Current Research in Parasitology & Vector-Borne Diseases 4 (2023) 100138

Contents lists available at ScienceDirect

Current Research in Parasitology & Vector-Borne Diseases

journal homepage: www.sciencedirect.com/journal/current-research-in-parasitology-and-vector-borne-diseases

Twitter trends in #Parasitology determined by text mining and topic modelling

John T. Ellis [a,*], Michael P. Reichel [b]

[a] School of Life Sciences, University of Technology Sydney, Broadway, NSW, Australia
[b] Department of Population Medicine & Diagnostic Sciences, College of Veterinary Medicine, Cornell University, Ithaca, NY, USA

ARTICLE INFO

Keywords:
Twitter
Keywords
Keyphrases
Topic model
Sentiment
E-Professionalism

ABSTRACT

This study investigated the emergence and use of Twitter, as of July 2023 being rebranded as X, as the main forum for social media communication in parasitology. A dataset of tweets was constructed using a keyword search of Twitter with the search terms 'malaria', '*Plasmodium*', '*Leishmania*', '*Trypanosoma*', '*Toxoplasma*' and '*Schistosoma*' for the period from 2011 to 2020. Exploratory data analyses of tweet content were conducted, including language, usernames and hashtags. To identify parasitology topics of discussion, keywords and phrases were extracted using KeyBert and biterm topic modelling. The sentiment of tweets was analysed using VADER. The results show that the number of tweets including the keywords increased from 2011 (for malaria) and 2013 (for the others) to 2020, with the highest number of tweets being recorded in 2020. The maximum number of yearly tweets for *Plasmodium*, *Leishmania*, *Toxoplasma*, *Trypanosoma* and *Schistosoma* was recorded in 2020 (2804, 2161, 1570, 680 and 360 tweets, respectively). English was the most commonly used language for tweeting, although the percentage varied across the searches. In tweets mentioning *Leishmania*, only ~37% were in English, with Spanish being more common. Across all the searches, Portuguese was another common language found. Popular tweets on *Toxoplasma* contained keywords relating to mental health including depression, anxiety and schizophrenia. The *Trypanosoma* tweets referenced drugs (benznidazole, nifurtimox) and vectors (bugs, triatomines, tsetse), while the *Schistosoma* tweets referenced areas of biology including pathology, eggs and snails. A wide variety of individuals and organisations were shown to be associated with Twitter activity. Many journals in the parasitology arena regularly tweet about publications from their journal, and professional societies promote activity and events that are important to them. These represent examples of trusted sources of information, often by experts in their fields. Social media activity of influencers, however, who have large numbers of followers, might have little or no training in science. The existence of such tweeters does raise cause for concern to parasitology, as one may start to question the quality of information being disseminated.

1. Introduction

Twitter is a popular online microblogging platform that allows people, including scientists, to express their thoughts, opinions and feelings through short text messages called tweets (Bik and Goldstein, 2013; Wolf, 2017). As recently as 10 years ago, Twitter was not considered an important tool for dissemination of scientific knowledge, despite recognition of its potential for the rapid communication of knowledge (Priem and Costello, 2010). This was partly caused by the limited number of scientists that were using Twitter at that time. Indeed, a 2014 study showed that only 3.7% of parasitology publications in PubMed were associated with Twitter activity (Haustein et al., 2014). Recently, we documented that Twitter is now the main form of social media communication being used in parasitology (Ellis et al., 2021).

Tweets represent short text messages (typically limited to 280 characters) that may contain hashtags, urls, images and @usernames (https://www.socialmediatoday.com/content/structure-perfect-tweet). In April 2023, under new management, a paid subscription to "Twitter Blue", allows users to write "tweets" up to 10,000 characters in length. The helpful links in tweets often connect to sites that include blogs, publications, conferences and other useful sources of information. Scientists have gone on record previously portraying the benefits of using Twitter as "learning things that are going on in the world of science and medicine", "it's another teaching tool", and "demonstrates a

* Corresponding author.
E-mail address: John.Ellis@uts.edu.au (J.T. Ellis).

https://doi.org/10.1016/j.crpvbd.2023.100138
Received 10 July 2023; Received in revised form 8 August 2023; Accepted 10 August 2023
Available online 12 August 2023

commitment to public outreach" (You, 2014). Despite these obvious benefits to scholarly activity, the spread of misinformation by the use of Twitter has been raised as a concern (Bermingham and Smeaton, 2011), as well as a means for the circulation of fake news (Bovet and Makse, 2019) and the uncivil discourse that may appear in tweets, often in an impulsive way (Ott, 2017).

Twitter is now increasingly being used in all sorts of aspects of science (Lee, 2019). These often include research related activities involving the processes of research collaboration and dissemination of research results, science outreach and communication, its use as an educational resource and in professional development, recruitment of staff and students and the advertisement of events such as conferences and the active participation in them (Shiffman, 2012; Collins et al., 2016; López-Goñi et al., 2016, 2019; Hull and Dodd, 2017; Cevik et al., 2019; Dol et al., 2019; Pizzuti et al., 2020). Recently, there have been tweets to also promote gender equity (Calvani et al., 2023) along with other current topics that have the potential to affect research in this discipline, such as climate change (Fownes et al., 2018). Many scientific journals facilitate research dissemination *via* logos embedded at the head of a published paper: Twitter, Facebook and e-mail logos are commonly used. Not only is Twitter the first link for sharing of the article, but alternative metrics (Altmetrics) for the article considers Twitter activity, and this may be used as a metric by which a researcher's work is rated at a professional level (Ellis et al., 2021).

Research impact and engagement may come in many forms and includes public and industry engagement, science communication, consultancy and commercialisation, citizen science type activities and much more. Along these lines, Twitter is considered as a method for engaging with the public and generating influence on public perceptions about science. A 2016 survey demonstrated that the main motivation behind a scientist's use of Twitter is scientific exchange (Collins et al., 2016); others however demonstrated that a scientist's followers on Twitter may include a diverse group of non-scientists, including those from media, members of the public with no known association with science and policy makers (Côté and Darling, 2018). It was therefore argued that building a social media profile is necessary for academics and scientists, as it represents a significant and important contribution to scientific outreach.

The topics found in tweets from scientists have also raised interest; research and teaching (scholarly communication) make up approximately one third of the content analysed, along with personal, social and political topics (Jünge and Fähnrich, 2020). Sinnenberg and colleagues pointed out that little is known about the use of Twitter in Public Health (Sinnenberg et al., 2017) while Twitter activity in relation to disease surveillance is well known. As an example, ProMED-mail was established in 1994 "as an email service to identify unusual health events related to emerging and re-emerging infectious diseases and toxins affecting humans, animals and plants" (Carrion and Madoff, 2017). ProMED now also distributes its information *via* social media including Twitter (@ProMED_mail). The subsequent development of real-time, internet-based biosurveillance methods is documented elsewhere, including the use of Twitter for surveillance of infectious diseases (Gomide et al., 2011; Broniatowski et al., 2013; Pollett et al., 2017; Oren et al., 2020). The recent COVID-19 pandemic highlights the positive and negative sides of Twitter activity in several ways within the wider science discipline (González-Padilla and Tortolero-Blanco, 2020). Twitter provides a rapid method of dissemination of knowledge worldwide; however, it is also associated with the spread of incorrect and misleading information (misinformation) (Krittanawong et al., 2020), which is known to contribute to panic (Depoux et al., 2020) and vaccination hesitancy (Rosenberg et al., 2020).

The use of Twitter in the parasitology discipline has not been investigated thoroughly. We recently documented the rise in use of social media associated with parasitology, including use of Twitter (Ellis et al., 2021). Others have discussed the role of Twitter in teaching and learning activities associated with parasitology (Jabbar et al., 2016); examples cited included promotion of parasitology case studies, journal clubs and following handles of reputable scientific journals for factual content (Cevik, 2019). In the study presented here, we investigated Twitter activity associated with six major research areas within parasitology, the types of content present in tweets and their sentiment, as well as the institutional affiliation of those heavily invested in tweeting about parasitology. We hope this will lead to an improved understanding of the impact and engagement by parasitologists for the greater good of the discipline.

2. Materials and methods

All analyses were conducted on a 64-bit HP Elitebook 840 G5 laptop with 16 GB of RAM and an Intel i7-8550U CPU. A schematic of the analyses conducted is shown in Fig. 1. This study was conducted as academic, non-commercial research to provide benefits to the discipline of parasitology. Several reviews have outlined the processes of Twitter data analyses in Python and provided a guide to the subsequent analyses performed here (Wisdom and Gupta, 2016; Alshammari and AlMansour, 2019).

2.1. Scraping and preprocessing of tweets

Twitter scraping was performed using the *Twint* package (https://github.com/twintproject/twint) with Python 3.8.3 run in a Jupyter notebook (v6.0.3). Scraping was performed using a keyword (e.g.

