



Full length article

What makes you tick? The psychology of social media engagement in space science communication



Yi-Ling Hwong^{a, *}, Carol Oliver^a, Martin Van Kranendonk^a, Claude Sammut^b,
Yanir Seroussi^c

^a Australian Centre for Astrobiology, School of Biological, Earth and Environmental Sciences, University of New South Wales (UNSW), Sydney, Australia

^b School of Computer Science Engineering, University of New South Wales (UNSW), Sydney, Australia

^c Independent Researcher

ARTICLE INFO

Article history:

Received 3 June 2016

Received in revised form

30 November 2016

Accepted 30 November 2016

Available online 1 December 2016

Keywords:

Science communication

Social media

Machine learning

Psychometrics

Facebook

Twitter

ABSTRACT

The rise of social media has transformed the way the public engages with science organisations and scientists. ‘Retweet’, ‘Like’, ‘Share’ and ‘Comment’ are a few ways users engage with messages on Twitter and Facebook, two of the most popular social media platforms. Despite the availability of big data from these digital footprints, research into social media science communication is scant. This paper presents a novel empirical study into the features of engaging science-related social media messages, focusing on space science communications. It is hypothesised that these messages contain certain psycholinguistic features that are unique to the field of space science. We built a predictive model to forecast the engagement levels of social media posts. By using four feature sets (n-grams, psycholinguistics, grammar and social media), we were able to achieve prediction accuracies in the vicinity of 90% using three supervised learning algorithms (Naive Bayes, linear classifier and decision tree). We conducted the same experiments on social media messages from three other fields (politics, business and non-profit) and discovered several features that are exclusive to space science communications: anger, authenticity, hashtags, visual descriptions—be it visual perception-related words, or media elements—and a tentative tone.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Every second more than 7000 tweets are sent out on Twitter ([Internet Live Stats, 2016](#)). If the users of the popular social network website Facebook were a country, it would be the most populous country in the world, with a citizenry of 1.65 billion people ([Facebook Inc., 2016](#)). This flood of information presents new opportunities for evidence-based research into science communication ([Brossard & Scheufele, 2013](#)). Scientific data is now accessible to anyone with an internet connection, granting the public spectator status to the making of science ([Jepson, 2014](#)). Social media enables fast-paced discussions on important topics between scientists and the public ([Bik & Goldstein, 2013](#)). Audiences also actively produce and curate content, while simultaneously serving as gatekeepers to their own community by evaluating and selectively disseminating information ([Sandu & Christensen, 2011](#), pp.

22–32).

In this paper, two studies are presented. They investigate the features of engaging space science-related social media messages using supervised learning (Study 1), and examine if these features are exclusive to the field of space science (Study 2). Our objective is to provide scientists and scientific institutions with insights into the ingredients of successful social media engagement in science communication (‘What makes your audience tick?’) that will ultimately help them effectively communicate their messages to the general public.

1.1. Space science and social media

Our study of social media science communication focuses on space science to allow us a more targeted approach in data collection and result interpretation. NASA’s Twitter account

* Corresponding author.

E-mail address: y.hwong@unsw.edu.au (Y.-L. Hwong).

commands a following of a staggering 16 million users¹; its Facebook page has close to 16 million likes.² Popular astrophysicist and science communicator Neil deGrasse Tyson's personal Twitter account has more than five million followers³ and his Facebook page has close to four million likes.⁴

Space science is intrinsically visually exciting because it unites cutting-edge technology with the visual splendour of our universe. Our brains are attuned to react to visual information (Delello & McWhorter, 2013). By making photo and video sharing an integral function, social media has become an indispensable and formidable tool for science communicators to capture the attention of their audience. In 2012, NASA landed the Curiosity rover on Mars. The accompanying social media activities generated 1.2 billion tweets, 17.4 million Facebook interactions and 36.4 million webcast streams (Pinholster & Ham, 2013). Three years later, the New Horizons spacecraft flew by Pluto. On 14 July 2015, NASA released the first ever close-up picture of Pluto's surface on its Instagram account. In the first 3 h, the image gathered more than 370,000 interactions (including shares and comments) and 142,000 likes (McCulloch, 2015); the associated Pluto Flyby messages reached 38.6 million people on Twitter and 29.9 million people on Facebook (NASA, 2015).

However, what precisely makes these messages so engaging is still poorly understood. The process and outcomes of science communication on social media have never been critically and systematically analysed (Brossard, 2013). Given the scarcity of systematic analysis into the mechanisms of social media science communication, current science communication practices are based on intuition and experiential rather than empirical evidence.

1.2. Machine learning and psychometrics

Digital traces and patterns emerging from the usage of social media platforms can be studied using sophisticated computational techniques such as machine learning to produce insights into audience behaviour (Lazer et al., 2009). Machine learning is a subdomain in computer science that focuses on constructing algorithms to analyse and learn the hidden patterns in data and make predictions based on these analyses (Bishop, 2006). By teaching machines how to learn, researchers no longer need to explicitly program computers to complete a particular task. The value of machine learning lies in its ability to uncover patterns and correlations from data sets that are large, diverse and fast changing—e.g. social media streams—and to create accurate predictive models to guide future actions (Bishop, 2006).

Retweet, like, share and comment are some of the commonest ways users engage with messages on Twitter and Facebook. Consequently, these metrics are a reflection of the interestingness or value of a post (Naveed, Gottron, Kunegis, & Alhadi, 2011). It is therefore important to know what are the ingredients of social media messages that prompt active user engagement.

Previous research has demonstrated that it is possible to use supervised learning—a type of machine learning that creates predictive models from labelled data—to predict the popularity of social media messages by using contextual (e.g. number of followers) and content-based (e.g. sentiment, URL, hashtags) features (Naveed et al., 2011; Petrovic, Osborne, & Lavrenko, 2011). The extent to which content-based features are used has mostly been limited to more 'obvious' type of features such as the presence of

URLs and hashtags. However, newest advances in psychometrics have opened up a way forward for researchers to conduct more in-depth investigations of social media mechanisms using psycholinguistic features.

1.3. Aim of study

The primary aim of this paper is to examine the features of engaging space science-related social media messages and investigate in particular the possibility of using content-based features—especially psycholinguistic features—to predict these messages. A secondary aim is to investigate whether these features are unique to the field of space science. Our project was divided into two studies to achieve these aims.

1.3.1. Study 1

The first study focused on using supervised learning to study the predictive power of content-related features in foretelling the engagement potential of space science-related social media messages. We formulated the following hypothesis:

H1. The success of space science-related social media messages, defined in terms of their engagement rate, can be predicted using only content-based features.

The related research questions are:

RQ1a. Do certain content-based features exist in engaging space science-related social media posts that set them apart from less engaging posts?

RQ1b. Can the engagement level (i.e. high or low) of a space science-related social media post be predicted?

1.3.2. Study 2

To investigate if there are any particular content-based features that are unique to space science, the same procedures were applied to data from three other fields: politics, business and non-profit. Results were compared with findings from Study 1 to test the following hypothesis:

H2. Engaging space science-related social media posts contain certain psycholinguistic features that are unique to space science.

The related research questions are:

RQ2a. What are the most prominent features of engaging space science-related social media posts?

RQ2b. Among the significant features of engaging space science-related social media posts, are there any features that are unique to space science?

RQ2c. Are there any similarities between the features found on different platforms, i.e. Facebook vs. Twitter?

2. Literature review

2.1. Predicting popularity and classification of social media posts

There is a large body of work that analyses and predicts resharing and reposting behaviour of social media users. Twitter is the focus of many such studies due to its popularity as a news sharing platform (Kwak, Lee, Park, & Moon, 2010). Retweet volume is influenced by many factors, one of which is an agency called 'informational value' (Rudat & Budar, 2015). Tweets with high informational value tend to receive larger number of retweets. The concept of informational value is derived and adapted from news

¹ <https://twitter.com/nasa>, accessed on 11 May 2016.

² <https://www.facebook.com/NASA/>, accessed on 11 May 2016.

³ <https://twitter.com/neiltyson>, accessed on 11 May 2016.

⁴ <https://www.facebook.com/neildegassetyson>, accessed on 11 May 2016.

value theory (Galtung & Ruge, 1965). When applied to contemporary contexts such as retweeting behaviour, high informational value is associated with controversy, negative consequences, relevance and unexpectedness, while aggression, personalisation, prominence and proximity are of low value (Rudat & Budar, 2015). Informational value is also suggested to have the potential to (1) affect a big audience and (2) have a psychological impact on its recipients and evoke behavioural change.

Machine learning has been applied in many social media studies to perform tasks such as network analysis and automatic classification (i.e. prediction) of texts (Aggarwal, 2011; Murthy, 2015). A study by Jenders, Kasneci, and Naumann (2013) used both 'obvious' features (e.g. number of followers, number of hashtags, tweet length, number of mentions, etc.) and 'latent' features (e.g. sentiment polarity and emotional divergence) to predict the virality of tweets. Their results showed that a combination of features covering structural, content-based and sentiment aspects leads to the best classifier performance (F1-score in the range of 91–96%). Elsewhere, it has been found that—among the 'obvious' features—social features such as number of followers and followees, favourites, times the user was listed and the user's verification status are reliable predictors of retweetability, while tweet features such as number of hashtags, URLs, tweet length, etc. contribute to the improvement of prediction accuracy (Petrovic et al., 2011). In this latter study where only 'obvious' features were used, prediction accuracies were lower (around 70%).

