer Science, Machine Learning

Keywords:
academics, researchers, perceptions, opinions, integrative research concepts, interdisciplinarity

Author for correspondence:
Charlotte Teresa Weber
e-mail: charlotte.t.weber@uit.no

THE ROYAL SOCIETY
PUBLISHING

Interdisciplinary Optimism? Sentiment Analysis of Twitter Data

C. T. Weber[1], S. Syed[2,3]

[1]Norwegian College of Fishery Science, UiT - The Arctic University of Norway, TromsÃÿ, Norway
[2]Department of Information and Computing Sciences, Utrecht University, Utrecht, the Netherlands
[3]Centre for Policy Modelling, Manchester Metropolitan University, Manchester, UK

Interdisciplinary research can face many challenges, from institutional and cultural, to practical ones, while it has also been reported as a "career risk" and even as "career suicide" for researchers pursuing such an education and approach. Yet, the propagation of the challenges and risks can easily lead to a feeling of anxiety and disempowerment in researchers, which we think is counterproductive to improving interdisciplinarity in practice. Therefore, in the search of 'bright spots', which are examples where people have had positive experiences with interdisciplinarity, this study assesses the perceptions of researchers on interdisciplinarity on the social media platform Twitter. The results of this study show researchers' many positive experiences and successes of interdisciplinarity, and as such document examples of bright spots. These bright spots can give reason for optimistic thinking, which can potentially have many benefits for researchers' well-being, creativity, and innovation, and may also inspire and empower researchers to strive for and pursue interdisciplinarity in the future.

1. Introduction

Interdisciplinarity involves activities that integrate more than one discipline with the aim to create new knowledge or solve a common problem. The interdisciplinary approach has gained popularity in science, education, and policy over the last years and it is often advocated for solving today's complex problems and societal issues, such as climate change, biodiversity loss, food and water security, and public health issues [1,2]. It is hoped that interdisciplinary research will help solve these problems and help create innovative solutions through coordinated approaches. Such coordinated approaches combine knowledge and enable a coalescence at the interfaces and frontiers of the different scientific disciplines. This bridging of disciplinary boundaries facilitates development and innovation [3]. However, successfully crossing and integrating diverse fields and disciplines is not an easy endeavor.

Interdisciplinarity can face many challenges, from institutional [4] and cultural [5] to practical challenges [6,7]. Interdisciplinary work has also been reported to have lower funding success [8], can be challenging to publish [9], and interdisciplinary journals are commonly perceived as less prestige compared to single-disciplinary ones [10]. As a result, young scholars following an interdisciplinary career-path now fear to "risk their careers", or even to "commit career suicide" [11,12]. Others have perceived their interdisciplinary experience as if they did not belong to a discipline, a research community, or a research group. They had to live without the comfort of expertise, while having to fight for identity, recognition, and legitimacy within their work environment and among their peers [13]. In part because their background was too diverse or too broad to belong to a single discipline or to be considered an 'expert'. Many of these negative experiences and challenges have been, and continue to be, reported in the literature.

We claim that the continued propagation of challenges is counterproductive to improving interdisciplinarity in practice. For example, negative wording as in "less funding success", and "career suicide" can easily create anxiety in (early career) scientists and lead to a feeling of disempowerment. While the study of such challenges and shortcomings is an important step when trying to improve interdisciplinary research in the future, we argue that we also need to study the 'bright spots' in order to harvest the full potential of interdisciplinarity. These bright spots are examples where people have had positive experiences with interdisciplinarity and success stories of interdisciplinary research (IDR)—despite its challenges and barriers. We believe that the documentation of such bright spots and success can propagate optimism (understood here as the generalized expectancy that one will experience good outcomes [14]), which can further unlock creativity and innovation in interdisciplinary individuals and teams.

Previous research has shown that the social media platform Twitter is generally used to broadcast thoughts and opinions [15]. Within academia, Twitter is also used to acquire and share real-time information, and to develop connections with others [16]. Previous research, furthermore, shows that Twitter plays a significant role in the discovery of scholarly information and cross-disciplinary knowledge spreading. Therefore, the aim of this study is to assess the perceptions of academics (also referred to as researchers or scientists) on interdisciplinarity on a larger scale, in the pursuit of such bright spots within people's experiences shared on Twitter with the ambition to create interdisciplinary optimism.

2. Material and Methods

(a) Defining the different modes of research

The terms *interdisciplinary*, *transdisciplinary*, and *multidisciplinary* all describe different modes of research that include a range of participants with a degree of disciplinary interaction. While many definitions already exist in the literature (and most point in a similar direction), sometimes the meaning of the terms can be unclear, especially if they are used interchangeably. For the purpose of this paper, the different research concepts are conceptually visualized in Fig. 1 and understood and defined as follows [17]:

- **Interdisciplinarity** refers to the integration of several unrelated academic disciplines that forces actors to cross boundaries with the goal to create integrated knowledge and theory;
- **Transdisciplinarity** involves the same process as in interdisciplinarity, but includes non-academic participants;

- **Multidisciplinarity** involves multiple disciplines researching a common theme in parallel, but without integration or the crossing of subject boundaries.

The two research concepts, interdisciplinary and transdisciplinarity, were included in this study because they can both be considered interdisciplinary due to their integrative nature. The concept of multidisciplinarity lacks the integrative process according to the above definition, a view further supported by other literature [18]. However, not all literature makes that clear distinction and as a result, multidisciplinarity has been described as a mode of research that allows for the integration of knowledge [19]. In addition, multidisciplinarity is often included in research addressing interdisciplinary activities and impact [8,20,21]. Therefore, the research concept of multidisciplinarity was also included in this study for the assessment of sentiment towards interdisciplinarity. In this study, the three concepts, interdisciplinary, transdisciplinary, and multidisciplinary, will be referred to as *modes of research* and *integrative research approaches*.

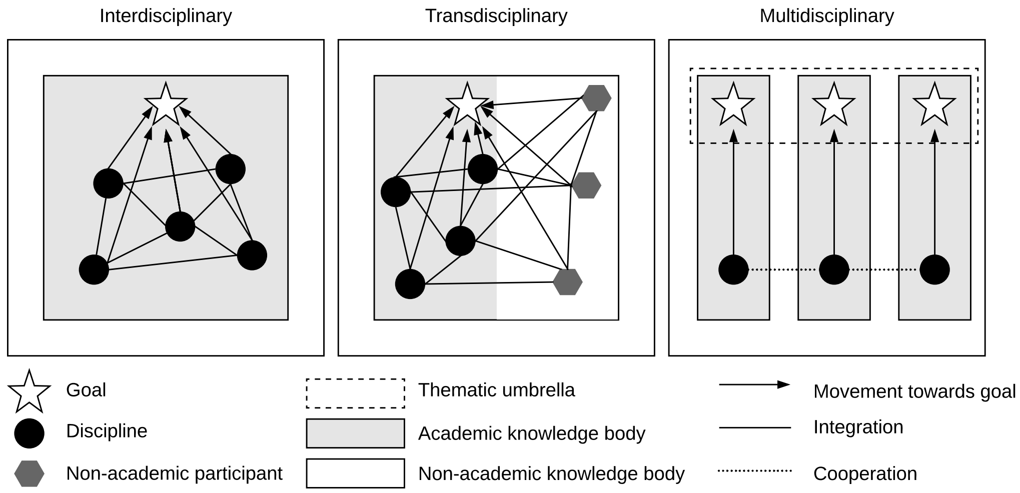


Figure 1: Overview of the three modes of research. Modified from [22].

(b) Dataset

The dataset of publicly available tweets related to the three modes of research was obtained by utilizing the Twitter Search API. The Twitter Search API returns tweet data, such as the tweet text (i.e., content) and ID, that matches a specified search query. We used the Python library *Tweepy*[1] to access the search API and query the tweets, which conveniently respects the Twitter rate limit of 900 tweets per 15 minute window. The used query strings for the three modes of research are listed in Table 1. The Twitter API automatically returns all hyphenated variants of the search words, such as inter-disciplinarity and inter-disciplinary, eliminating the need to include such variations within the search queries. Since the Twitter API only returns tweet data not older than 7 days, we collected tweets from week 32 (2017) up to week 33 (2018), a time frame of 53 weeks. During data collection, Twitter rolled out their expanded 280 character limit—previously 140 characters—which resulted in a dataset with 140 and 280 character limit tweets. It is important to note that the Twitter API is not an exhaustive source of tweets, as not all tweets are indexed or available via the search interface. The full dataset of all collected tweets is made available in the electronic supplementary material S1 Data.

[1]http://www.tweepy.org/

Table 1: Overview of search queries used to retrieve tweets related to the three modes of research.

Mode of research	Twitter search query
Interdisciplinary	'Interdisciplinary OR #Interdisciplinary OR Interdisciplinarity OR #Interdisciplinarity'
Transdisciplinary	'Transdisciplinary OR #Transdisciplinary OR Transdisciplinarity OR #Transdisciplinarity'
Multidisciplinary	'Multidisciplinary OR #Multidisciplinary OR Multidisciplinarity OR #Multidisciplinarity'

(c) Audience of the Dataset

Within this study, our aim is the make inferences regarding tweet sentiments associated with an academic or research domain. To identify tweets originating from a research domain setting, we filtered the dataset of all publicly available tweets for tweets from individuals who identify themselves as scientists (including all variations thereof, such as researchers or academics). To enable this, we used an adaptation of the systematic approach to identifying scientists on Twitter proposed by [23]. Our filtering process essentially aimed to match occupational classifications to the description field associated with the user account of the tweet. The description field is an optional field of maximum 160 characters where the user can describe herself, colloquially known as the user's bio.

We utilized the list of 322 scientific occupations (e.g., biologist, computer scientist, political anthropologist) compiled by [23]. This list was constructed by selecting the scientific occupations from: (i) the 2010 Standard Occupational Classification[2] (SOC) system released by the Bureau of Labor Statistics, United States Department of Labor; (ii) Wikipedia's list of scientific occupations[3]; and (iii) the authors choice of adding generic occupations such as 'scientist' and 'researcher'.

We furthermore augmented the list of 322 occupations by obtaining all the synsets (i.e., synonyms that share a common meaning) for each occupation from an on-line lexical reference system called WordNet [24]. Including the synsets and excluding duplicate entries resulted in a total list of 430 occupations related to a scientific or academic profession (see electronic supplementary material, Data S2). We then used regular expressions to match occupations with the user description field. This filtering approach identifies, for example, tweets from a user describing himself as 'senior lecturer in human geography at university of Liverpool' as valid for inclusion, and a description of a user describing herself as 'costume design & visual arts' as valid for exclusion. A random sample of 5,000 tweets were manually examined to assess the inclusion and exclusion criteria, and adjustments were made to the regular expressions to enhance the filtering process (for instance to capture American vs British spelling). We furthermore excluded tweets that contained no description text. On the one hand, the exclusion of tweets with no description text might negatively affect the recall of relevant tweets. On the other hand, it positively affects the precision. In other words, we might not be able to include all the tweets from an academic or research setting (i.e. lower recall), but we can be more sure that the included tweets are all from the correct audience (i.e., higher precision).