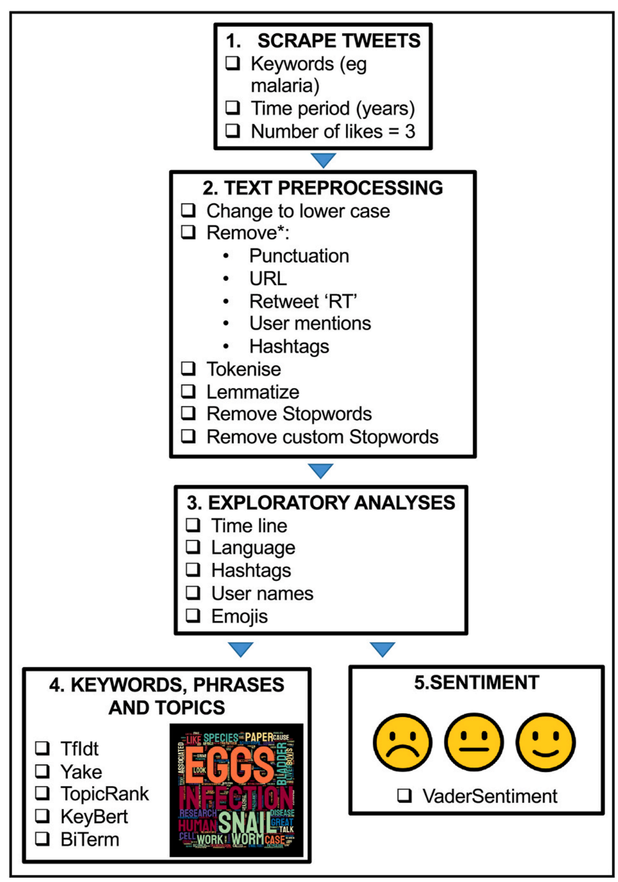

Fig. 1. Summary of the workflow and analyses used in this study. Tweets were scrapped, their text preprocessed, and analysed for keywords, phrases and topics. Analyses were conducted in a Jupyter notebook running Python. *Note*: Level of preprocessing depends on the analyses conducted. The main tools used in the analyses and visualisation are named (e.g. NLTK, Natural Language Tool Kit). Stopwords and custom stopwords (e.g. malaria) were incorporated into the preprocessing as necessary.

'malaria'), restricted by dates (e.g. 2020-01-01 to 2020-12-31), and a minimum number of likes ($n = 3$). The choice of the keywords used in the search were based on previous analyses that showed malaria, *Plasmodium*, *Toxoplasma*, *Leishmania*, *Trypanosoma* and *Schistosoma* were among the most species referred to in tweets (Ellis et al., 2021). Searches using the term 'malaria' were also conducted in order to determine how it differed from searches using '*Plasmodium*'. This scraping strategy generated datasets (in json format) containing viable numbers of tweets for analysis.

Twitter data were loaded into a Python pandas (v1.2.4) dataframe for further preprocessing and analyses (https://pandas.pydata.org/) (McKinney, 2010). Each column of the dataframe contains unique metadata associated with a tweet (e.g. text, username, number of retweets) and rows represent individual tweets. Columns not needed in further analyses were dropped, mainly because they contained little useful data (e.g. 'id', 'conversation_id', 'created_at', 'timezone', 'user_id', 'cashtags', 'place','quote_url', 'near', 'geo', 'source', 'user_rt_id', 'user_rt', 'retweet_id', 'retweet_date', 'translate', 'trans_src', 'trans_dest', 'video', 'retweet'). As tweets contain text and embedded URLs, pictures, usernames, emoticons, abbreviated and misspelt words, the text data of the tweet was subject to preprocessing. All text was changed to lower case. Python regular expressions were used to remove unwanted punctuation, URLs ("http://url") as well as tags related to retweet (RT), user mentions (@) and hashtags (#). The text was tokenized and lemmatized using the Natural Language Took Kit (NLTK, https://www.nltk.org/) word_tokenize and WordNetLemmatizer (Bird et al., 2009).

The removal of stopwords, representing common words of little meaning to topic modelling such as "the" and "and", represents an important step in the preprocessing of text. Common English stopwords are included in NLTK, but a wide range of others exist including domain-related terms (Sarica and Luo, 2021). In order to identify these later terms, FreqDist () in NLTK was used to determine and rank the frequency of terms in the preprocessed data, and selected terms were later used as custom_stopwords (see below). In order to expand the NLTK list of stopwords, we also merged those from elsewhere (Sarica and Luo, 2021) including those known as the Glasgow, SMART, USPTO and Technical stopwords. This provided a list of 652 unique, common terms as stopwords. Custom-stopwords representing words either related to the search terms or present at very high frequency were also removed from text as follows for specific analyses: (a) Malaria ('malaria', 'parasite', 'parasites', 'plasmodium', 'falciparum', 'vivax'); (b) *Plasmodium* ('parasite', 'parasites', 'plasmodium', 'falciparum', 'vivax', 'malaria', 'slime mold', 'slime', 'plasmodium malaria', 'fungus', 'molds', 'mold', 'sporangia'); (c) *Leishmania* ('leishmania', 'leishmaniasis', 'leishmaniosis', 'parasite', 'parasites', 'donovani'); (d) *Trypanosoma* ('chagas', 'disease', 'parasite', 'parasites', 'whitehelmets', 'tryp', 'trps', 'trypanosoma', 'trypanosome', 'brucei', 'cruzi', 'gambiense', 'rhodesiense', 't. cruzi', 't.brucei', 't.gambiense', 't.rhodesiense'); (e) *Toxoplasma* ('gondii', 't.gondii', 'toxoplasma', 'toxo', 'cat', 'toxoplasmainfected', 'cats', 'aaah', 'footballs', 'mice', 'parasite', 'parasites', 'infected', 'foodthought', 'foodsafety', 'foods', 'foodhygiene', 'foodbornezoonoses', 'foodborne', 'food', 'fomc'). Stopwords and custom_stopwords were removed from texts during preprocessing.

Finally, words were filtered by their length and words containing 4 or more characters were retained. The processed text was saved as a text file.

2.2. Exploratory data analyses

The features (columns) of the Twitter data were investigated by the standard methods of exploratory data analyses using in-built Python methods associated with the pandas dataframe, such as value_counts and unique (Chen, 2018). Other than for analyses of tweet language, analyses were restricted (using groupby) to tweets in the English language.

2.3. Keyword and phrase extraction from tweets

Keywords and phrases were extracted from preprocessed tweets using the Python pke pipeline (https://github.com/boudinfl/pke) and either the TfIdf, YAKE or TopicRank models (Boudin, 2016). These models represent commonly used approaches to define keyphrases and topics found in short text based on either statistical (TfIdf and YAKE) (Campos et al., 2020) or graph-based methods (TopicRank) (Bougouin et al., 2013). The github workflows were used to select the 30 highest scoring words or phrases. The logical relationship amongst the three lists generated was investigated using Venn diagrams (https://pypi.org/project/matplotlib-venn/). Word identities were produced using Python's intersection function. As the arrays created using this pipeline for the malaria dataset were very large these data were not analysed in this way.

Since single words on their own lack context, the content of the tweets was also analysed using KeyBert with maximal marginal relevance (MMR) (https://github.com/MaartenGr/KeyBERT) (Grootendorst, 2020). 4-gram phrases (4-g) were extracted, and MMR was set at 0.2 to encourage selection of phrases with associated diversity. This approach generates a series of short lists (tuples) of related, overlapping keyphrases associated with keywords. Wordclouds containing the most frequent 200 words present in the phrases identified by KeyBert were produced using Stylecloud v0.5.2 (https://github.com/minimaxir/stylecloud).

2.4. Biterm topic modelling

Topic modelling was also used to identify groups of words representing the topics discussed in the tweet datasets. The biterm topic model was used because of its well-known performance using short text such as Twitter data (Yan et al., 2013). This model is "inspired by the idea that topics are groups of correlated words where the correlation is revealed by word co-occurrence" (Jónsson and Stolee, 2015). A Python implementation of the Biterm Topic Model (BTM, https://github.com/markoarnauto/biterm) was used. Preprocessing of the tweet text column in pandas was performed as described above. Stopwords were removed, along with common words related to the search terms as described above. The dataframe was then filtered for tweets in the English language, and the English text converted to a list of lists of words, representing lists of words from the tweets. CountVectoriser (from scikit-learn v0.24.2 (Pedregosa et al., 2011), https://scikit-learn.org/stable/) was used to produce a matrix of word counts. A topic model was generated using the BTM as described (https://github.com/markoarnauto/biterm). Visualisation of topic models was performed using the Python library pyLDAvis (https://pyldavis.readthedocs.io/en/latest/index.html) which is a port of LDAvis (Sievert and Shirley, 2014).

2.5. Emoji and sentiment analyses using VADER

A merged dataset was constructed from the initial tweet data by merging all the search results across the species. This merged dataset was preproccessed for emoji analyses using ekphrasis, a tool specifically designed for preprocessing text from social media (https://github.com/cbaziotis/ekphrasis) (Baziotis et al., 2017). The SocialTokenizer was used to tokenize the text and the segmenter and corrector used was Twitter. For each tweet the emojis were counted and summarised across the entire dataset.

Sentiment analysis used the following workflow on the tweet text data. As punctuation, capitalisation of words, various stopwords and emoticons can influence sentiment in a tweet, only a minimal preprocessing of the text described above was performed and these features were retained (Vashishtha and Susan, 2019; Pano and Kashef, 2020). All text was changed to lower case and preprocessing was limited to removing unwanted punctuation, URLs ("http://url") as well as tags related to retweet (RT), user mentions (@) and hashtags (#). Tweets were restricted to the English language for sentiment analyses. Python's VADER (v3.3.2) package (Hutto and Gilbert, 2014) was used for analysis of sentiments. Histograms were plotted using the seaborn (v0.11.0) histplot function (Waskom, 2021). VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and rule-based sentiment analysis tool specifically used for analysis of sentiments found in social media text such as tweets (Elbagir and Yang, 2020). VADER provides positive, neutral and negative scores that are ratios for the proportions of text that fall into each category (and so the total adds up 1). VADER also provides a compound score that is a normalized, weighted composite score derived by summing the valence scores of each word in the lexicon, adjusted according to the rules, and then normalized to be between −1 (most extreme negative) and +1 (most extreme positive). Further details can be found at the Github page (https://github.com/cjhutto/vaderSentiment).

3. Results

3.1. Exploratory data analyses

As a first preliminary investigation of Twitter content on parasitology, a dataset was constructed using keyword searches (malaria, *Plasmodium*, *Toxoplasma*, *Leishmania*, *Trypanosoma* or *Schistosoma*), restricted by year (2011–2020) and likes (greater than or equal to 3). A summary of these data is shown in Fig. 2. The choice of these keywords and timeline was guided by our previous studies that identified these terms as being predominant in Twitter activity (Ellis et al., 2021). We note here that the use of abbreviated terms such as 'toxo', 'schisto' and 'tryp' are used just as frequently (if not more) in tweets compared to the full taxon term; for 2020 the use of these abbreviated terms in a search returned 2096, 1574 and 1529 tweets, respectively.

