Characterizing COVID-19 Misinformation Communities Using a Novel Twitter Dataset

Shahan Ali Memon
samemon@cs.cmu.edu

Kathleen M. Carley
kathleen.carley@cs.cmu.edu

Carnegie Mellon University

Abstract

From *conspiracy theories* to *fake cures* and *fake treatments*, COVID-19 has become a hotbed for the spread of misinformation online. It is more important than ever to identify methods to debunk and correct false information online. In this paper, we present a methodology and analyses to characterize the two competing COVID-19 misinformation communities online: (i) *misinformed users* or users who are actively posting misinformation, and (ii) *informed users* or users who are actively spreading true information, or calling out misinformation. The goals of this study are twofold: (i) collecting a diverse set of annotated COVID-19 Twitter dataset that can be used by the research community to conduct meaningful analysis; and (ii) characterizing the two target communities in terms of their network structure, linguistic patterns, and their membership in other communities. Our analyses show that COVID-19 misinformed communities are denser, and more organized than informed communities, with a possibility of a high volume of the misinformation being part of disinformation campaigns. Our analyses also suggest that a large majority of misinformed users may be anti-vaxxers. Finally, our sociolinguistic analyses suggest that COVID-19 informed users tend to use more narratives than misinformed users.

1 Introduction

With the emergence of COVID-19 pandemic, the political and medical misinformation has elevated to create what is being commonly referred to as the *global infodemic*. False information has hampered proper communication, and affected the decision-making process [BE+20]. This makes debunking of false information vitally important. According to one study [TLC15], if left undisputed, misinformation can in fact exacerbate the spread of the epidemic itself. Process of debunking misinformation, however, is complex and one that is not completely understood [CJHJA17]. This is because in order to conduct any intervention, it is first imperative to be able to identify the misinformation, as well as the misinformed communities. Because of the scarcity of data, and diversity of misinformation themes, this is already a challenging task in itself, but is also not enough. A second, and arguably a more important aspect of an intervention is to be able to *correct* and *change* the beliefs of the misinformed communities. To be able to do this, it is important to understand how different communities interact, which sub-communities they belong to, and what are their preferences. In this paper, we characterize the COVID-19 misinformation communities on Twitter in terms of their network structure, linguistic patterns, and membership in other misinformation and disinformation sub-communities. In the process, we also design and collect a large annotated dataset with a comprehensive codebook that we make available for the community to use for further analysis and models for misinformation detection.

2 Background

2.1 COVID-19 Datasets

In the short amount of time, many COVID-19 datasets have been released. Most of these datasets are generic, and lack annotations or labels. Examples include multilingual corpus on a wide variety of topics related

Title of the Proceedings: "Proceedings of the CIKM 2020 Workshops October 19-20, Galway, Ireland"
Editors of the Proceedings: Stefan Conrad, Ilaria Tiddi

to COVID-19 [CLF20, AMEP+20, HJB+20], longitudinal Twitter chatter dataset [BTW+20], multilingual dataset with location information of the users [QIO20], Twitter dataset for Arabic tweets [AAA20], Twitter dataset for popular Arabic tweets [HHSE20], and dataset for identification of stance, replies, and quotes [VCKBC20]. Most of these datasets either have no annotations at all, employ automated annotations using transfer learning or semi-supervised methods, or are not specifically designed for misinformation.

In terms of datasets collected for COVID-19 misinformation analysis and detection, examples include CoAID [CL20] which contains automatic annotations for tweets, replies, and claims for fake news; ReCOVery [ZMFZ20] is a multimodal dataset annotated for tweets sharing reliable versus unreliable news, annotated via distant supervision; FakeCovid [SN20] is a multilingual cross-domain fake news detection dataset with manual annotations; and [DSW20] is a large-scale Twitter dataset also focused on fake news. A survey of the different COVID-19 datasets can be found in [LUM+20] and [SAAA20].

In terms of the diversity of the classes, and the size of the dataset, the most relevant dataset is by Alam et al. [ASN+20] who, like our study, present a comprehensive codebook to annotate tweets on a finer granularity. Their dataset, however, is limited to a few hundred tweets, and our dataset is much more diverse in the range of topics covered. Dharawat et al. [DLMZ20] present a similar dataset with focus on the severity of the misinformation. However, their dataset does not consider the different "types" of misinformation. Finally, Song et al. present a dataset in [SPJ+20] which contains a diverse set of 10 categories, but still is not as large, and contains fewer categories in relation to the dataset collected within our study.