(d) Preprocessing Tweets

Tweeting, the process of publishing a tweet, proceeds in the form of free text, often in combination with special characters, symbols, emoticons, and emoji. This, in combination with a character limit, make tweeters creative and concise in their writing, favoring brevity over readability to convey their message—even more so with the 140 characters limit. Thus tweet data is highly

[2]http://www.bls.gov/soc/
[3]http://en.wikipedia.org/wiki/Scientist#By_field

idiosyncratic and several preprocessing steps were necessary (described below) to make the dataset suitable for sentiment analysis.

Retweets and duplicate tweets We removed retweets, identified by the string 'RT' preceding the tweet, as they essentially are duplicates of the initial or first tweet. Additionally, duplicate tweets that were identical in their content were also excluded.

Non-English tweets We focused our analysis on English tweets only and excluded all non-English tweets according to the 'lang' attribute provided by the Twitter API.

User tags and URLs For the purpose of sentiment analysis, the user tags (i.e., mentioning of other Twitter user accounts by using @) and URLs (i.e., a link to a specific website) convey no specific sentiment and were therefore replaced with a suitable placeholder (e.g. `USER`, `URL`). As a result, the presence and frequency of user tags and URLs were retained and normalized.

Hashtags Hashtags are an important element of Twitter and can be used to facilitate a search while simultaneously convey opinions or sentiments. For example, the hashtag #love reveals a positive sentiment or feeling, and tweets using the hashtag are all indexed by #love. Twitter allows users to create their own hashtags and poses no restrictions in appending the hashtag symbol (i.e., #) in front of any given text. Following the example of the #love hashtag, we preprocessed hashtags by removing the hash sign, essentially making #love equal to the word *love*.

Contractions and repeating characters Contractions, such as *don't* and *can't*, are a common phenomenon in the English spoken language and, generally, less common in formal written text. For tweets, contractions can be found in abundance and are an accepted means of communication. Contractions were preprocessed by splitting them into their full two-word expressions, such as *do not* and *can not*. In doing so, we normalized contractions with their "decontracted" counterparts. Another phenomenon occurring in tweets is the use of repeating characters, such as *I loveeeee it*, often used for added emphasis. Words that have repeated characters are limited to a maximum of two consecutive characters. For example, the word *loveee* and *loveeee* are normalized to *lovee*. In doing so, we maintained some degree of emphasis.

Lemmatization and uppercase words For grammatical reasons, different word forms or derivationally related words can have a similar meaning and, ideally, we would want such terms to be grouped together. For example, the words *like*, *likes*, and *liked* all have similar semantic meaning and should, ideally, be normalized. Stemming and lemmatization are two NLP techniques to reduce inflectional and derivational forms of words to a common base form. Stemming heuristically cuts off derivational affixes to achieve some kind of normalization, albeit crude in most cases. We applied lemmatization, a more sophisticated normalization method that uses a vocabulary and morphological analysis to reduce words to their base form, called lemma. It is best described by its most basic example, normalizing the verbs *am*, *are*, *is* to *be*, although such terms are not important for the purpose of sentiment analysis. Additionally, uppercase and lowercase words were grouped as well.

Emoticons and Emojis Emoticons are textual portrayals of a writer's mood or facial expressions, such as :-) and :-D (i.e., smiley face). For sentiment analysis, they are crucial in determining the sentiment of a tweet and should be retained within the analysis. Emoticons that convey a positive sentiment, such as :-), :-], or ;), were replaced with the positive placeholder word `EM_POS`; in essence, grouping variations of positive emoticons with a common word. Emoticons conveying a negative sentiment, such as :-(, :c, or :-c, were replaced by the negative placeholder word `EM_NEG`. A total of 47 different variations of positive and negative emoticons were replaced. A similar approach was performed with emojis that resemble a facial expression and convey a positive or negative sentiment. Emojis are graphical symbols that can represent an idea, concept or mood expression, such as the graphical icon of a happy face. A total of 40 emojis with positive and negative facial expressions were replaced by the placeholder word `EM_POS` and `EM_NEG`, respectively. Replacing and grouping the positive and negative emoticons and emojis will result in the sentiment classification algorithm learning an appropriate weight factor for the corresponding sentiment class. For example, tweets that have been labeled as conveying a negative sentiment (by a human annotator for instance) and predominantly containing negative emoticons (e.g., :-(), can result in the classification algorithm assigning a higher probability or weight to the negative sentiment class for such emoticons. Note that this only holds when the neutral and positively labeled tweets do not predominantly contain negative emoticons; otherwise their is no discriminatory power behind them (see also Section (e)).

Numbers, punctuation, and slang Numbers and punctuation symbols were removed, as they typically convey no specific sentiment. Numbers that were used to replace characters or syllables of words were retained, such in the case of *see you l8er*. We chose not to convert slang and abbreviations to their full word expressions, such as *brb* for *be right back* or *ICYMI* for *in case you missed it*. The machine learning model, described later, would correctly handle most common uses of slang, with the condition that they are part of the training data. As a result, slang that is indicative of a specific sentiment class (e.g. positive or negative) would be assigned appropriate weights or probabilities during model creation.

Input features Each tweet was tokenized, the process of obtaining individual words from sentences. Furthermore, we represented tweets as count vectors with and without inverse document frequency (IDF) weighting [25]. Different variations of tokenization were explored, such as 1-word (unigram), 2-word (bigrams), 3-word (trigrams), and 4-word (n-gram) combinations. Bi-grams are especially important to capture negation of words combinations, such as *not good* or *not great*, that would not be captured when using 1-word (unigram) features alone.

(e) Creating the Machine Learning Classifier

This paper employs a supervised machine learning approach to predict positive, neutral, and negative sentiments from the tweets related to the three modes of research. Supervised machine learning essentially learns a sentiment classification model, called a classifier, from labeled tweet data, that is, tweets that have been labeled as positive, neutral, and negative by human annotators. With the use of labeled data, the machine learning classifier learns that certain words convey, for example, positive sentiments when they more frequently occur in positively labeled tweets. The word *happy*, generally speaking, is used to convey a positive sentiment or feeling and tweets containing the word might be assigned a higher probability for the positive sentiment class. This is a somewhat basic and straightforward example but the classifier learns to assign every word—technically called a feature—a probability for each of the three sentiment classes. The tweet is thus a combination of features with corresponding probabilities and, ultimately, the classifier assigns the tweet a probability for the positive, neutral, and negative class. The class with the highest probability is the inferred sentiment class. In essence, a supervised machine learning classifier is built or trained from labeled data and is applied to unlabeled data to predict or infer their label.

Several online repositories are available that contain human annotated tweet data. We combined several of such online repositories that serve as input data, called training data, to create or train the machine learning classifier. A total of seven different repositories were used which contained a total of 71,239 labeled tweets, with 22,081 positive, 31,423 neutral, and 17,735 negative tweets. Table 2 shows an overview of the datasets used to train the classifier, together with the frequency of tweets for the three sentiment classes, the domain or subject of the tweets, the number of human annotators used to label the tweets, and a selection of research studies that have used the dataset. We provide descriptions of the seven datasets in the electronic supplementary material, Text S1. Note that the Twitter terms of service do not permit direct distribution of tweet content and so the tweet IDs (references to the original tweets) with their respective sentiment labels are often made available without the original tweet text and associated meta-data. As a result, we used the Twitter API to retrieve the full tweet content, the tweet text and meta-data, by searching for the tweet ID. Some tweets appeared not to be available from the Twitter API and this, in some cases, resulted in the training datasets having fewer tweets than originally included in the published datasets.

Table 2: Overview of training datasets. For a full description of the datasets, see the electronic supplementary material Text S1

Dataset	Positive	Neutral	Negative	Total	Domain	Annotators	Study
Sanders	424	1,996	475	2,895	Apple, Google, Microsoft, Twitter	1	[26–28]
OMD	704	-	1,192	1,896	#tweetdebate, #current,#debate08	3-7	[29–31]
Stanford Test	182	139	177	498	consumer products, companies, and people	1	[32–34]
HCR	537	337	886	1,760	#hcr	1	[31]
SemEval-2016	3,918	2,736	1,208	7,889	100 different topics	5	[35]
SS	1,252	1,952	861	4,066	major events	1	[36,37]
CLARIN-13	15,064	24,263	12,936	52,263	1% public available tweets	1-9	[38,39]
Total	22,081	31,423	17,735	71,239			

(f) Machine Learning Model Selection

The labeled training datasets serve as input for building the machine learning classifier (i.e., learning a model to classify tweets into positive, neutral, and negative sentiments). The tweets with their corresponding sentiment label enable the classifier to extract features that best predict the sentiment of a tweet. Typically, with supervised machine learning, one would need sufficient data for each sentiment class to make good predictions on new tweets. We obtained a training dataset containing 71,239 labeled tweets, which can be considered a sufficiently large dataset for sentiment analysis.

Several (supervised) machine learning algorithms are suitable for the purpose of creating a sentiment classifier from labeled tweet data. Unfortunately, no consensus exists on what classification algorithm to use since different studies have different datasets, perform different pre-processing steps, use different features, have incompatible performance measures, or simply have different use cases. Thus, adopting one strategy that worked for a particular use case might not work for another. The current state-of-the-art for sentiment analysis typically use algorithms based on neural networks [40,41]—also referred to as deep learning models—as can be seen from top-ranking teams during the SemEval 2017 competition [42]. The downsides of such winning entries are complexity, computational cost, and the fact that they are highly tuned and optimized to achieve a high score on the task's performance measure. Besides neural network models, more traditional machine learning classifiers have also shown high accuracy and performance on sentiment classification tasks. They include Support Vector Machines (SVM) [43–45], logistic regression [26,45], Naive and Multinomial Naive Bayes [26,31,44,46], and Conditional Random

Fields (CRF) [42]. Less complex neural networks, such as the Multi-layer Perceptron, have also been explored [46].

The aim of this paper is not to exhaustively explore the full suite of algorithms available but to use one that accurately predicts sentiments from tweets with reasonable complexity and computational time. Though complexity and computational time are hard to define concretely [47], we limit complexity to the basic machine learning and ensemble classification algorithms found in the Python library *Scikit-Learn* [48]. Additionally, we included a basic neural network, thus excluding very deep models and convolutional or sequence models. In terms of computational time, all selected models could be trained in reasonable amount of time (10-30 minutes of wall clock time per model) on an Apple MacBook Pro with i7 Processor and 16GB of internal memory. For example, the support vector machine with Gaussian kernel was not explored since it was too time consuming to train a single model. A total of seven different supervised machine learning algorithms were considered: (1) Support Vector Machine (SVM) Linear Kernel; (2) Logistic Regression; (3) Multinomial Naive Bayes; (4) Bernoulli Naive Bayes; (5) Decision Trees; (6) ADA boost; and (7) the Multi-Layer Perceptron.