For all searches, the number of tweets including the keywords increased yearly from 2011 (for malaria) and 2013 (for the others) to 2020, with the highest number of tweets being recorded in 2020. Over the period analysed, the number of tweets mentioning malaria ranged from ~30 to 110,000, while the other terms were far less popular. The maximum number of yearly tweets in 2020 for the terms *Plasmodium*, *Leishmania*, *Toxoplasma*, *Trypanosoma* and *Schistosoma* was 2804, 2161, 1570, 680 and 360 tweets, respectively.

After preprocessing of the search text (without custom_stopwords), the identification of the most frequent words used was investigated using FreqDist. As an example, in the malaria tweets, the word "malaria" was used nearly 200,000 times and represented the most frequent, dominant term; the next five words were drug, people, covid, health and fight. Based on the relevance of these terms to parasitology, custom_stopwords were restricted to malaria-related terms. For the other texts, the five most frequent words were as follows: *Plasmodium* ("Plasmodium", "malaria", "falciparum", "parasite", "vivax"; *Leishmania* ("leishmania", "parasite", "work", "leishmaniasis", "parasites"; *Trypanosoma* ("trypanosome", "cruzi", "brucei", "disease", "parasite"; *Toxoplasma* ("toxoplasma", "gondii", "parasite", "infected", "cats"; *Schistosoma* ("schistosoma", "mansoni", "haematobium", "parasite", "schistosomiasis"). Custom_stopwords were identified from these, related and other frequent terms, and were refined throughout the study.

3.2. Language

For each search, the language of the tweets was determined from the language metadata associated with each tweet (Table 1). English was the most commonly used language for tweeting, although the percentage varied across the searches. For example, in tweets mentioning *Leishmania*, only ~37% were in English, with Spanish being more common. Across all the searches, Portuguese was another common language found. The profile of the languages found in the individual searches varied significantly amongst the searches.

3.3. Usernames

To identify the main influencers in parasitology from Twitter data, usernames were analysed according to amount of tweet activity. Table 2 shows the affiliation of the top 25 usernames for each search; there are several points worth raising from these data. First, academics and non-government organisations are the most common group of tweeters in these lists. 80 academics feature in these lists (not shown); for example, @wjsullivan (Professor Bill Sullivan) is a researcher on *Toxoplasma*, @pauljbrindley (Professor Paul Brindley) studies schistosomiasis, @ariel_lab (Professor Ariel Silber) studies trypanosomiasis and @ rayner_julian (Professor Julian Rayner) studies malaria (from the *Plasmodium* search). Secondly, malaria Twitter activity is dominated by non-government organisations. The username lists for malaria and *Plasmodium* also differ; the malaria list identifies usernames associated with the global "fight against malaria" and "end malaria". Thirdly, the usernames from each search are quite unique with little overlap suggesting users

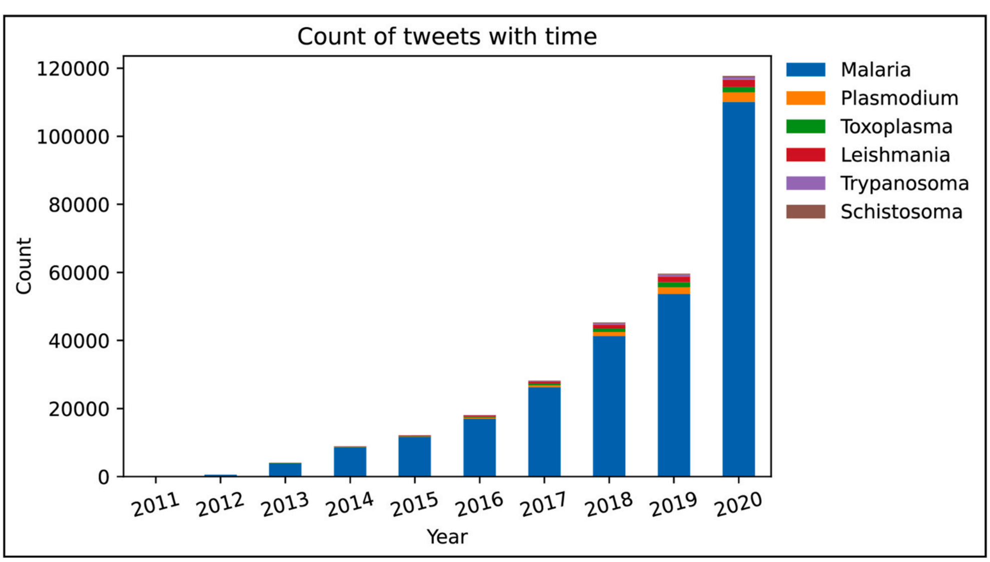

Fig. 2. Change in Twitter activity with time. Search results for tweets associated with the keywords malaria, *Plasmodium*, *Toxoplasma*, *Leishmania*, *Trypanosoma* and *Schistosoma* are shown.

Table 1
Languages used in tweets about malaria, *Plasmodium*, *Leishmania*, *Trypanosoma*, *Toxoplasma* or *Schistosoma* over the time period studied to 2020. The top five languages from each search are highlighted in bold (excluding those that were undetermined).

Language	Country code[a]	Malaria	*Plasmodium*	*Leishmania*	*Trypanosoma*	*Toxoplasma*	*Schistosoma*
English	en	**201,232**	**5137**	**2359**	**1204**	**3830**	**725**
Spanish	es	**33,610**	**1056**	**2764**	**295**	**465**	**76**
Portuguese	pt	**15,214**	**121**	103	**126**	**117**	**49**
Indonesian	in	**5537**	**63**	15	10	**125**	4
Italian	it	**4092**	23	**480**	9	12	14
French	fr	3161	**114**	20	7	**184**	12
German	de	2328	18	8	4	82	5
Tagalog	tl	1888	26	13	14	7	**32**
Dutch	nl	1095	4	20	0	10	5
Catalan	ca	1028	33	**267**	17	39	11
und[b]	und	663	91	55	25	45	53
Hindi	hi	598	1	2	0	0	0
Swedish	sv	547	8	1	0	9	1
Polish	pl	365	2	0	4	7	0
Japanese	ja	230	25	23	16	64	3
Finish	fi	219	1	4	0	3	0
Romanian	ro	188	41	38	48	15	18
Haitian	ht	177	1	0	0	0	1
Danish	da	169	4	0	1	0	1
Estonian	et	159	3	2	0	2	1
Norwegian	no	116	0	1	0	2	1
Turkish	tr	109	22	122	32	78	7
Arabic	ar	93	29	9	13	41	61
Basque	eu	59	2	1	0	4	0
Tamil	ta	31	0	8	0	0	0
Urdu	ur	30	0	12	0	0	0
Sum		272,938	6825	6327	1825	5141	1080
% English		73.7	75.3	37.3	66.0	74.5	67.1
Unique users		127,662	2748	1876	1029	2728	755

[a] ISO 639 country code.
[b] Undetermined.

Table 2
Total number of tweets by category of tweeter for those tweeting in parasitology from the selected searches for malaria, *Plasmodium*, *Leishmania*, *Trypanosoma*, *Toxoplasma* and *Schistosoma*.

Category	Malaria	*Plasmodium*	*Toxoplasma*	*Leishmania*	*Trypanosoma*	*Schistosoma*
Academic (Academic/scientist/professional)	1965	928	342	132	125	116
Academic Institution	1295	147	177	116	164	5
Government	2505	0	0	0	0	0
Individual (not identifiable as academic category)	1086	72	267	208	0	0
Scientific Journal	473	277	41	40	89	51
Non-government organisation (NGO)	8732	0	82	1283	9	16
Organisation/group	0	0	40	0	0	8
Undeterminable	0	414	54	460	61	0
Unique usernames (N)	127,622	2748	2728	1876	1029	755
Total: Academic/scientist/professional	3260	1075	519	248	289	121
Total: Organisations/government/journals	11,710	277	163	1323	98	75
Total: Individuals and Undeterminable	1086 (6.8%)	486 (26.4%)	321 (32.0%)	668 (29.8%)	61 (13.6%)	8 (3.9%)

tend to tweet and follow others working in the same discipline area (not shown). There are few usernames appearing across the lists, although @hanniepower is of note for tweeting across the many different areas of parasitology (but not on *Leishmania* in Spanish). Finally, it is noticeable that scientific societies such as the International Society of Protistologists and Journals (Microbes & Infection, Trends in Parasitology, PLoS Neglected Tropical Diseases) are represented. It is interesting to note the inclusion of "generalist" tweeters, such as @Fact (Interesting facts about life); @factsinyourface (The BEST FACTS on Twitter straight in your face!) and @UberFacts (The most unimportant things you'll never need to know) in the *Toxoplasma* username list, which is unusual.

3.4. Hashtags

The main hashtags appearing in the Twitter data (as frequency counts) within each search were determined (Table 3). In addition to finding the search terms and diseases related to the search terms, several other trends were identified. Reference to conferences (e.g. mpm20, mam2020) and world days (malaria, mosquito, chagas, parasite) are readily identifiable. Discussion on publishing topics can also be identified from hashtags such as open access, biorxiv, and several PLoS journals.

3.5. Keywords and phrases present in tweets

Keyphrase extraction and biterm topic modelling was performed to identify groups of words in tweets representing areas of discussion. These approaches were used because of the very short length of text in a tweet and their suitability of use. Detailed results of keywords and phrases identified are provided in Supplementary file S1.

The intersection of the top 100 keywords obtained from TfIdf, Yake and TopicRank were determined (Table 4). This provided 127 unique terms in the lists, but only two terms were present in each list; these were "great" and "host". Even at this simple level, several trends were obvious

Table 3
Summary of main hashtags (and their frequency of use) found in tweets from the selected searches for malaria, *Plasmodium*, *Leishmania*, *Trypanosoma*, *Toxoplasma* and *Schistosoma*. Examples highlighted in bold include world days, conferences, and publishing topics.