2.2 Misinformation Analysis

A plethora of research has already been conducted for analysing COVID-19 misinformation online. Some examples include categorization and identification of misinformed users based on their home countries, social identities, and political affiliation [HC20, SSM+20], characterization of different types conspiracy theories propagated by Twitter bots [Fer20], characterization of the prevalence of low-credibility information related to COVID-19 [YTLM20], exploratory analysis of the content of COVID-19 tweets [OPR20, SDM20], understanding the types, sources, and claims of COVID-19 misinformation [BSHN20], and comparison of the credibility of COVID-19 tweets to datasets pertaining to other health issues [BKF+20]. To the best of our knowledge none of the studies have characterized COVID-19 misinformation communities in terms of their sociolinguistic patterns. In this study, we do not characterize the misinformation *content* directly. Instead, we conduct a set of analysis to understand and characterize the competing COVID-19 *communities* through their content, and content-sharing behaviors and interactions.

3 Methodology

3.1 Data Collection

To collect Twitter dataset, we use Twitter search API using a diverse set of keywords as shown in table 1 to collect data. We collected our data on three days: 29th March 2020, 15th June 2020, and 24th June 2020. Each of these collections extracted a set of tweets from their corresponding week. For the annotation process, tweets were randomly sampled from that set.

Table 1: This table shows the hashtags, and keywords we used in conjunction with "coronavirus" and "covid" to collect data from Twitter

Type	Terms
Keywords	*bleach, vaccine, acetic acid, steroids, essential oil, saltwater, ethanol, children, kids, garlic, alcohol, chlorine, sesame oil, conspiracy, 5G, cure, colloidal silver, dryer, bioweapon, cocaine, hydroxychloroquine, chloroquine, gates, immune, poison, fake, treat, doctor, senna makki, senna tea*
Hashtags	*#nCoV20199, #CoronaOutbreak, #CoronaVirus, #CoronavirusCoverup, #CoronavirusOutbreak, #COVID19, #Coronavirus, #WuhanCoronavirus, #coronaviris, #Wuhan*

3.2 Data Annotation

Our annotation task aims to determine the category to which a given tweet belongs to. After many discussions and revisions, we identify 17 categories that a particular tweet could classify to. These 17 categories are defined in table 2. These categories are defined in further detail along with their definitions and examples in our codebook which we make available for the public to use.

Based on these categories, tweets were randomly and uniformly sampled from the data collection to maintain diversity in terms of topics covered. In the first phase around 4573 tweets were annotated by a single annotator. Table 2 shows the distribution of the data in terms of the different categories as annotated by the first annotator. In the second phase, 651 of these annotated tweets were assigned randomly to 6 other annotators.

Table 2: This table describes the categories we identified to classify/annotate tweets along with the distribution of annotations as identified by Annotator 1 in the first phase.

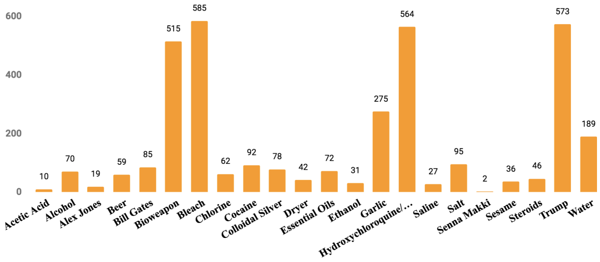
Category	Count
Irrelevant	131
Conspiracy	924
True Treatment	0
True Prevention	175
Fake Cure	141
Fake Treatment	34
False Fact or Prevention	321
Correction/Calling out	1331
Sarcasm/Satire	476
True Public Health Response	163
False Public Health Response	3
Politics	512
Ambiguous/Difficult to Classify	143
Commercial Activity or Promotion	37
Emergency Response	17
News	95
Panic Buying	70

4 Data Description

Our data collection strategy is different from others in two main aspects: (i) we have a diverse set of categories taking into consideration different types of information and misinformation online; and (ii) our dataset is one of the very few, if not the only one, with emphasis on informed communities with categories such as "True Prevention", "Calling out/correction", "True Public Health Response", and "Sarcasm". We believe this is necessary as building models requires not just the annotation of false information, but as well as complementary true information categories.