The dataset of 71,239 labeled tweets was partitioned into two parts. The first part, called the training set, contained 80% randomly selected tweets used to train and validate the seven different algorithms. The second part, a random sample of 20% of the data called the test set, was used to test the performance of the algorithms on tweet data that was not used during training. All seven different algorithms were applied to the training set with 10-fold cross validation, which is a standard approach in machine learning [38]. During 10-fold cross-validation, the training set is partitioned into 10 parts, called folds, and training is done on 9 folds with the remaining fold used to test the performance of the algorithm. This process is repeated 10 times, essentially creating 10 different classification models in which each model is tested against the remaining fold. Partitioning the data into folds is done on a stratified random basis, preserving the percentages of samples for each sentiment class. Additionally, we used a standard grid-search approach to establish the optimal performing parameter values for each of the seven algorithms. Since algorithms are parameterized and regularized by a set of parameters or hyper-parameters, finding the best performing values of these parameters can be obtained by trying out different combinations of values, called a grid-search. Other approaches, such as a random grid-search or Bayesian optimization [49] can also be considered, but were not employed in this study. The grid-search approach was combined with the 10-fold cross validation method. For example, a grid search that tries out four different values for two separate parameters, combined with 10-fold cross validation for a single algorithm results in 4 x 4 x 10 = 160 different sentiment classification models. The different hyper-parameters and their parameter values that were explored are listed in Table 3. The model that achieves the highest performance score (described in Section (h)) is validated against the test set (i.e., remaining 20% of the data) to assess the performance of the model and its parameters on hold-out data; a way to measure the model's generalizability to unseen data.

(g) Suitability Training Data

It is important that the training data from which we train the supervised machine learning classifier can appropriately infer sentiment classes for the (unseen) tweets containing the three modes of research. By drawing on 71,239 training tweets (described in Table 2), we captured a wide array of sentiment expressions. However, specific sentiment indicators associated with the three modes of research can be absent from the training data, making accurate classification a challenging task. To mitigate this risk, we manually labeled a random subset of the tweets related to the three modes of research. A total of 1,000 tweets, stratified by mode of research, were labeled positive, negative or neutral. To have a common understanding of what a positive, negative or neutral tweets constitutes, we utilized the sentiment description text provide by Amazon Mechanical Turk's documentation for setting up a sentiment annotation project[4]. Amazon

[4]https://docs.aws.amazon.com/AWSMechTurk/latest/RequesterUI/Create-Sentiment-Project.html

Table 3: Overview of explored hyper-parameter values when performing a cross-validated grid search to obtain the machine learning classification model with best classification performance (F1-score). The value 'x' indicates that the hyper-parameter was used to explore different variations of the algorithm. Not all hyper-parameters are possible for all explored models, these are indicated by the absence of an 'x'. A full description of the hyper-parameters can be found in the Scikit-learn documentation at https://scikit-learn.org/stable/documentation.html

Hyper-parameter	**Value Range**	**SVM**	**Logisitic Regression**	**Multinomial NB**	**Bernoulli NB**	**Descision Trees**	**ADA Boost**	**ML-Perceptron**
n-grams	1–4	x	x	x	x	x	x	x
min-df	0,5,10,15	x	x	x	x	x	x	x
max-df	1.0, 0.95, 0.90	x	x	x	x	x	x	x
IDF	Yes, No	x	x	x	x	x	x	x
sublinear-TF	Yes, No	x	x	x	x	x	x	x
C (penalty term)	0.001, 0.01, 0.5, 0.1, 1, 5, 10, 15, 20, 100	x	x					
fit prior	Yes, No			x	x			
alpha	0.001, 0.01, 0.5, 0.1, 1, 5, 10, 15, 20, 100			x	x			
splitter	best, random					x		
criterion	gini, entropy					x		
max features	auto, sqrt, log2, None					x		
num. estimators	100,200,300,400,500						x	
algorithm	SAMME, SAMME.R						x	
neural architectures	(100,50,20),(200, 100, 100),(300, 50, 50, 50),(50, 40, 30, 10),(20, 30, 50, 50),(70, 50, 40, 30)							x
Cross-validation	10-Fold	x	x	x	x	x	x	x
	Total Models (x1000)	19.2	19.2	38.4	38.4	30.7	19.2	11.5

Mechanical Turk is typically used as a crowd sourcing platform to annotate tweets for their sentiments [29,50,51]. Positive tweets embodied a happy, excited or satisfied emotion; negative tweets embodied an angry, upsetting, or negative emotion; and neutral tweets did not embody much negative nor positive emotion. The 1,000 labeled tweets containing interdisciplinarity, multidisciplinarity, and transdisciplinarity content were added to training data to build the classification model. Additionally, we utilized Laplace smoothing—for the algorithms that allow smoothing—to mitigate some risk of misclassification related to absence of sample features in the training data [52]. Additionally, the preprocessing steps described in Section (d) were similarly applied on the training tweets and target tweets so that the learned features of the training data can be applied onto the target tweets.

(h) Calculating Classification Performance

Typically for classification purposes, the performance of a model is assessed by the number of correctly predicted tweet sentiments in relation to incorrectly predicted tweet sentiments. We can discern three evaluation metrics when classifying tweets into positive, neutral, and negative sentiments: (1) precision, (2) recall, and (3) F1. Precision measures how many of the tweets predicted to belong to a certain sentiment (e.g. positive) are actually positive. Precision, thus, measures how precise the predictions are. Recall measures how many of the e.g. positive tweets are captured by all of the predicted positive tweets. Recall can be seen as a metric to evaluate if the classification model is able to identify all the positive tweets from the complete dataset. There is a trade-off between optimizing recall and optimizing precision and a summarized measure between the two is captured by the F1 score, a harmonic mean of precision and recall. We equally care about precision and recall and thus optimize these to achieve a high F1 score, which is an appropriate performance metric when having imbalanced classes (i.e., classes that are not represented equally). The model with the highest F1 score was ultimately used to predict tweet sentiments for the three modes of research (i.e., interdisciplinary, transdisciplinary, and multidisciplinary tweets).

Additionally, after utilizing the best performing classification algorithm on the target tweets (i.e., tweets related to the three modes of research), we created a random stratified dataset of 1,000 target tweets (in addition to the 1,000 manually labeled tweets reported in section (g)), stratified by mode of research and the inferred sentiment label. This dataset was manually labeled for the true sentiment class, and the performance of the classification algorithm was measured against it. In doing so, the classification performance on the target tweets could be measured. We report the precision, recall, and F1 measure as previously described.

(i) Inspecting Tweets

For the analysis of the content of the tweets (to understand what people felt positive, neutral or negative about), a random set of tweets (n=2,000 or less, depending on the number of tweets classified into a specific sentiment) for each sentiment class and for each mode of research was manually examined for the content and context by one examiner. Additionally, word clouds were constructed to summarize each class and each mode of research (9 word clouds) for the most dominant words (here dominant measured by frequency). Each word cloud contained all the available tweets for that sentiment class and mode of research.

3. Results

(a) Classifier Performance

Table 4 shows the evaluation metrics (i.e., precision, recall, F1) for the seven different classifiers that were applied to the training (80%) and test (20%) partitions of the labeled tweet data. Ideally, the performance results between the training and test data should be similar; an indication that the model is not overfitting. A model that overfits the training data performs, generally, worse on the test data as it is unable to generalize to new unseen data. The results within the test set columns, and specifically, the reported F1 metric show that the SVM classifier performs best. As such, we applied the SVM algorithm, created from 80% of the labeled tweets (see Table 2 for an overview) to infer sentiment labels from the tweets related to the three modes of research. The hyper-parameter values that resulted in the classification model with best classification performance are listed in Table 5. The classification performance on the target tweets, thus the tweets related to the three modes of research, are shown in Table 6. The SVM classifier achieves an F1 score of 0.83 on the target tweets.

Table 4: Sentiment classification precision, recall and F1 for the six different classifiers. The Training set constitutes 80% of all labeled tweet data used to train the classification models. The remaining 20%, the Test set, is used to measure how well the classifiers performs on new unseen tweets.

	Training Set	**Test Set**		
Algorithm	**F1**	**Precision**	**Recall**	**F1**
SVM	0.66	0.67	0.67	**0.67**
Logisitic Regression	0.66	0.66	0.66	0.66
Multinomial NB	0.64	0.65	0.65	0.65
Bernoulli NB	0.63	0.64	0.64	0.64
Descision Trees	0.55	0.56	0.56	0.56
ADA Boos	0.62	0.64	0.64	0.63
ML-Perceptron	0.59	0.60	0.60	0.60

Table 5: Hyper-parameter values that resulted in highest F1 score for the seven explored classification algorithms.

Hyper-parameter	**SVM**	**Logisitic Regression**	**Multinomial NB**	**Bernoulli NB**	**Descision Trees**	**ADA Boost**	**ML-Perceptron**
n-grams	1–4	1–4	1–3	1–2	1–4	1–4	1–3
min-df	0	0	5	5	5	5	5
max-df	1	1	1	1	1	1	1
IDF	no	no	n	yes	no	no	no
sublinear-TF	yes	yes	yes	yes	yes	yes	yes
C	0.5	10					
fit-prior			yes	no			
alpha			0.1	1			
splitter					random		
criterion					gini		
max-features					None		
n-estimator						400	
algorithm						SAMME.R	
neural architecture							200,100,100

Table 6: Performance of the selected SVM Classifier on the target tweets.

Algorithm	**Dataset**	**Precision**	**Recall**	**F1**
SVM	1,000 int/mult/trans-disciplinary tweets	0.84	0.83	0.83

(b) Sentiment Analysis

The largest set of data, over 47,000 tweets, was collected for the interdisciplinary mode of research, followed by multidisciplinary, and transdisciplinary with the least tweets (Figure 2). All three modes of research contained more positive than negative tweets. Interdisciplinary tweets contained the largest percentage of positive tweets with almost half of the tweets being positive (45%), while the transdisciplinary and multidisciplinary modes of research contained less than 20% positive tweets.

The percentage of negative tweets was relatively low, with similar percentages between two and three percent amongst all the different modes of research. Neutral was the most common sentiment in all three modes, including more than 50% of the interdisciplinary tweets, and around 80% of the transdisciplinary and multidisciplinary tweets (Figure 2).

The full dataset, including all assigned sentiment labels is made available in the supplementary material in Data S2.

