Fake Cures: User-centric Modeling of Health Misinformation in Social Media

AMIRA GHENAI, University of Waterloo, Canada

YELENA MEJOVA, ISI Foundation, Italy

Social media's unfettered access has made it an important venue for health discussion and a resource for patients and their loved ones. However, the quality of the information available, as well as the motivations of its posters, has been questioned. This work examines the individuals on social media that are posting questionable health-related information, and in particular promoting cancer treatments which have been shown to be ineffective (making it a kind of misinformation, willful or not). Using a multi-stage user selection process, we study 4,212 Twitter users who have posted about one of 139 such "treatments", and compare them to a baseline of users generally interested in cancer. Considering features capturing user attributes, writing style, and sentiment, we build a classifier which is able to identify users prone to propagate such misinformation at an accuracy of over 90%, providing a potential tool for public health officials to identify such individuals for preventive intervention.

1 INTRODUCTION

Searching and sharing health information online is becoming a common practice. A 2011 survey indicated that as many as 62% of adult Internet users in U.S. used social network sites for health related topics, from following friends' health experiences or updates to gathering health related information [30]. Some manage their health via general platforms such as PatientsLikeMe, while others join condition-specific communities like TuDiabetes, yet others share their experiences in general-purpose social media [22, 50]. For instance, Paul et al. [71] showed that a significant number of personal and health-related questions were being asked on the microblogging platform Twitter, which is becoming the top destination for both patients and healthcare professionals [4]. The social component of such interactions is especially important in educating the population on health matters, as individual opinions can be strongly biased by their social network [53].

While social media use for health management is growing, so do the concerns over the lack of accountability, dubious quality and loose confidentiality [40, 66]. With few legal constraints

Authors' addresses: Amira Ghenai, University of Waterloo, 200 University Ave W, Waterloo, ON, N2L 3G1, Canada, aghenai@uwaterloo.ca; Yelena Mejova, ISI Foundation, Via Chisola, 5, Torino, TO, 10126, Italy, yelenamejova@acm.org.

imposed on often profit-seeking websites, social media provides a dynamic forum for propagating possible medical misinformation [31]. Recent rise in vaccine hesitancy has been linked to an active movement on Twitter, promoting conspiratorial thinking and mistrust in the government [65]. Image sharing platforms such as Flickr and Instagram have become battlegrounds between the pro-anorexia movement and physicians attempting to intervene [13, 92]. Uncertainty surrounding infectious disease outbreaks, such as the Zika epidemic of 2016, yielded rumors and speculations about its causes, preventive measures, and consequences [23, 34].

In this study we turn to the individuals sharing questionable medical information on Twitter, in particular cancer treatments which have been medically proven to be ineffective. Having around 336 million monthly active users in the first quarter of 2018[1], Twitter is one of the largest social media websites expressly dedicated to the sharing of information, including that on cancer. Compiling hundreds of thousands of tweets on 139 queries spanning acupuncture, cinnamon, reflexology, and vitamin C, we apply strict selective criteria employing human/organization classification [61], name dictionaries, usage thresholds, and crowdsourced relevance refinement resulting in 4,212 users, which we then compare to those mentioning cancer in general from a previous study [70]. Employing previous research on rumor detection, we characterize these users in multi-faceted feature spaces, encompassing user attributes, linguistic style, sentiment, and post timing. We find users who have a more sophisticated language, who are interested in cancer, but who are not personally involved with the illness. We build a logistic regression model which, out of Twitter users mentioning cancer, is able to identify those who will eventually post a piece of misinformation with a high level of accuracy.

Misinformation on social media is an urgent issue, and even more so in the health field. This paper is one of the first to look into the characteristics of users propagating unverified "cures" of cancer on Twitter as a case study of tracking health misinformation outside crisis communication management domain. The identification of potential sources of such misinformation would allow public health officials to monitor social media discourse, characterize the deficiencies in current communication strategy, and detect new misinformation before it causes serious harm.

2   RELATED WORK

The present work marries two burgeoning directions of social media research: the tracking of misinformation and the measurement of health-related attitudes and activities. Below we describe the latest developments in both topics, as well as the recent attempts at tracking health misinformation in particular.

Misinformation tracking. In 2017, the term "fake news" has been named Collins' Word of the Year, referring to "false, often sensational, information"[2]. In the context of news, rumors and misinformation have been associated with political sphere, with the latest works proposing data mining solutions [81] which encompass opinionated language detection [17], visual feature extraction [42] and user group characteristics [59]. In fact, users play an important role, as misinformation sometimes originates from automated accounts working in synchrony – bot nets – with [80] claiming millions of political tweets spread thusly during and following the 2016 U.S. Presidential election. Tools to detect such bots include Hoaxy [79] to track the spread of claims and Botometer [20] to detect social bots. Beyond the automated accounts, case studies of incidents such as the Ukrainian conflict [49] and the mass shootings in US [82] examine human reactions to questionable information online. The provenance and motivation behind such information has increasingly become a contentious issue, as speculations rose that important political decisions, including the

---

[1]https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/
[2]https://ind.pn/2AnI2Bw

United Kingdom vote to leave the European Union and the election of Donald Trump to the U.S. Presidency, have been potentially swayed by forces outside those nations [45], posing a danger to democracy itself.

Health-related attitudes on social media. Beyond politics, social media also provides ample resources for health-related decision making, capturing behaviors and attitudes impacting individual health. Automated methods have been devised for tracking marijuana use on Twitter [36], and to capture attitudes toward legal drugs including Xanax and Adderall [78]. Forum threads have been analyzed by Wu et al. [90] to discover adverse drug effects and drug interactions, using association mining. Likewise, behaviors related to lifestyle diseases such as diabetes type 2 and obesity have been tracked using Twitter (a microblogging platform) [1], Instagram (a photo sharing platform) [64], and Facebook (a social network) [5], along with attitudes toward food and diet [62]. A study of a community promoting anorexia on Flickr (another photo sharing platform) [92] showed that attempts of the anti-anorexia programs to infiltrate the community with intervention messages tagged with pro-anorexia tags was counterproductive in the long run (with users exposed to such remaining in the group longer). Further uses of social media to gauge the efficacy of health communication includes a recent study of breast cancer mammography advisory on Twitter [67] which found many users to be confused by it than to approve of it. Finally, public awareness of health-related topics has been recently gauged through the advertising platforms provided by these social media – for instance using Facebook Advertising Manager to estimate the number of Facebook users interested in diabetes-related topics [6]. Thus, as social media adoptions increases, so does the health-related discussion and information seeking on these platforms, allowing for large-scale analysis and tracking.

Health misinformation online. For over a decade, medical sociologists have studied Internet as a new component of health ecosystems, with especially cancer patients utilizing it to collect information and make treatment decisions [16]. The patient reliance on the Web has resulted in a patient–Web–physician "triangulation" [86] with benefits such as more efficient use of clinical time and additional support from online support groups, coupled with potential harms like the dangers posed by the variable quality of information, unnecessary visits to a physician, and exacerbating existing socioeconomic health disparities. Early on, the use of Internet by patients has been shown to be problematic [77]. For instance, [25] find that the quality of Web search results varies for differently worded medical queries. Democratization of content publishing may also be exacerbating quality concerns, as YouTube videos have been found to contain instances of public display of harmful or unhealthy behaviors, promotion of tobacco to consumers, and distorting policy and research funding agendas [54]. Moreover, concentrated efforts promoting doubt of medical establishment, such as the "anti-vaxxer" movement on Twitter, play into the larger skepticism of government and conspiratorial thinking [65].

Recent attempts to track health misinformation include "VAC Medi+ board", an online interactive visualization framework integrating heterogeneous real-time data streams with Twitter data [51], which tracks the spread of vaccine related information on Twitter and the sources of information spread. Such social media content promoting vaccine hesitancy has been shown to impact future opinions of its users, and potentially their subsequent behaviors [24]. It is also possible to study the impact of the interventions designed to change such behaviors, such as in the previously mentioned study by [92] on the potentially unsuccessful attempts to infuse anti-anorexia content into pro-anorexia communities on Flickr. Dynamic nature of social media also allows for fast spread of misinformation during an ongoing epidemic, and machine learning techniques have been deployed to track such content, for instance during the 2016 Zika outbreak [23, 34]. To help social media users, tools are being developed to ease the verification of health claims via natural language processing and retrieval of necessary medical literature [75].