At the end, we have 4573 annotated tweets, comprising of 3629 users with an average of 1.24 tweets per user. Our annotated data not only covers a wide range of categories as observed in table 2, but also covers a wide range of topics as can be seen in figure 1. We call this dataset *CMU-MisCOV19* [MC20]. In adherence to the FAIR principles, the database and the codebook has been uploaded to Zenodo and is accessible with the following link: http://doi.org/10.5281/zenodo.4024154. In adherence to the Twitter's terms and conditions, we do not provide the full tweet JSONs, but provide the tweet IDs so that the tweets can be rehydrated. We also provide the annotations, and the date of creation for each tweet for the reproduction of the results of our analyses. The annotated tweets are included in a CSV file with the following fields: *status_id* (tweet id of the tweet), *status_created_at* (timestamp of the creation of the tweet), *annotation1* (annotated class of the tweet by the first annotator), and *annotation2* (annotated class of the tweet by the second annotator, if exists).

Figure 1: This chart shows the frequency of each identified topic across all the tweets. Note: Some tweets may have more than one topic.



5 Analysis and Discussion

5.1 Identifying Communities

Conducting analyses for a competing set of communities requires identifying those communities first. Because we have already annotated data across a set of true and false information categories, we identify the membership of the users by assigning a valence of +1 to the categories *True Treatment, True Prevention, Correction/Calling Out, Sarcasm/Satire, and True Public Health Response*, and a valence of -1 to the categories *Conspiracy, Fake Cure, Fake Treatment, False Fact or Prevention, and False Public Health Response.* Note that we assign the valence to the categories (or annotations) and not the tweets themselves. This is so that we can leverage the annotations from multiple annotators. At the end, we compute the valence of each user as a weighted sum of the valence of the annotations assigned to their tweets. Then we use the valence assigned to each user to identify their membership i.e. if valence is greater than 0, the user is assigned to the *informed* group, and if the valence is less than 0, the user is assigned to the *misinformed* group. Out of 3629 users, the community detection process assigns 47% (1697) of the users to the informed group, 29% (1043) of the users to the misinformed group, and 24% (889) of the users to ambiguous or irrelevant category[1].

5.2 Data Augmentation

Because our goal is to characterize communities and their behaviors, once we identify the two communities, we collect the timelines of users in each community to augment our data. Our hypothesis is that these additional posts can be used to mitigate survivorship bias [BGIR92] within our analyses. To conduct network analysis, bot analysis, and sociolinguistic analysis, we first extract only the COVID-19 related tweets from

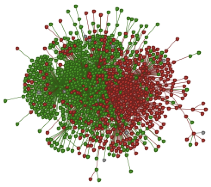
[1] Irrelevant users are users who have only posted tweets within other categories such as "Politics" or "Emergency". Because these categories do not have an assigned valence related to misinformation, they are not relevant for the purposes of this study.

the timelines of each user. We do this by filtering all the tweets by the case-insensitive keywords *"corona"* and *"covid"*. This yields a total of 330609 tweets with an average of 91 tweets per user.

5.3 Network Analysis

To conduct network analysis, we extract the retweet, mention, and reply networks of the two target communities, and combine those networks together. We then compute the *network density* for each of the two groups. As described in [MTMC20], network density is defined as the ratio of actual connections and potential connections. In dense networks, conformity of the ideas is highly encouraged, and difference of opinions is discouraged. We also use ORA-PRO [CRC, ACR17, ACR18] to plot the network graph as shown in figure 2

Figure 2: Retweet+Mention+Reply network with informed users (in green) and misinformed users (in red) created using ORA-PRO [ACR18, Car17]. Note: Users with unidentified or ambiguous membership have been removed from the graph for simplicity.



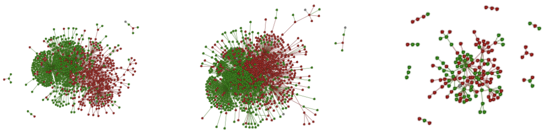
We note that both the informed and misinformed users display echo-chamberness with misinformed sub-communities being much denser than the informed sub-communities as shown in table 3. We do, however, notice some two-way communication from both sides.

Table 3: This table shows the number of nodes, links, and the network density for the two target sub-communities.