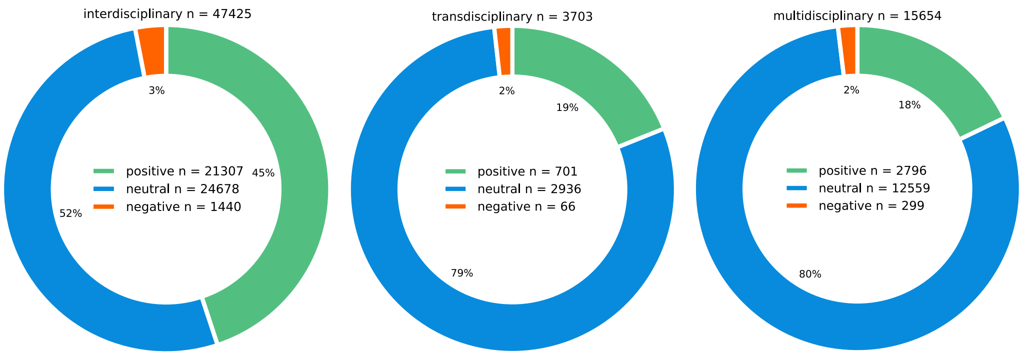


Figure 2: Frequency of tweets by sentiment for the three modes of research.

The different modes of research showed a high number of unique user names, which refers to the number of individual user accounts that tweeted within our dataset (Table 7). The ratio for the number of tweets per unique user is between one or 1.6 for all research modes and sentiments, which shows that the number of unique user IDs is close to the number of tweets posted (i.e. between one and 1.6 tweets per user). Within all three modes of research, neutral tweets showed the highest ratio, and negative tweets the lowest ratio.

Table 7: Number of unique user names for each mode of research and sentiment class and the ratio for the number of tweets per unique user.

Mode of research	**Sentiment**	**Unique users**	**Ratio**
Interdisciplinary	Negative	1,259	1.13
	Neutral	15,466	1.58
	Postive	15,022	1.41
Transdisciplinary	Negative	56	1.18
	Neutral	1,967	1.49
	Postive	584	1.20
Multidisciplinary	Negative	275	1.08
	Neutral	8,741	1.43
	Postive	2,445	1.14

The absolute number of tweets per week changed slightly over the course of the study period for both interdisciplinary and multidisciplinary tweets (Figure 3). The interdisciplinary tweets show that almost half of the tweets had a positive sentiment at times. Transdisciplinary tweets showed the highest fluctuation in numbers over time. All three modes of research show a drop in the number of tweets around week 52 in 2017 during what is typically called the winter holidays for everyone on the Northern hemisphere. The number of tweets also drops around week 23 in 2018 because tweets were not collected for 3 days due to server downtime. The sentiment and the proportions of positive, neutral and negative tweets stayed relatively proportional to each other over the study period.

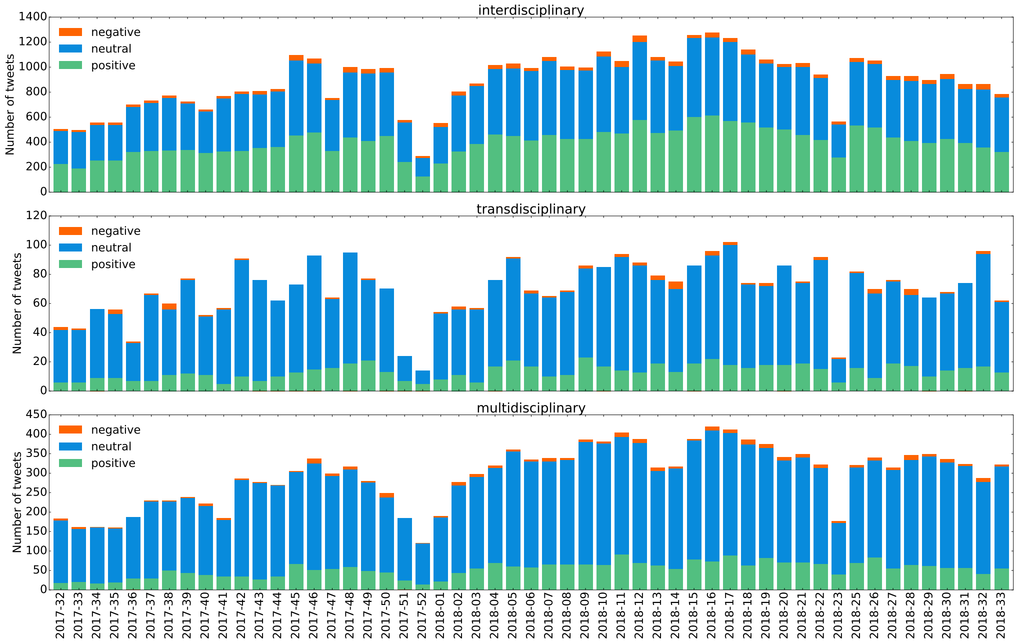


Figure 3: Sentiment over time

The manual inspection of the 2,000 randomly selected tweets provided detailed insights into the content of the tweets, and the reasons for a particular sentiment. Word clouds for each mode of research and sentiment summarize the most frequently used terms in each tweet class (Figure 4).

Negative tweets. Several of the negative tweets within all three modes of research contained explicit language. Negative tweets had a higher frequency in the use of user tags (@), compared to the use of URLs and emojis, but show similar results for each of the three modes of research (Figure 5). Negativity is associated with the multi, inter- or trans-disciplinarity itself, where researchers explicitly state that they do not enjoy the approach. Additionally, tweets reflect the hardship that is associated with integrative research in practice and being an integrative scholar. Additionally, negativity is expressed by criticizing the people or the system that discourages integrated research approaches, rather than the mode of research.

Negative interdisciplinary tweets most often discuss challenges of interdisciplinary research (IDR), such as rejections by peer-review (Figure 6a), a lack of integration, communication problems across disciplines, and difficulties to secure funding. Also criticism to the existing institutional system of disciplinary departments was mentioned. The need for more IDR was repeatedly mentioned, while the lack of acknowledgment and respect for IDR was also a re-occurring topic.

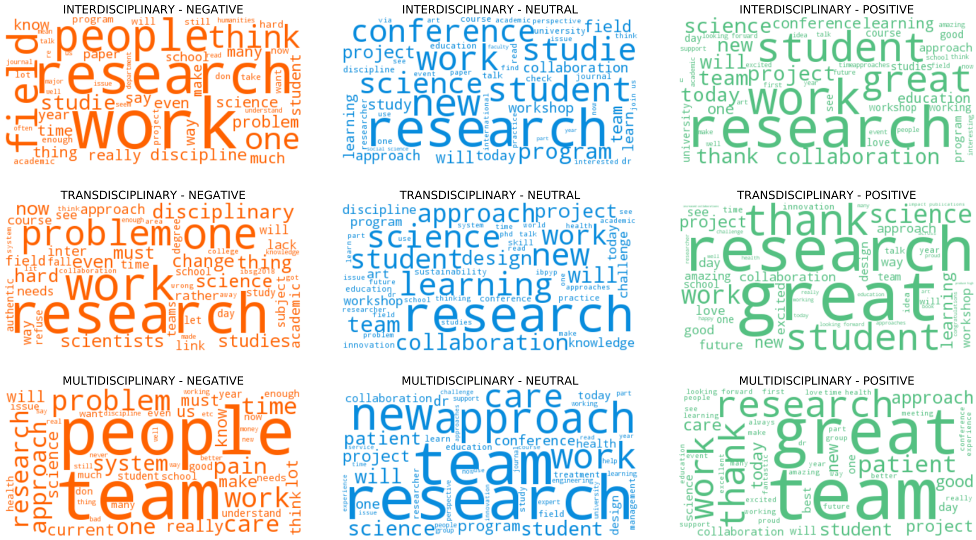


Figure 4: Word cloud displaying the most frequent words per sentiment class for the three modes of research.

There were only very few *negative transdisciplinary* tweets (n=66). However, the tweets related mostly to challenges within transdisciplinary work, being a transdisciplinary scholar, and funding concerns (Figure 6b).

Negative multidisciplinary tweets often related to the challenges and problems in health care, patient care, and treatment. Also education and teaching were a re-occurring theme, in which the lack of multidisciplinary perspectives and teaching approaches was criticized. Challenges in publishing were also mentioned.

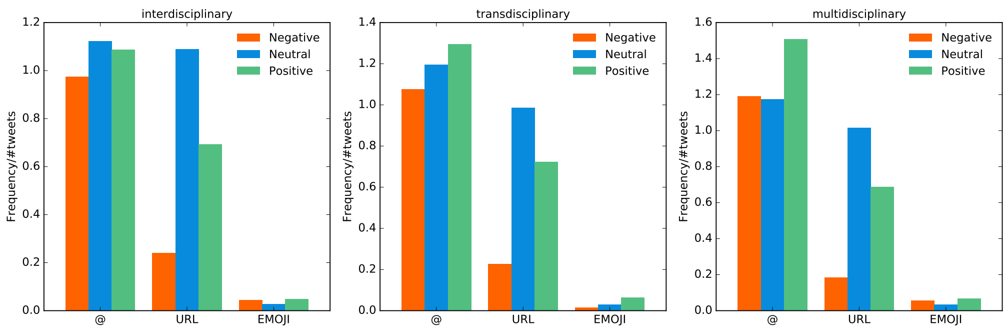
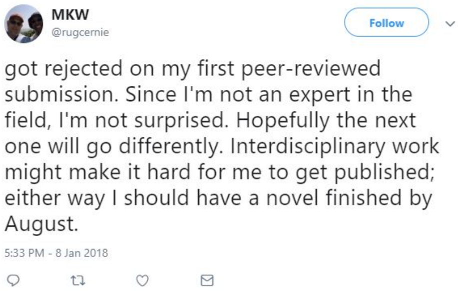
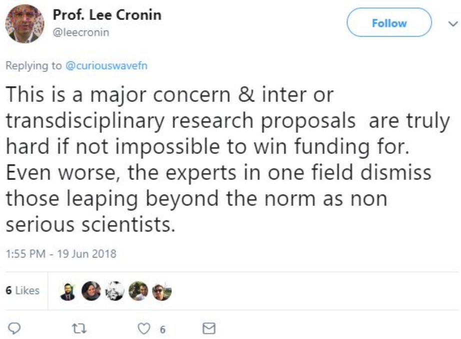


Figure 5: Relative counts of user tags (@), URLs, and emojis within the tweets of the three modes of research

Neutral tweets. The neutral tweets are mostly informative and many of them share and advertise publications, websites, or job announcements. The informative character of the neutral tweets is also apparent in the high frequency of URLs in the tweets (Figure 5). *Neutral interdisciplinary* tweets frequently share URLs referring to job postings, PhD positions, open calls for applications, news, blog posts, pod-casts, discussions, or researchers announcing a



(a) negative



(b) negative

(c) neutral

(d) neutral

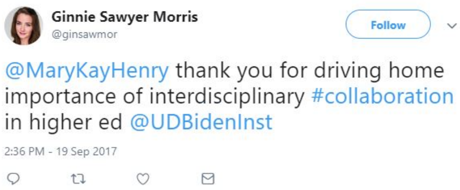
Figure 6: Examples of negative and neutral tweets.

paper publication (Figure 6c and 6d). A similar pattern is visible in the *neutral transdisciplinary* tweets, with many URLs referring to informative topics such as job posting, articles, and news publications. Similar results can also be found within the *neutral multidisciplinarity* tweets, which mainly refer to websites, news, events, articles, paper publications, and job postings.