Malaria	*Plasmodium*	*Leishmania*	*Trypanosoma*	*Toxoplasma*	*Schistosoma*
Unknown	Unknown	Unknown	Unknown	Unknown	Unknown
malaria	plasmodium, malaria	leishmania	usmle	toxoplasma	schistosoma
worldmalariaday	malaria	lennon	trypanosoma	toxoxv	schistosomiasis, schistosoma
covid19	plasmodium	oro	chagas	**biorxiv**	beatntds
sleepunderthenetug	malaria, plasmodium	blanca	chagas, trypanosoma	toxoplasma, protists	schistosomiasis
endmalaria	**biorxiv**	valencia, rt	**kmcb2019**	**mpm2019, toxoplasma**	gipath
malaria, endmalaria	**mam2020**	marrón	trypanosoma, chagas	**mpm2020**	ihpe
malariamustdie	**worldmalariaday**	socungos10, etnecessito, elsqueninguvol, ungoscomqualsevolaltre	protists	protists	bspman2019
bbnaija	emblmalaria, plasmodium, malaria	leishmaniasis, leishmania	**biorxiv**	**openaccess**	schistosoma, schistosomiasis
coronavirus	emblmalaria	adopta	**mpm2019, trypanosoma**	toxoplasma, toxoplasmosis	**plosntds**
malaria, worldmalariaday	**wrc2018**	fiona, navidad	**plosntds**	toxoplasma, parasite	**schistosomes, plosbiology**
stepupthefight	**mpm2018, womeninparasitology**	valencia, ayuda	**openaccess**	parasite	**openaccess**
plasmodium, malaria	protists, malaria	leishmaniasis	trypanosoma, protists	**plospathogens**	immunology
malaria, malariamustdie	womeninmalaria, plasmodium	4voiceless	**trypanosoma, mpm2019**	usmle	schistosoma, immunology
worldmalariaday, malaria	**mpm2020**	adopcion, adoptanocompres	**worldchagasday**	midweekmicro	schistosoma, bladder, cancer, proteomics
malaria, globalhealth	**mpm2018**	guinda	laporkarantina	**toxoplasma, cats, plosbiology**	asmclinmicro, mayoclinmicro, parasites
worldmosquitoday	plasmodium, malaria, parasitology	madrid	parasitesinisolation	protists, toxoplasma	cytopath
malaria, defeatmalaria	copaplasmodium	rudolf	parasitesinisolation, ntd	toxoxv, toxoplasma	pathbugs, crittersontwitter
venezuela	malaria, protists	pirata	microrounds, asmclinmicro, idtwitter	**mpm2016**	**sicb2019**
malaria, covid19	malaria, endmalaria	vida	repasomd	**parasiteday2019**	vaccine
hydroxychloroquine	**openaccess**	etnecessito, ambtuserebo	cd8_t_cell_immunity	**mpm2018**	hugotmedtech
defeatmalaria	womeninmalaria, plasmodium, malaria	urgent, ungoscomqualsevolaltre	**plosbiology**	**msphere**	**parasiteday2019**
womeninmalaria, malaria	micromoocsem	leishmania, leishmaniasis	trypanosoma, sleepingsickness	**mpm2020, toxoplasma**	**plospathogens**
covid19, malaria	**worldmosquitoday**	duca	chagas, chagas	nachtschicht	**tropmed2020**
zeromalariastartswithme	plasmodium, malaria, womeninmalaria	protists	trypanosoma_brucei, 3d_genomearchitecture, localchromatinconformation, antigenicvariation	**mpm2019**	schistosoma, mansoni

from these lists. For example, *Toxoplasma* contained keywords relating to mental health including depression, anxiety and schizophrenia. The *Trypanosoma* list referenced drugs (benznidazole, nifurtimox) and vectors (bugs, triatomines, tsetse), while the *Schistosoma* list referenced areas of biology including pathology, eggs and snails. These results are informative in that they identify popular areas of parasitology that are discussed on Twitter. Several celebratory words were present in the lists, such as “happy”, “congratulations”, “amazing”, and “great”, that suggests a positive sentiment present in the tweets. The inclusion of the word “today” in four of the lists in Table 4, attests to the currency of the information being circulated around Twitter.

Consideration of the 4-g keywords extracted using KeyBert expands on these topics (Supplementary file S1). For example, searching the *Trypanosoma* 4-g for “genome” shows expansion to pacbio and nanopore sequencing and “significantly improves genome assembly”. A similar search of the *Leishmania* 4-g with “genome” reveals discussion about genome assembly and diversity (on a variety of species) and CRISPR. The wordclouds in Fig. 3 show the most frequent words present in the KeyBert phrases, which helps in their interpretation. The term “infection” is a dominant word (and arguably could have been included as a stopword) and is linked to a wide variety of topics in the KeyBert Phrases. Another of the most common words appearing in all the searches is the term “paper”, and searches of the KeyBert phrases shows a great deal of discussion is occurring about published papers. From a science communication viewpoint, the term “talk” also features.

If we consider each of the searches individually, it is possible to investigate the main topics of discussion in these discipline areas. For example, the dominant word in *malaria* is drug, and searching the KeyBert phrases shows phrases relating to drug resistance, drug target, drug development and fake drugs. For *Plasmodium* (and we eliminate potential stopwords human and infection) mosquito is the dominant word, linking to discussion on malaria transmission. The *Leishmania* search (excluding another potential stopword “work”) shows a wide variety of topics; worthy of mention is reference to “student” which expands in phrases to celebratory comments, workshop and a variety of student activities. The focus on mental health in *Toxoplasma* is clearly visible in Fig. 3E and the importance of eggs and snails to *Schistosoma* is evident in Fig. 3F.

Biterm topic modelling was used to identify longer groups of words that were commonly used in tweets (Supplementary file S2). The results obtained reinforce those observations reported above; *Toxoplasma* topics were very much focussed on brain, while many of the *Schistosoma* topics were focussed on biology related to eggs and snails. However, across the datasets we see reference to “paper”, which indicates a great

Table 4
The top 100, most common, keywords obtained from TfIdf, YAKE and TopicRank for the Twitter data about *Plasmodium*, *Leishmania*, *Trypanosoma*, *Toxoplasma* or *Schistosoma*. Words identified by all three algorithms are presented.

Plasmodium (176)[a]	*Leishmania* (180)[a]	*Trypanosoma* (170)[a]	*Toxoplasma* (178)[a]	*Schistosoma* (168)[a]
antimalarial	life	protozoan	toxoxv	thanks
liver	parasitology	parasitology	thanks	liver
life	story	tsetse	chronic	life
vaccine	clinical	cardiomyopathy	life	male
female	plasmodium	plasmodium	parasitology	vaccine
invasion	skin	today	invasion	cancer
humans	cells	resistance	mitochondria	skin
species	species	congolense	plasmodium	species
today	today	drug	cells	bladder
talk	resistance	toxoplasma	humans	worm
review	crispr	intracellular	health	human
drug	infantum	nice	today	drug
toxoplasma	drug	cell	talk	fluke
genetic	infection	heart	crispr	infection
cell	flies	nifurtimox	mouse	praziquantel
host	trypanosoma	host	infection	cell
happy	host	sickness	anxiety	ihpe
molecular	sandfly	usmle	mind	water
immunity	molecular	molecular	latest	host
session	malaria	america	depression	happy
disease	visceral	seminar	fear	molecular
cycle	leishmania	collaboration	poop	intestinal
stage	disease	protists	cool	atlas
genome	people	leishmania	host	eggs
anopheles	macrophages	cycle	immune	pathology
mosquito	biology	mitochondrial	malaria	disease
transmission	genome	bugs	rage	people
knowlesi	home	biology	proteins	gipath
scientists	amazing	genome	disease	transmission
friends	sand	metabolic	schizophrenia	cercariae
treatment	treatment	trypanosomiasis	people	urine
great	aneuploidy	transmission	rats	beatntds
week	cutaneous	american	mitochondrial	amazing
genes	great	triatomine	biology	mass
virus	congratulations	benznidazole	student	treatment
blood	week	treatment	apicomplexan	diagnostics
protein	virus	african	great	snails
	dogs	great	congratulations	urinary
		protein		spine
		blood		great
				protein
				blood

[a] Total number of unique words in the TfIdt, YAKE and TopicRank set of 300 words.

deal of discussion is occurring about publications *via* Twitter. This is particularly noticeable across the *Leishmania*, *Trypanosoma* and *Plasmodium* topic models. The coherence scores provided for the topics in Supplementary file S2 reflect how coherently the words in the topic describe the topic. For *Toxoplasma*, the topic with the highest coherence score, contains words (brain, cells, immune, cyst, formation, innate, important, required, synthase) suggestive of cyst formation in response to host innate immunity; in the case of *Plasmodium* mosquito transmission of malaria dominates. For *Trypanosoma*, the topic words (bugs, transmitted, triatomine, treat, nifurtimox, benznidazole, cardiomyopathy, usmle, dilated, kissing) suggest discussion on the United States Medical Licensing Examination (USMLE) is occurring with a focus on the treatment, pathology and biology on Chagas' disease. For *Schistosoma*, the topic words (study, development, people, infect, insights, work, biology, worldwide, million, schistosomes) suggest focus on the worldwide effort into research in this area; and for *Leishmania*, the topic of highest coherence is focussed on "save the galgo" which is a Spanish greyhound of conservation interest (https://www.galgonews.com/symptoms-of-leishmania-in-dogs.html).