Measure	Overall	Informed	Misinformed
Nodes	2477	1515	923
Links	2947	1489	826
Network Density	4.8e-4	6.5e-4	9.7e-4

We also plot the retweet, mention and reply network separately as shown in figure 3. While retweet, and mention network show little to no two-way communication, we can observe that the reply network, while small in size, does in fact have much more intergroup engagement. We hypothesize that this is likely a consequence of the "corrective" or "calling-out" behavior.

Figure 3: Retweet (left), mention (middle), and reply network (right) with informed users (in green) and misinformed users (in red) created using ORA-PRO. [ACR18, Car17]



5.4 Bot Detection

To understand the role of bots within the two competing groups, we used Bot-Hunter [BC18b, BC18a, BCB$^+$18, BC20], which has a precision of .957 and a recall of .704, to identify potential bot-like accounts. We use the probability of greater than or equal to .75 as our confidence threshold to identify bots. We use a two-sample z-test for the difference of proportions ($\alpha = 0.05$) to test the difference in proportion of bots between the two competing groups of users. The results of our analyses can be found in table 4.

Table 4: This table shows the number and percentage of bots within each of the two competing groups

Measure	Overall	Informed	Misinformed
Number of Users	3629	1697	1043
Number of Bots	505	184	202
Percentage of Bots	14%	11%	19%

We observe that from a total of 3629 users, 14% (505) of the users are identified as bots. The percentage of bots within identified misinformed users, however, is much higher (19%) than within identified informed users (11%). We find our results to be statistically significant ($p < 0.001$; $z = -6.23$). This indicates that more than 1/5th of the misinformation related posts in our dataset are potentially a result of disinformation campaigns related to COVID-19.

5.5 Sociolinguistic Analysis

To understand the linguistic differences between the two competing communities, we conduct a linguistic analysis based on the tweets of the two groups by using the Linguistic Inquiry and Word Count (LIWC) program [PBJB15]. LIWC is a text analysis tool which looks at the different lexical categories each of which is psychologically meaningful. For a given text, LIWC calculates the percentage of each LIWC categories. All of these categories are based on word counts.

We run the LIWC program on the timelines of all the members for each of the two competing groups. We only use tweets relevant to COVID-19. We also remove users identified as bots. Because some users

Table 5: This table shows the summary of our analyses across all the linguistic dimensions described above using LIWC. The first column shows the lexical category. The second and third columns show the test statistic (M_1) as the mean of the LIWC indices for informed and misinformed communities respectively. The fourth and fifth columns display the z-score and p-value for the independent z-test for the difference in means.

Lexical Category	M_1 (Informed)	M_1 (Misinformed)	z-score (Z_1)	p-value (Z_1)
function	33.90	29.32	7.25	$< .001$
pronoun	7.97	6.53	4.89	$< .001$
ipron	3.26	3.03	1.23	.2
ppron	4.71	3.49	5.39	$< .001$
Analytic	69.83	76.01	-4.82	$< .001$
social	6.49	5.05	5.45	$< .001$
family	.34	.20	2.24	.03
friend	.17	.17	-.03	.97
Authentic	25.12	16.43	6.78	$< .001$
Tone	35.42	37.59	-1.45	.15
informal	4.89	5.16	-1.63	.10
swear	.51	.34	1.86	.06

may be more active than others, using the results of the program as is may introduce biases in our analyses. To account for those biases, we first normalize the percentages by the size of the data for each user. We use the mean of the normalized LIWC indices of tweets of individual users for a given lexical category as our test statistic. We use an independent z-test for the difference in means to establish statistical significance. For all our tests, $\alpha = 0.05$. Our analyses are summarized in table 5.

For this part, we focus on investigating three linguistic dimensions, each of which, along with its linguistic correlates, is described below.

5.5.1 Narrative Discourse Structure

Narratives play a central role in how individuals process information, communicate, and reason [Ves17]. We set to test the differences in the usage of narratives or anecdotes between the two COVID-19 misinformation communities. The LIWC correlates for narrative discourse structure include high usage of function words, pronouns, analytic summary dimension, and authenticity. High usage of function words and pronouns happens more often when expressing feelings and behaviors which tends to happen frequently in narratives [Pen11]. Moreover, low analytical thinking also suggests narrative language [PBJB15]. Furthermore, authentic individuals tend to be more personal, humble, and vulnerable [PBJB15]. Therefore, we use all of these as proxies to identify variation in the use of narratives across communities.