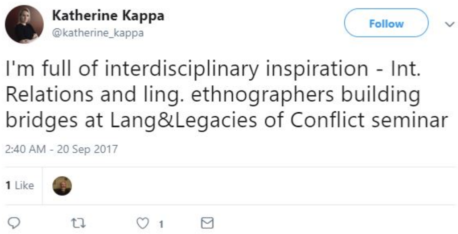
Positive tweets. The content of the positive tweets within all three modes of research showed enthusiasm and excitement by complimenting and praising different studies, lectures, approaches, and discussions. In addition, all three modes of research contained a high usage of the user tag (at least one per tweet). Emoji's were less frequently used, but appeared to be slightly more used compared to the neutral and negative tweets (Figure 5).

Many of the *positive interdisciplinary* tweets were positive about conferences, seminars, symposia, and workshops, in which researchers felt excited about it and thought that these events were interesting and useful. In particular, researchers were positive about meeting like-minded researchers and listening to inspiring talks, discussions and thoughts during conferences, seminars, and workshops. Researchers described their participation in such events as inspiring, exciting, and they felt lucky to have participated. In many of the tweets, researchers also described having fun learning, enjoyed listening to others, and appreciated critical interdisciplinary discussions. In many tweets, researchers were also thankful to the organizers and participants of these events.

The tweets also expressed positivity towards research communities, teams and collaboration (Figure 7a). Researchers were happy and excited about effective, successful, and inspiring team work, collaboration, and cooperation in integrative research projects and studies. Collaboration was also enjoyed during paper writing, seminars and workshops. Many expressed appreciation



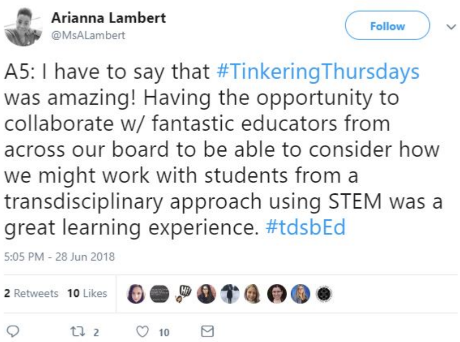
(a) importance of collaboration



(b) inspiring seminar

(c) funding acquisition

(d) work appraisal



(e) collaboration and education

(f) appreciation

(g) excitement

(h) value of collaboration

Figure 7: Examples of positive tweets referring to all three modes of research (interdisciplinary 7a–7d, transdisciplinary 7e–7g, and multidisciplinary 7h).

for cooperation and collaboration, and described how interdisciplinary collaboration can have an added value to advance research and understanding. The term 'Bridge building between disciplines' was often mentioned as a goal to strive for. Many of the positive tweets reported overall positive experiences, feelings, or praised interdisciplinary work and results (Figure 7b and 7d).

Researchers expressed their appreciation, the value and importance of integrative work, and were often in support of integrative research, as from their perspective, it can provide important solutions and promising results in different fields. For example, the fields of neuroscience, cancer research, political science, fisheries science, engineering, gender studies, cognitive science, computer science, and archaeology were mentioned. Researchers also expressed their explicit support for a certain interdisciplinary approach, conveyed their excitement about a study, and highlighted how they believe that a certain approach can address a particular challenge or solve a certain complex problem. Many described interdisciplinary work as impactful, excellent, creative and innovative, with potential for new discoveries. Others highlighted the strengths of integrative approaches, and how integrative research could potentially benefit sustainability and human well-being, such as patient care and mental health.

Tweets about a successful acquisition of funding and research grants for interdisciplinary research by individuals, teams, and research groups were also shared several times (Figure 7c). Some tweets showed excitement about having an interdiciplinary job, getting a new job within an interdisciplinary field, or the successful completion of an interdisciplinary PhD. Many described their work and research as fun and rewarding, feeling proud for their achievements and accomplishments. Researchers also appraised interdisciplinary universities and the value and importance of interdisciplinary education and training, and the benefits thereof. For example, researchers described the benefits of newly learned abilities and skills through interdisciplinary work, the interdisciplinary training that functioned as an eye-opener towards other fields, and how an interdisciplinary perspective can provide additional food-for-thought. Interdisciplinary training was also described as a way for researchers to open up alternative career paths.

Researchers also commonly shared their excitement and happiness about the successful publication of an interdisciplinary paper, a book, a book chapter, a news article, or blog post about interdisciplinary research or experiences. Some were also joyful about sharing news over the recognition of their interdisciplinary research and teaching through awards and prices.

Positive transdisciplinary tweets express positive sentiment about very similar topics found in interdisciplinary tweets, such as publication and funding successes, positive experiences from conferences and discussion, and the value of transdisciplinary teaching and education (Figure 7e). Also, positive experiences from the involvement into transdisciplinary research projects were shared and highlighted the value and importance of the work (Figure 7f and 7g).

Many of the *positive multidisciplinary* tweets referred to healthcare and medical topics (Figure 7h). Other topics included, similarly to inter- and transdisciplinary tweets, good experiences from teams and team work, positive conference experiences, a general positive attitude and excitement towards multidisciplinary work and success stories, such as winning an award.

Key Successes and Positive Experiences. In summary, the positive tweets demonstrate a number of key successes, and document positive experiences for the three modes of research.
Key successes included:

- Attaining advanced skills
- Successful publications (papers, books, etc.)
- Acquisition of funding and research grants
- Production of creative, innovative, and impactful research
- Improved research practice and research results through effective and successful team work, collaboration, and cooperation
- Research results provide useful solutions and can be of value to science and society

- Recognition of scientific work through awards and prizes

Descriptions of positive inter-, trans-, and multidisciplinary experiences can be summarized as follows:

- Conferences, seminars, and workshops provide valuable insights and inspiring talks
- Meetings with other inter-, trans- and multidisciplinary scholars provide thought-provoking and inspiring discussions
- Work and research is fun, exciting, and rewarding
- Training and learning opportunities are interesting and valuable

4. Discussion

We used a sentiment analysis to explore the opinions of researchers towards different modes of research (including inter-, trans-, and multidisciplinarity) communicated via the social media platform Twitter. Twitter provides an immense resource of text-based data covering a very large group of people [53], with an average of 328 million active monthly users who broadcast their thoughts and opinion as tweets [15]. As such, Twitter offers an abundance of easily-accessible data which has resulted in the rise and development of machine learning techniques that enabled a sentiment analysis of tweets [54]. To date, sentiment analysis on Twitter data has been conducted across a variety of disciplines and topics, ranging from computer science to environmental and medical sciences [55–59]. In addition, Twitter data has also been used to e.g. save lives during earth quakes, while organizations such as the United Nations collaborate with Twitter to achieve the sustainable development goals[5]. For example to monitor outbreaks of diseases [60]. Hence, it is increasingly being recognized that tweets can provide valuable information and insights into peoples' lives, health, and opinions.

In our study, over 70,000 tweets were collected for the different modes of research over the time span of 53 weeks. The large number of unique users (Table 7) producing a large amount of tweets (Figure 2) demonstrate that there is an active scientific community that is interested in the discourse of integrative research concepts. Proportionally, interdisciplinarity appears to be the most popular research concept with the largest number of collected tweets, which could be related to the general interdisciplinary research 'break out' over the last couple of years [61].

Negative tweets. The negative tweets highlight the challenges of integrative research that people have experienced. It is often the integrative nature of these approaches that gives rise to challenges for the researchers involved, because disciplinary boundaries have to be crossed, which can introduce institutional and cultural issues [4,5]. The detection of negative opinions within the inter- and trans-disciplinary tweets was therefore foreseen. In contrast, multidisciplinary tweets, reflecting a non-integrative research concept, were expected to have less negative tweets because multidisciplinarity is often perceived as being 'easier'. Interestingly, this hypothesis could not be confirmed in this study nor in a similar study by [62].

However, only very few of the tweets were classified as negative (2–3%), which stands in contrast to the larger literature where often the challenges and difficulties of integrative research are propagated and discussed [4,5,7,8]. Therefor, this study highlights how a sentiment analysis can offer new insights into researchers' opinions, and has the ability to identify perspectives on a larger scale that contrast the common (negative) perception and experiences typically found in the literature. However, the articles and publications in the literature are usually published by selected individuals and do not necessarily reflect the opinion of the wider research community, whereas our study captured opinions of thousands of different individuals (Table 7). In addition, not all researchers may want to share their negative experiences with their friends and colleagues on Twitter, as it is often easier to share and disseminate one's successes rather than one's failures. Yet, from a scientific perspective, it may be perceived as more valuable to identify and analyze

[5]https://developer.twitter.com/en/developer-terms/more-on-restricted-use-cases

challenges within a research practice, with the aim to understand and overcome these challenges. However, human brains tend to have a tendency towards a negative bias [63], which means that people usually have a higher sensitivity to negative information. Despite the description of positive examples and experiences in the literature, e.g. [64,65], the negative ones might be more likely to be remembered by researchers.

Neutral tweets. The majority of tweets found in this study were neutral and mostly informative, which did not reflect any particular perception or sentiment. In these cases, Twitter is being accessed as a dissemination and advertising platform, rather than a way to express an opinion, since it is generally a cheap and easy way to spread information. This is not surprising, when considering that the Twitter platform is increasingly used as a source of real-time information from news channels, politics, business, science, and entertainment, but also the personal use by individuals and organizations lies at the core of Twitter's utility to express thoughts, share information and connect with friends [66].

Communication and dissemination is additionally facilitated by being able to tweet information with just a single click through the small 'Tweet button' displayed on the website to enable viewers to share the content on Twitter [67]. Also, the latest website plug-ins allow for new posts (e.g. blog posts) to be automatically shared on a user's Twitter account [68].

Typically, scholars, which are the population of this study, tweet rather neutral information, resources, and media [66] and this explains the high amount of neutral tweets with URLs found in this study (Figure 2 and 5). Twitter is also increasingly used as a teaching and communication platform by instructors [69], who have been found to have a higher credibility among students when their tweets are professional [70]. Such educational tweets most often express no sentiment but are of a neutral nature. There has also been a gradual shift within the scientific community towards increased communication and dissemination of research and scientific results. A possible explanation could be the increasing demands for dissemination by funding agencies such as the European Commission [71]. For researchers, the Internet has become a useful way to disseminate and promote events and publications because Twitter offers a quick and easy option to disseminate scientific results, to contribute to a discussion and increase visibility via hashtags [66]. The use of Twitter by researchers for these purposes is also apparent in our results.

Positive tweets. The positive tweets demonstrated experiences of success stories of the known challenges of interdisciplinarity, such as successful funding acquisitions, interdisciplinary publications, successful and positive team work experiences, and successful interdisciplinary projects (Figure 7). Thereby, demonstrating real-life examples of how also the positive opposite of what is commonly feared is possible and attainable within interidsciplinary research.