Applying similar demarcation between feature types, Naveed et al. (2011) divided tweet features into low-level (e.g. presence of URLs, hashtags, mentions and specific punctuations) and high-level ones (e.g. topic and sentiment of a tweet). The authors used logistic regression to compute the probability of a tweet being retweeted and made the observation that 'bad news travel fast', postulating that tweets with negative sentiments are more likely to be retweeted. Additionally, punctuations were found to be useful predictors of retweet likelihood: tweets that end with a question mark are more likely to be retweeted than the ones that end with an exclamation mark. Interestingly, the authors did not find high-level features to be more discriminative than low-level ones.

Another large-scale study on factors influencing the number of retweets found the presence of URLs and hashtags to have a significant impact on retweetability (Suh, Hong, Pirolli, & Chi, 2010). The authors collected 74 million tweets and used statistical methods such as Principal Component Analysis (PCA) and a generalised linear model to examine the impact of both content and contextual features. Apart from URLs and hashtags, the age of the account, the number of followers and followees are also useful indicators of retweetability. These results are consistent with the findings we have discussed so far.

Retweet prediction has also been framed as a propagation analysis problem (Hong, Dan, & Davison, 2011). As such, the 'social' aspect of social media weighs heavily in the feature selection process. Structural information contained in a user's social graph (e.g. degree distribution) is a useful feature in retweet prediction. When employed alongside features related to message content, temporal properties, metadata of messages and users, these social graph features contribute significantly to the overall performance.

Cheng, Adamic, Dow, Kleinberg, and Leskovec (2014) focused on the prediction of reshare cascades on Facebook. The authors analysed cascades of Facebook photo reshares and found that temporal features (properties related to the speed of the cascade) and structural features (social ties between the initial reshares) are the biggest predictors of the eventual cascade size. They also found that it is the initial breadth, rather than depth, in a cascade that is a better indicator of larger cascades.

The application of machine learning in automatic classification

tasks is not limited to the prediction of message popularity. Supervised learning has been used to predict the emergence of trending topics on Twitter by examining the features of hashtags associated with these topics (Ma, Sun, & Cong, 2013). Both content features of the hashtag (e.g. if it contains a digit, hashtag sentiment, etc.) and contextual features based on the characteristics of the Twitter community who adopted the hashtag (e.g. if a user replied or retweeted a tweet) were found to be useful in hashtag popularity prediction, with contextual features yielding the bigger impact. Another application of supervised learning is in the automatic classification of opinions expressed in tweets. Coteló, Cruz, Enríquez, and Troyano (2016) integrated features based on both textual content (i.e. bag of words) and structural information derived from graph-based representations of communities to predict political opinions exhibited in tweets. Using a multiple pipeline method and additional pre-processing and feature selection procedures, the authors were able to achieve cross-validation accuracies in the ranges of 80%. Credibility prediction is another area where supervised learning has found useful applications (Castillo, Mendoza, & Poblete, 2011). Information credibility of tweets can be forecasted—with accuracies of up to 92%—by using a combination of message-based (e.g. tweet length, sentiment), user-based (e.g. number followers and followees), topic-based (e.g. presence of hashtags and URLs) and propagation-based (e.g. depth of retweet tree) features.

2.2. Psychometric assessment of online phenomena

Words and language lie at the heart of human communications and are the medium by which personalities, motivations and cognitive processes can be studied and understood (Pennebaker, Booth, Boyd, & Francis, 2015). We are living in an era where technological advances have enabled researchers to relate word use to a wide range of human behaviours (Tausczik & Pennebaker, 2010). The Linguistic Inquiry and Word Count (LIWC) is a computer-based text analysis program that has been developed based on the psychometrics of language (Pennebaker, Boyd, Jordan, & Blackburn, 2015). It uses an internal dictionary and statistical methods to classify words into psychologically relevant categories (Lee, Mahmud, Chen, Zhou, & Nichols, 2015). There are over 90 categories (e.g. anger, sadness, anxiety), covering a broad array of structural, emotional and cognitive components of human communications. For the summary dimension 'Clout', for example, the authors explored the role of pronouns in revealing the relative rank among individuals in a group (Kacwicz, Pennebaker, Davis, Jeon, & Graesser, 2013). They found that individuals occupying higher status seldom use first-person singular but frequently use first-person plural and second-person singular pronouns, indicating 'other-focus' as opposed to 'self-focus'. For another summary dimension, 'Analytical thinking', the authors proposed that function words (e.g. articles, prepositions) contain valuable psychological information pertaining *how* people think (rather than *what* people think) (Pennebaker, Chung, Frazee, Lavergne, & Beaver, 2014). Their analysis of college admission essays revealed that submissions that obtained higher grades displayed greater use of articles and prepositions (indicating analytical thinking) while essays with lower grades contained more auxiliary verbs, adverbs and conjunctions (indicating a narrative style).

LIWC has been shown to have high word coverage: it captures over 86% of the words people use in writings and speeches (Pennebaker, Boyd et al., 2015). It has been used in hundreds of studies using verbal and written texts to analyse various psychological processes with promising results (Veltri & Atanasova, 2015). One recent study by Davalos, Merchant, and Rose (2016) used big data analysis to study complex psychological constructs (e.g.

nostalgia) exhibited on social media. The authors analysed Facebook posts using LIWC and found the deep and bittersweet nature of nostalgia to manifest itself in social media posts in the following ways: these posts are more reflective, emotional and regularly express both positive and negative sentiments. By analysing individuals' word use on Twitter, the psychological disparities among people with differing political orientations can be examined (Sylwester & Purver, 2015). Findings are largely consistent with previous research, showing liberals to place more emphasis on uniqueness, use more swear words, anxiety- and feeling-related words, while conservatives value group association and their tweets contain more achievement and religion-related words. Veltri and Atanasova (2015) combined a variety of automatic text analysis methods to study climate change conversations on Twitter. They analysed the psychological processes underlying these tweets using LIWC and found that emotionally arousing texts are more likely to be shared. They also suggested that this type of investigation, which goes beyond simple positive-negative sentiment analysis, offers a more refined understanding of communication mechanisms and opens up opportunities for new theoretical premises to be drawn.

Psycholinguistic features derived using LIWC have been used to investigate how people with different personalities differ in their reply and retweet patterns on Twitter (Mahmud, Chen, & Nichols, 2014). Users whose tweets are found to be more communicative, social and positive are more active in replying and retweeting other people's messages. LIWC has also been used to characterise affective components in tweets with the goal of examining the role of sentiments in reply and retweet behaviour. In political communications, for example, the presence of negative words has been found to lead to more replies and retweets (Kim & Yoo, 2012). Different types of negative emotions (e.g. anger vs. anxiety) have varying degrees of impact on the number of retweets and replies. A similar study that also used LIWC affective features to examine politically relevant tweets found that tweets containing *either* positive or negative emotions (as opposed to neutral sentiments) receive more retweets (Stieglitz & Dang-Xuan, 2012). These findings suggest that, when it comes to political tweets, emotive words are more likely to inspire actions on Twitter when compared with neutral words.

As vehicles for Twitter ideas, hashtags are highly relevant to the study of idea dissemination on Twitter (Tsur & Rappoport, 2012). Using a combination of hashtag content features—including their cognitive dimensions derived using LIWC—the spread of hashtags in a given time frame can be predicted. The cognitive attributes of hashtags were found to improve prediction accuracies, confirming that word choices have an impact on driving greater user involvement in online discourses. Also focusing on hashtags, a study conducted by Maity, Gupta, Goyal, and Mukherjee (2015) used a range of features to predict the popularity of Twitter idioms (a type of Twitter hashtags, e.g. #10ThingsAboutMe) and found that after content-related features of the hashtags, cognitive dimensions of the hashtags are the most discriminative features.

A better understanding of the psychological meaning of words in social media communications has many practical implications, for example in the study of mental health issues (Harman, 2014). LIWC has been used to analyse mental health phenomena in publicly available Twitter data. Researchers applied machine learning to train an analytic model, using the proportion of swear words, words related to anxiety, anger, positive and negative sentiments as features. Results showed a statistically significant difference between the depression and control group regarding their word use, attesting to the potential of automatic content analysis programs in uncovering mental health related linguistic signals on social media. In another study, Lee et al. (2015) addressed the problem of actively

identifying the right strangers on Twitter to retweet information within a certain time frame. They used a feature-based model that includes people's exhibited social behaviour calculated using LIWC to predict if an individual is likely to respond to a stranger's retweet request. Their results showed that users exhibiting features such as 'inclusive', 'achievement' and 'sadness' are more likely to respond positively to a stranger's retweet request.

3. Methods

3.1. Study 1

For Study 1, we focused on building a predictive model using supervised learning for space science-related social media messages to answer RQ1a and RQ1b. The models that were used are described in detail in Section 3.1.7.