However, little attention has been paid to the modeling and understanding of the spread of cancer "complementary and alternative medicine" (CAM) on the Internet. A 2008 survey of 80 cancer patients found that, when going online, respondents dealt with emotional stress of being reminded about their prognosis, and were seeking second opinion of a doctor before using CAM promoted online [11]. However, more recent surveys find internet to be increasingly important source of information on CAM, with around half of patients using the alternative medicines, as well as their relatives, getting their health advise online [26, 47]. Thus, in this work we focus on the kinds of individuals who are susceptible to propagating unverified information about cancer treatments which have been found to be ineffective at treating cancer. Guided by previous literature on rumors (see [74] for an overview), we examine writing style, personal involvement, and other interests, as described below, in order to model individuals particularly vulnerable to such misinformation.

3 DATA COLLECTION

Dataset used in this work consists of tweets belonging to two groups of users: (1) a "rumor" group who have posted content promoting one of 139 cancer "treatments" which have been proven ineffective, and (2) a "control" group who posted generally about cancer, but not on any of the above topics. The initial data gathering, and multiple steps of user selection and relevance refinement are described below.

3.1 Health Rumor and Control Data Collection

As the focus of this study is the behavior of users who post on social media health content of questionable nature, we begin by compiling a set of purported cancer "cures" which have been shown by experimentation and medical professionals to be ineffective. Four of such dubious cancer treatments come from White & Hassan [88] where authors judged and reached a consensus about the medical treatments' efficacy by reading the corresponding Cochrane Review [18, 27]. Next, we collect nine rumor topics from David Colquhoun (Professor of Pharmacology at University College London) blog[3]. Professor David's blog focuses particularly on alternative medicine such as homoeopathy, traditional Chinese medicine and herbal medicine. Finally, we collect 126 unproven cancer treatments listed in the "List of unproven and dis-proven cancer treatments"[4] Wikipedia page which was refereed by Cancer Research UK[5]. The selection of these unproven cancer treatments is then supervised by a trained oncologist (acknowledged below) in order to validate the ground truth of the treatments' efficacy, making sure all collected "treatments" are indeed ineffective. This process results in a total of 139 cancer treatment-related topics[6]. Henceforth we will call these topics cancer treatment rumors, or simply rumor topics. Note that some of the above treatments may be effective in alleviating some of the symptoms of cancer, but do not actually affect the underlying progression of cancer (see Discussion for more).

Considering Twitter users posting about the above topics as the "rumor" group, we turn to existing research on health discussions for the "control" group. These would be people talking in general about cancer such as cancer causes, prevention, symptoms, and awareness or sharing personal experiences with the medical condition. For this purpose, we use Paul & Dredze [70] public health topics dataset which consists of 144 million tweets that are related to a selection of health topics gathered during the period of 01 August 2011 - 28 February 2013. As the focus of this

---

[3]http://www.dcscience.net/
[4]https://en.wikipedia.org/wiki/List_of_unproven_and_disproven_cancer_treatments#Ineffective_treatments
[5]http://scienceblog.cancerresearchuk.org/2014/03/24/dont-believe-the-hype-10-persistent-cancer-myths-debunked/#superfoods
[6]The topics, along with the keyword queries are available at https://tinyurl.com/y78mkg6s

study is cancer, we focus on the 676,236 users who have posted 969,259 tweets in this dataset (for a summary of user selection process, see Figure 1).

Next, we turn back to the Rumor group and collect tweets on rumor topics that span the same time period as the control. For every rumor topic, we hand craft a query and expand it using general domain tools such as Google search and Google keyword planner[7] as well as medical domain tools including Mayo clinic[8], Merriam-Webster dictionary[9] and SNOMED CT BioPortal which is a repository of biomedical ontologies [87]. For instance, below is an expanded query for topic *shark cartilage*, which has been shown to have no effect on survival rate or quality of life for cancer patients [57]:

```
''Shark cartilage'' OR ''AE-941'' OR ''Marine Collagen'' OR ''Marine Liquid Cartilage'' OR ''MSI-1256F''
OR ''Neovastat'' OR ''Sphyrna lewini'' OR ''Squalus:acanthias'') AND cancer
```

It includes a typical way to refer to the topic, as well as more technical version of the treatment, and related products such as Neovastat, a shark cartilage extract[10]. Once again, the extended queries were verified by an oncologist for correctness and completeness. Using the Twitter Streaming Application Program Interface (API), we collect a total of 215,109 tweets about these rumor topics (see Figure 1) spanning 2011-2013 and 39,675 users.

3.2 User Selection

For both rumor and control tweets, we aim at eliminating users that are not human such as bots, organizations, or whose tweets do not refer to the actual topics of interest (but were picked up due to a faulty or ambiguous keyword matching). We perform several steps to raise the likelihood the selected users meet the above criteria.

- We apply the *Humanizr* tool [61] to the tweets, which was shown to have an accuracy of 94.1% predicting whether a Twitter user is an organization account. In this step, we remove 161 accounts from the rumor user set and 615 from the control set.
- Next, we compile a (human) name dictionary with associated genders by combining names extracted from a large collection of Google+ accounts [60] with baby names published by National Records of Scotland[11] and United States National Security[12], resulting in a dictionary containing 106,683 names. After matching this dictionary to user names, as well as applying heuristics (such as having "Mrs." or "Mr."), we keep only users with a matching name or identifier, excluding 15,164 (38.2%) users from rumor and 207,394 (30.6%) from control sets. As illustrated in Table 1, name matched accounts are more often verified accounts, have fewer overall tweets, followers and following users than the non-gendered users, indicating they are less active than those not matching a name in our dictionary.
- Finally, we compute the average tweeting rate for every user as the ratio of total number of lifetime tweets over the number of days since the account was created. To exclude what are likely to be automated accounts in both rumor and control datasets, we retain users with an average tweeting rate less than or equal to 24 tweets per day (following posting activity thresholds such as in [43, 68]). Applying this criteria, we discard 6,463 (26%) users from rumor and 144,904 (31%) users from control sets.

---

[7]https://adwords.google.com/ko/KeywordPlanner

[8]http://www.mayoclinic.org/

[9]https://www.merriam-webster.com/

[10]https://www.cancer.gov/publications/dictionaries/cancer-drug?cdrid=42021

[11]https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/vital-events/names/babies-first-names/full-lists-of-babies-first-names-2010-to-2014

[12]https://www.ssa.gov/oact/babynames/limits.html

Fig. 1. Data collection and refinement process.

Table 1. Average statistics of users whose names were found in name dictionary versus those not found.

Name match?	Followers	Followees	Tweets	Verified
Control				
yes	3,566	841	28,459	1.24%
no	5,594	1,011	20,012	0.88%
Rumor				
yes	5,306	1,559	25,347	1.17%
no	10,163	1,761	35,850	0.80%

For the remaining user accounts in both sets, we use the Twitter API user endpoint to collect the most recent 3,200 tweets, synchronizing the time spans for the two datasets to span Paul & Dredze timeline in 2012-2013.

3.3 Relevance Refinement

3.3.1 Human Labeling. As thus far the data has been gathered using keyword matching, we refine the document (and thus, user) inclusion criteria by employing crowdsourced labeling and

machine learning. In particular, we take this opportunity to make sure our data is on topic using CrowdFlower[13] crowdsourcing platform to label a subset of the data, which then we use to build topical classifiers to determine the labels for the rest. Note that instead of assessing the the veracity of the claims, we are now interested in making sure the text of the tweets indeed contains the cure claims, requiring basic lay language understanding, as is reflected in the task description below.

We begin by sampling the datasets. To ensure representativeness, we stratify the sample of rumor dataset such that at least 10 tweets from each topic are present, and the rest of the larger topics are sampled until a maximum of 100 tweets. This results in 4,152 tweets (which were de-duplicated by cleaned text). Similarly, we sample 4,000 tweets from control set for labeling.

To ensure high quality of annotations, for each subset, 30 tweets were labeled and used as a "gold-standard". Using these tweets with known labels, the annotators are first given a quiz, and thereafter tested in each task (wherein gold standards are hidden among other tweets). The annotator must pass the quiz and maintain at least 70% accuracy throughout the labeling process for their work to be accepted. A minimum of three independent labels were collected for each tweet to achieve a majority decision, and trial tasks of 100 tweets each were first run. A total of 184 annotators were selected by CrowdFlower, contributing a minimum of 60 annotations each. Due to this large number of participants we report %-age agreement instead of Fleiss' kappa.