3.6. *Emoji and sentiment analyses*

A total of 489 different emoji were present in the merged dataset created from malaria, *Plasmodium*, *Leishmania*, *Trypanosoma*, *Toxoplasma* and *Schistosoma*. The 25 most popular emoji in the merged data were 👉 (backhand index pointing right, $n = 601$), ♡ (heart, $n = 594$), 🙏 (folded hands, $n = 567$), 😂(face with tears of joy, $n = 506$), 🐾(paw prints, $n = 396$), 📧(e-mail, $n = 382$), 🔬(microscope, $n = 318$), 🐶(dog face, $n = 285$), 🏻(light skin tone, $n = 266$), 😍(smiling face with heart-eyes, $n = 255$), 💕(two hearts, $n = 224$), ➡(right arrow, $n = 222$), ✅check mark, $n = 221$), 😭(loudly crying face, $n = 215$), 👏(clapping hands, $n = 204$), 👇(backhand index pointing down, $n = 162$), 💔(broken heart, $n = 161$), ♥(heart suit, $n = 148$), 😊(smiling face with smiling eyes, $n = 144$), 🏼(medium-light skin tone, $n = 142$), 😀(grinning face, $n = 135$), ⚠ (warning, $n = 130$), 💜(purple heart, $n = 126$), 💪(flexed biceps, $n = 116$), 🐕(dog, $n = 112$), ‼(double exclamation mark, $n = 106$), 🐈(cat, $n = 102$). Many of these are associated

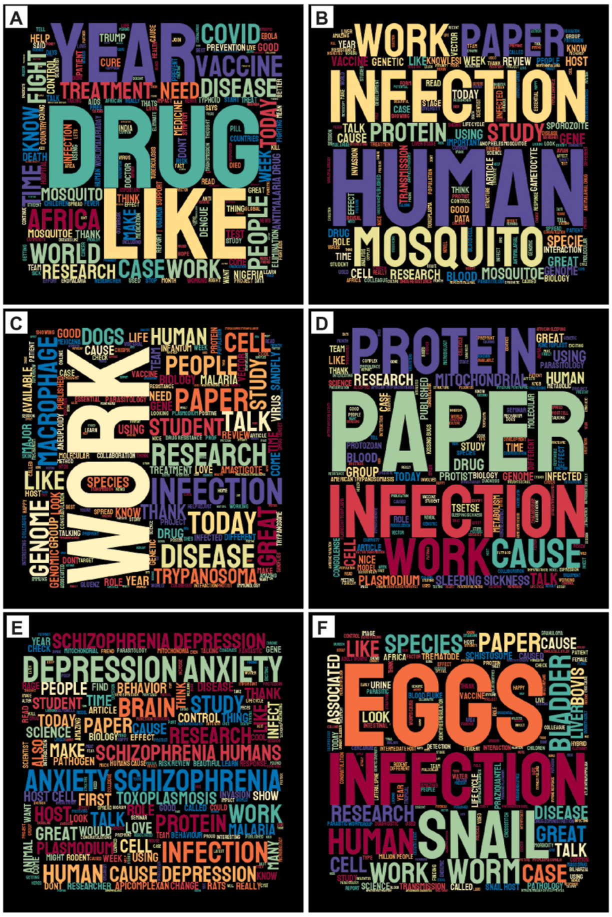

Fig. 3. Word clouds representing the most frequently used words in tweets resulting from searches on malaria (**A**), *Plasmodium* (**B**), *Leishmania* (**C**), *Trypanosoma* (**D**), *Toxoplasma* (**E**) and *Schistosoma* (**F**). Removal of stopwords were refined throughout the analyses leading to these word clouds.

with a positive sentiment, such as love, funny, pleasing, uncontrollable laughter, pride or overwhelming joy, adoration and positive intentions. Two skin tone modifiers, namely light and medium-light skin tones, represent the most common modifiers applied to human emoji characters, which is surprising because of current concerns associated with racism.

Sentiment in the Twitter data was analysed in detail using the VADER package. Histograms showing the distributions of the sentiments in the Twitter data are present in Fig. 4 and summary data are shown in Table 5. All the datasets show very similar trends. The use of emojis/emoticons in tweets is common and contributes to the sentiment in a number of ways. For example, in the *Schistosoma* dataset, 172 different emojis were used, including winking face with tears of joy which was the most commonly used (frequency of 61). Other faces showing emotion were used much less frequently. A wide range of animal species are represented. Overall, the VADER sentiment analyses shows that the majority of the tweets are neutral in emotion, with smaller numbers showing both positive and negative sentiment. Positive tweets are characterised by the inclusion of relevant emojis (e.g. hearts, smiley faces, party poppers) and often associated with a celebratory attitude and statements. It is noticeable that research students completing PhDs are often the subject of the most positive tweets (see examples in Supplementary file S3), as well as sharing papers amongst colleagues ("I'm excited to share our paper on the regulation of chronic *Toxoplasma* infection!" and "Must-read for all techie pathogen lovers!") or praising talks of others ("… gives marvelous talk on *Toxoplasma* invasion"). Negative tweets appear characterised by their choice of words such as war, death, killed, violence and dangerous. An example from the *Toxoplama* dataset says "This is my new strategy when I go to non-parasitology conferences: I scare people that they might have #toxoplasmosis so they listen my #toxoplasma research". Although intended to be humorous, the choice of words (e.g. scare) gives the tweet a negative sentiment. A peak of negative sentiment surrounding *Toxoplasma* at a compound score of −0.85 promotes a dislike of cats "About 80% of all cats are infected with *Toxoplasma gondii*, a parasite that can cause depression, anxiety, and schizophrenia in humans, so I don't love cats at all".

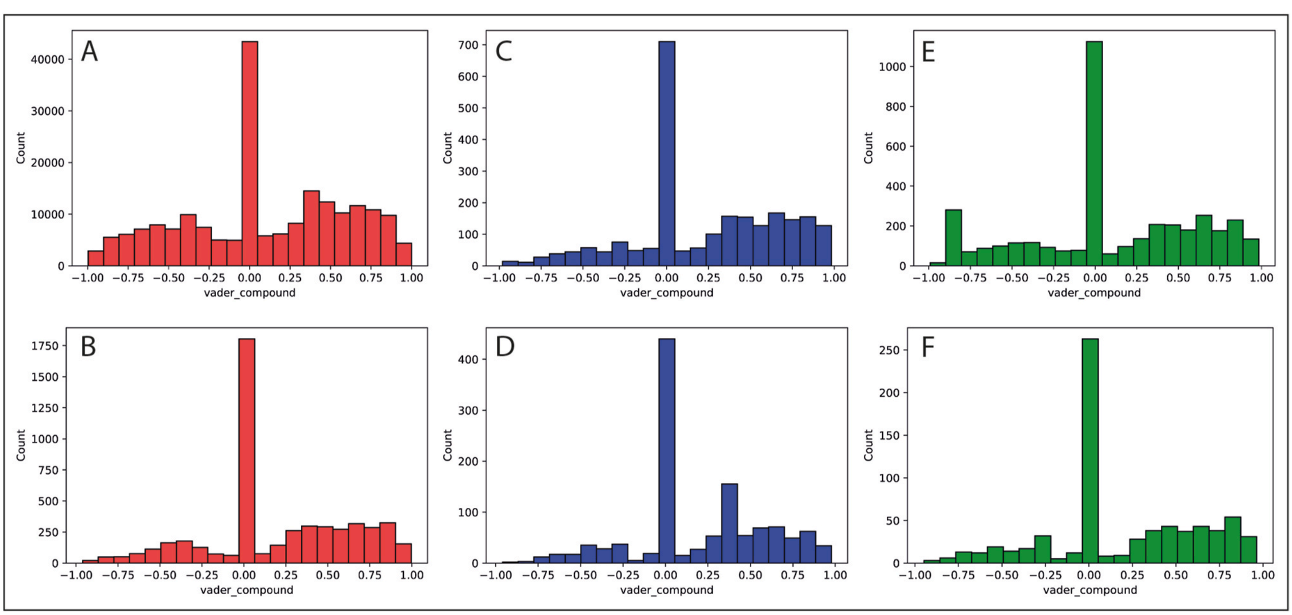

Fig. 4. Sentiment analyses of Twitter data by VADER. The compound score is normalized to be between −1 (most extreme negative sentiment) and +1 (most extreme positive sentiment). **A** Malaria. **B** *Plasmodium*. **C** *Leishmania*. **D** *Trypanosoma*. **E** *Toxoplasma*. **F** *Schistosoma*.

Table 5
Summary of VADER sentiment compound scores derived from tweets about malaria, *Plasmodium*, *Leishmania*, *Trypanosoma*, *Toxoplasma* or *Schistosoma*.

	Malaria	*Plasmodium*	*Leishmania*	*Trypanosoma*	*Toxoplasma*	*Schistosoma*
Count[a]	201,232	5137	2359	1204	3830	725
Mean	0.0894	0.193	0.232	0.201	0.102	0.188
SD	0.497	0.427	0.435	0.388	0.501	0.431
Min	−0.996	−0.961	−0.980	−0.959	−0.992	−0.947
25%	−0.296	0	0	0	−0.128	0
50%	0	0	0.178	0	0	0
75%	0.494	0.557	0.611	0.493	0.526	0.556
Max	0.999	0.996	0.983	0.977	0.983	0.964

Note: VADER sentiment compound scores are rounded to three significant figures. The range of the VADER sentiment scores possible is −1 to +1.
Abbreviations: SD, standard deviation; Min, minimum; Max, maximum.
[a] Number of tweets analysed.

4. Discussion

Previously the rise in the use of social media associated with parasitology was documented, including the use of Twitter (Ellis et al., 2021). The focus of the study presented here was the use of Twitter for discussion on malaria/*Plasmodium*, *Leishmania*, *Trypanosoma*, *Toxoplasma* and *Schistosoma*, which were several of the dominant areas identified in our previous study. It is reassuring from a discipline perspective that the Elsevier platform SciVal, which is often used to benchmark research performance (https://www.scival.com/home) contains a Parasitology research section that identifies and reinforces the six areas studied here (malaria, *Plasmodium*, *Leishmania*, *Trypanosoma*, *Toxoplasma* and *Schistosoma*) as predominant areas of research in parasitology (Fig. 5). This observation provides additional reassurance for our selection of the six key areas of study in parasitology.