3.1.1. Data collection

Data was collected from Facebook and Twitter. For Facebook, we used Facebook's Graph API and the Facebook SDK for Python library to collect posts from 50 public pages. The pages were selected based on the following criteria:

- The pages are publicly accessible
- The pages post about space science-related content
- The pages are actively maintained

We ran the script on 26 February 2016 to collect the latest 2000 posts from each page. Due to the young age of some pages, which means they do not yet have 2000 posts since their establishment, we did not manage to collect the same number of posts for all pages. The final Facebook dataset contained around 66 K public Facebook posts.

For Twitter, we used Twitter's REST API and Python's Tweepy library to collect Tweets from the user timeline of 60 Twitter users. Similar selection criteria were applied: the accounts are public, about space science and actively maintained. We ran the script on 30 March 2016 to collect the latest 2000 tweets from each user. Tweets that are retweeted were excluded, as they are not original content created by the user. Tweets that are replies to another user were also excluded, as they are usually conversations between two users and not, strictly speaking, meant for the public. We did not manage to collect the same number of tweets for all accounts for the same reason mentioned above. The final tweet dataset contained around 77 K public tweets.

3.1.2. Psycholinguistic scores of social media messages

To obtain the psycholinguistic scores for the messages in our data sets we used LIWC, a widely used computerised text analysis program, to derive the psychological meaning behind the words in a message. The LIWC software takes text messages as input and for every message it computes a score for each of the LIWC categories. There are a total of 95 LIWC psycholinguistic categories, including linguistic processes (e.g. pronouns, prepositions, articles), affective processes (e.g. 'anxiety', 'anger', 'sadness'), perceptual processes (e.g. 'see', 'hear', 'feel'), drives (e.g. 'affiliation', 'achievement', 'power'), informal language (e.g. swear words, netspeak, nonfluencies) and what the authors termed 'summary dimensions' ('Analytical thinking', 'Authenticity', 'Clout', 'Emotional tone').

We selected 32 categories that were considered to be relevant for the purpose of this study. We ran the messages in our data sets through the LIWC program and summed up the scores for each category. We then calculated the mean score for every category, for Facebook and Twitter respectively. To benchmark our results, we compared the mean scores we computed with the baseline scores

provided by the developers (Pennebaker, Boyd et al., 2015). These baseline scores are given in the form of grand means and are the result of analyses conducted on several corpora of texts (including tweets), covering over 231 million words (Pennebaker, Boyd et al., 2015).

3.1.3. Defining engagement rate and class labels

Engagement in a general sense means ‘establishing and sustaining relationships, while developing a level of trust that makes people comfortable enough to do business with you’ (Gattiker, 2011). On social media, engagement is measured in interactions with others and can take several forms, including having a conversation with other users through commenting or actions that demonstrate support, such as liking, sharing, or retweeting (Drula, 2012; Leander, 2011). The engagement rate for a social media post is generally recognised as the most important metric and measure of success (Niciporuc, 2014). Different formulas have been used to measure engagement rate, but they are usually calculated as the number of interactions a message receives divided by the total number of page likes (for Facebook) or followers (for Twitter).

Facebook defines post engagement rate as ‘the percentage of people who saw a post that reacted to, shared, clicked or commented on it’ (Facebook Help Centre, 2016). Twitter defines tweet engagement rate as ‘number of engagements divided by impressions’, where engagement includes all clicks anywhere in the tweet (including hashtags, links, username, etc.), retweet, replies, follows and likes (previously known as ‘favourites’) (Twitter Help Centre, 2016).

Not all the metrics that Facebook and Twitter use to define engagement rate are publicly available through their APIs. The number of impressions per post and reach per tweet, for example, are only available through Facebook’s Insight and Twitter’s Analytics dashboards, which only account administrators have access to. Therefore we have come up with our own definition for engagement rate, using only metrics that are obtainable through the APIs. Generally, we defined engagement rate (ER) for a message as follows:

$$ER = \frac{1 \cdot p + 0.75 \cdot r + 0.5 \cdot a}{nf} \quad (1)$$

where p is the number of propagative interactions, r is the number of written remarks expressing a reaction, a is the number of appreciative or supportive interactions, and nf is the number of fans of the account under consideration. For Facebook, p is then the number of shares, r the number of comments, a the number of likes a post received and nf the number of fans that liked the Facebook page. For Twitter, p is the number retweets, r the number of replies, a the number of likes a tweet received and nf the number of followers of the Twitter account. Unfortunately Twitter API does not provide a way to retrieve the number of replies to a tweet. Our Twitter engagement rate definition therefore does not include tweet reply count. This is a limitation that was imposed by Twitter API and which we have not overcome in this study. However, the inclusion of retweets and likes is likely sufficient to provide an adequate representation of the engagement levels of tweets, given the importance of these two Twitter interactions (Boyd, Golder, & Lotan, 2010; Suh et al., 2010).

As different types of interactions show varying degrees of engagement, we assigned more weight to metrics that represent higher degrees of engagement. On Facebook, for example, when users like or comment on a post, the post will appear on the news feed of their Facebook friends. When users share a post, apart from appearing on their friends’ news feed, the post will also appear on their personal profile page (i.e. their ‘wall’). Additionally, users can add their own commentary when they share a post. Intuitively, the

engagement levels displayed by an action of ‘liking’, ‘commenting’ and ‘sharing’ are not equal. Certain types of interactions are considered more valuable and important than others and indeed, Facebook’s Edgerank algorithm account for this difference by assigning different weights to them (Bucher, 2012). We adopted the hierarchy employed by Vadivu and Neelamalar (2015) to rank the importance of the three different interactions: propagative interactions (p) are given the highest value, followed by written remarks (r), and then appreciative interactions (a).

We labelled our training datasets in the following way: the top 5% of Facebook posts and top 11% of Tweets with the highest engagement rates are labelled as HER (High Engagement Rate), while the lowest ones are labelled as LER (Low Engagement Rate). The higher percentage used for Twitter is necessitated by the brevity of tweets: more data is needed for optimal machine learning performance. We strived to achieve the same number of data points for each class to avoid running into problems associated with imbalanced datasets in machine learning (Provost, 2000). Table 1 gives an overview of the final datasets and the boundary values used for the two classes.

3.1.4. Principal Component Analysis

We conducted Principal Component Analysis (PCA) on our training data as an initial exploration of the features in the data. PCA is a statistical method that converts a group of variables containing features that possibly correlate with each other into a smaller number of variables that are independent of each other (Abdi & Williams, 2010). The result of this data reduction technique is called principal components, where the first principal component explains the biggest amount of variance in the data.

This technique is frequently applied to uncover hidden patterns or underlying structure in big datasets. Using PCA, we could get an inkling of the features that are important in each class. The PCA function that comes with scikit-learn, a Python machine learning library, was used.

3.1.5. Feature sets

Feature selection is the process of extracting discriminating and relevant attributes that characterise the messages in a dataset. These features are used to train the supervised learning model. For this project, we focused only on content-based feature. Four sets of features were identified: n-grams, psycholinguistic features, grammatical features and social media features.

n-grams are individual or sequence of words. In this feature set, the features are coded as contiguous n number of words (called n-grams), with n ranging from one to three, i.e. unigram, bigram and trigram. When n is one, the messages are represented as a bag of words. Grammar and the sequence in which the words appear are disregarded. This feature set considers the discriminative power of word(s) to distinguish between more engaging and less engaging content. The word(s) that are weighted heaviest by our model are most indicative of an engaging message.

Psycholinguistic features (psy) consider the psychological meaning of the words in a message. The selection of this feature set is motivated by the intuition that the attitude and mental characteristics conveyed by social media messages have the potential to impact how well the messages are received. We ran the messages in our labelled datasets through the LIWC software. For each message, LIWC outputs a score for each category. These scores denote the percentage of words in the message that belongs to that category (except for the summary dimensions ‘Analytical thinking’, ‘Clout’, ‘Authenticity’ and ‘Emotional tone’ which are computed using proprietary algorithms). The scores were used to determine if a message contains certain psycholinguistic features. For example ‘anger_yes’ is set to 1 for a message (i.e. ‘anger’ is a feature) if its

Table 1
Facebook and Twitter datasets and engagement rates.

Platform/field	Total # of post	Engagement rate (ER)		Engagement rate boundary values (BV)				Selected posts for classification	
		Mean	SD	BV for high ER	# of HER post	BV for low ER	# of LER post	Total # of selected post	% of total
Facebook									
Space	65,252	0.120	0.415	0.390	3520	0.002	3501	7021	10.8%
Politics	68,307	0.393	1.622	1.250	4249	0.006	4256	8505	12.5%
Business	49,431	0.025	0.091	0.089	3068	6.4e-05	3077	6145	12.4%
Non-profit	85,458	0.120	0.588	0.400	4537	0.003	4535	9072	10.6%
Twitter									
Space	76,954	0.059	0.362	0.101	8767	0.002	8754	17,521	22.7%
Politics	98,852	0.045	0.205	0.098	10,241	0.002	10,193	20,434	20.6%
Business	52,793	0.014	0.056	0.024	6019	0.001	6004	12,023	22.7%
Non-profit	97,420	0.011	0.040	0.023	9963	0.001	9956	19,919	20.4%

Table 2
Overview of feature sets.