The tasks themselves differed slightly between the data sources. For control we ask the workers to label each tweet on (i) whether it is about cancer, and if so, (ii) whether there is a personal (or friend/family) experience, (iii) whether there is a claim that something cures cancer, or (iv) whether some other cancer-related information is present. For rumor (recall these tweets also mention some treatment or remedy) we ask whether the tweet (i) is about some cancer remedy, and if so, whether there is (ii) a claim it helps with treating or curing cancer, (iii) prevents cancer, or (iv) debunks such a claim. Given the tasks had multiple selections, the agreement was relatively high at 78.7% for control and 82.0% for rumor. Note, as discussed earlier, in the instructions to the labelers we emphasized looking for a claim that the remedy *treats or cures cancer*, not just a symptom, with several illustrative examples for clarity.

The results of the labeling task for the control tweets were as follows: 2,890 were labeled as having information about cancer whereas 1,110 tweets were labeled as non-related to the cancer topic. From the 2,890 cancer related tweets, 1,632 (40%) were about personal experience, 98 (2%) were about cancer cure and 1,1160 (29%) were about other cancer-related information (symptoms, awareness, prevention, causes, etc.). The results of the labeling task for rumor tweets were as follows: 2,564 tweets were about a cancer cure and 1,587 were not about a cancer cure. From the 2,564 tweets about the remedy, 1,791 (43%) tweets claimed that the suggested treatment helped to cure cancer (claimed a rumor), 564 (13%) tweets were about prevention and 209 (5%) tweets were debunking the claim.

3.3.2 Classification. Next, we train several logistic regression classifiers on the labeled tweets using 1,2,3-grams as features. We train the classifiers on the labeled tweets, which we then apply to the rest to characterize each user's behavior. Summaries of selection for the two datasets are below:

- Rumor: (1) is the tweet about a cancer cure? yes: 12,685, no: 7,872. Out of cancer cure tweets, (2) what kind of information does it have? claiming a cure: 9,549, prevention: 2,850, debunking claims of cure: 285. We define Rumor users as *users who claim a cure is helpful for curing or treating cancer and **not** users who talk about prevention or debunking*, resulting in 12,046 tweets for 7,221 users.

---

[13]http://crowdflower.com/

Table 2. Top rumor topics by number of unique users contributing tweets matching the expanded query.

#	Topic	Users	Tweets	Expanded Query
70	Juicing	6,656	13,083	juice OR juicing OR "juice diet" OR "juice plus" OR "juice +" OR "fruit vegetable juice"
11	Apitherapy	3,330	7,905	apitherapy OR honey OR pollen OR "bee bread" OR "propolis" OR "royal jelly" OR "bee venom" OR "bee sting"
52	Ginger	3,113	8,928	ginger
10	Antioxidants	2,908	5,671	antioxidant
121	Urine therapy	2,532	4,686	urine OR urinotherapy OR uropathy OR "auto-urine therapy" OR shivambu
9	Antineoplaston therapy	2,365	7,889	antineoplaston OR burzynski
81	Magnetic therapy	2,327	30,789	magnetic OR magnet OR magnets OR magnotherapy
124	Walnuts	2,013	5,474	walnut OR walnuts OR "Juglans regia" OR akhrot OR "wall nut"
4	Acupuncture	1,817	5,359	acupuncture OR accupuncture
103	Poly-MVA	1,705	7,252	"lipoic acid mineral complex" OR "poly-mva" OR "poly mva" OR "minerals vitamins and amino acids" OR vitalzym OR curcumin OR ahcc OR essiac

- Control: (1) is the tweet about cancer? yes: 339,047, no: 50,670. Out of cancer tweets, (2) which include a personal experience? yes: 197,608, no: 141,439. Further, (3) is the tweet is suggesting a cure? (Applying *Synthetic Minority Oversampling Technique* [15] to balance classes) cure: 2,252, not 336,794. We define Control users as *users who post at least once about cancer, but **not** about a cancer cure*, resulting in 341,157 tweets for 270,622 users without and 199,343 tweets for 163,261 users with personal experience with cancer.

The overall process of user selection is summarized in Figure 1, with resulting 16M tweets for 7,221 users in Treatment and 506M tweets for 443,883 users in Control datasets. Note that we do not make the distinction in the Rumor set between personal and non-personal experiences, as in a separate crowdsourced evaluation we find only 4% to be about personal experiences.

4 RESULTS

4.1 Rumor Topics

Table 2 shows the "treatments" (or "rumors") which have the most user membership, along with the expanded queries which were used to collect the tweets. The most popular is *juicing*, followed by similar widely available remedies, *honey* and *ginger*, as well as the *antioxidant* keyword (which is often applied to a range of foods). We find a wide variety of claims surround foods and drinks. Some make bold claims outright: *"[...] University show that the soursop fruit kills cancer cells effectively, particularly prostate cancer cells, pancreas and lung"*, others speculate *"Can ginger help cure ovarian cancer? Since 2007, the University of [...] has been studying GINGER... <url>"*, yet others invoke religious backing: *"RT @<user>: Islamic backed #cancer cure: Prophetic medicine cures woman of cancer using [...]: <url>"*. More unusual topics include *Antineoplaston therapy* available in Dr. Burzynski clinic (for more, see Discussion), and *urine therapy*. Note that the keyword queries returned both outrageous claims as well as debunking, such as *"RT Dr. Burzynski He has the cure for cancer, the FDA want to shut him down <url>"* and on the other side *"Burzynski Clinic libel threat to silence critics of fake cancer treatment <url> @<user>"*[14]. In Data section, we describe how we apply supervised machine learning to sort out actual claims of purported cures (from unrelated content or rumor debunking, for instance), and use them to identify the users engaged in misinformation. Overall, of the 139 topics collected, median number of tweets collected was 269.5, with a minimum of 11.

---

[14]Tweets have been slightly re-phrased to preserve user's privacy.
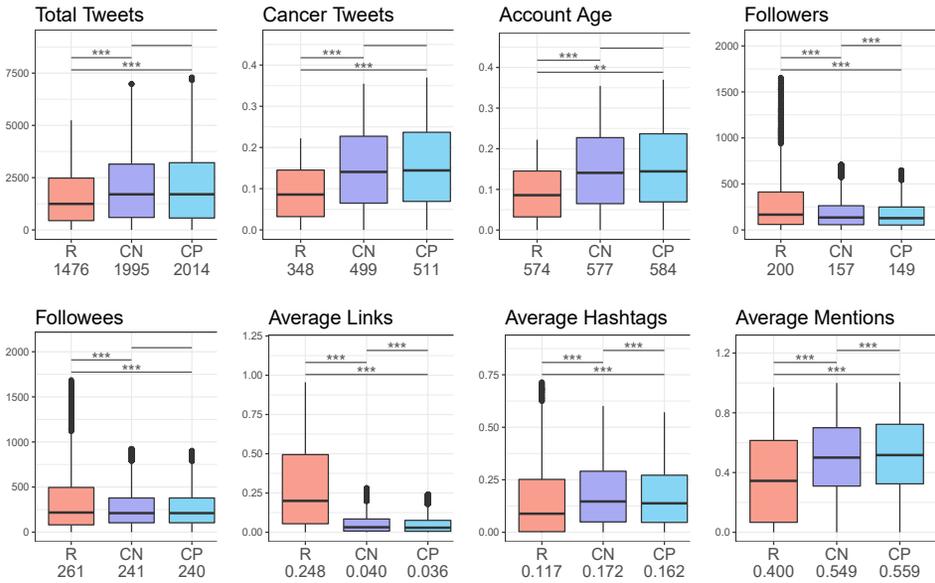
Fig. 2. Summary of characteristics of Rumor, Control Non-personal and Control Personal user groups. For each characteristic a box plot (excluding outliers outside 90th percentile) is shown with median values under the title. Differences in medians are tested using Mann-Whitney U test, for which p-values, Bonferroni adjusted for multiple hypothesis testing, are shown on the corresponding lines spanning the two variables being compared: $p < 0.0001$ ***, $p < 0.001$ **, $p < 0.01$ *.

These topics exemplify the breadth of the subjects covered in this dataset, as well as indicate the alternative cancer medicines popular on social media.