There are several important conclusions that can be reached from the analyses provided here. First, the use of Twitter is global in its outreach; this of course implies that although English is a predominant language, tweets are being distributed in a wide range of languages. Portuguese and Spanish are notable in their use and our analyses highlight the importance of Spanish for tweeting about *Leishmania*. Language barriers are known to have serious consequences in all walks of life and a recent study on non-native English speakers in science showed (and we quote from Amano et al., 2023) “that non-native English speakers, especially early in their careers, spend more effort than native English speakers in conducting scientific activities, from reading and writing papers and preparing presentations in English, to disseminating research in multiple languages. Language barriers can also cause them not to attend, or give oral presentations at, international conferences conducted in English”. Our results might hence be biased by the disadvantages encountered by non-native English speakers, and this might disproportionately affect some speciality fields of parasitology, such as *Leishmania*.

We documented here some of the main influencers on Twitter associated with parasitology Twitter activity, and it is interesting to note that academics, individuals, institutions and non-government organisations are the major tweeters in the areas studied. For the parasitology discipline, this is reassuring as it is a known fact that most tweets are from individuals not associated with scholarly activity (Haustein et al., 2014). Scientific journals publishing in the parasitology arena regularly tweet and blog about publications and professional societies promote activity and events that are important to them. In this context, publishers are using social media for increasing awareness of their journals amongst the community, which may have a positive influence on the publishers brand (Dwivedi et al., 2021). These tweeters all represent examples of trusted sources of information, suggesting that the quality of the information circulating around Twitter on parasitology is likely to be accurate. Social media activity, including Twitter, is known to contain influencers who have large numbers of followers, regularly tweet on topics and so regularly contribute to the content but have little or no training in science. The existence of such tweeters on parasitology will raise cause for concern to this discipline, as one may then start to question the quality of information being disseminated.

The Kardashian (K) Index was proposed as a way of identifying individuals who are famous on social media but contribute little to the published literature and citations. This alternative H (Hall) Index was perceived by the author as a “tongue-in-cheek” approach (Hall, 2014), that subsequently was used to highlight the significance and importance of science communication and outreach (You, 2014). Whilst acknowledging the limited scientific value of the K-index it was used to highlight social media activity in neurology (Vilanilam et al., 2020) and

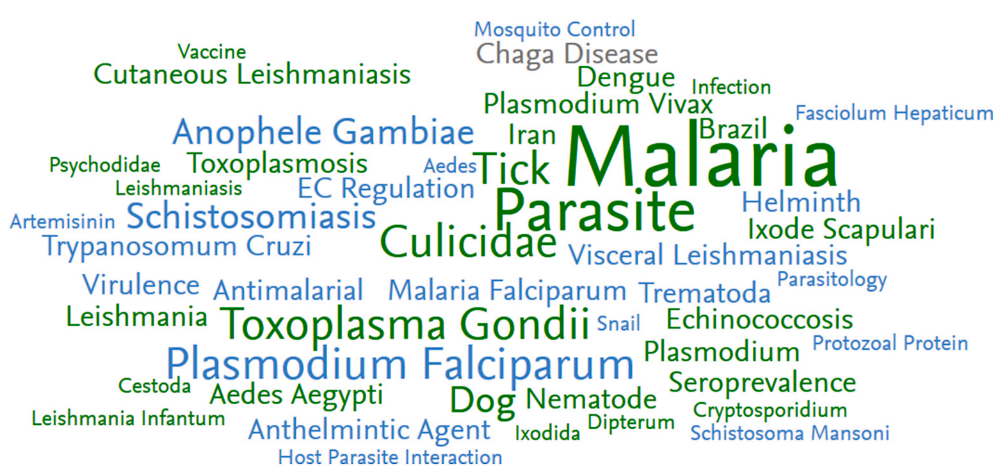

Fig. 5. Top 50 keyphrases derived from 88,451 publications (2011–2020) in the parasitology research area of SciVal (accessed 24/9/2021). SciVal uses the Elsevier Fingerprint Engine to extract keyphrases within the research area based on a modified inverse document frequency calculation. The word cloud presented is from this automated engine and shows the most common words determined by the engine.

cardiology (Khan et al., 2020) and the presence of a limited number of Kardashians within the associated 'Twitteratis'. For obvious reasons, we refrain here from identifying the Kardashians of the parasitology community; however, we emphasise that we consider the communication of parasitology information as a positive outcome at any level (if it is accurate). A K-index calculator can be found here (https://theinformationalturn.net/kardashian-index/) for those wishing to review their own status.

We also used several approaches to determine topics of discussion on Twitter. Analyses of hashtags show that dissemination of news about events in parasitology such as conferences is occurring, as well as details on published articles and publishing strategy including open access. Analyses of keywords identified topics of importance present in tweets; examples include *Toxoplasma* and mental health (Flegr and Horáček, 2020), *Trypanosoma cruzi* treatment by benznidazole and nifurtimox (Villar et al., 2019), aneuploidy in *Leishmania* (Dumetz et al., 2017) and the role of eggs in *Schistosoma* pathology (Costain et al., 2018). From a technological viewpoint, improvements in genome sequencing (Díaz-Viraqué et al., 2019) and the progress in the *Schistosoma* single cell atlas project (Wendt et al., 2020) are worthy of noting. The analyses of keyphrases extends these topics significantly to malaria drug resistance (Blasco et al., 2017) and drug development (Ashley and Phyo, 2018) as well as transmission (Gonçalves et al., 2017). The topics on Twitter however are not limited to these few and are extensive in nature. Supplementary files S1 and S2 are provided for reference; it is reassuring that the approaches adopted for identifying keywords and phrases, KeyBert and biterm topic modelling all provide similar results.

The study of the semantic content of text is called sentiment analysis, where the aim is to investigate how people feel about a topic, which is categorised in the form of a neutral, positive or negative viewpoint. The difficulties and limitations of using tweets in such analyses relates to their short length (280 characters or Unicode glyphs such as emojis or emoticons, although 10,000 characters or Unicode glyphs are available to Twitter/X users with a paid subscription to 'Twitter Blue'), presence of misspelt words and abbreviations, as well as punctuation. A variety of approaches have been developed and used for sentiment analyses including those for analyses of Twitter data (Bhuta et al., 2014; Zimbra et al., 2018; Ruz et al., 2020), which also include the influence of emojis and punctuation (Kralj Novak et al., 2015; Shiha and Ayvaz, 2017). Sentiment analyses indicate that the vast majority of tweets are neutral in their sentiment, although we note the need for the careful choice of words in tweets that can alter sentiment in a non-intended way (Ott, 2017).

Studies have shown that scholars in higher education use Twitter in several ways. Central to this process is the establishment of a network of contacts, often with shared interests. Sharing information, resources, and media relating to professional practice clearly represents the main feature of this network. This also includes asking for assistance and providing suggestions to others. Engagement with this network may also include social commentary on topics considered important to that network. Finally, from a teaching perspective, scholars may share information with their students (Veletsianos, 2012). The ongoing development of a digital identity needs to be successfully managed within the context of the professional role that academics hold (Veletsianos, 2012; Ruan et al., 2020).

The type of social media activity used in medicine has been analysed, and separated into 'push', 'engagement' and 'blended (both push and engagement)' style activities. Push strategies are commonplace in parasitology and include dedicated journal Twitter accounts and informational blogs. True engagement leading to impact is emphasised by hashtags, although engagement can go so much further than that and is likely to be individual or organisation focussed. An excellent example are the weekly case studies in parasitology provided by @ParasiteGal (Professor Bobbi Pritt). A recent strategic review of how Twitter is used by medical journals, summarises the four main methods used by publishers for knowledge dissemination; the basic tweet, infographics, podcasts, and hosting monthly internet-based journal clubs. All these strategies were perceived to bring benefits in promoting journal articles. Of note was the observation that only about one third of medical journals have a Twitter profile and it was argued that the cause for "this underutilization is the lack of evidence-based best practices" (Erskine and Hendricks, 2021). More recent data indicate only 22.2% of scientific journals have Twitter accounts. However, new accounts are being created every 1.5 days, and so the Twitter activity from journals will continue to grow (Nishikawa-Pacher, 2023).

Despite their extensive use, best practices for use of social media including Twitter have not been thoroughly examined. Others have pointed out that no knowledge translation model exists for social media, other than the four C's typically used in online marketing: content, context, connections leading to conversations (Elliott et al., 2020). Nevertheless, case studies are starting to identify recommendations for best practice that scientists can use and incorporate into their scholarly activities, including those engaged in parasitology. A recent example are the recommendations for live tweeting at scientific conferences (Power, 2022).

The developments behind the idea of e-Professionalism represent important steps forward in professional practice and is indicative of the thought processes going into best practices in the use of social media (Cain and Romanelli, 2009). Best practices for use of Twitter in science were suggested to include using an image, tagging people and journals, and using hashtags. Suggestions also extended to a wider range of features associated with social media use, including content accuracy and accountability (Lu et al., 2021). Many of these features are found in parasitology-related tweets. Further, we note that others have put forward a series of recommendations on key practices for knowledge translation and scholarly debate for health researchers using social media (Elliott et al., 2020). These include defining the intended audience (e.g. the Twitter network involved), timing, frequency, and duration of messaging, and a need to consider the requirement for necessary ethical approvals.

There are several limitations of this study to consider. For example, there are a range of methods available for preprocessing of text data and the approach used may influence study outcomes (Baziotis et al., 2017). In addition, the use and identification of stopwords and custom_stopwords is essential and there are several alternative approaches to determine them (Gerlach et al., 2019; Sarica and Luo, 2021). Similarly, there are additional algorithms for keyword and keyphrase extraction and sentiment analyses not used in this study (Siddiqi and Sharan, 2015; Thelwall, 2017; Papagiannopoulou and Tsoumakas, 2020; Sun et al., 2020). The strategies used here were based on simplicity and the nature of the short text available for analyses in tweets, although we note more complex approaches are available for their analyses (Edo-Osagie et al., 2020). Finally, Twitter provides enormous opportunity for the study of parasitology text, and so we point out that our analyses are literally the "tip of the iceberg" behind future studies.