In the past [MTMC20], it has also been suggested that misinformed communities (eg. anti-vaxxers) tend to use many more pronouns suggesting highly narrative discourse structure. In this analysis, however, we find that informed users in the COVID-19 discourse use significantly more pronouns, more functional words, mention more family-related keywords, are less analytical, and more authentic and honest in comparison to misinformed users. All of these suggest that informed users may use many more narratives than misinformed users. This is an interesting finding as it presents a dichotomy between the different misinformation communities (eg. anti-vaxxers and COVID-19 misinformed community). In hindsight, this is also an intuitive result, as our informed group is obtained from corrective discourse where users present their stories of family members or friends suffering from COVID-19 to call out conspiracies and false information. Because the two communities still seem to have less two-way communication, this also suggests that just the content and framing of the message (i.e. narratives) may not be enough, and perhaps there is a need to connect the two groups by identifying an effective medium of communication.

5.5.2 Tone

Tone describes how positive a given text is. According to the definition by LIWC, the higher the tone index, the more positive the tone. Indices less than 50 typically suggest a more negative tone. While we do not see significant differences in the emotional tone of the competing groups, we find both the communities to be highly negative.

5.5.3 Linguistic formality

Formality of the language has often been considered as one of the most important dimensions for stylistic variation. In [GMC+14], authors define linguistic formality as a style of writing that is meant to be precise, coherent, articulate and convincing to an educated audience, as opposed to informal discourse which is filled with deictic references (eg. here, there), pronouns, and narration. The LIWC correlates to this dimension are swear words (swear), and informal language (informal). Informal language in LIWC is computed on the bases of swear words, netspeak (eg. btw, lol), nonfluencies (eg. err, hmm), assents (eg. agree, OK), and fillers (eg. youknow).

From table 5, it can be observed that misinformed users tend to be more informal than informed users, though informed users tend to use more swear words than misinformed users. This is intuitive as many of our informed users post corrective or sarcastic tweets to call out misinformation. However, our results are not significant, and, hence inconclusive.

5.6 Vaccination Stance

To understand the interplay between the different kinds of misinformation themes and communities, we identify the vaccination-related stance of the members of the misinformed sub-community. To do that, we first identify the subset of misinformed community who have posted at least one tweet related to "vaccines" in the past. We then collect the user-to-hashtag co-occurrence network. We use the valence of the vaccination hashtags obtained via the label propagation-based method mentioned in the study in [MTMC20] to identify the stance of each member (pro vs. anti) based on the weighted sum of the valences of the hashtags. If the weighted sum is greater than 0, we identify the member as pro-vaxxer, and if the weighted sum is less than 0, we identify the member as anti-vaxxer. The distribution of the pro- and anti-vaxxers within the COVID-19 misinformed group is as shown in table 6.

Table 6: This table shows the number and percentage of pro- and anti-vaxxers within the misinformed group.

Measure	Value
Users w/ vaccine-related tweets	2750 (out of 3629)
Misinformed users	1027 (37%)
Anti-vaxxers	423 (41%)
Pro-vaxxers	224 (22%)
Ambiguous	380 (37%)
Misinformed pro-vaxxer bots	37 (17%)
Misinformed anti-vaxxer bots	82 (22%)

We observe that from 1027 COVID-19 misinformed users in our dataset, 41% of the members are identified as anti-vaxxers, whereas only 22% of the members are identified as pro-vaxxers. The difference between the proportions of the two communities is significantly high. We also identify the proportion of bots within each of the two groups: *misinformed pro-vaxxers*, and *misinformed anti-vaxxers*. As shown in table 6, 17% of the misinformed pro-vaxxers are bots, which is significantly lower than the proportion of bots within the misinformed anti-vaxxers. The first thing this suggests is that a big chunk of COVID-19 misinformation online may in fact be *disinformation*, and hence, intentional. The existence of bots within both the informed and misinformed communities also suggests that much of the disinformation online may be an organized effort to amplify the COVID-19 debate to create discord in the communities as seen in the past with Twitter bots and Russian trolls [BJQ+18].