4.2 Modeling Rumormongering

We begin by comparing the users who have posted on these and other topics. Figures 2 show box plots of behavioral statistics for the three kinds of users identified above (with outliers beyond 90th percentile excluded for clarity), such that the median is shown graphically as the bold line in each box, and also shown numerically under the label. The datasets are compared using using Mann-Whitney U test – a non-parametric test that is more appropriate for highly skewed data for which normality cannot be assumed – in the bars above the plots, with p-value level indicated symbolically. We find the Rumor user set to be quite different from the other two sets of users, having fewer total account lifetime tweets (1,476 compared to around 2,000 for Control), as well as cancer-related tweets in our dataset (more than 100 fewer), more followers and followees (some users being vastly more popular, note the long tails), and sharing more links (however fewer hashtags and mentions). Interestingly, in some behaviors there is a significant difference between personal and non-personal control tweets, with users having personal interactions with cancer having fewer followers, sharing fewer links and hashtags, but posting more mentions than non-personal control.

To examine the user behavior more deeply, we characterize the content which may be predictive of rumormongering behavior. In particular, we are interested in examining the tweets *before* a user started posting about a rumor, not necessarily the claims themselves. Thus, for Rumor users we select the tweets before the first rumor post, and for Control we sample such a date from a normal distribution having mean and variance of first rumor posts of Rumor data. This way we aim to

Table 3. User and aggregated tweet features.

Scope	Feature	Description
User	FOLLOWING	The number of people the user is following
	FOLLOWERS	The number of people following the user
	STATUS_COUNT	The number of tweets at posting time
	ACCOUNT AGE	The time passed since the author registered his/her account, in days
	VERIFIED	Whether account has been verified by Twitter
Sentiment	SENTIMENT SCORE	Sentiment score value [58]
	POSITIVE/NEGATIVE WORDS	The number of positive/negative words in text
	EMOTICONS POS/NEG	Count total number of positive and negative emoticons in text
Linguistic	IS RETWEET	Is a retweet; contains RT
	HAS MENTIONS	Mentions a user, eg: @CNN
	HAS HASHTAG	Contains hash_tags
	URLS COUNT	Count total number of URLs in text
	HASHTAG COUNT	Count total number of hashtags
	MENTION COUNT	Count total number of mentions
	WORD COUNT	Count total number of words in text
	CHAR COUNT	Count total number of characters in text
	UPPER COUNT	Count total number of upper case letters
	COUNT SENTENCES	Count number of sentences
	QUESTION MARK	Contains question mark '?'
	EXCLAMATION MARK	Contains exclamation mark '!'
	PERCENTAGE UPPER/LOWER	The percentage of upper and lower case characters
	MULTIPLE QUES/EXCL	Contains multiple questions or exclamation marks
	COUNT NOUN	Count total number of nouns in text
	COUNT ADVERB	Count total number of adverbs in text
	COUNT ADJECTIVE	Count total number of adjectives in text
	COUNT VERB	Count total number of verbs in text
	COUNT PRONOUN	Count total number of pronouns in text
	HAS PRONOUN 1	Contains a personal pronoun in 1th person
	HAS PRONOUN 2	Contains a personal pronoun in 2nd person
	HAS PRONOUN 3	Contains a personal pronoun in 3rd person
	LIWC	73 categories from psycholinguistic resource LIWC
Readability	COMPLEX WORDS	Count total number of complex words in text
	READABILITY SCORES	Automated, Flesch_Kincaid, Gunning, and SMOG [29]
	COUNT NOT WORD2VEC	Count total number of words not in "word2vec" Google News vocabulary
	AVG SYLLABLES	The Average number of syllables per word in text
Medical	MEDICAL_DOMAINS	Refers to URL from a known medical organization, from (anonymized)
Timing	INTERVAL ENTROPY	Entropy of hour intervals between tweets from [35]

avoid biasing the selection to different time periods which may trivially differentiate users. This selection allowed for the analysis of at least 100 posts for 4,212 Rumor users.

Building on our previous work (anonymized), and use multifaceted behavior and content features which in the literature have been linked to credibility assessment of social media content. The features, listed in Table 3, span user-specific statistics, as well as aggregated (via averaging) tweet-specific metrics. User features encompass proxies of popularity (number of followers and followees), as well as productivity (number of posts up to date). Tweet features can be grouped into surface and linguistic forms of the tweet and well as semantically enriched ones including sentiment extracted from words and special characters, readability indices, and number of domains known to come from medical organization (anonymized). We also include a measure of entropy of the intervals between posts, which has been used to measure the predictability of retweeting patterns [35]. Finally, we include the psycholinguistic resource LIWC[15], which has been shown to relate to user mindset [21].

---

[15]https://liwc.wpengine.com/

Table 4. Logistic regression with LASSO regularization model, predicting whether a user posts about a rumor, with forward feature selection. For each feature, coefficient (unstandardized), standard error, and accompanying p-value are shown. Significance levels: $p < 0.0001$ ***, $p < 0.001$ **, $p < 0.01$ *, $p < 0.05$.

variable	coefficient	std. error	p-value
(Intercept)	-6.160	1.405	***
Avg syllables per word	17.120	0.660	***
Is verified	-40.310	42310	
Percentage uppercase / lowercase	-0.201	0.018	***
Word count	1.491	0.131	***
SMOG readability score	-0.753	0.123	***
Percentage uppercase	0.191	0.019	***
Character count	-0.163	0.024	***
Number of cancer tweets	0.001	1.9E-04	***
LIWC48: ingest	1.839	0.722	*
Negative word count	-1.460	0.262	***
URL count	3.364	0.505	***
Is retweet	4.947	0.790	***
word2vec count	-0.634	0.165	***
LIWC55: focuspast	-1.636	0.567	**
LIWC37: tentat	2.531	0.859	**
Number of sentences	-0.610	0.205	**
LIWC32: male	-1.820	1.000	
Interval entropy	0.508	0.105	***
Account age	-0.001	2.7E-04	***
LIWC23: posemo	-0.490	0.384	
LIWC61: time	-1.431	0.378	***
LIWC13: adverb	1.758	0.536	**
LIWC20: number	2.936	1.317	*
Statuses count	7.1E-05	2.6E-05	**
LIWC42: hear	-4.742	1.799	**
Has 1st person pronoun	-1.504	0.662	*
LIWC62: work	1.591	0.665	*
LIWC40: percept	1.217	0.754	

We then turn to examining the relationship between these variables and the tendency of the user to post about a cancer treatment rumor. To mitigate class imbalance, we under-sample Control group by randomly sampling users to achieve a one to one balance with Rumor set. We then apply logistic regression with LASSO regularization, as the predicted class is binary and LASSO performs variable regularization and selection. However, as data has a large number of potentially collinear features, we also use forward feature selection method which employs Akaike Information Criterion (AIC) to select features contributing the most to the performance of the model [84] (note the significant features remain largely the same, but the selection process assists in ranking most prominent ones). The resulting model is shown in Table 4, such that the features selected first are at the top. The McFadden R^2, the alternative to the R^2 of linear regression, is 0.925, indicating a good fit to the data. We also perform a matched experiment wherein we match Rumor to Control

users on the number of followers, such that for each Rumor user the closest match in Control is picked, resulting in McFadden R^2 of 0.906. Examining the features, we can observe:

- We find **readability** to be of importance, with the average number of syllables per word and SMOG readability score at the top, as well as other style-related features.
- The fact whether or not account is **verified** is also important, however due to sparsity it is not statistically significant, indicating that such policing by the social media website may be of limited value.
- The top LIWC category is "ingest", one dealing with **eating and drinking**, echoing user's interest in topics potentially related to some of the most popular remedies we found (juices, superfoods, supplements, etc.).
- These users are also more **prolific in writing about cancer**, with the number of cancer tweets being positively associated with posting a rumor (however each individual tweet counts little toward overall probability, with coefficient at 0.001).
- They are also more likely to use **tentative language** (LIWC category 37), possibly speculating about topics other than the rumors captured in this data.
- Besides other LIWC categories pointing to speaking less positively and male and using more adverbs and numbers (as well as sharing more URLs), we find a weak negative relationship between using **first person pronouns** ("I","we"), indicating those engaging in posting about these rumors are not likely to be personally involved (remember also that we did not find many personal statements in Rumor set during labeling as well).
- The positive relationship of posting **interval entropy** [35] means the higher inter-tweeting time entropy – and the less regular (not bot-like) is the posting behavior – more likely the user to post about a rumor, pointing to a largely "human" cohort.