Features set	Description	Number of features	Example
n-grams	Individual or sequence of n number of words	Variable (depends on length of message)	'jupiter', 'rosetta', 'plutoflyby', 'nasa science', 'international space station'
Psycholinguistic (psy)	The psychological meaning underlying words in a message, encoded by LIWC	32	'anger', 'anxiety', 'authentic', 'sadness', 'tentative'
Grammatical (gr)	The structural aspect of language, such as punctuations	4	Words per sentence, presence of exclamation mark, presence of question mark
Social media (sm)	Properties which are typical of social media communications	6 (Facebook), 5 (Twitter)	Presence of hashtags, URLs, photos, videos

anger score is higher than a threshold value.

Grammatical features (gr) consider the structural attributes of language. For each message, we computed the number of words per sentence, the percentage of words that is longer than six letters (sixltr), the presence of certain punctuation marks such as exclamation and question marks. The 'words per sentence' and 'sixltr' features encapsulate the 'complexity' aspect of a message: longer sentences and messages with many words longer than six letters tend to be more complex. We included the presence of exclamation and question marks as features as these punctuation marks have the potential to drastically alter the tone of a message, possibly influencing users' propensity to engage with them.

Social media features (sm) take into account the elements that are characteristic of social media communications. For Facebook messages, we encoded the post type (i.e. photo, video, link or event) and the presence of hashtags and URLs as features. For tweets, this feature set includes the presence of hashtags, URLs, and media elements such as photos, videos and animated GIFs. As mentioned in Section 2.1, these features have been found to be highly informative when it comes to predicting the popularity of a message.

An overview of the feature sets is given in Table 2. We ran our supervised learning algorithms using each feature set in isolation, and also a combination of them. The first combination (combination A) combined psycholinguistic, grammatical and social media features. The second combination (combination B) included all four feature sets.

3.1.6. Data pre-processing

To create the input data for the n-gram feature set, various pre-processing steps were applied:

- All characters were transformed to lowercase letters
- Contractions were expanded, e.g. don't becomes 'do not'.
- Stopwords were removed, e.g. a, an, the
- URLs and hashtags were removed
- Punctuations and special symbols were removed

These pre-processing steps are only necessary to create the input data for the analysis of n-gram features. For input to the LIWC program, raw tweets and Facebook posts were used.

3.1.7. Supervised learning algorithms

We used supervised learning to build our predictive models. Supervised learning is a type of machine learning algorithm that learns a mapping function f from a set of labelled training data to map from an input x to an output y . The goal is to derive a mapping function (i.e. model) that can predict—with as little error as possible—the output variables y when given new input data. The input variables are represented by a feature vector $x = (x_1, x_2, \dots, x_n)$, where n is the number of features. In a classification task, y is hence the 'class' to be predicted,

$$y = f(x_1, x_2, \dots, x_n) \quad (2)$$

For our study, the class that we want to predict is 'HER' or 'LER' and the input variables are the features described in Section 3.1.5. Concretely,

$$y = f(\text{see}, \text{clout}, \text{authentic}, \dots) \quad (3)$$

We used Python's scikit-learn machine learning library to implement our supervised learning procedures (Pedregosa et al., 2011). Three models were selected for the purpose of this study. The first model is based on a multinomial Naive Bayes (NB) classifier. NB is a popular classifier for text classification tasks, mainly due to its speed and ease of use. The Naive Bayes model applies Bayes' theorem with the 'naive' assumption that all features are independent of each other (Scikit-learn, 2016a).

The second model we used is a linear classifier with Stochastic Gradient Descent (SGD) training. Given a set of training data, the goal of this model is to learn a linear function $f(x) = \theta^T x + c$ with model parameters θ and intercept c by minimising the training error (Scikit-learn, 2016b). By altering the model's 'loss' parameter, it becomes either a linear support vector machine or a logistic regression classifier.

The third model is a version of the decision tree classifier called extra-tree (ET). It trains a number of randomised decision trees on sub-samples of the whole dataset and averages the results with the goal of deriving better predictive accuracy and avoid overfitting, a common problem associated with machine learning (Hawkins, 2004). Generally, a model based on decision tree learns a set of decision rules derived from the features in the training data in order to predict the class of a target variable. The model parameters are selected to minimise impurity, which can be measured in terms of Entropy, Gini or classification error (Scikit-learn, 2016c).

Scikit-learn's train-test-split module was used to randomly split the input data into training and test sets. Model training is performed on the training set and the resulting model is used to predict the labels of data in the test set. To improve performance and reduce computing time, feature selection methods were implemented to select the most discriminating features based on their Chi-Square scores. Scikit-learn's GridSearch module was used to search for the best hyperparameters for our classifiers, e.g. the 'loss' parameter for the SGD classifier. The parameters are selected using the best cross-validation scores. Finally a pipeline was implemented to streamline the successive steps in building a predictive model: tokenisation and normalisation of input data, feature selection and model training.

To evaluate model performance, we conducted 10-fold cross-validation to assess how the result of our model training generalises to an independent data set. Cross-validation also helps us avoid overfitting. Prediction results are evaluated using precision, recall, F1-score and accuracy. Mathematically,

$$Precision = \frac{tp}{tp + fp} \tag{4}$$

$$Recall = \frac{tp}{tp + fn} \tag{5}$$

$$F1\ score = 2 \cdot \frac{precision \cdot recall}{precision + recall} \tag{6}$$

$$Accuracy = \frac{tp + tn}{tp + tn + fp + fn} \tag{7}$$

where *tp* is the number of true positives, *fp* is the number of false positives, *tn* is the number of true negatives and *fn* is the number of false negatives.

3.1.8. Feature analysis

The features of a model and their importance were obtained by querying the attributes of the trained classifier. The relative importance of a feature is given by its weight or coefficient (depending on the classifier). To enable comparisons between classifiers and across fields, these values were normalised to a number between 0 and 1. The normalised feature scores were used to rank the features in order of importance - the higher the score, the more important the feature.

3.2. Study 2

Study 2 focused on the comparison between the features of engaging social media messages in space science and other fields. To this end, the methods described in Sections 3.1.1 to 3.1.8 were applied on social media messages from three other fields—politics, business and non-profit—and the results were compared with the results from Study 1 to answer RQ2a, RQ2b and RQ2c.

To test H2, which postulates that engaging space science-related

social media posts contain certain psycholinguistic features that are unique to space science, we repeated our feature analysis experiment for each classifier and performed a one-way Analysis of Variance (ANOVA) test with feature scores as dependent variables and fields as independent factors. For each feature, pairwise contrast tests were conducted to compare the differences between the mean feature scores for space science and the three other fields, i.e. space-politics, space-business, space-non-profit.

4. Results

4.1. Psycholinguistic features of social media communication

Apart from a few exceptions, the LIWC scores computed for all four fields generally do not show significant differences to the baseline scores. A number of noteworthy observations are as follows (Figs. 1 and 2; scores for Facebook and Twitter are similar, for simplicity only scores for Facebook are listed):

- 'Analytic thinking' scores for all four fields are higher than the baseline score (M = 56.34, SD = 17.58). Among the four fields, space science has the highest 'Analytic' score (M = 90.36, SD = 16.91).
- Non-profit has the highest 'Clout' score (M = 80.55, SD = 20.03), which is also significantly higher than baseline score (M = 57.95, SD = 60.94).
- Space science has the highest 'Authentic' scores (M = 41.65, SD = 33.65) while non-profit has the lowest (M = 28.44, SD = 30.01).
- Space science has the lowest score for 'Tone' (M = 49.35, SD = 43.20).
- Space science scores the highest for the 'See' category (M = 2.74, SD = 3.94).
- Space science exhibits the lowest 'Certainty' score among all four fields (M = 0.58, SD = 1.82). Space science's score in this category is also significantly lower than baseline score (M = 1.43, SD = 0.70).
- Politics and non-profit show the highest scores for 'Affiliation' (politics: M = 4.04, SD = 5.18; non-profit: M = 4.63, SD = 5.09), which are significantly higher than baseline score (M = 2.05, SD = 1.28).

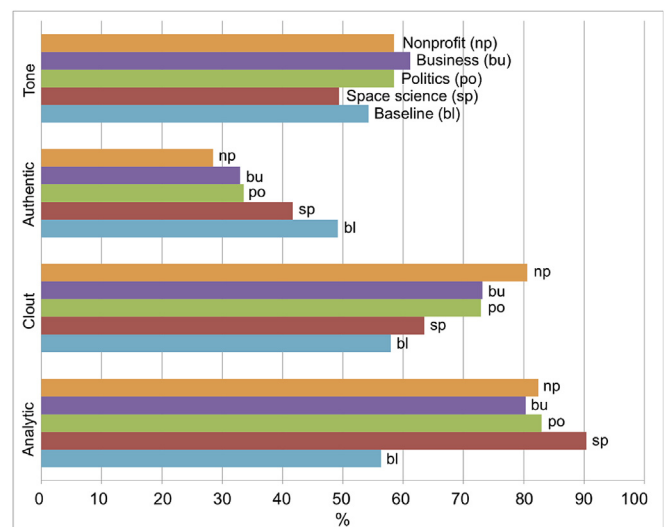


Fig. 1. Comparison of mean scores for the LIWC summary dimensions 'Analytic thinking', 'Clout', 'Authentic', and 'Tone'.