6 Limitations

The first important limitation pertaining to our work is that most of our analyses are based on the data that has been annotated by only 1 annotator. We try to mitigate this by having more than 1/7th of our annotations annotated by a second annotator, and taking into account all those annotations while computing the membership for each user. Another limitation to our work is that all our analyses are correlational in nature, and do not depict causation. A limitation pertaining to our data collection strategy is that we collect our data across a period of three weeks, augment our data with timelines of users, and update our list of hashtags to account for new themes. We then sample a subset of this data for annotation process. Because of the way data was collected, it cannot be used for assessing change over time. Moreover, while this ensures the diversity of misinformation-related topics and agents, it may limit our ability to estimate the actual extent to which the different types of stories are more or less present. Another limitation related to our bot analysis is that we use a second-level inference from a trained model. We try to mitigate this by using labels with probability greater than or equal to .75 to ensure high quality labels. Finally, unlike "vaccination" related discourse, COVID-19 does not have a clear definition of the "stance" of the users. This is because there are many sub-topics associated to COVID-19 each of which could have its own stance. In this work, we categorize users based on misinformation. However, the relationship between misinformation and stance vis-a-vis issues is complex, and one that needs to be understood. In the future work, we hope to explore this relationship to create a systematic way of characterizing communities both in terms of misinformation, and the difference stances of the users.

7 Conclusion

In this paper, we present a methodology to characterize the competing COVID-19 misinformation communities by comparing them in terms of their network structure, sociolinguistic variation, and membership in disinformation campaigns and in other health-related misinformation communities such as anti-vaxxers. We find that even though COVID-19 is a recent event, misinformation related to it has created a set of polarized communities with high echo-chamberness. Misinformed communities are observed to be denser than informed communities which is in line with previous studies such as [MTMC20]. We find that bots exist in both the informed and misinformed groups, but the percentage of bots in misinformed users is significantly higher suggesting the prevalence of disinformation campaigns. Our sociolinguistic analysis suggests that both the target communities depict negative emotional tone in their posts, with signals that informed users use many more narratives than misinformed users. Finally, we discover that many misinformed users may be anti-vaxxers. Our analyses suggest that misinformation communities are much more complex as they are highly organized, and tend to be highly analytical. Unlike previous suggestions [SOC19], they may not be responsive to narrative correctives, and hence, a "one size fits all" generic messaging intervention for debunking misinformation may not be a feasible solution. A successful intervention may require to identify, and ban the disinformation campaigns. It may also be useful to identify the right medium of communication to connect the two groups. This can be achieved by identifying users in misinformed communities who are not *rebroadcasting*, or have high betweenness centrality to be messengers for disseminating factual information. It may also be useful to further understand the linguistic patterns and preferences of these communities to create an effective *content* and *framing* of the messaging.

7.0.1 Acknowledgements

This work was partially supported by a fellowship from Carnegie Mellon University's Center for Machine Learning and Health to Shahan A. Memon. We thank David Beskow for access to his Bot-Hunter model for bot analysis. We also thank members of CMU's Center for Computational Analysis of Social and Organizational Systems (CASOS) for insightful comments and discussions related to the data codebook and its revisions.

References

[AAA20] Sarah Alqurashi, Ahmad Alhindi, and Eisa Alanazi. Large arabic twitter dataset on covid-19. *arXiv preprint arXiv:2004.04315*, 2020.

[ACR17] Neal Altman, Kathleen M Carley, and Jeffrey Reminga. Ora user's guide 2017. *Carnegie-Mellon Univ. Pittsburgh PA Inst of Software Research International, Tech. Rep.*, 2017.

[ACR18] Neal Altman, Kathleen M Carley, and Jeffrey Reminga. Ora user's guide 2018. *Carnegie-Mellon Univ. Pittsburgh PA Inst of Software Research International, Tech. Rep.*, 2018.

[AMEP+20] Muhammad Abdul-Mageed, Abdel-Rahim Elmadany, Dinesh Pabbi, Kunal Verma, and Rannie Lin. Mega-cov: A billion-scale dataset of 65 languages for covid-19. *arXiv preprint arXiv:2005.06012*, 2020.

[ASN+20] Firoj Alam, Shaden Shaar, Alex Nikolov, Hamdy Mubarak, Giovanni Da San Martino, Ahmed Abdelali, Fahim Dalvi, Nadir Durrani, Hassan Sajjad, Kareem Darwish, et al. Fighting the covid-19 infodemic: Modeling the perspective of journalists, fact-checkers, social media platforms, policy makers, and the society. *arXiv preprint arXiv:2005.00033*, 2020.