Thus, we find (likely non-bot) users who have a more sophisticated language, who are interested in cancer, and whose language already contains speculations (besides the rumor), but who are not personally involved with the illness.

To examine the language of these groups of users in more detail, in Table 5 we summarize the top 20 words, with stopwords removed, in all historical tweets by control users (left), all historical tweets of rumor users (center), and only rumor tweets (right). The frequency list on the right shows some of the main trends in the tweets explicitly endorsing a "treatment". Again, we find juices and antioxidants to be popular, and prominent mentions of "help", "cure", and "treatment" (with "cure" being the more popular keyword than "treatment"). The center and the left lists show words in non-rumor tweets of Rumor users (center) and Control (left). Note that although both groups of users are in our dataset because at some point they have mentioned cancer, Rumor users are more focused on health, even when they are not explicitly talking about rumors, with these top 20 words containing 5 health-related words for Rumor users, and none in Control. Thus, we find an encouraging sign that propensity for posting cancer treatment misinformation can be modeled and predicted automatically. Next, we discuss ramifications of this observation.

5 DISCUSSION

This study expands the misinformation research prominent in Social Computing, which has been largely focused on the political domain [17, 42, 59, 79, 81], to healthcare – where erroneous beliefs and actions may cause serious bodily damage. Complementing HCI literature on human computer use and its sociocultural implications, this study extends current work on monitoring social media during crises and pandemics [41] as well as on the tracking of specific behaviors within a community of interest (such as in [3, 62, 63]). Below, we elaborate on the ecosystem of health communication and monitoring, possible application of our model, its theoretical contributions, and limitations.

Table 5. Word frequency tables summarizing the top 20 most popular terms, excluding stopwords, in all historical tweets by control users (left), all historical tweets of rumor users (center), and only rumor tweets (right).

Control History				Rumor History				Rumor Misinformation			
love	1.95%	night	0.66%	good	1.01%	video	0.54%	cancer	1.43%	cells	0.50%
good	1.55%	life	0.63%	health	1.00%	food	0.54%	juice	0.81%	out	0.48%
day	1.34%	happy	0.60%	day	0.96%	back	0.50%	RT	0.77%	healthy	0.45%
time	1.22%	ill	0.59%	love	0.85%	work	0.46%	breast	0.73%	diabetes	0.44%
people	1.00%	hope	0.58%	time	0.78%	work	0.45%	risk	0.61%	prostate	0.44%
lol	0.99%	feel	0.55%	great	0.73%	diet	0.44%	help	0.58%	antioxidant	0.42%
today	0.96%	haha	0.51%	people	0.71%	healthy	0.40%	health	0.55%	pain	0.40%
back	0.94%	follow	0.51%	today	0.68%	post	0.38%	helps	0.54%	chronic	0.37%
great	0.73%	home	0.49%	news	0.62%	weight	0.38%	cure	0.54%	patients	0.37%
work	0.70%	man	0.47%	life	0.57%	blog	0.36%	treatment	0.53%	study	0.36%

Context and Case Studies. Internet has long contributed to the ongoing "deprofessionalization" of medical practice. As Michael S. Goldstein writes in *Persistence and Resurgence of Medical Pluralism*, "Health information on the Internet enhances the autonomy of those who are ill, demystifies the knowledge and practices of doctors, and increases overall skepticism about medicine" [38]. In the larger history of antipathy toward professionals and their monopoly on knowledge, social media presents a new venue for patients, consumers, and concerned citizens to network and share their experiences and knowledge. It brings many of the remedies traditionally associated with home and family to a social domain. In such networked setting, knowledge may spread and evolve. For instance, [53] found that people tended to change their belief about a health topic when it did not concur with a majority of others. These findings emphasize the potential power of social and word-of-mouth (WOM) marketing, which could benefit both from positive messages and from controversy around the product [52].

Such marketing of lay health information may be especially effective in vulnerable populations, those having difficulty accessing medical care, or having poor medical literacy. For instance, a drug called "Laetrile", also known as "Amygdalin" or "Vitamin B17", is popularly promoted in India as an anti-cancer remedy [9]. Despite a ban on its marketing as such by Food and Drug Administration (FDA) in the United States in 1979, it remains popular in India [44], and in our data we found 2,417 mentions of it in the context of cancer. Currently, Laetrile is promoted on YouTube[16] and other social media – media to which general public has much more access than scientific literature – allowing for an international audience to be reached. An exciting future research direction lies in enriching our dataset with geo-location in order to track the supporters of these treatments across the world.

However, an opposite reaction can also be possible. A clinic purporting to cure its patients of cancer by its founder Stanislaw R. Burzynski, MD has for several decades been a subject of scientific renunciation [39]. Recently, bloggers and activists which have been criticizing Burzynski for "disturbing business and research practices" have been attacked personally on social media, which encouraged these "skeptics" to organize and educate the public about the unproven nature of Burzynski's treatments [10]. In 2017, Dr. Burzynski was placed on probation for five years by the Texas Medical Board and ordered to pay a total of $60,000 in fines and restitution for not adequately informing patients about the treatments that they were receiving [14]. We find social media and Twitter specifically to be a new battleground for the health claims of different parties, some of which may be businesses set to lose profit if their message is contested. In our dataset, we found 7,889 tweets mentioning Burzynski or "Antineoplaston therapy" he proposes. Case studies of such

---

[16]A YouTube search for "laetrile" on April 15, 2018 has resulted in a top video titled "Learn How Laetrile Kills Cancer Cells!".

two-sided interactions provide a window into the consequences of increased plurality in voices aiming to spread health-related information.

Applications. As beliefs are strongly linked to behavior, honing internet-enabled communication with patients and public at large is important in improving interventions of health behaviors. In this context, present work contributes to the Social Computing community in proposing a tool for monitoring health misinformation on a large scale. Specifically, the model built in this study exemplifies specialized tools that can help address the spread of health misinformation on social media, mainly in (i) automatically detecting Twitter users who may be likely to post questionable information, (ii) attempting to change those individuals' view of the topic, and (iii) quickly identifying and limiting the spread of misinformation.

First, in the age of personalization, such models can be employed to target individuals potentially susceptible to follow questionable accounts, consume poor quality health information, and propagate it. Early identification of rumor-prone individual accounts allows for refining traditional broad-spectrum information campaigns via personalized employment of *persuasive technologies* which offer a way to tailor content to the individual and track individual progress [8]. In particular, such technologies attempt to nudge the user to change his or her attitude or behavior through persuasion or social influence (but not misinformation or coersion), and have already been applied to health coaching and communication [19, 91].

Secondly, multi-faceted features proposed in this study provide a foundation for examining both behavioral characteristics and interests of those prone to rumormongering. Such interest lists can be expanded beyond LIWC to specialized topical lexicons, such as those on vaccination hesitancy [24], eating disorders [33, 92], antibiotics [76], etc.

Finally, further monitoring of suspect accounts will allow for timely identification of new potentially questionable content before it has a chance to propagate through the network, alerting public health officials of new waves of content or public interest. This content then can be automatically pre-assessed for credibility using approaches such as in [12], notifying officials if the content passes a certain threshold.

Note that, automated tools will not be able to replace expert knowledge, but instead contribute to a fruitful human-expert-in-the-loop paradigm which has been proposed for research and machine learning processes [37, 46]. In particular, we describe a pipeline for training the model for the tracking of discussions around "complementary and alternative medicines" for cancer, and we show that it achieves a high McFadden R^2 in fitting the data, however the pipeline can be applied to any other healthcare topic. Further, in order to remain relevant in the changing discourse, it must be periodically re-trained with fresh data in order to ameliorate "concept drift" (for which streaming solutions are also being developed [32]).