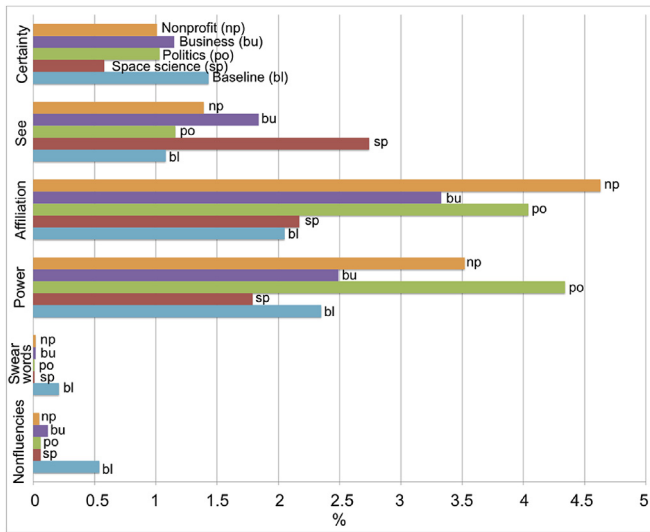


Fig. 2. Comparison of mean scores for a selection of LIWC categories.

- Politics has the highest score for ‘Power’ (M = 4.34, SD = 4.85), also significantly higher than baseline score (M = 2.35, SD = 1.12).
- Scores for ‘Swear words’ and ‘Nonfluencies’ for all four fields are generally lower than baseline scores (Swear words: M = 0.21, SD = 0.37; Nonfluencies: M = 0.54, SD = 0.49).

4.2. PCA

For Facebook, principal component #1 (which accounts for the biggest variance in the data) accounts for 8.2% of the variance in the data, whereas for Twitter, it accounts for 10.5%. The top five features encompassed in principal component #1 for Facebook are ‘authentic_yes’, ‘type_photo’, ‘tone_yes’, ‘low_wps’ (low word per sentence), ‘posemo_yes’, and for Twitter they are ‘has_photo’, ‘has_hashtag’, ‘high_wps’, ‘see_yes’ and ‘analytic_yes’.

4.3. Predicting engagement for space science (RQ1a, RQ1b)

For Facebook, we achieved overall prediction accuracies between 88% and 90% when all feature sets were used (combination B). Performance for Twitter is slightly better, obtaining prediction accuracies between 90% and 93%. The prediction results (precision, recall, F1-score and overall accuracy) and their confusion matrices

are given in Table 3.

4.4. Classifier cross-validation scores for all fields

The best performance is achieved with combination B (all features), followed by n-gram, combination A (psy + gr + sm), social media features and psycholinguistic features (Table 4). Using all features we managed to achieve cross-validation scores between 86% and 92% for Facebook, and 86% and 93% for Twitter. The worst performance came from using grammatical features in isolation, although this feature set still achieved scores between 54% and 60% for Facebook, and 55% and 60% for Twitter. Classifier performances for space science are generally better than other fields, particularly for social media and combination A feature sets.

4.5. Important words in space science

Using only n-gram features, we were able to achieve prediction accuracies in the range of 82%–86%, indicating that the words in social media messages carry strong prediction power of its engagement potential. We queried the feature weights and ranked them according to their importance: the heavier their weights, the more important they are. The top five words (in order of importance) for the most engaging space science-related Facebook posts are ‘yearinspace’, ‘principia’, ‘philae’, ‘sls’ (space launch system) and ‘station’; for Twitter they are ‘exomars’, ‘pluto’, ‘plutoflyby’, ‘67p’ and ‘cosmos’.

4.6. Important features in social media communications (RQ2a, RQ2c)

The most significant features for each field were normalised and ranked so that inter-field and inter-platform comparisons can be drawn. The top five features for highly engaging Facebook posts for space science, in order of importance, are ‘type_photo’, ‘anger_yes’, ‘see_yes’, ‘posemo_yes’ (positive emotion) and ‘anx_yes’ (anxious); whereas Twitter is dominated by features related to visual elements: ‘has_photo’, ‘has_animated_gif’, ‘has_video’, ‘has_hashtag’ and ‘certain_yes’.

Table 5 lists the 10 most important features for all four fields. The grey cells represent features that are common between Facebook and Twitter (on a per field basis).

4.7. Feature comparison between space science and other fields (RQ2b)

We found four space science features on Facebook whose feature scores are significantly different (higher) than all other

Table 3

Classification results for engaging space science-related social media messages (grey cells represent messages that are correctly predicted).

Platform	Classifier	Class	Confusion matrices		Results				
			HER	LER	Precision	Recall	F1-score	Accuracy	SD
Facebook	NB	HER	625	62	0.89	0.91	0.90	0.90	0.008
		LER	79	639	0.91	0.89	0.90		
	SGD	HER	628	59	0.89	0.91	0.90	0.90	0.008
		LER	81	637	0.92	0.89	0.90		
	ET	HER	625	62	0.85	0.91	0.88	0.88	0.009
		LER	107	611	0.91	0.85	0.88		
Twitter	NB	HER	1611	203	0.92	0.89	0.91	0.90	0.005
		LER	131	1560	0.88	0.92	0.90		
	SGD	HER	1667	147	0.95	0.92	0.93	0.93	0.004
		LER	89	1602	0.92	0.95	0.93		
	ET	HER	1648	166	0.93	0.91	0.92	0.92	0.005
		LER	127	1564	0.90	0.92	0.91		

Table 4
Supervised learning algorithm cross-validation results for Facebook (SD between 2 and 4%).

	n-grams (%)				Psycholinguistics (psy) (%)				Grammar (gr) (%)				Social media (sm) (%)				psy + gr + sm (%)				All features (%)			
	Sp	Po	Bu	Np	Sp	Po	Bu	Np	Sp	Po	Bu	Np	Sp	Po	Bu	Np	Sp	Po	Bu	Np	Sp	Po	Bu	Np
NB	83	86	86	86	67	71	63	62	59	54	55	56	82	69	70	71	82	78	73	75	90	89	91	91
SGD	84	87	86	86	68	72	63	64	60	54	56	56	83	73	69	71	83	81	73	76	90	92	90	90
ET	82	86	85	84	67	71	63	62	59	54	54	56	83	73	70	71	83	80	72	75	88	86	87	87

Table 5
Top features of engaging social media messages for all fields (grey cells represent platform-common features).

Ranking	Space	Politics	Business	Non-profit
Facebook				
1	type_photo	type_photo	type_photo	type_photo
2	anger_yes	certain_yes	cause_yes	sad_yes
3	see_yes	anx_yes	has_exclam	analytic_no
4	posemo_yes	has_hashtag	has_url	differ_yes
5	anx_yes	assent_yes	clout_no	risk_yes
6	authentic_yes	risk_yes	certain_yes	posemo_yes
7	certain_yes	analytic_no	tone_no	certain_yes
8	has_hashtag	negemo_yes	analytic_yes	discrep_yes
9	discrep_yes	posemo_yes	has_hashtag	has_url
10	reward_yes	sad_yes	tentat_yes	clout_no
Twitter				
1	has_photo	has_photo	has_photo	has_photo
2	has_animated_gif	has_animated_gif	has_animated_gif	has_animated_gif
3	has_video	has_video	has_video	risk_yes
4	has_hashtag	differ_yes	low_wps	anger_yes
5	certain_yes	discrep_yes	assent_yes	certain_yes
6	see_yes	certain_yes	analytic_no	negemo_yes
7	nonflu_yes	power_yes	certain_yes	discrep_yes
8	clout_no	anger_yes	differ_yes	differ_yes
9	few_sixltr	negemo_yes	has_url	power_yes
10	anger_yes	analytic_no	has_hashtag	tone_no

fields: 'type_photo', 'anger_yes', 'see_yes' and 'authentic_yes'. For Twitter, we also found four such features: 'has_photo', 'has_hashtag', 'see_yes' and 'nonflu_yes' (nonfluencies) (Table 6). Our ANOVA contrast test results showed that the differences are significant enough to suggest that these features are characteristic of engaging social media messages in, and only in, the field of space science

($p < 0.01$).

5. Discussions

Our research took a data-centric approach to address the problem of understanding audience engagement in social media

Table 6
Unique features of engaging space science-related social media messages.