[BC18a] David M Beskow and Kathleen M Carley. Bot conversations are different: leveraging network metrics for bot detection in twitter. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 825–832. IEEE, 2018.

[BC18b] David M Beskow and Kathleen M Carley. Bot-hunter: a tiered approach to detecting & characterizing automated activity on twitter. In *Conference paper. SBP-BRiMS: International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation*, 2018.

[BC20] David Beskow and Kathleen M Carley. *Social Cybersecurity*. Springer, 2020.

[BCB+18] David Beskow, Kathleen M Carley, Halil Bisgin, Ayaz Hyder, Chris Dancy, and Robert Thomson. Introducing bothunter: A tiered approach to detection and characterizing automated activity on twitter. In *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and*

Behavior Representation in Modeling and Simulation. Springer, 2018.

[BE+20] Darrin Baines, RJ Elliott, et al. Defining misinformation, disinformation and malinformation: An urgent need for clarity during the covid-19 infodemic. *Discussion Papers*, pages 20–06, 2020.

[BGIR92] Stephen J Brown, William Goetzmann, Roger G Ibbotson, and Stephen A Ross. Survivorship bias in performance studies. *The Review of Financial Studies*, 5(4):553–580, 1992.

[BJQ+18] David A Broniatowski, Amelia M Jamison, SiHua Qi, Lulwah AlKulaib, Tao Chen, Adrian Benton, Sandra C Quinn, and Mark Dredze. Weaponized health communication: Twitter bots and russian trolls amplify the vaccine debate. *American journal of public health*, 108(10):1378–1384, 2018.

[BKF+20] David A Broniatowski, Daniel Kerchner, Fouzia Farooq, Xiaolei Huang, Amelia M Jamison, Mark Dredze, and Sandra Crouse Quinn. The covid-19 social media infodemic reflects uncertainty and state-sponsored propaganda. *arXiv preprint arXiv:2007.09682*, 2020.

[BSHN20] J Scott Brennen, Felix Simon, Philip N Howard, and Rasmus Kleis Nielsen. Types, sources, and claims of covid-19 misinformation. *Reuters Institute*, 7, 2020.

[BTW+20] Juan M Banda, Ramya Tekumalla, Guanyu Wang, Jingyuan Yu, Tuo Liu, Yuning Ding, and Gerardo Chowell. A large-scale covid-19 twitter chatter dataset for open scientific research–an international collaboration. *arXiv preprint arXiv:2004.03688*, 2020.

[Car17] Kathleen M Carley. Ora: A toolkit for dynamic network analysis and visualization., 2017.

[CJHJA17] Man-pui Sally Chan, Christopher R Jones, Kathleen Hall Jamieson, and Dolores Albarracín. Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation. *Psychological science*, 28(11):1531–1546, 2017.

[CL20] Limeng Cui and Dongwon Lee. Coaid: Covid-19 healthcare misinformation dataset. *arXiv preprint arXiv:2006.00885*, 2020.

[CLF20] Emily Chen, Kristina Lerman, and Emilio Ferrara. Tracking social media discourse about the covid-19 pandemic: Development of a public coronavirus twitter data set. *JMIR Public Health and Surveillance*, 6(2):e19273, 2020.

[CRC] L Richard Carley, Jeff Reminga, and Kathleen M Carley. Ora & netmapper.

[DLMZ20] Arkin R Dharawat, Ismini Lourentzou, Alex Morales, and ChengXiang Zhai. Drink bleach or do what now? covid-hera: A dataset for risk-informed health decision making in the presence of covid19 misinformation. 2020.

[DSW20] Enyan Dai, Yiwei Sun, and Suhang Wang. Ginger cannot cure cancer: Battling fake health news with a comprehensive data repository. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14, pages 853–862, 2020.

[Fer20] Emilio Ferrara. What types of covid-19 conspiracies are populated by twitter bots? *First Monday*, 2020.

[GMC+14] Arthur C Graesser, Danielle S McNamara, Zhiqang Cai, Mark Conley, Haiying Li, and James Pennebaker. Coh-metrix measures text characteristics at multiple levels of language and discourse. *The Elementary School Journal*, 115(2):210–229, 2014.