Thus, contributing to CSCW community's interest in application-driven computer supported systems, our tool is a potential component in the public health communication and monitoring ecosystem. For example, the largest regulator of possible cancer treatments in the US is the Food and Drug Administration (FDA), which solicits reports concerning "defects in the quality or safety" of some product via Safety Reporting Portal [83]. Such reporting scheme relies on parties existing who are concerned about a particular product, who to study it and file a report. Such system allows in the meanwhile for potentially dangerous health products and advise to affect the public. Outside official channels, social media is being explored by thinktanks in order to track adverse drug side effects [7], however to the best of our knowledge currently no official FDA social media surveillance systems exist. Early detection of harmful information is a necessary precondition to timely deployment of corrective communication, if not to the sources of misinformation, then to their immediate audience. As such, early detection systems contribute to the discussion of computer-supported emergency communication in CSCW community [55, 69]. Thus as outlined above, automated detection of

social media accounts susceptible to propagating health misinformation proposed here has great potential for expanding the capacity for preventative monitoring and communication of current healthcare communication ecosystem.

Theory of Rumormongering. More generally, psychological underpinnings of believing and propagating rumors have been studied in the context of larger societal impact. In post WWII Europe, Gordon W. Allport and Leo Postman [2] studied rumors out of concern about the damage to morale and national safety. In the ample literature following, rumors have been defined as public communications that are infused with private hypotheses about how the world works, in particular to help us cope with our anxieties and uncertainties [73]. Hypothesizing about the nature of rumor, Allport and Postman postulated that the strength of a rumor (R) will vary with the importance of the subject to the individual concerned (i) times the ambiguity of the evidence pertaining to the topic at hand (a), or $R \approx i \times a$ [74]. Contributing to this theoretical work, our empirical analysis of social media health misinformation shows that those engaged in spreading unproven cancer treatments largely are not personally involved in the matter. Thus, we propose to extend the definition of "importance" i to other motivating factors beyond the personal, which may be at play in the public sphere of Twitter. Further analysis is required to reveal motivating factors in such health misinformation spread, as cancer fraud has been acknowledged in Oncology literature [85].

Lessons learned from the empirical studies of psychological drives in rumormongering are essential in building more effective policies on communicating scientific information and managing public opinion on issues of medicine and related policies. Unlike in political domain where bots can hijack the conversation [20, 80], we find that posting interval entropy (measuring irregularity of post timings) was positively related with a rumormongering behavior, pointing to a more "human" trait. This finding emphasizes the importance of public education and communication campaigns as preventive measures targeting public beyond social media. For instance, after a public outcry in Italy to legalize a stem cell-based treatment for neurological diseases unsupported by published evidence, researchers and public health officials called for an improvement in guidelines for its media on communicating scientific information to the public [48]. Special care needs to be taken to promote clarifications and retractions, as it has been shown that these are not as popular as, for example, the original wrong news stories [28].

Limitations. Studying health misinformation on social media has several important limitations. Social media adoption and use differs widely between population segments: for instance, close to half (45%) of 18- to 24-year-olds in U.S. use Twitter, compared to 24% of all adults, as reported by Pew Research Center in March 2018 [72]. Detecting legitimate personal accounts (as opposed to bots or organizational accounts) remains a challenge, which we attempted to address using existing tools like Humanizr and baby name dictionaries, which undoubtedly introduce their own biases potentially excluding certain minorities. In particular, the Social Security name database includes all names registered at least 5 times in a year, dating back to 1880, capturing a large majority of names used. However, more resources could have been used to include the names of minorities, such as the Register of Liberated Africans[17]. The results of this study should be taken in the light of this limitation, as we may have failed to detect misinformation in some communities. Improvement in the detection of real humans (versus bot or organizational accounts) will allow for a more accurate account selection for studying individuals.

Accessibility issues also bias the view of the populations having, for instance, visual impairment [89] or other constraints to using the medium. Further, some health conditions and personal topics are associated with a social stigma which limit their discussion on social forums. For example, [22] found some illnesses to be searched more often than discussed in social media – a bias which may

---

[17]http://liberatedafricans.org/

affect our selection of certain cancers. Finally, attitudes toward self-expression and trust in publicly available information may differ wildly between cultural subgroups, such as in case of Hong Kong youths, who were found to be significantly more likely to disclose personal health issues with peers online compared to their U.S. counterparts [56]. Hong Kong youths also held the highest level of trust towards health-related information on social media, again pointing to the need of personalized approach to health communication sensitive to the culture of the participants. Finally, observations in this study concern exclusively treatment claims of cancer, and may not generalize to other illnesses, especially if they have different societal stigmas. The model proposed in this work inherits the above limitations, thus any integration of such automated tracking must be closely monitored for bias and topic drift, and regularly updated in order to capture latest developments in social media norms.

Finally, as any technology, the proposed analytical pipeline may be misused when applied within faulty policies. The expert definition of misinformation must be subjected to ethical constraints of medical and public health standards. How the information that is flagged by the system is handled must also avoid discouraging public discourse and information seeking.

Privacy. This study used Twitter posts which were publicly available at the time of data collection, with no private messages or messages deleted by the time of the collection included. Also, accounts which have been deleted since Paul & Dredze collection have not been included in the study. Furthermore, the sharing of this dataset will be done according to Twitter's Terms of Use.

6 CONCLUSION

In this paper we present a case study of health misinformation on social media by examining Twitter users involved in propagating alternative medicines claiming to treat or cure cancer. Through a multi-stage process including machine learning, crowdsourcing, and heuristics, we select users who are likely to be real people, and who post on one of 139 such topics. We find that these users are likely to use more sophisticated language, and circulate in health domain prior to posting a rumor, but are not likely to be personally involved in the illness. Our findings suggest that cancer treatment misinformation may be spread not by patients, but by other actors. More research needs to be done to ascertain motivations, tactics, and impact of such accounts. Finally, the dataset collected for this study presents a highly-curated resource for the research community's future studies on the topic of health misinformation.

7 ACKNOWLEDGEMENTS

REFERENCES

[1] Sofiane Abbar, Yelena Mejova, and Ingmar Weber. 2015. You tweet what you eat: Studying food consumption through twitter. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 3197–3206.

[2] Gordon W Allport and Leo Postman. 1947. The psychology of rumor. (1947).

[3] Teresa Almeida, Rob Comber, and Madeline Balaam. 2016. HCI and Intimate Care as an Agenda for Change in Women's Health. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2599–2611.

[4] Marjolijn L Antheunis, Kiek Tates, and Theodoor E Nieboer. 2013. Patients' and health professionals' use of social media in health care: motives, barriers and expectations. *Patient education and counseling* 92, 3 (2013), 426–431.

[5] Matheus Araújo, Yelena Mejova, Ingmar Weber, and Fabricio Benevenuto. 2017. Using Facebook Ads Audiences for Global Lifestyle Disease Surveillance: Promises and Limitations. In *Proceedings of the 2017 ACM on Web Science Conference*. ACM, 253–257.

[6] Matheus Lima Diniz Araújo, Yelena Mejova, Michael Aupetit, and Ingmar Weber. 2017. Visualizing Health Awareness in the Middle East.. In *ICWSM*. 725–726.

[7] Joanna Belbey. 2016. The FDA And Social Listening For Adverse Effects. *Forbes* (2016). https://www.forbes.com/sites/joannabelbey/2016/07/15/the-fda-and-social-listening-for-adverse-effects/

[8] Shlomo Berkovsky, Jill Freyne, and Harri Oinas-Kukkonen. 2012. Influencing Individually: Fusing Personalization and Persuasion. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 2, 2 (2012), 9.

[9] Aakriti Bhatnagar, Samta Mittal, and Achint Garg. 2017. Laetrile: A wonder drug or farce? *International Journal of Applied Dental Sciences* (2017).

[10] Robert Blaskiewicz. 2016. Skeptic Activists Fighting for Burzynski's Cancer Patients. *Skeptical Inquirer* (2016).

[11] Alex Broom and Philip Tovey. 2008. The role of the Internet in cancer patients' engagement with complementary and alternative treatments. *Health:* 12, 2 (2008), 139–155.

[12] Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. 2011. Information Credibility on Twitter. In *Proceedings of the 20th International Conference on World Wide Web*. ACM, 675–684.

[13] Stevie Chancellor, Jessica Annette Pater, Trustin Clear, Eric Gilbert, and Munmun De Choudhury. 2016. # thyghgapp: Instagram content moderation and lexical variation in pro-eating disorder communities. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 1201–1213.

[14] Julie Chang. 2017. Texas Medical Board sanctions controversial cancer doctor Burzynski. https://www.mystatesman.com/news/texas-medical-board-sanctions-controversial-cancer-doctor-burzynski/L9lDsfNTbBOuaWWqLEDBUI/

[15] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. 2002. SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research* 16 (2002), 321–357.