A large amount of the tweets were positive, in which users expressed positive experiences and perceptions. The high frequency of user tags (@) (Figure 5) implies discussions between people, projects and institutions, e.g. within a circle of friends, between co-workers, or in connection with a shared field of research or project. Participants from integrative projects are often found to have positive experiences based on the team work and collaboration with other participants [22,62], which is also indicated in our results (Figure 7a). The positive tweets are likely to originate from people who are actively involved in interdisciplinary projects themselves (Figure 7g), which is also the group of researchers that has been found to describe their work as positive most of the time [62].

We hypothesize that it is many of the younger generations of scientists within our dataset that perceive interdisciplinarity as mainly positive and beneficial, more intellectually interesting, and more practically important. This is due to the fact that younger researchers have been found to show higher rates of interdisciplinarity when compared to tenure track researchers and professor [12]. The perceptions of those early career researchers could possibly be dominating the positive discourse on Twitter due to the fact that 24 to 35 year olds make up the largest age group of Twitter users [72]. Women might also make up a larger proportion of the Twitter users sampled within this study because they have a higher preference for interdisciplinary work than men [73], and might therefore be among the researchers tweeting about it. Even though the integrative research path has been described as a career risks for early researchers, interdisciplinary PhD graduates have shown higher likelihood of academic employment and higher publication productivity [74]. These scholars are likely to share their enthusiasm and success through a positive attitude and discourse on integrative research. However, age and gender distributions were not investigated in this study and cannot be confirmed at this stage.

Reason for Optimism? Overall, our study revealed that researchers have mostly positive perceptions about multi-, inter-, and trandisciplinarity (Figure 2). This study also demonstrates many examples of positive experiences that were created through successful funding, accepted publications, interesting research outcomes, and effective teamwork and collaboration, besides other aspects. This highlights that there are indeed many of the 'good experiences' and 'bright spots' to be found within these research practices. It also shows the value of this Twitter analysis, as some of these experiences may not be shared to the same extent within the literature as such. For example, publications seldom cover success stories regarding the acquisition of funding for an interdisciplinary project or the experiences of conference participants. Hence, this Twitter sentiment analysis is able to capture and quantify interdisciplinary experiences from a different perspective.

We believe that the findings of this study demonstrate and document positive experiences and opinions, and as such, give reason for optimism within integrative research approaches. Hence, this study is a first step towards building interdisciplinary optimism. The continued documentation and propagation of such 'bright spots' and successes could further increase optimistic thinking about integrative research, which may have many potential benefits.

Optimism can increase people's psychological and physical well-being [14], and facilitate and increase creativity in individuals and teams [75]. Creativity is also closely linked and thought to play an important role in people's innovation capacity [76,77], which makes it a key aspect for integrative research approaches, which are often hoped to show high innovation potential. Positive thinking and optimism are also beneficial to team work—a crucial part of most integrative approaches—and can have positive effects on team level cooperation, collaboration, and overall team outcomes [78]. Hence, the findings of this study could potentially contribute to the future success of integrative research through their propagation of optimism.

We, therefore, believe that it is important to make these results visible to the wider research community through publication and dissemination, and that additional positive experiences as well as studies of bright spots should be shared and propagated. Thereby, interdisciplinary researchers are encouraged to follow this example and to participate in the interdisciplinary discourse and the study of bright spots to support integrative research practices in the future from an optimistic perspective.

(a) Limitations and future work

An inherent limitation of Twitter is that it is not representative of the whole population and our study could be expanded to compare and contrast our results with other communication media in the future. In addition, we provided only a snapshot in time and therefore, we would encourage the study of the long-term trends of public sentiment towards integrative research approaches, along with additional investigations on the distribution of age and gender among Twitter users. We included tweets exclusively in English because it is the most common language found on Twitter but this potentially excludes all non-English discourse on interdisciplinarity.

The detection of sarcasm and irony remains a difficult and challenging task, and is a limitation of this study. However, this limitation is somewhat mitigated by drawing on a large dataset of over 70,000 tweets. In addition, we assumed people to be truthful in their tweets. There is also a risk that the use of the words multidisciplinary, interdisciplinary, and transdisciplinary may not be based on a solid understanding of these terms. In addition, it is extremely difficult to assess whether a person has a correct understanding of the terms. However, by drawing on a large dataset, which was targeting tweets from scientists and researchers, this risk is reduced. Also, tweets in our dataset referred to definitions and the differences between the modes of research, which indicates that (at least some of) the researchers are familiar with the terminology and the different meanings of the three modes of research. In addition, we assume that there are tweets in which the content refers to the three different modes of research but does not mention the terms explicitly. These tweets were not capture in our dataset. Yet, such tweets are difficult to capture and require a manual analysis and context interpretation, which was not feasible within the bounds of this study.

Another limitation includes that 'transdisciplinary' can potentially have two different meanings, i.e. (i) the inclusion of non-academic participants in interdisciplinary research, and (ii) transcending disciplinary boundaries through the development of new methods from two or more scientific disciplines, which could potentially lead to a new discipline. However, based on the manual inspection of the content of tweets, it is assumed that transdisciplinarity is most commonly used with meaning (i) within the dataset.

The current state-of-the-art in sentiment analysis typically employs neural network models [42]. Specifically, the recurrent neural network model is better designed to handle sequence data, such as text, compared to, for instance, the Multi-layer Perceptron. Recurrent neural networks have the advantage of taking word order into account, and advanced implementations such as the LSTM model [41] achieve top ranking results in sentiment classification competitions. Additionally, tweet data is, increasingly, represented as word embeddings [79], a high dimensional vector representation of words, and have shown to boost performance of sentiment classification tasks [42,80]. However, such state-of-the-art methods and architectures come with a high degree of complexity and training these models can be challenging and time consuming. It would be an interesting approach to study in future research, although the purpose of this analysis was not to build top ranking sentiment classification systems. More importantly, the employed classifier, the support vector machine, achieves near to state-of-the-art results if properly parameterized [42,80], and would adequately provide a sense of sentiments for the three modes of research studied in this paper.

5. Conclusion

For the success of integrative research approaches, it is important to foster positive thinking and optimism through the study of 'bright spots'. Bright spots can help to harvest the full potential of integrative research and to enable a feeling of empowerment among researchers engaging in these approaches. This study identified such 'bright spots' by analyzing the sentiment of tweets on inter-, trans- and multidisciplinarity where researchers expressed dominantly positive opinions (excluding neutral tweets). Positive opinions were created through and based on positive experiences and successes within integrative research, such as accepted publications, the acquisition of funding and effective team work. As such, this study demonstrates and documents positive thinking within integrative research and gives reason for optimism. The continued study of bright spots and propagation of optimism can potentially have many benefits for integrative research, and maybe hopefully inspire and empower scientists to continue and strive for integrative research in the future.

6. Supplementary material

Data S1. Full dataset of all tweets that were collected and analyzed in this study. Contains tweet IDs for each modes of research and assigned sentiment labels.
Data S2. List of 430 occupations related to a scientific or academic profession.
Text S1. Description of the training datasets.

Data Accessibility. The dataset of all extracted tweets are made available in the supporting information (Data S1).

Authors' Contributions. C.T.W. and S.S. retrieved the data. S.S. processed the data and prepared the figures. S.S. analyzed the data. C.T.W. and S.S. designed the research, interpreted the results, wrote the manuscript and gave their final approval for publication.

Competing Interests. The authors declare no competing interests.

Funding. This research was funded by the project SAF21 – Social Science Aspects of Fisheries for the 21st Century, a project financed under the EU Horizon 2020 Marie Skłodowska-Curie (MSC) ITN-ETN Program; project number: 642080.

References

1. Rice M. 2013 Spanning disciplinary, sectoral and international boundaries: a sea change towards transdisciplinary global environmental change research?. *Current Opinion in Environmental Sustainability* **5**, 409–419.
2. Ledford H. 2015 How to solve the world's biggest problems. *Nature* **525**, 308–311.
3. Klein JT. 1990 *Interdisciplinarity: History, Theory, and Practice*. Detroit: Wayne State University Press.
4. Buanes A, Jentoft S. 2009 Building bridges: Institutional perspectives on interdisciplinarity. *Futures* **41**, 446–454.
5. Chiu CY, Kwan LYY, Liou S. 2013 Culturally motivated challenges to innovations in integrative research: Theory and solutions. *Social Issues and Policy Review* **7**, 149–172.
6. Lang DJ, Wiek A, Bergmann M, Stauffacher M, Martens P, Moll P, Swilling M, Thomas CJ. 2012 Transdisciplinary research in sustainability science: practice, principles, and challenges. *Sustainability Science* **7**, 25–43.
7. Pischke EC, Knowlton JL, Phifer CC, Gutierrez Lopez J, Propato TS, Eastmond A, de Souza TM, Kuhlberg M, Picasso Risso V, Veron SR, Garcia C, Chiappe M, Halvorsen KE. 2017 Barriers and Solutions to Conducting Large International, Interdisciplinary Research Projects. *Environmental Management* **60**, 1011–1021.
8. Bromham L, Dinnage R, Hua X. 2016 Interdisciplinary research has consistently lower funding success. *Nature* **534**, 684–687.
9. Huston AC. 2012 A Path to Inderdisciplinary Scholarship. In Campbell L, Loving TJ, editors, *Interdisciplinary research on close relationships: The case for integration* , pp. 253–272. Washington DC: American Psychological Association.
10. McLeish T, Strang V. 2016 Evaluating interdisciplinary research: the elephant in the peer-reviewers' room. *Palgrave Communications* **2**, 1–8.
11. Fischer EV, Mackey KRM, Cusack DF, DeSantis LRG, Hartzell-Nichols L, Lutz JA, Melbourne-Thomas J, Meyer R, Riveros-Iregui DA, Sorte CJB, Taylor JR, White SA. 2012 Is pretenure interdisciplinary research a career risk?. *Eos Transactions American Geophysical Union* **93**, 311–312.
12. Rhoten D, Parker A. 2004 Risks and rewards of an interdiciplinary research path. *Science* **306**, 2046.
13. Pfirman S, Begg M. 2012 Perspective: Troubled by Interdisciplinarity?. *Science*.
14. Scheier MF, Carver CS. 1992 Effects of optimism on psychological and physical well-being: Theoretical overview and empirical update. *Cognitive Therapy and Research* **16**, 201–228.
15. Agarwal A, Xie B, Vovsha I, Rambow O, Passonneau R. 2011 Sentiment analysis of Twitter data. In *Proceedings of the Workshop on Languages in Social Media* pp. 30–38.
16. Mohammadi E, Thelwall M, Kwasny M, Holmes KL. 2018 Academic information on Twitter: A user survey. *PLOS ONE* **13**, e0197265.
17. Tress B, Tress G, Fry G. 2005 Defining concepts and the process of knowledge production in integrative research. In Tress B, Tress G, Fry G, Opdam P, editors, *From landscape research to landscape planning: aspects of integration, education and application* , pp. 13–26. Springer Netherlands.
18. Klein JT. 2010 A taxonomy of interdisciplinarity. In Frodeman R, editor, *The Oxford Handbook of Interdisciplinarity* , . Oxford, UK: Oxford University Press 1 edition.
19. Feller I. 2007 Interdisciplinarity: Paths Taken and Not Taken.. *Change* **39**, 46–51.
20. Borrego M, Newswander L. 2010 Definitions of Interdisciplinary Research: Toward Graduate-Level Interdisciplinary Learning Outcomes. *The Review of Higher Education* **34**, 61–84.
21. Hall P, Weaver L. 2001 Interdisciplinary education and teamwork: a long and winding road. *Medical Education* **35**, 867–875.
22. Tress B, Tress G, Fry G. 2005 Researchers' Experiences, Positive and Negative, in Integrative Landscape Projects. *Environmental Management* **36**, 792–807.
23. Ke Q, Ahn YY, Sugimoto CR. 2017 A systematic identification and analysis of scientists on Twitter. *PLOS ONE* **12**, e0175368.