Finally, we focus on the limitations of Twitter. One of the most important is associated with the sheer volume of information being communicated and the number of users involved. Declarations on conflicts of interest are normally missing or ignored but should be declared, as is the case of any business activity. e-Professionalism dictates the need for media literacy training by scientists that supports opportunities for professional development. Separation of personal from professional life on social media is also recommended, and that should extend to separate Twitter accounts and activities.

5. Conclusions

The present study documents, for the first time, the use of Twitter (X as it is now known) for discussion of dominant areas of parasitology research, such as malaria/*Plasmodium*, *Leishmania*, *Trypanosoma*, *Toxoplasma* and *Schistosoma*. The study demonstrates the global reach of social media use on the topics, especially using the English (but also other) language and the utilisation of the platform by academia, governments and scientific journals for the promulgation of their science (news, events, conferences and publications). It provides a snapshot of the *status quo* (to 2020) and trends of social media usage, using the example of Twitter,in this subject area of science and should be repeated at appropriate intervals (say, every five or ten years) to monitor for changes in the usage and use patterns.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Ethical approval

This study was approved by the UTS Human Research Ethics Committee under reference number ETH22-7184.

CRediT authorship contribution statement

John T. Ellis: Conceptualization, Methodology, Data curation, and, Formal analysis, Project administration, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Michael P. Reichel:** Conceptualization, Investigation, Writing – original draft, Writing – review & editing, Visualization.

Declaration of competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. Neither of the authors are Kardashians of the parasitology community based on their K-index.

Data availability

The data supporting the conclusions of this article are included within the article and its supplementary files.

Acknowledgements

This study would not be possible without Twitter. We thank Bethany Ellis for her thoughts and assistance with this study and two conscientious referees for their valuable feedback.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.crpvbd.2023.100138.

References

Alshammari, N.F., AlMansour, A.A., 2019. State-of-the-art review on Twitter sentiment analysis. In: 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS), pp. 1–8. https://ieeexplore.ieee.org/document/8769465.

Amano, T., Ramírez-Castañeda, V., Berdejo-Espinola, V., Borokini, I., Chowdhury, S., Golivets, M., et al., 2023. The manifold costs of being a non-native English speaker in science. PLoS Biol. 21, e3002184.

Ashley, E.A., Phyo, A.P., 2018. Drugs in development for malaria. Drugs 78, 861–879.

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. https://aclanthology.org/S17-2126/.

Bermingham, A., Smeaton, A., 2011. On using Twitter to monitor political sentiment and predict election results. In: Proceedings of the Workshop on Sentiment Analysis where AI Meets Psychology, pp. 2–10. SAAIP 2011). https://doras.dcu.ie/16670/1/saaip2011.pdf.

Bhuta, S., Doshi, A., Doshi, U., Narvekar, M., 2014. A review of techniques for sentiment analysis of Twitter data. In: 2014 International Conference on Issues and Challenges in Intelligent Computing Techniques. ICICT), pp. 583–591. https://ieeexplore.ieee.org/abstract/document/6781346.

Bik, H.M., Goldstein, M.C., 2013. An introduction to social media for scientists. PLoS Biol. 11, e1001535.

Bird, S., Loper, E., Klein, E., 2009. Natural language processing with Python. O'Reilly Media Inc., Sebastopol, USA. https://www.nltk.org/book/.

Blasco, B., Leroy, D., Fidock, D.A., 2017. Antimalarial drug resistance: Linking *Plasmodium falciparum* parasite biology to the clinic. Nat. Med. 23, 917–928.

Boudin, F., 2016. Pke: An open source Python-based keyphrase extraction toolkit. In: Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: System Demonstrations, pp. 69–73. https://aclanthology.org/C16-2015.

Bougouin, A., Boudin, F., Daille, B., 2013. Topicrank: Graph-based topic ranking for keyphrase extraction. In: International Joint Conference on Natural Language Processing (IJCNLP), pp. 543–551. https://hal.science/hal-00917969/.

Bovet, A., Makse, H.A., 2019. Influence of fake news in Twitter during the 2016 US presidential election. Nat. Commun. 10, 7. https://doi.org/10.1038/s41467-018-07761-2.

Broniatowski, D.A., Paul, M.J., Dredze, M., 2013. National and local influenza surveillance through Twitter: An analysis of the 2012–2013 influenza epidemic. PLoS One 8, e83672.

Cain, J., Romanelli, F., 2009. E-professionalism: A new paradigm for a digital age. Curr. Pharm. Teach. Learn. 1, 66–70.

Calvani, N.E.D., Verissimo, C.D.M., Cantacessi, C., Clark, E., Kanduma, E., 2023. Herminthology: Promoting gender equity in science and parasitology. Trends Parasitol. 39, 73–79.

Campos, R., Mangaravite, V., Pasquali, A., Jorge, A., Nunes, C., Jatowt, A., 2020. YAKE! Keyword extraction from single documents using multiple local features. Inf. Sci. 509, 257–289.

Carrion, M., Madoff, L.C., 2017. ProMED-mail: 22 years of digital surveillance of emerging infectious diseases. Int. Health 9, 177–183.

Cevik, M., 2019. Social media to engage, communicate and interact. Clin. Microbiol. Inf. 25, 1165–1166.

Cevik, M., Ong, D.S.Y., Mackenzie, G., 2019. How scientists and physicians use Twitter during a medical congress. Clin. Microbiol. Inf. 25, 1561.e1567-1561.e1512.

Chen, D.Y., 2018. Pandas for everyone. Addison-Wesley Data & Analytics Series, Pearson Education, Inc., Boston, USA. https://www.pearson.com/store/p/pandas-for-everyone-python-data-analysis/P200000009429/9780134546933.

Collins, K., Shiffman, D., Rock, J., 2016. How are scientists using social media in the workplace? PLoS One 11, e0162680.

Costain, A.H., MacDonald, A.S., Smits, H.H., 2018. Schistosome egg migration: Mechanisms, pathogenesis and host immune responses. Front. Immunol. 9, 3042.

Côté, I.M., Darling, E.S., 2018. Scientists on Twitter: Preaching to the choir or singing from the rooftops? Facets 3, 682–694.

Depoux, A., Martin, S., Karafillakis, E., Preet, R., Wilder-Smith, A., Larson, H., 2020. The pandemic of social media panic travels faster than the COVID-19 outbreak. J. Trav. Med. 27.

Díaz-Viraqué, F., Pita, S., Greif, G., De Souza, R.D.C.M., Iraola, G., Robello, C., 2019. Nanopore sequencing significantly improves genome assembly of the protozoan parasite *Trypanosoma cruzi*. Genome Biol. Evol. 11, 1952–1957.

Dol, J., Tutelman, P.R., Chambers, C.T., Barwick, M., Drake, E.K., Parker, J.A., et al., 2019. Health researchers' use of social media: Scoping review. J. Med. Internet Res. 21, e13687.

Dumetz, F., Imamura, H., Sanders, M., Seblova, V., Myskova, J., Pescher, P., et al., 2017. Modulation of aneuploidy in *Leishmania donovani* during adaptation to different *in vitro* and *in vivo* environments and its impact on gene expression. mBio 8, e00599-00517.

Dwivedi, Y.K., Ismagilova, E., Hughes, D.L., Carlson, J., Filieri, R., Jacobson, J., et al., 2021. Setting the future of digital and social media marketing research: Perspectives and research propositions. Int. J. Inf. Manag. 59, 102168.

Edo-Osagie, O., De La Iglesia, B., Lake, I., Edeghere, O., 2020. A scoping review of the use of Twitter for public health research. Comput. Biol. Med. 122, 103770.

Elbagir, S., Yang, J., 2020. Sentiment analysis on Twitter with Python's natural language toolkit and VADER sentiment analyzer. In: IAENG Transactions on Engineering Sciences: Special Issue for the International Association of Engineers Conferences 2019, pp. 63–80. https://doi.org/10.1142/9789811215094_0005.

Elliott, S.A., Dyson, M.P., Wilkes, G.V., Zimmermann, G.L., Chambers, C.T., Wittmeier, K.D., et al., 2020. Considerations for health researchers using social media for knowledge translation: Multiple case study. J. Med. Internet Res. 22, e15121.

Ellis, J., Ellis, B., Tyler, K., Reichel, M.P., 2021. Recent trends in the use of social media in parasitology and the application of alternative metrics. Curr. Res. Parasitol. Vector-Borne Dis. 1, 100013.

Erskine, N., Hendricks, S., 2021. The use of Twitter by medical journals: Systematic review of the literature. J. Med. Internet Res. 23, e26378.

Flegr, J., Horáček, J., 2020. Negative effects of latent toxoplasmosis on mental health. Front. Psychiatr. 10, 1012.

Fownes, J.R., Yu, C., Margolin, D.B., 2018. Twitter and climate change. Soc. Compass 12, e12587.

Gerlach, M., Shi, H., Amaral, L.A.N., 2019. A universal information theoretic approach to the identification of stopwords. Nat. Mach. Intell. 1, 606–612.

Gomide, J., Veloso, A., Meira Jr., W., Almeida, V., Benevenuto, F., Ferraz, F., Teixeira, M., 2011. Dengue surveillance based on a computational model of spatio-temporal locality of Twitter. In: Proceedings of the 3rd International Web Science Conference, pp. 1–8. https://doi.org/10.1145/2527031.2527049.

Gonçalves, B.P., Kapulu, M.C., Sawa, P., Guelbéogo, W.M., Tiono, A.B., Grignard, L., et al., 2017. Examining the human infectious reservoir for *Plasmodium falciparum* malaria in areas of differing transmission intensity. Nat. Commun. 8, 1133.