Platform	Feature	Contrast	Difference between means	Std. error	p-value
Facebook	type_photo	space, politics	-0.5956	0.0532	0.000
		space, business	-0.5185	0.0507	0.000
		space, non-profit	-0.3531	0.0503	0.000
	anger_yes	space, politics	-0.3307	0.0995	0.007
		space, business	-0.3808	0.0954	0.003
		space, non-profit	-0.4086	0.0905	0.002
	see_yes	space, politics	-0.2488	0.0308	0.000
		space, business	-0.1875	0.0403	0.000
		space, non-profit	-0.2488	0.0308	0.000
	authentic_yes	space, politics	-0.1992	0.0240	0.000
		space, business	-0.1170	0.0319	0.002
		space, non-profit	-0.2003	0.0240	0.000
Twitter	has_photo	space, politics	-0.5531	0.0728	0.000
		space, business	-0.4097	0.0839	0.000
		space, non-profit	-0.4731	0.0652	0.000
	has_hashtag	space, politics	-0.0914	0.0123	0.000
		space, business	-0.0879	0.0188	0.001
		space, non-profit	-0.0711	0.0134	0.000
	see_yes	space, politics	-0.0442	0.0092	0.001
		space, business	-0.0442	0.0092	0.001
		space, non-profit	-0.0442	0.0092	0.001
	nonflu_yes	space, politics	-0.0378	0.0064	0.000
		space, business	-0.0378	0.0064	0.000
		space, non-profit	-0.0241	0.0064	0.001

space science communication—a problem that, as far as we know, has never been directly addressed in previous research—and provided empirical evidence for the biggest factors impacting engagement. We expected a significant difference in terms of language use between ‘effective’ and ‘ineffective’ messages and our results confirmed our speculations. From the two studies we found that (1) engagement levels of space science-related social media messages can be predicted with accuracies in the vicinity of 90% using only content-based features, and (2) there are certain features that are unique to the field of space science and they are anger, authenticity, hashtags, visual descriptions—be it visual perception-related words, or media elements—and a tentative tone.

We formulated a novel way to define engagement rates, which takes into account the different degrees of engagement demonstrated through various user actions on social media. By assigning weights to different types of interactions, our definition offers greater granularity and tunability in measuring user engagement on social media. We acknowledge that the metrics used may not fully define engagement (whose definition is debatable anyway). However, by assuming audience engagement is linked to these passively collected digital metrics we were able to consider the perceived interestingness or value of social media science communication in an unobtrusive way, removing observer-expectancy effects and other biases associated with sampling and surveys (Mayer-Schönberger & Cukier, 2013).

5.1. Predicting engaging social media messages

The predictive model we built in Study 1 achieved high prediction accuracies (in the ranges of 90%), confirming that it is possible to forecast the success of social media messages, thereby answering RQ1b. Best performances were achieved using the linear classifier with SGD training. The high F1-scores (the weighted average of precision and recall) demonstrate strong model performance in terms of both sensitivity and specificity (Table 3). The success of machine learning algorithms is highly dependent on the way data is presented, i.e. the features that were used (Domingos, 2012). The strong model performance indicates that the features that we selected are highly discriminative and relevant features that can be used to distinguish between high and low engagement messages, thereby answering RQ1a.

The same experiment conducted on data from three other fields achieved similarly strong performance, demonstrating the robustness of our model. Strongest performances were achieved when all features were used. These results are consistent with the observations of several previous machine learning experiments on social media messages (e.g. Jenders et al., 2013; Ma et al., 2013): combining features representing different aspects of social media communication tends to lead to better prediction accuracies. It is also worth noting that although prediction accuracies (and cross-validation scores) for the n-gram feature group is relatively high (over 80% for all fields), these features are not very useful in terms of generalisability across fields. The words that are predictive of high engagement in the field of space science are usually not very relevant for any other fields.

5.2. Features of social media communications

One of the goals of our project was to exploit the opportunities offered by advancements in psychometrics to draw more nuanced insights into the linguistics of engagement. The results of a general comparison between the LIWC mean scores from all four fields offer several interesting insights. Space science has the highest score for ‘Analytical thinking’. A high score for this dimension

reveals a more methodical and logical thought process, while a low score illustrates thinking patterns that are more casual, personal and narrative in style. Space science also scores the highest for the ‘Authenticity’ category, while non-profit scores the lowest. High scores in this category reflect a more ‘honest, personal, and disclosing’ communication style, while low scores are associated with a more ‘guarded and distanced’ style (Pennebaker, Booth et al., 2015).

Space science has the lowest score in the ‘Clout’ category, where a high score reflects high confidence while a low one suggests a more humble and tentative style. Space science also scores the lowest in the ‘Tone’ category, where high scores suggest upbeat and positive moods and low scores more negative sentiments. The low scores in ‘Tone’ are corroborated by space science’s low scores for ‘Positive emotions’ (2.51 for space science vs. 3.67 for baseline). Space science’s high score for the ‘See’ category reveals a frequent occurrence of words such as ‘see’ and ‘view’. Another noteworthy observation is space science’s low score for the ‘Certainty’ category, which encompasses words such as ‘always’ and ‘never’. A low score suggests a less assured style, which is characteristic of the questioning nature of science.

We also note the low score for ‘Swear words’ and ‘Nonfluencies’ for all four fields compared with the baseline scores. These results can be explained by the fact that the messages in our datasets are mostly from official social media accounts of established organisations and individuals. These users might be more cautious in language use when compared with the private users whose communications were analysed to create the baseline scores.

The results from the preliminary PCA analysis also gave us an initial idea of the important features in engaging space science-related social media messages. To recall, PCA converts groups of possibly correlated features into principal components, where the first principal component explains the biggest variance in the data and is therefore the most important one. The results provide insight into the significant features in space science-related social media communications. According to our PCA analysis, ‘Authenticity’ and ‘Photo’ are two of the most prominent features for Facebook while for Twitter they are ‘Photo’ and ‘Hashtags’. As will be seen in Section 5.3, these results turned out to be useful indications for our feature importance analysis.

We examined the words that are important features in engaging social media messages. As mentioned previously, these features are usually not generalisable to other fields. Nonetheless, they portray the zeitgeist of the period the data was collected. They provide insight into the topics that are discussed in a particular field and the type of content that attracts higher user engagement, thus making them worthy of closer inspection. Among these words, ‘yearinspace’ is the hashtag popularised by NASA and astronaut Scott Kelly for his year-long mission on the International Space Station; ‘philae’ is the spacecraft of European Space Agency (ESA)’s Rosetta mission that landed on the comet 67p; ‘exomars’ is ESA and the Russian Federal Space Agency’s astrobiology mission to study the habitability of the planet Mars; ‘pluto’ and ‘plutoflyby’ are associated with the close encounter with Pluto by NASA’s New Horizon spacecraft, the first ever space vehicle to explore the dwarf planet up close. These words are highly revealing of the ‘hot topics’ in space science in the previous year and their inclusion is evidence of the general public’s enthusiasm towards these topics. It is also noteworthy that hashtag use (e.g. ‘yearinspace’ and ‘plutoflyby’) is a prominent feature of engaging space science-related social media messages.

5.3. Comparing important features from different fields

Study 2 investigated the second hypothesis, which posits that

engaging space science-related social media messages contain certain psycholinguistic features that are exclusive to space science. The feature importance analysis revealed the most significant features for space science, politics, business and non-profit (Table 5). The top features for space science-related Facebook posts are 'Photo', 'Anger', 'See', 'Positive emotions' and 'Anxiety'. For Twitter they are 'Photo', 'Animated GIF', 'Video', 'Hashtags' and 'Certainty'. These results answered RQ2a.

Several interesting conclusions can be drawn from these results. First, it is apparent that visual elements are important features in engaging space science-related social media messages. This observation is supported by the presence of various media elements and the feature 'See' in the top ten list for both Facebook (3rd) and Twitter (6th). Space science is inherently visually arresting. Our results suggest that audience is more engaged when a message evokes responses pertaining to our visual perception and thus posts containing visual components—be it in the form of words or media elements—are highly effective.

Second, anger and anxiety appear to be strong features of engaging space science messages. Space science is linked to some of the most pressing global issues such as climate change and natural disasters (see for example, NASA Climate, 2016; ESA CCI, 2016). Given the urgency of these matters and the perceived inactivity in the political sphere to tackle them, it is little wonder that messages expressing anger and anxiety would resonate strongly with the public. Furthermore, a global consciousness borne out of having seen Earth from space may prompt astronauts to lament about the state of the world. Indeed, a closer look at the messages that contain these two features revealed that they often relate to natural disasters, climate change and frustrations over the world's political affairs (Table 7).

Third, the inclusion of the features 'Positive emotions' and 'Certainty' appears somewhat contradictory to our observations so far. As discussed in Section 5.2, space science has the lowest mean score in the 'Certainty' category. However, it seems that messages that are certain and assertive are popular among the general public, perhaps due to the confidence displayed by the authors. Similarly, space science's mean score for 'Positive emotions' is the lowest among all four fields but messages that display upbeat and cheerful emotions are highly engaging. These results indicate that although confident and positive posts are relatively rare in social media space science communication, messages that do display these emotions are usually very well-received by the audience.

The features from space science were then compared with features from other fields. The results of our ANOVA contrast tests revealed a number of features that are unique to space science ($p < 0.01$). For Facebook, these unique features are 'Photos', 'Anger', 'See' and 'Authenticity'. For Twitter, they are 'Photo', 'Hashtags', 'See' and 'Nonfluencies'. These results answered RQ2b.

Most of these features have already been discussed, except for 'Authenticity', 'Hashtags' and 'Nonfluencies'. As discussed previously, space science has the highest mean score for 'Authenticity' among all fields. With the discovery of this feature as a unique space science feature, we can conclude that authenticity is not only a common feature of space science communication on social media, it is also a very effective one when it comes to audience engagement. We have also discussed the presence of hashtags in popular space science messages, e.g. 'yearinspace' and 'plutoflyby'. It appears that hashtags—although also frequently used in other fields—are markedly more effective when employed in space science communication, especially on Twitter. At first glance, non-fluency might seem like an odd feature of science communication. However, upon closer inspection of the messages that are tagged with this feature, we discovered that they are usually tentative and jesting in tone, and often also contain the 'Authenticity' feature,

suggesting a more down to earth and genuine style of communication typical of science (Table 7). Nonfluency is also a Twitter-specific feature.