[16] Xueyu Chen and Lillian L Siu. 2001. Impact of the media and the internet on oncology: survey of cancer patients and oncologists in Canada. *Journal of Clinical Oncology* 19, 23 (2001), 4291–4297.

[17] Yimin Chen, Niall J Conroy, and Victoria L Rubin. 2015. Misleading online content: Recognizing clickbait as false news. In *Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection*. ACM, 15–19.

[18] A Cipriani, TA Furukawa, and C Barbui. 2011. What is a Cochrane review? *Epidemiology and psychiatric sciences* 20, 03 (2011), 231–233.

[19] Brian Cugelman. 2013. Gamification: what it is and why it matters to digital health behavior change developers. *JMIR Serious Games* 1, 1 (2013).

[20] Clayton Allen Davis, Onur Varol, Emilio Ferrara, Alessandro Flammini, and Filippo Menczer. 2016. Botornot: A system to evaluate social bots. In *Proceedings of the 25th International Conference Companion on World Wide Web*. International World Wide Web Conferences Steering Committee, 273–274.

[21] Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. 2013. Predicting Depression via Social Media. *International Conference on Web and Social Media (ICWSM)* 13 (2013), 1–10.

[22] Munmun De Choudhury, Meredith Ringel Morris, and Ryen W White. 2014. Seeking and sharing health information online: comparing search engines and social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1365–1376.

[23] Mark Dredze, David A Broniatowski, and Karen M Hilyard. 2016. Zika vaccine misconceptions: A social media analysis. *Vaccine* 34 (2016), 3441–3442.

[24] Adam G Dunn, Julie Leask, Xujuan Zhou, Kenneth D Mandl, and Enrico Coiera. 2015. Associations between exposure to and expression of negative opinions about human papillomavirus vaccines on social media: an observational study. *Journal of medical Internet research* 17, 6 (2015).

[25] Christopher J Dy, Samuel A Taylor, Ronak M Patel, Alison Kitay, Timothy R Roberts, and Aaron Daluiski. 2012. The effect of search term on the quality and accuracy of online information regarding distal radius fractures. *Journal of Hand Surgery* 37, 9 (2012), 1881–1887.

[26] Marie-Desirée Ebel, Ivonne Rudolph, Christian Keinki, Andrea Hoppe, Ralph Muecke, Oliver Micke, Karsten Muenstedt, and Jutta Huebner. 2015. Perception of cancer patients of their disease, self-efficacy and locus of control and usage of complementary and alternative medicine. *Journal of cancer research and clinical oncology* 141, 8 (2015), 1449–1455.

[27] J. P. T. Higgins (Ed.). 2008. *Cochrane Handbook for Systematic Reviews of Interventions*. Vol. 5. The Cochrane Collaboration. www.cochrane-handbook.org.

[28] Gunther Eysenbach and Per Egil Kummervold. 2005. "Is cybermedicine killing you?" -the story of a Cochrane disaster. *Journal of Medical Internet Research* 7, 2 (2005).

[29] Lijun Feng, Martin Jansche, Matt Huenerfauth, and Noémie Elhadad. 2010. A comparison of features for automatic readability assessment. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*. Association for Computational Linguistics, 276–284.

[30] Susannah Fox. 2011. *The social life of health information, 2011*. Pew Internet & American Life Project Washington, DC.

[31] Yael Frish and Dov Greenbaum. 2017. Is Social Media a Cesspool of Misinformation? Clearing a Path for Patient-Friendly Safe Spaces Online. *The American Journal of Bioethics* 17, 3 (2017), 19–21.

[32] Adel Ghazikhani, Reza Monsefi, and Hadi Sadoghi Yazdi. 2014. Online neural network model for non-stationary and imbalanced data stream classification. *International Journal of Machine Learning and Cybernetics* 5, 1 (2014), 51–62.

[33] Jannath Ghaznavi and Laramie D Taylor. 2015. Bones, body parts, and sex appeal: An analysis of #thinspiration images on popular social media. *Body Image* 14 (2015), 54–61.

[34] Amira Ghenai and Yelena Mejova. 2017. Catching Zika Fever: Application of Crowdsourcing and Machine Learning for Tracking Health Misinformation on Twitter. In *Proceedings - 2017 IEEE International Conference on Healthcare Informatics, ICHI 2017*. https://doi.org/10.1109/ICHI.2017.58

[35] Rumi Ghosh, Tawan Surachawala, and Kristina Lerman. 2011. Entropy-based classification of 'retweeting' activity on twitter. *Proceedings of KDD workshop on Social Network Analysis (SNA-KDD)* (2011).

[36] Antonio A Ginart, Sanmay Das, Jenine K Harris, Roger Wong, Hao Yan, Melissa Krauss, and Patricia A Cavazos-Rehg. 2016. Drugs or Dancing? Using Real-Time Machine Learning to Classify Streamed "Dabbing" Homograph Tweets. In *Healthcare Informatics (ICHI), 2016 IEEE International Conference on*. IEEE, 10–13.

[37] Dominic Girardi, Josef Küng, Raimund Kleiser, Michael Sonnberger, Doris Csillag, Johannes Trenkler, and Andreas Holzinger. 2016. Interactive knowledge discovery with the doctor-in-the-loop: a practical example of cerebral aneurysms research. *Brain informatics* 3, 3 (2016), 133.

[38] Michael S Goldstein. 2004. The persistence and resurgence of medical pluralism. *Journal of health politics, policy and law* 29, 4 (2004), 925–945.

[39] Saul Green. 1992. 'Antineoplastons': an unproved cancer therapy. *JAMA* 267, 21 (1992), 2924–2928.

[40] Jeremy A Greene, Niteesh K Choudhry, Elaine Kilabuk, and William H Shrank. 2011. Online social networking by patients with diabetes: a qualitative evaluation of communication with Facebook. *Journal of General Internal Medicine* 26, 3 (2011), 287–292.

[41] Xinning Gui, Yubo Kou, Kathleen H Pine, and Yunan Chen. 2017. Managing Uncertainty: Using Social Media for Risk Assessment during a Public Health Crisis. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 4520–4533.

[42] Aditi Gupta, Hemank Lamba, Ponnurangam Kumaraguru, and Anupam Joshi. 2013. Faking sandy: characterizing and identifying fake images on twitter during hurricane sandy. In *Proceedings of the 22nd international conference on World Wide Web*. ACM, 729–736.

[43] Maria Han Veiga and Carsten Eickhoff. 2016. A cross-platform collection of social network profiles. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. ACM, 665–668.

[44] E Sheehan Helen. 2008. Cancer, access to investigational drugs, and patient rights in the USA and India. *Indian Journal of Medical Ethics* 5, 4 (2008).

[45] Alex Hern. 2017. How social media filter bubbles and algorithms influence the election. *The Guardian* (2017). https://www.theguardian.com/technology/2017/may/22/social-media-election-facebook-filter-bubbles

[46] Andreas Holzinger. 2016. Interactive machine learning for health informatics: when do we need the human-in-the-loop? *Brain Informatics* 3, 2 (2016), 119–131.

[47] Jutta Huebner, Oliver Micke, Ralph Muecke, Jens Buentzel, Franz Josef Prott, Ulrich Kleeberg, Bianca Senf, Karsten Muenstedt, et al. 2014. User rate of complementary and alternative medicine (CAM) of patients visiting a counseling facility for CAM of a German comprehensive cancer center. *Anticancer research* 34, 2 (2014), 943–948.

[48] Kalina Kamenova and Timothy Caulfield. 2015. Stem cell hype: media portrayal of therapy translation. *Science Translational Medicine* 7, 278 (2015), 278ps4–278ps4.

[49] Irina Khaldarova and Mervi Pantti. 2016. Fake news: The narrative battle over the Ukrainian conflict. *Journalism Practice* 10, 7 (2016), 891–901.

[50] Holly Korda and Zena Itani. 2013. Harnessing social media for health promotion and behavior change. *Health Promotion Practice* 14, 1 (2013), 15–23.

[51] Patty Kostkova, Vino Mano, Heidi J Larson, and William S Schulz. 2016. VAC Medi+ board: Analysing Vaccine Rumours in News and Social Media. In *Proceedings of the 6th International Conference on Digital Health Conference*. ACM, 163–164.