González-Padilla, D.A., Tortolero-Blanco, L., 2020. Social media influence in the COVID-19 Pandemic. Int. Braz J. Urol. 46, 120–124.
Grootendorst, M., 2020. KeyBERT: Minimal keyword extraction with BERT. Zenodo. https://doi.org/10.5281/zenodo.4461265.
Hall, N., 2014. The Kardashian index: A measure of discrepant social media profile for scientists. Genome Biol. 15, 424.
Haustein, S., Peters, I., Sugimoto, C.R., Thelwall, M., Larivière, V., 2014. Tweeting biomedicine: An analysis of tweets and citations in the biomedical literature. J. Assoc. Inf. Sci. Technol. 65, 656–669.
Hull, K., Dodd, J.E., 2017. Faculty use of Twitter in higher education teaching. J. Appl. Res. High Educ. 9, 91–104.
Hutto, C.J., Gilbert, E.E., 2014. VADER: a parsimonious rule-based model for sentiment analysis of social media text. In: Eighth International Conference on Weblogs and Social Media (ICWSM-14). Association for the Advancement of Artificial Intelligence, Ann Arbor, MI, USA, pp. 216–225. www.aaai.org.
Jabbar, A., Gasser, R.B., Lodge, J., 2016. Can new digital technologies support parasitology teaching and learning? Trends Parasitol. 32, 522–530.
Jónsson, E., Stolee, J., 2015. An evaluation of topic modelling techniques for Twitter. University of Toronto, Canada. https://api.semanticscholar.org/CorpusID:53680644.
Jünge, J., Fähnrich, B., 2020. Does really no one care? Analyzing the public engagement of communication scientists on Twitter. New Media Soc. 22, 387–408.
Khan, M.S., Shahadat, A., Khan, S.U., Ahmed, S., Doukky, R., Michos, E.D., Kalra, A., 2020. The Kardashian index of cardiologists. JACC Case Rep 2, 330–332.
Kralj Novak, P., Smailović, J., Sluban, B., Mozetič, I., 2015. Sentiment of emojis. PLoS One 10, e0144296.
Krittanawong, C., Narasimhan, B., Virk, H.U.H., Narasimhan, H., Hahn, J., Wang, Z., Tang, W.H.W., 2020. Misinformation dissemination in Twitter in the COVID-19 era. Am. J. Med. 133, 1367–1369.
Lee, J.-S.M., 2019. How to use Twitter to further your research career. Nat. Career Column. https://www.nature.com/articles/d41586-019-00535-w.
López-Goñi, I., Giner-Lamia, J., Álvarez-Ordoñez, A., Benitez-Páez, A., Claessen, D., Cortesao, M., et al., 2019. #EUROmicroMOOC: Using Twitter to share trends in microbiology worldwide. FEMS Microbiol. Lett. 366, fnz141.
López-Goñi, I., Martínez-Viñas, M.J., Antón, J., Cid, V.J., González, A.M., Brown-Jaque, M., et al., 2016. Twitter as a tool for teaching and communicating microbiology: The #microMOOCSEM Initiative. J. Microbiol. Biol. Educ. 17, 492–494.
Lu, D., Ruan, B., Lee, M., Yilmaz, Y., Chan, T.M., 2021. Good practices in harnessing social media for scholarly discourse, knowledge translation, and education. Persp. Med. Educ. 10, 23–32.
McKinney, W., 2010. Data structures for statistical computing in Python. In: Proceedings of the 9th Python in Science Conference (SCIPY), Austin, TX, USA, pp. 51–56. In: https://conference.scipy.org/proceedings/scipy2010/pdfs/mckinney.pdf.
Nishikawa-Pacher, A., 2023. The Twitter accounts of scientific journals: A dataset. UKSG Insights 36, 1. https://insights.uksg.org/articles/10.1629/uksg.593.
Oren, E., Martinez, L., Hensley, R.E., Jain, P., Ahmed, T., Purnajo, I., et al., 2020. Twitter communication during an outbreak of hepatitis A in San Diego, 2016–2018. Am. J. Publ. Health 110, S348–S355.
Ott, B.L., 2017. The age of Twitter: Donald J. Trump and the politics of debasement. Crit. Stud. Media Commun. 34, 59–68.
Pano, T., Kashef, R., 2020. A complete VADER-based sentiment analysis of bitcoin (BTC) tweets during the era of COVID-19. Big Data Cogn. Comput. Times 4, 33. https://doi.org/10.3390/bdcc4040033.
Papagiannopoulou, E., Tsoumakas, G., 2020. A Review of Keyphrase Extraction, vol. 10. Wiley Interdisc. Rev. Data Mining Knowledge Discovery, e1339.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al., 2011. Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.
Pizzuti, A.G., Patel, K.H., McCreary, E.K., Heil, E., Bland, C.M., Chinaeke, E., et al., 2020. Healthcare practitioners' views of social media as an educational resource. PLoS One 15, e0228372.
Pollett, S., Althouse, B.M., Forshey, B., Rutherford, G.W., Jarman, R.G., 2017. Internet-based biosurveillance methods for vector-borne diseases: Are they novel public health tools or just novelties? PLoS Neglected Trop. Dis. 11, e0005871-e0005871.
Power, B.J., 2022. How to use Twitter at a scientific conference. mSphere 7, e00121-00122. https://doi.org/10.1128/msphere.00121-22.
Priem, J., Costello, K.L., 2010. How and why scholars cite on Twitter. Proc. Am. Soc. Information Sci. Technol. 47, 1–4.
Rosenberg, H., Syed, S., Rezaie, S., 2020. The Twitter pandemic: The critical role of Twitter in the dissemination of medical information and misinformation during the COVID-19 pandemic. Can. J. Emerg. Med. 22, 418–421.
Ruan, B., Yilmaz, Y., Lu, D., Lee, M., Chan, T.M., 2020. Defining the Digital Self: A qualitative study to explore the digital component of professional identity in the health professions. J. Med. Internet Res. 22, e21416.
Ruz, G.A., Henríquez, P.A., Mascareño, A., 2020. Sentiment analysis of Twitter data during critical events through Bayesian networks classifiers. Future Generat. Comput. Syst. 106, 92–104.
Sarica, S., Luo, J., 2021. Stopwords in technical language processing. PLoS One 16, e0254937.
Shiffman, D.S., 2012. Twitter as a tool for conservation education and outreach: What scientific conferences can do to promote live-tweeting. J. Environ. Stud. Sci. 2, 257–262.
Shiha, M., Ayvaz, S., 2017. The effects of emoji in sentiment analysis. Int. J. Comput. Electron Eng. (IJCEE) 9, 360–369.
Siddiqi, S., Sharan, A., 2015. Keyword and keyphrase extraction techniques: A literature review. Int. J. Comput. Appl. 109, 18–23.
Sievert, C., Shirley, K., 2014. LDAvis: A method for visualizing and interpreting topics. In: Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces, pp. 63–70. https://aclanthology.org/W14-3110.pdf.
Sinnenberg, L., Buttenheim, A.M., Padrez, K., Mancheno, C., Ungar, L., Merchant, R.M., 2017. Twitter as a tool for health research: A systematic review. Am. J. Publ. Health 107, e1–e8.
Sun, C., Hu, L., Li, S., Li, T., Li, H., Chi, L., 2020. A review of unsupervised keyphrase extraction methods using within-collection resources. Symmetry 12, 1864.
Thelwall, M., 2017. The heart and soul of the web? Sentiment strength detection in the social web with SentiStrength. In: Holyst, J. (Ed.), Cyberemotions. Understanding Complex Systems. Springer, Cham. https://doi.org/10.1007/978-3-319-43639-5_7.
Vashishtha, S., Susan, S., 2019. Fuzzy rule based unsupervised sentiment analysis from social media posts. Expert Syst. Appl. 138, 112834.
Veletsianos, G., 2012. Higher education scholars' participation and practices on Twitter. J. Comput. Assist. Learn. 28, 336–349.
Vilanilam, G.K., Wadhwa, V., Purushothaman, R., Rohilla, M., Radvany, M.G., 2020. The Kardashian index of interventional neuroradiologists: Measuring discrepant social media influence. NeuroRadiol. J. 33, 525–527.
Villar, J.C., Herrera, V.M., Carreño, J.G.P., Herrera, E.V., Domínguez, Y.Z.C., Vásquez, S. M., et al., 2019. Nifurtimox *versus* benznidazole or placebo for asymptomatic *Trypanosoma cruzi* infection (Equivalence of Usual Interventions for Trypanosomiasis-EQUITY): Study protocol for a randomised controlled trial. Trials 20, 431.
Waskom, M.L., 2021. Seaborn: Statistical data visualization. J. Open Source Softw. 6, 3021.
Wendt, G., Zhao, L., Chen, R., Liu, C., O'Donoghue, A.J., Caffrey, C.R., et al., 2020. A single-cell RNA-seq atlas of *Schistosoma mansoni* identifies a key regulator of blood feeding. Science 369, 1644–1649.
Wisdom, V., Gupta, R., 2016. An introduction to Twitter data analysis in Python. Artigence Inc. https://www.researchgate.net/profile/Vivek-Wisdom/publication/308371781_An_introduction_to_Twitter_Data_Analysis_in_Python/links/57e24cf708ae1f0b4d95b409/An-introduction-to-Twitter-Data-Analysis-in-Python.pdf.
Wolf, J.M., 2017. The multipurpose tool of social media: Applications for scientists, science communicators, and educators. Clin. Microbiol. Newsl. 39, 75–79.
Yan, X., Guo, J., Lan, Y., Cheng, X., 2013. A biterm topic model for short texts. In: Proceedings of the 22nd International Conference on World Wide Web. Association for Computing Machinery, Rio de Janeiro, Brazil, pp. 1445–1456. https://doi.org/10.1145/2488388.2488514.
You, J., 2014. Who are the science stars of Twitter? Science 345, 1440-1144.
Zimbra, D., Abbasi, A., Zeng, D., Chen, H., 2018. The state-of-the-art in Twitter sentiment analysis: A review and benchmark evaluation. ACM Trans. Manag. Inf. Syst. (TMIS) 9, 1–29.