24. Miller GA. 1995 WordNet: A Lexical Database for English George A. Miller. In *Communications of the Acm* pp. 39–41.
25. Salton G, McGill MJ. 1982 *Introduction to Modern Information Retrieval*. New York, NY, USA: McGraw-Hill, Inc.
26. Bravo-Marquez F, Mendoza M, Poblete B. 2013 Combining strengths, emotions and polarities for boosting Twitter sentiment analysis. In *Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining - WISDOM '13* pp. 1–9 New York, NY, USA. ACM Press.
27. Liu KL, Li WJ, Guo M. 2012 Emoticon Smoothed Language Models for Twitter Sentiment Analysis. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence* pp. 1678–1684 Toronto, Ontario, Canada. AAAI Press.
28. Deitrick W, Hu W. 2013 Mutually Enhancing Community Detection and Sentiment Analysis on Twitter Networks. *Journal of Data Analysis and Information Processing* **01**, 19–29.
29. Hu X, Tang L, Tang J, Liu H. 2013 Exploiting social relations for sentiment analysis in microblogging. In *Proceedings of the sixth ACM international conference on Web search and data mining - WSDM '13* pp. 537–546 New York, New York, USA. ACM Press.
30. Speriosu M, Sudan N, Upadhyay S, Baldridge J. 2011 Twitter Polarity Classification with Label Propagation over Lexical Links and the Follower Graph. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* pp. 53–56 Stroudsburg, PA, USA. Association for Computational Linguistics.
31. Saif H, He Y, Alani H. 2012 Semantic Sentiment Analysis of Twitter. In *Proceedings of the 11th international conference on The Semantic Web* vol. 1 pp. 508–524 Berlin, Heidelberg. Springer-Verlag.
32. Kalchbrenner N, Grefenstette E, Blunsom P. 2014 A Convolutional Neural Network for Modelling Sentences. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* pp. 655–665 Stroudsburg, PA, USA. Association for Computational Linguistics.
33. Saif H, He Y, Alani H. 2011 Semantic smoothing for Twitter sentiment analysis. In *Proceeding of the 10th International Semantic Web Conference (ISWC)* pp. 23–27 Bonn, Germany. Springer.
34. Barbosa L, Feng J. 2010 Robust Sentiment Detection on Twitter from Biased and Noisy Data. In *Proceedings of the 23rd International Conference on Computational Linguistics* pp. 36–44 Beijing, China. Association for Computational Linguistics.
35. Xu S, Liang H, Baldwin T. 2016 UNIMELB at SemEval-2016 Tasks 4A and 4B: An Ensemble of Neural Networks and a Word2Vec Based Model for Sentiment Classification. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)* pp. 183–189 Stroudsburg, PA, USA. Association for Computational Linguistics.
36. Zheludev I, Smith R, Aste T. 2015 When Can Social Media Lead Financial Markets?. *Scientific Reports* **4**, 4213.
37. Garas A, Garcia D, Skowron M, Schweitzer F. 2012 Emotional persistence in online chatting communities. *Scientific Reports* **2**, 402.
38. Mozetič I, Grčar M, Smailović J. 2016 Multilingual Twitter Sentiment Classification: The Role of Human Annotators. *PLOS ONE* **11**, e0155036.
39. Mozetič I, Torgo L, Cerqueira V, Smailović J. 2018 How to evaluate sentiment classifiers for Twitter time-ordered data?. *PLOS ONE* **13**, e0194317.
40. Baziotis C, Pelekis N, Doulkeridis C. 2017 DataStories at SemEval-2017 Task 4: Deep LSTM with Attention for Message-level and Topic-based Sentiment Analysis. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)* pp. 747–754 Vancouver, Canada. Association for Computational Linguistics.
41. Cliche M. 2017 BB_twtr at SemEval-2017 Task 4: Twitter Sentiment Analysis with CNNs and LSTMs. In *Proceedings of the 11th International Workshop on Semantic Evaluations (SemEval-2017)* pp. 573–580 Vancouver, Canada. Association for Computational Linguistics.
42. Rosenthal S, Farra N, Nakov P. 2017 SemEval-2017 Task 4: Sentiment Analysis in Twitter. In *Proceedings of the 11th International Workshop on Semantic Evaluations (SemEval-2017)* pp. 502–518 Vancouver, Canada. Association for Computational Linguistics.

43. Mohammad SM, Kiritchenko S, Zhu X. 2013 NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets. In *Proceedings of the seventh international workshop on Semantic Evaluation Exercises (SemEval-2013)* pp. 321–327.
44. Go A, Bhayani R, Huang L. 2009 Twitter Sentiment Classification using Distant Supervision. Technical report Stanford.
45. Balikas G, Amini MR. 2016 TwiSE at SemEval-2016 Task 4: Twitter Sentiment Classification. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)* pp. 85–91 San Diego, California. Association for Computational Linguistics.
46. Thelwall M, Buckley K, Paltoglou G, Cai D, Kappas A. 2010 Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology* **61**, 2544–2558.
47. Kearns MJ. 1990 *The Computational Complexity of Machine Learning*. Cambridge, MA, USA: MIT press.
48. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay É. 2011 Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **12**, 2825–2830.
49. Bergstra J, Yamins DLK, Cox DD. 2013 Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures. In *Proceedings of the 30th International Conference on Machine Learning* pp. 115–123 Atlanta, Georgia, USA. JMLR.org.
50. Shamma Da, Kennedy L, Churchill EF. 2009 Tweet the debates. In *Proceedings of the first SIGMM workshop on Social media* pp. 3–10 Beijing, China. ACM Press.
51. Mohammad SM, Turney PD. 2013 Crowdsourcing a word-emotion association lexicon. In *Computational Intelligence* vol. 29 pp. 436–465.
52. Ramadhani RA, Indriani F, Nugrahadi DT. 2017 Comparison of Naive Bayes smoothing methods for Twitter sentiment analysis. In *2016 International Conference on Advanced Computer Science and Information Systems, ICACSIS 2016* pp. 287–292.
53. O'Connor B, Balasubramanyan R, Routledge BR, Smith NA. 2010 From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. In *Proceeding of the Fourth International AAAI Conference on Weblogs and Social Media* p. 8.
54. Pang B, Lee L. 2008 *Opinion mining and sentiment analysis*. Now Publishers Inc. vol 2 edition.
55. Froehlich HE, Gentry RR, Rust MB, Grimm D, Halpern BS. 2017 Public perceptions of aquaculture: Evaluating spatiotemporal patterns of sentiment around the world. *PLoS ONE* **12**, e0169281.
56. Kharde VA, Sonawane S. 2016 Sentiment Analysis of Twitter Data: A Survey of Techniques. *International Journal of Computer Applications* **139**, 5–15.
57. Cody EM, Reagan AJ, Mitchell L, Dodds PS, Danforth CM. 2015 Climate change sentiment on Twitter: An unsolicited public opinion poll. *PLoS ONE* **10**, e0136092.
58. Oscar N, Fox PA, Croucher R, Wernick R, Keune J, Hooker K. 2017 Machine Learning, Sentiment Analysis, and Tweets: An Examination of Alzheimer's Disease Stigma on Twitter. *The Journals of Gerontology Series B Psychological Sciences and Social Sciences* **72**, 742–751.
59. Moh M, Moh TS, Peng Y, Wu L. 2017 On adverse drug event extractions using twitter sentiment analysis. *Network Modeling Analysis in Health Informatics and Bioinformatics* **6**, 18.
60. Odlum M, Yoon S. 2015 What can we learn about the Ebola outbreak from tweets?. *American Journal of Infection Control* **43**, 563–571.
61. Gewin V. 2014 Interdisciplinary research: Break out. *Nature* **511**, 371–373.
62. Bruun H, Hukkinen J, Huutoniemi K, Klein JT. 2005 *Promoting interdisciplinary research: The case of the academy of Finland* vol. 8.
63. Ito TA, Larsen JT, Smith NK, Cacioppo JT. 1998 Negative information weighs more heavily on the brain: The negativity bias in evaluative categorizations.. *Journal of Personality and Social Psychology* **75**, 887–900.
64. Thompson MA, Owen S, Lindsay JM, Leonard GS, Cronin SJ. 2017 Scientist and stakeholder perspectives of transdisciplinary research: Early attitudes, expectations, and tensions. *Environmental Science & Policy* **74**, 30–39.

65. Ommer RE. 2018 Curiosity, interdisciplinarity, and giving back. *ICES Journal of Marine Science* **fsy022**.
66. Weller K, Bruns A, Burgess J, Mahrt M, Puschmann C, editors. 2014 *Twitter and Society*. New York: Peter Lang 89 edition.
67. Twitter Inc.. 2017 Tweet Button. .
68. WordPress. 2018 Social Tools: Twitter. .
69. Carpenter JP, Krutka DG. 2014 How and why educators use Twitter: A survey of the field. *Journal of Research on Technology in Education* **46**, 414–434.
70. DeGroot JM, Young VJ, VanSlette SH. 2015 Twitter Use and its Effects on Student Perception of Instructor Credibility. *Communication Education* **64**, 419–437.
71. European IPR Helpdesk. 2015 Fact Sheet: The Plan for the Exploitation and Dissemination of Results in Horizon 2020. Technical report.
72. Statista. 2018 Distribution of Twitter users in the United States as of December 2016, by age group. .
73. Rhoten D, Pfirman S. 2007 Women in interdisciplinary science: Exploring preferences and consequences. *Research Policy* **36**, 56–75.
74. Millar MM. 2013 Interdisciplinary research and the early career: The effect of interdisciplinary dissertation research on career placement and publication productivity of doctoral graduates in the sciences. *Research Policy* **42**, 1152–1164.
75. Rego A, Sousa F, Marques C, Cunha MPe. 2012 Optimism predicting employees' creativity: The mediating role of positive affect and the positivity ratio. *European Journal of Work and Organizational Psychology* **21**, 244–270.
76. Zhang X, Bartol KM. 2010 Linking Empowering Leadership and Employee Creativity: The Influence of Psychological Empowerment, Intrinsic Motivation, and Creative Process Engagement. *Academy of Management Journal* **53**, 107–128.
77. Tang M, Werner CH. 2017 An interdisciplinary and intercultural approach to creativity and innovation: Evaluation of the EMCI ERASMUS intensive program. *Thinking Skills and Creativity* **24**, 268–278.
78. West BJ, Patera JL, Carsten MK. 2009 Team level positivity: investigating positive psychological capacities and team level outcomes. *Journal of Organizational Behavior* **30**, 249–267.
79. Mikolov T, Sutskever I, Chen K, Corrado G, Dean J, Chen K, Dean J, Mikolov T, Chen K. 2013 Distributed Representations of Words and Phrases and their Compositionality.. In *Proceedings of the 26th International Conference on Neural Information Processing Systems* vol. 2 pp. 3111–3119 Lake Tahoe, Nevada. Curran Associates Inc.
80. Nakov P, Ritter A, Rosenthal S, Stoyanov V, Sebastiani F. 2016 SemEval-2016 Task 4: Sentiment Analysis in Twitter. In *Proceedings of the 10th International Workshop on Semantic Evaluation* pp. 1–18 San Diego, California. Association for Computational Linguistics.