[52] Robert V Kozinets, Kristine De Valck, Andrea C Wojnicki, and Sarah JS Wilner. 2010. Networked narratives: Understanding word-of-mouth marketing in online communities. *Journal of Marketing* 74, 2 (2010), 71–89.

[53] AYS Lau, TMY Kwok, and E Coiera. 2011. How online crowds influence the way individual consumers answer health questions. *Applied clinical informatics* 2, 02 (2011), 177–189.

[54] Annie YS Lau, Elia Gabarron, Luis Fernandez-Luque, and Manuel Armayones. 2012. Social media in health - what are the safety concerns for health consumers? *Health Information Management Journal* 41, 2 (2012), 30–35.

[55] Alex Leavitt and John J Robinson. 2017. The Role of Information Visibility in Network Gatekeeping: Information Aggregation on Reddit during Crisis Events.. In *CSCW*. 1246–1261.

[56] Wan-Ying Lin, Xinzhi Zhang, Hayeon Song, and Kikuko Omori. 2016. Health information seeking in the Web 2.0 age: Trust in social media, uncertainty reduction, and self-disclosure. *Computers in Human Behavior* 56 (2016), 289–294.

[57] Charles L Loprinzi, Ralph Levitt, Debra L Barton, Jeff A Sloan, Pam J Atherton, Denise J Smith, Shaker R Dakhil, Dennis F Moore, James E Krook, Kendrith M Rowland, et al. 2005. Evaluation of shark cartilage in patients with

advanced cancer. *Cancer* 104, 1 (2005), 176–182.

[58] Will Lowe, Kenneth Benoit, Slava Mikhaylov, and Michael Laver. 2011. Scaling policy preferences from coded political texts. *Legislative studies quarterly* 36, 1 (2011), 123–155.

[59] Jing Ma, Wei Gao, Zhongyu Wei, Yueming Lu, and Kam-Fai Wong. 2015. Detect rumors using time series of social context information on microblogging websites. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management.* ACM, 1751–1754.

[60] Gabriel Magno and Ingmar Weber. 2014. International gender differences and gaps in online social networks. In *International Conference on Social Informatics.* Springer, 121–138.

[61] James McCorriston, David Jurgens, and Derek Ruths. 2015. Organizations Are Users Too: Characterizing and Detecting the Presence of Organizations on Twitter. In *ICWSM.* 650–653.

[62] Yelena Mejova, Sofiane Abbar, and Hamed Haddadi. 2016. Fetishizing Food in Digital Age:# foodporn Around the World. In *ICWSM.* 250–258.

[63] Yelena Mejova, Youcef Benkhedda, and Khairani. 2017. #Halal Culture on Instagram. *Frontiers in Digital Humanities* 4 (2017), 21.

[64] Yelena Mejova, Hamed Haddadi, Sofiane Abbar, Azadeh Ghahghaei, and Ingmar Weber. 2015. Dietary Habits of an Expat Nation: Case of Qatar. In *Healthcare Informatics (ICHI), 2015 International Conference on.* IEEE, 57–62.

[65] Tanushree Mitra, Scott Counts, and James W Pennebaker. 2016. Understanding Anti-Vaccination Attitudes in Social Media. In *ICWSM.* 269–278.

[66] S Anne Moorhead, Diane E Hazlett, Laura Harrison, Jennifer K Carroll, Anthea Irwin, and Ciska Hoving. 2013. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. *Journal of medical Internet research* 15, 4 (2013).

[67] Anthony Nastasi, Tyler Bryant, Joseph K Canner, Mark Dredze, Melissa S Camp, and Neeraja Nagarajan. 2017. Breast cancer screening and social media: a content analysis of evidence use and guideline opinions on Twitter. *Journal of Cancer Education* (2017), 1–8.

[68] Alexandra Olteanu, Onur Varol, and Emre Kiciman. 2017. Distilling the Outcomes of Personal Experiences: A Propensity-scored Analysis of Social Media. In *CSCW.* 370–386.

[69] Alexandra Olteanu, Sarah Vieweg, and Carlos Castillo. 2015. What to expect when the unexpected happens: Social media communications across crises. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing.* ACM, 994–1009.

[70] Michael J Paul and Mark Dredze. 2014. Discovering health topics in social media using topic models. *PloS one* 9, 8 (2014), e103408.

[71] Sharoda A Paul, Lichan Hong, and Ed H Chi. 2011. Is Twitter a Good Place for Asking Questions? A Characterization Study. In *ICWSM.*

[72] "Pew Research Center". 2018. *Social Media Use in 2018.* Technical Report. http://www.pewinternet.org/2018/03/01/social-media-use-in-2018/

[73] Ralph L Rosnow. 1991. Inside rumor: A personal journey. *American Psychologist* 46, 5 (1991), 484.

[74] Ralph L Rosnow and Eric K Foster. 2005. Rumor and gossip research. *Psychological Science Agenda* 19, 4 (2005), 1–2.

[75] Hamman Samuel and Osmar Zaıane. 2018. MedFact: Towards Improving Veracity of Medical Information in Social Media using Applied Machine Learning. *Canadian Conference on Artificial Intelligence* (2018).

[76] Daniel Scanfeld, Vanessa Scanfeld, and Elaine L Larson. 2010. Dissemination of health information through social networks: Twitter and antibiotics. *American journal of infection control* 38, 3 (2010), 182–188.

[77] Katja Schmidt and Edzard Ernst. 2004. Assessing websites on complementary and alternative medicine for cancer. *Annals of Oncology* 15, 5 (2004), 733–742.

[78] Iris Seaman and Christophe Giraud-Carrier. 2016. Prevalence and Attitudes about Illicit and Prescription Drugs on Twitter. In *Healthcare Informatics (ICHI), 2016 IEEE International Conference on.* IEEE, 14–17.

[79] Chengcheng Shao, Giovanni Luca Ciampaglia, Alessandro Flammini, and Filippo Menczer. 2016. Hoaxy: A platform for tracking online misinformation. In *Proceedings of the 25th International Conference Companion on World Wide Web.* International World Wide Web Conferences Steering Committee, 745–750.

[80] Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Alessandro Flammini, and Filippo Menczer. 2017. The spread of fake news by social bots. *arXiv preprint arXiv:1707.07592* (2017).

[81] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake News Detection on Social Media: A Data Mining Perspective. *ACM SIGKDD Explorations Newsletter* 19, 1 (2017), 22–36.

[82] Kate Starbird. 2017. Examining the Alternative Media Ecosystem Through the Production of Alternative Narratives of Mass Shooting Events on Twitter. In *ICWSM.* 230–239.

[83] US Food and Drug Administration. 2018. The Safety Reporting Portal. https://www.safetyreporting.hhs.gov/

[84] WN Venables and BD Ripley. 2002. Modern Applied Statistics with S. Springer, New York, NY. (2002).

[85] Wendy H Vogel. 2011. Internet Oncology: Cure Seekers Beware! *Journal of the Advanced Practitioner in Oncology* 2, 6 (2011), 409–412.

[86] Hedy S Wald, Catherine E Dube, and David C Anthony. 2007. Untangling the Web - The impact of Internet use on health care and the physician–patient relationship. *Patient education and counseling* 68, 3 (2007), 218–224.

[87] Patricia L Whetzel, Natalya F Noy, Nigam H Shah, Paul R Alexander, Csongor Nyulas, Tania Tudorache, and Mark A Musen. 2011. BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. *Nucleic acids research* 39, suppl_2 (2011), W541–W545.

[88] Ryen W White and Ahmed Hassan. 2014. Content bias in online health search. *ACM Transactions on the Web (TWEB)* 8, 4 (2014), 25.

[89] Shaomei Wu and Lada A Adamic. 2014. Visually impaired users on an online social network. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*. ACM, 3133–3142.

[90] Haodong Yang and Christopher C Yang. 2013. Harnessing social media for drug-drug interactions detection. In *Healthcare Informatics (ICHI), 2013 IEEE International Conference on*. IEEE, 22–29.

[91] Lucy Yardley, Leanne Morrison, Katherine Bradbury, and Ingrid Muller. 2015. The person-based approach to intervention development: application to digital health-related behavior change interventions. *Journal of medical Internet research* 17, 1 (2015).

[92] Elad Yom-Tov, Luis Fernandez-Luque, Ingmar Weber, and Steven P Crain. 2012. Pro-anorexia and pro-recovery photo sharing: a tale of two warring tribes. *Journal of medical Internet research* 14, 6 (2012).