Additionally, we would like to mention the 'low word-per-sentence' feature for Facebook, which almost made the cut to be included as a unique feature ($p = 0.053$ for space-politics, $p < 0.05$ for space-business and space-non-profit). Perhaps due to the potential complexity in describing or explaining various space-related phenomena, when it comes to communicating space science on social media, audiences seem to prefer brevity to verbosity.

We conducted inter-platform comparisons to examine if there are any overlaps between the top ten features on Facebook and Twitter (grey cells in Table 5). We found moderate similarities, with space science and non-profit having five platform-common features, while politics and business have four, thereby answering RQ2c. The five platform-common features for space science are 'Photo', 'Anger', 'See', 'Certainty' and 'Hashtags', all of which have been discussed so far, demonstrating, yet again, the important role they play in social media space science communication.

5.4. Limitations

Our studies provided a comprehensive view into the ingredients of engaging social media messages. However, a number of limitations must be considered. First, the datasets only included 50 users from each field for Facebook and 60 users for Twitter. These users do not necessarily represent their entire field, despite efforts to include as much diversity as possible in our user selection (e.g. by selecting users from both genders, wide age range, a wide range of follower numbers, etc.). Second, we limited our study of science communication to the field space science. This was a decision motivated by both practicality and performance. Due to our positions as researchers in the closely related field of astrobiology, we have broader access to and knowledge of space science. Additionally, our preliminary experiments, which included messages from mixed disciplines within science, returned poorer results when compared with a more targeted study focusing only on space science. This observation is supported by the superior classifier performance for space science when compared with the three other fields, where no such subdomain-filtering has been carried out during data collection (discussed in Section 4.4). Third, due to limitations imposed by Facebook and Twitter API, we only managed to collect data from a certain time period. Consequently, our data might be dominated by the popular talking points during this period (e.g. Pluto fly-by). Although this may reduce the generalisability of our results, it provided valuable insights into the type of content that drew the most audience attention. Nonetheless, additional research should be conducted for other time periods (and for that matter, other science disciplines) to validate our findings and further understand the phenomena we observed.

Fourth, in using a program such as LIWC to interpret the psychological meaning behind words, we ignored certain important components of human communications such as context, irony, sarcasm and idioms. Like many computerised text analysis programs, LIWC is a probabilistic system and is prone to misinterpretation when complex linguistic features such as sarcasm are present (Tausczik & Pennebaker, 2010). Fifth, our engagement rate calculations assumed constant follower and page-like counts, as no historical data of these two parameters can be obtained through the APIs. This posed restrictions to the extent to which our engagement rates can be accurately calculated, leading to potentially lower engagement rates for older messages. However, the effect is attenuated by the fact that the accumulative number of shares, likes, comments and retweets also increases over time, i.e. both the numerator and denominator in our formulas increase.

Table 7
Examples of Facebook posts and tweets and their features.

Platform	Message	Features
Facebook	Celestial Valentine! Generations of stars can be seen in this infrared portrait of a wispy star-forming region. Details: http://go.nasa.gov/1SOMgWq	see_yes, type_photo
	Curtains on an era continue to draw to a close. -NDTyson "You develop an instant global consciousness a people orientation an intense dissatisfaction with the state of the world and a compulsion to do something about it. From out there on the moon international politics look so petty. You want to grab a politician by the scruff of the neck and drag him a quarter of a million miles out and say 'Look at that you son of a bitch.'" RIP: Edgar Mitchell (1930–2016) – Apollo 14 Astronaut	anger_yes, see_yes, type_photo
	Hurricane Patricia from Space About an hour ago NASA astronaut Scott Kelly tweeted this image of Hurricane Patricia approaching Mexico from his vantage point on the International Space Station. "It's massive. Be careful!" he said. The storm's record-breaking winds low central pressure and unusually rapid intensification has meteorologists using words like "stunning historic mind-boggling and catastrophic" to describe the storm [...] Read about how climate change may be strengthening hurricanes at [URL]	anx_yes, see_yes, type_photo
	We are so proud to mark 35 years as The Planetary Society today. Cheers to space fans everywhere. Let's keep exploring.	posemo_yes, authentic_yes, low_wps,
	Countdown. Let's take this 16 sunsets at a time. 8 days to go tomorrow! Good night from the International Space Station! #YearInSpace	authentic_yes, low_wps, posemo_yes, has_hashtag
Twitter	From one distant world to another: Pluto as spied from Saturn during the #PlutoFlyby	has_photo, has_hashtag
	If you never make mistakes then you are not on the frontier of discovery for there is where mistakes are made all the time.	certain_yes, authentic_yes
	There's nothing you can ever tell scientists about the natural world that will hurt their feelings	certain_yes
	Since #Perseid meteors hit Earth could their parent comet hit us too? (Uh oh–answer is yes.)	nonflu_yes
	Hey @MarsCuriosity – you're really making tracks now! Where are you headed? Somewhere fun? And um how long will you be away?	nonflu_yes, authentic_yes
	Things that are still wrong in 2016: Astrology Homeopathy Climate change denial Anti-vaxxers Young-Earth creationism Fear mongering	anx_yes, authentic_yes
	@CassiniSaturn shows how our view of the Saturn system has changed since my 1980 flyby http://t.co/vx5WzPVAMg	has_animated_gif, see_yes

Sixth, we only considered 'overt' form of engagements such as shares and likes. We do not have access to more latent parameters such as number of clicks on URLs, amount of time spent reading an article, etc. These parameters are not accessible through the APIs and are only available to account administrators. Further studies should be performed to include these deeper forms of engagement. Seventh, we limited our data to social media messages written in the English language. Therefore, our conclusions may not be applied to communications in other languages.

6. Conclusions

Social media has become an indispensable component of individual, corporate and organisational communications. In the field of space science, Facebook and Twitter have been enthusiastically embraced by space agencies and astronomers alike to communicate and engage with the public. However, research into the processes and outcomes of these communication efforts are lacking. Without a deeper understanding of the factors that influence audience engagement, social media communication cannot be effective. Our study is an initial step toward filling this gap.

We applied supervised learning algorithms on Facebook and Twitter messages from the field of space science and three other fields (politics, business and non-profit) to predict highly engaging posts and investigated the features that make these posts engaging. We used LIWC, a widely used text analysis program based on psychometrics, to examine the psychological aspects of word use. We hypothesised that (1) highly engaging space science-related

social media messages can be predicted using only content-based features, and (2) engaging space science-related social media messages contain certain psycholinguistic features that are unique to space science. Both hypotheses were rigorously tested and confirmed. We achieved prediction accuracies in the ranges of 90% using four feature sets that explored different content aspects, providing evidence for the ingredients of social media engagement. The top features for engaging messages in space science are photos, anger, visual description, positivity and anxiety for Facebook. For Twitter the top features are visual elements (photos, GIFs, videos), hashtags and certainty. Our uniqueness test uncovered several features that are exclusive to space science: photos, anger, visual description and authenticity for Facebook, and photos, hashtags, visual description and nonfluency for Twitter.

Based on these results, we conclude that there are significant and quantifiable differences between engaging and non-engaging space science-related social media messages. The unique psycholinguistic features that make space science captivating are anger, authenticity, visual descriptions and a tentative tone. Future research should consider a more refined definition of engagement by for example taking into account deeper forms of interaction (e.g. click through rates), and examining data from alternative time periods and science disciplines to validate this paper's findings.

References

- Abdi, H., & Williams, L. J. (2010). *Principal Component Analysis*. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4), 433–459.
- Aggarwal, C. C. (2011). *Social network data analytics, Chapter an introduction to social*