Supplementary Text S1. Descriptions of the seven datasets used to train the sentiment classifier.

Sanders The Sanders dataset consists out of 5,513 hand classified tweets related to the topics Apple (@Apple), Google (#Google), Microsoft (#Microsoft), and Twitter (#Twitter). Tweets were classified as positive, neutral, negative, or irrelevant; the latter referring to non-English tweets which we discarded. The Sanders dataset has been used for boosting Twitter sentiment classification using different sentiment dimensions [1], combining automatically and hand-labeled twitter sentiment labels [2], and combining community detection and sentiment analysis [3]. The dataset is available from `http://www.sananalytics.com/lab/`.

Obama-McCain Debate (OMD) The Obama-McCain Debate (OMD) dataset contains 3,238 tweets collected in September 2008 during the United States presidential debates between Barack Obama and John McCain. The tweets were collected by querying the Twitter API for the hash tags #tweetdebate, #current, and #debate08 [4, 5]. A minimum of three independent annotators rated the tweets as positive, negative, mixed, or other. Mixed tweets captured both negative and positive components. Other tweets contained non-evaluative statements or questions. We only included the positive and negative tweets with at least two-thirds agreement between annotators ratings; mixed and other tweets were discarded. The OMD dataset has been used for sentiment classification by social relations [6], polarity classification [7], and sentiment classification utilizing semantic concept features [8]. The dataset is available from `https://bitbucket.org/speriosu/updown`.

Stanford Test The Stanford Test dataset contains 182 positive, 139 neutral, and 177 negative annotated tweets [9]. The tweets were labeled by a human annotator and were retrieved by querying the Twitter search API with randomly chosen queries related to consumer products, company names and people. The Stanford Training dataset, in contrast to the Stanford Test dataset, contains 1.6 million labeled tweets. However, the 1.6 million tweets were automatically labeled, thus without a human annotator, by looking at the presence of emoticons. For example, tweets that contained the positive emoticon :-) would be assigned a positive label, regardless of the remaining content of the tweet. Similarly, tweets that contained the negative emoticon :-(would be assigned a negative label. Such an approach is highly biased [10] and we choose not to include this dataset for the purpose of creating a sentiment classifier from labeled tweets. The Stanford Test dataset, although relatively small, has been used to analyze and represent the semantic content of a sentence for purposes of classification or generation [11], semantic smoothing to alleviate data sparseness problem for sentiment analysis [12], and sentiment detection of biased and noisy tweets [13]. The dataset is available from `http://www.sentiment140.com/`.

Health Care Reform (HCR) The Health Care Reform (HCR) dataset was created in 2010 – around the time the health care bill was signed in the United States – by extracting tweets with the hashtag #hcr [7]. The tweets were manually annotated by the authors by assigning the labels positive, negative, neutral, unsure, or irrelevant. The dataset was split into training, development and test data. We combined the three different datasets that contained a total of 537 positive, 337 neutral, and 886 negative tweets. The tweets labeled as irrelevant or unsure were not included. The HCR dataset was used to improve sentiment analysis by adding semantic features to tweets [8]. The dataset is available from `https://bitbucket.org/speriosu/updown`.

SemEval-2016 The Semantic Analysis in Twitter Task 2016 dataset, also known as SemEval-2016 Task 4, was created for various sentiment classification tasks. The tasks can be seen as challenges where teams can compete amongst a number of sub-tasks, such as classifying tweets into positive, negative and neutral sentiment, or estimating distributions of sentiment classes. Typically, teams with better classification accuracy or other performance measure rank higher. The dataset consist of training, development, and development-test data that combined consist of 3,918 positive, 2,736 neutral, and 1,208 negative tweets. The original dataset contained a total of 10,000 tweets – 100 tweets from 100 topics. Each tweet was labeled by 5 human annotators and only tweets for which 3 out of 5 annotators agreed on their sentiment label were considered. For a full description of the dataset and annotation process see [14]. The dataset is available from `http://alt.qcri.org/semeval2016/task4/`.

Sentiment Strength (SS) The Sentiment Strength (SS) dataset was used to detect the strength of sentiments expressed in social web texts, such as tweets, for the sentiment strength detection program SentiStrength [15]. The dataset was labeled by human annotators and each tweet was rated on a scale from 1 to 5 for both positive and negative sentiment, i.e. a dual positive-negative scale. For the purpose of this paper, we re-labeled the tweets into positive, negative and neutral tweets as follows. Tweets were considered positive if the positive score was at least 1.5 times larger than the negative score; a positive score of 4 and a negative score of 1 would result in a positive label. Tweets that have a negative score of 1.5 times larger than the positive score were considered negative. A similar score on the positive and negative scale would result in a neutral tweet, such when the positive score is 2 and the negative score 2. A similar re-labeling process was performed by [10]. A total of 1,252 positive, 1,952 neutral, and 861 negative tweets were used. SentiStrength has been used to quantify and statistically validate trading assets from social media data [16], and analyzing emotional expressions and social norms in online chat communities [17]. The dataset is available from `http://sentistrength.wlv.ac.uk/documentation/`

CLARIN 13-Languages The CLARIN 13-languages dataset contains a total of 1.6 million labeled tweets from 13 different languages, the largest sentiment corpus made publicly available [18]. We used the English subset of the dataset since we restricted our analysis to English tweets. Tweets were collected in September 2013 by using the Twitter Streaming API to obtain a random sample of 1% of all publicly available tweets. The tweets were manually annotated by assigning a positive, neutral, or negative label by a total of 9 annotators; some tweets were labeled by more than 1 annotator or twice by the same annotator. For tweets with multiple annotations, only those with two-third agreement were kept. The original English dataset contained around 90,000 labeled tweets. After recollection, a total of 15,064 positive, 24,263 neutral, and 12,936 negative tweets were obtained. The dataset is available from `http://hdl.handle.net/11356/1054`.

References

[1] Bravo-Marquez F, Mendoza M, Poblete B. 2013 Combining strengths, emotions and polarities for boosting Twitter sentiment analysis. In *Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining - WISDOM '13* pp. 1–9 New York, NY, USA. ACM Press.

[2] Liu KL, Li WJ, Guo M. 2012 Emoticon Smoothed Language Models for Twitter Sentiment Analysis. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence* pp. 1678–1684 Toronto, Ontario, Canada. AAAI Press.

[3] Deitrick W, Hu W. 2013 Mutually Enhancing Community Detection and Sentiment Analysis on Twitter Networks. *Journal of Data Analysis and Information Processing* **01**, 19–29.

[4] Shamma Da, Kennedy L, Churchill EF. 2009 Tweet the debates. In *Proceedings of the first SIGMM workshop on Social media* pp. 3–10 Beijing, China. ACM Press.

[5] Diakopoulos NA, Shamma DA. 2010 Characterizing debate performance via aggregated twitter sentiment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* pp. 1195–1198 Atlanta, Georgia, USA. ACM Press.

[6] Hu X, Tang L, Tang J, Liu H. 2013 Exploiting social relations for sentiment analysis in microblogging. In *Proceedings of the sixth ACM international conference on Web search and data mining - WSDM '13* pp. 537–546 New York, New York, USA. ACM Press.

[7] Speriosu M, Sudan N, Upadhyay S, Baldridge J. 2011 Twitter Polarity Classification with Label Propagation over Lexical Links and the Follower Graph. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* pp. 53–56 Stroudsburg, PA, USA. Association for Computational Linguistics.

[8] Saif H, He Y, Alani H. 2012 Semantic Sentiment Analysis of Twitter. In *Proceedings of the 11th international conference on The Semantic Web* vol. 1 pp. 508–524 Berlin, Heidelberg. Springer-Verlag.

[9] Go A, Bhayani R, Huang L. 2009 Twitter Sentiment Classification using Distant Supervision. Technical report CS224N Project Report, Stanford.

[10] Saif H, Fernandez M, He Y, Alani H. 2013 Evaluation datasets for Twitter sentiment analysis a survey and a new dataset, the STS-Gold. In *Proceedings of the 1st Workshop on Emotion and Sentiment in Social and Expressive Media (ESSEM)* vol. 1096 pp. 9–21 Turin, Italy. CEUR.

[11] Kalchbrenner N, Grefenstette E, Blunsom P. 2014 A Convolutional Neural Network for Modelling Sentences. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* pp. 655–665 Stroudsburg, PA, USA. Association for Computational Linguistics.

[12] Saif H, He Y, Alani H. 2011 Semantic smoothing for Twitter sentiment analysis. In *Proceeding of the 10th International Semantic Web Conference (ISWC)* pp. 23–27 Bonn, Germany. Springer.

[13] Barbosa L, Feng J. 2010 Robust Sentiment Detection on Twitter from Biased and Noisy Data. In *Proceedings of the 23rd International Conference on Computational Linguistics* pp. 36–44 Beijing, China. Association for Computational Linguistics.

[14] Nakov P, Ritter A, Rosenthal S, Stoyanov V, Sebastiani F. 2016 SemEval-2016 Task 4: Sentiment Analysis in Twitter. In *Proceedings of the 10th International Workshop on Semantic Evaluation* pp. 1–18 San Diego, California. Association for Computational Linguistics.

[15] Thelwall M. 2017 The Heart and Soul of the Web? Sentiment Strength Detection in the Social Web with SentiStrength. In Holyst J, editor, *Cyberemotions. Understanding Complex Systems* , pp. 119–134. Switzerland: Springer International Publishing.

[16] Zheludev I, Smith R, Aste T. 2015 When Can Social Media Lead Financial Markets?. *Scientific Reports* **4**, 4213.

[17] Garas A, Garcia D, Skowron M, Schweitzer F. 2012 Emotional persistence in online chatting communities. *Scientific Reports* **2**, 402.

[18] Mozetič I, Grčar M, Smailović J. 2016 Multilingual Twitter Sentiment Classification: The Role of Human Annotators. *PLOS ONE* **11**, e0155036.