- network data analytics (Vol. 13). Hawthorne, NY 10532: IBM TJ Watson Research Center.
- Bik, H. M., & Goldstein, M. C. (2013). *An introduction to social media for scientists*. Bishop, C. M. (2006). *Pattern recognition and machine learning* (1st ed.). New York: Springer.
- Boyd, D., Golder, S., & Lotan, G. (2010). Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In *System sciences (HICSS), 2010 43rd Hawaii international conference on* (pp. 1–10). IEEE.
- Brossard, D. (2013). New media landscapes and the science information consumer. *Proceedings of the National Academy of Sciences*, 110(3), 14096–14101.
- Brossard, D., & Scheufele, D. A. (2013). Science, new media, and the public. *Science*, 339(6115), 40–41.
- Bucher, T. (2012). Want to be on the top? Algorithmic power and the threat of invisibility on Facebook. *New Media & Society*, 14(7), 1164–1180.
- Castillo, C., Mendoza, M., & Poblete, B. (2011). Information credibility on twitter. In *Proceedings of the 20th international conference on World wide web* (pp. 675–684). ACM.
- Cheng, J., Adamic, L., Dow, P. A., Kleinberg, J. M., & Leskovec, J. (2014). Can cascades be predicted?. In *Proceedings of the 23rd international conference on World wide web* (pp. 925–936). ACM.
- Cotelo, J. M., Cruz, F. L., Enríquez, F., & Troyano, J. A. (2016). Tweet categorization by combining content and structural knowledge. *Information Fusion*, 31, 54–64.
- Davalos, S., Merchant, A., & Rose, G. (2016). Using big data to study psychological constructs: Nostalgia on facebook. *Journal of Psychology & Psychotherapy*, 2015.
- Delello, J. A., & McWhorter, R. R. (2013). New visual social media for the higher education classroom. *The social classroom: Integrating social network use in education*, 368–393.
- Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, 55(10), 78–87.
- Drula, G. (2012). Social and online media research—data, metrics and methods. *Review of Applied Socio-Economic Research*, 3(1), 77–86.
- ESA CCI. (2016). *Space for our climate*. ESA Climate Change Initiative. Retrieved 11 May 2016 from http://www.esa.int/Our_Activities/Observing_the_Earth/Space_for_our_climate/ESA_s_Climate_Change_Initiative_CCI.
- Facebook Help Centre. (2016). *How is engagement rate defined?*. Retrieved 11 May 2016 from <https://www.facebook.com/help/178043462360087>.
- Facebook Inc. (2016). *Company info — Stats*. Retrieved 11 May 2016 from <http://newsroom.fb.com/company-info/>.
- Galtung, J., & Ruge, M. H. (1965). The structure of foreign news the presentation of the Congo, Cuba and Cyprus Crises in four Norwegian newspapers. *Journal of peace research*, 2(1), 64–90.
- Gattiker, U. E. (2011). *Measuring Facebook engagement: What is good? Commetrics.com — tools for benchmarking social media*. Retrieved 11 May 2016 from <http://commetrics.com/articles/2011-engagement-metrics-1-social-media-marketing/#comments>.
- Harman, G. C. M. D. C. (2014). *Quantifying mental health signals in twitter, 2014* p. 51). ACL.
- Hawkins, D. M. (2004). The problem of overfitting. *Journal of chemical information and computer sciences*, 44(1), 1–12.
- Hong, L., Dan, O., & Davison, B. D. (2011). Predicting popular messages in twitter. In *Proceedings of the 20th international conference companion on World wide web* (pp. 57–58). ACM.
- Internet Live Stats. (2016). *In 1 second, each and every second, there are*. Retrieved 11 May 2016 from <http://www.internetlivestats.com/one-second/>.
- Jenders, M., Kasneci, G., & Naumann, F. (2013). Analyzing and predicting viral tweets. In *Proceedings of the 22nd international conference on world wide web companion* (pp. 657–664). International World Wide Web Conferences Steering Committee.
- Jepson, K. M. (2014). *An examination of the effects of digital media on the communication of science*. Doctoral dissertation. Montana State University Bozeman.
- Kacewicz, E., Pennebaker, J. W., Davis, M., Jeon, M., & Graesser, A. C. (2013). Pronoun use reflects standings in social hierarchies. *Journal of Language and Social Psychology*, 33(2), 125–143.
- Kim, J., & Yoo, J. (2012). Role of sentiment in message propagation: Reply vs. retweet behavior in political communication. In *Social informatics (SocialInformatics), 2012 international conference on* (pp. 131–136). IEEE.
- Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media?. In *Proceedings of the 19th international conference on World wide web* (pp. 591–600). ACM.
- Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A. L., Brewer, D., ... Alstyn, M. V. (2009). Computational social science. *Science*, 323, 721.
- Leander, M. (2011). *What is a good engagement rate on a Facebook page? Here is a benchmark for you*. Retrieved 11 May 2016 from <http://www.michaelleander.me/blog/facebook-engagement-rate-benchmark/>.
- Lee, K., Mahmud, J., Chen, J., Zhou, M., & Nichols, J. (2015). Who will retweet this? Detecting strangers from Twitter to retweet information. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 6(3), 31.
- Mahmud, J., Chen, J., & Nichols, J. (2014). *Why are you more Engaged? Predicting social engagement from word use*. arXiv preprint arXiv:1402.6690.
- Maity, S. K., Gupta, A., Goyal, P., & Mukherjee, A. (2015). A stratified learning approach for predicting the popularity of twitter idioms. In *Proc. of AAAI international conference on web and social media (ICWSM)*.
- Ma, Z., Sun, A., & Cong, G. (2013). On predicting the popularity of newly emerging hashtags in twitter. *Journal of the American Society for Information Science and Technology*, 64(7), 1399–1410.
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.
- McCulloch, A. (2015). *NASA makes dwarf planet a MegaStar on Instagram*. *Socialbakers.com*. Retrieved 11 May 2016 from <http://www.socialbakers.com/blog/2436-nasa-makes-dwarf-planet-a-megastar-on-instagram>.
- Murthy, D. (2015). Critically engaging with social media research methods. In *An end to the crisis of empirical Sociology?: trends and challenges in social research*, 81.
- NASA. (2015). *New horizons #PlutoFlyby social media statistics [NASA presentation]*. Retrieved 11 May 2016 from <http://images.spaceref.com/news/2015/nhsocialmedia.pdf>.
- NASA. Climate. (2016). *NASA's role. Taking a global perspective on Earth's climate*. Retrieved 11 May 2016 from http://climate.nasa.gov/nasa_role/.
- Naveed, N., Gottron, T., Kunegis, J., & Alhadi, A. C. (2011). Bad news travel fast: A content-based analysis of interestingness on twitter. In *Proceedings of the 3rd international web science conference* (p. 8). ACM.
- Nicporuc, T. (2014). Comparative analysis of the engagement rate on Facebook and Google Plus social networks. In *Proceedings of international academic conferences (No. 0902287)*. International Institute of Social and Economic Sciences.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12(Oct), 2825–2830.
- Pennebaker, J. W., Booth, R. J., Boyd, R. L., & Francis, M. E. (2015). *Linguistic inquiry and word count: LIWC2015*. Austin, TX: Pennebaker Conglomerates (www.LIWC.net).
- Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. UT Faculty/Researcher Works.
- Pennebaker, J. W., Chung, C. K., Frazee, J., Lavergne, G. M., & Beaver, D. I. (2014). When small words foretell academic success: The case of college admissions essays. *PLoS one*, 9(12), e115844.
- Petrovic, S., Osborne, M., & Lavrenko, V. (2011). *RT to win! Predicting message propagation in twitter*. ICWSM.
- Pinholster, G., & Ham, B. (2013). Science communication requires time, trust, and Twitter. *Science*, 342(6165), 1464–1464.
- Provost, F. (2000). Machine learning from imbalanced data sets 101. In *Proceedings of the AAAI'2000 workshop on imbalanced data sets* (pp. 1–3).
- Rudat, A., & Buder, J. (2015). Making retweeting social: The influence of content and context information on sharing news in Twitter. *Computers in Human Behavior*, 46, 75–84.
- Sandu, O., & Christensen, L. L. (2011). *Outrageous outreach—Unconventional ways of communicating science*.
- Scikit-learn. (2016a). *1.9. Naive Bayes — Scikit-learn 0.18.1 documentation*. Retrieved 23 November 2016 from http://scikit-learn.org/stable/modules/naive_bayes.html.
- Scikit-learn. (2016b). *1.5. Stochastic Gradient Descent — Scikit-learn 0.18.1 documentation*. Retrieved 23 November 2016 from <http://scikit-learn.org/stable/modules/sgd.html>.
- Scikit-learn. (2016c). *1.10 Decision Trees — Scikit-learn 0.18.1 documentation*. Retrieved 23 November 2016 from <http://scikit-learn.org/stable/modules/tree.html>.
- Stieglitz, S., & Dang-Xuan, L. (2012). Political communication and influence through microblogging—An empirical analysis of sentiment in Twitter messages and retweet behavior. In *System science (HICSS), 2012 45th Hawaii international conference on* (pp. 3500–3509). IEEE.
- Suh, B., Hong, L., Pirolli, P., & Chi, E. H. (2010). Want to be retweeted? large scale analytics on factors impacting retweet in twitter network. In *Social computing (socialcom), 2010 IEEE second international conference on* (pp. 177–184). IEEE.
- Sylwester, K., & Purver, M. (2015). Twitter language use reflects psychological differences between democrats and republicans. *PLoS one*, 10(9), e0137422.
- Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology*, 29(1), 24–54.
- Tsur, O., & Rappoport, A. (2012). What's in a hashtag?: content based prediction of the spread of ideas in microblogging communities. In *Proceedings of the fifth ACM international conference on web search and data mining* (pp. 643–652). ACM.
- Twitter Help Centre. (2016). *Tweet activity dashboard — definition*. Retrieved 11 May 2016 from <https://support.twitter.com/articles/20171990>.
- Vadivu, V. M., & Neelamalar, M. (2015). Digital brand management—a study on the factors affecting customers' engagement in Facebook pages. In *Smart technologies and management for computing, communication, controls, energy and materials (ICSTM), 2015 international conference on* (pp. 71–75). IEEE.
- Veltri, G. A., & Atanasova, D. (2015). Climate change on Twitter: Content, media ecology and information sharing behaviour. *Public Understanding of Science*. <http://dx.doi.org/10.1177/09636625151613702>.