[HC20] Binxuan Huang and Kathleen M Carley. Disinformation and misinformation on twitter during the novel coronavirus outbreak. *arXiv preprint arXiv:2006.04278*, 2020.

[HHSE20] Fatima Haouari, Maram Hasanain, Reem Suwaileh, and Tamer Elsayed. Arcov-19: The first arabic covid-19 twitter dataset with propagation networks. *arXiv preprint arXiv:2004.05861*, 2020.

[HJB+20] Xiaolei Huang, Amelia Jamison, David Broniatowski, Sandra Quinn, and Mark Dredze. Coronavirus twitter data: A collection of covid-19 tweets with automated annotations, 2020.

[LUM$^+$20] Siddique Latif, Muhammad Usman, Sanaullah Manzoor, Waleed Iqbal, Junaid Qadir, Gareth Tyson, Ignacio Castro, Adeel Razi, Maged N Kamel Boulos, Adrian Weller, et al. Leveraging data science to combat covid-19: A comprehensive review. 2020.

[MC20] Shahan Ali Memon and Kathleen M. Carley. Cmu-miscov19: A novel twitter dataset for characterizing covid-19 misinformation, Sep 2020.

[MTMC20] Shahan Ali Memon, Aman Tyagi, David R Mortensen, and Kathleen M Carley. Characterizing sociolinguistic variation in the competing vaccination communities. *arXiv preprint arXiv:2006.04334*, 2020.

[OPR20] Catherine Ordun, Sanjay Purushotham, and Edward Raff. Exploratory analysis of covid-19 tweets using topic modeling, umap, and digraphs. *arXiv preprint arXiv:2005.03082*, 2020.

[PBJB15] James W Pennebaker, Ryan L Boyd, Kayla Jordan, and Kate Blackburn. The development and psychometric properties of liwc2015. Technical report, 2015.

[Pen11] James W Pennebaker. The secret life of pronouns. *New Scientist*, 211(2828):42–45, 2011.

[QIO20] Umair Qazi, Muhammad Imran, and Ferda Ofli. Geocov19: a dataset of hundreds of millions of multilingual covid-19 tweets with location information. *SIGSPATIAL Special*, 12(1):6–15, 2020.

[SAAA20] Junaid Shuja, Eisa Alanazi, Waleed Alasmary, and Abdulaziz Alashaikh. Covid-19 open source data sets: A comprehensive survey. *medRxiv*, 2020.

[SDM20] Gautam Kishore Shahi, Anne Dirkson, and Tim A Majchrzak. An exploratory study of covid-19 misinformation on twitter. *arXiv preprint arXiv:2005.05710*, 2020.

[SN20] Gautam Kishore Shahi and Durgesh Nandini. Fakecovid–a multilingual cross-domain fact check news dataset for covid-19. *arXiv preprint arXiv:2006.11343*, 2020.

[SOC19] Angeline Sangalang, Yotam Ophir, and Joseph N Cappella. The potential for narrative correctives to combat misinformation. *Journal of communication*, 69(3):298–319, 2019.

[SPJ$^+$20] Xingyi Song, Johann Petrak, Ye Jiang, Iknoor Singh, Diana Maynard, and Kalina Bontcheva. Classification aware neural topic model and its application on a new covid-19 disinformation corpus. *arXiv preprint arXiv:2006.03354*, 2020.

[SSM$^+$20] Karishma Sharma, Sungyong Seo, Chuizheng Meng, Sirisha Rambhatla, and Yan Liu. Covid-19 on social media: Analyzing misinformation in twitter conversations. *arXiv preprint arXiv:2003.12309*, 2020.

[TLC15] Andy SL Tan, Chul-joo Lee, and Jiyoung Chae. Exposure to health (mis) information: Lagged effects on young adults' health behaviors and potential pathways. *Journal of Communication*, 65(4):674–698, 2015.

[VCKBC20] Ramon Villa-Cox, Sumeet Kumar, Matthew Babcock, and Kathleen M Carley. Stance in replies and quotes (srq): A new dataset for learning stance in twitter conversations. *arXiv preprint arXiv:2006.00691*, 2020.

[Ves17] Marcela Veselková. Narrative policy framework: Narratives as heuristics in the policy process. *Human Affairs*, 27(2):178, 2017.

[YTLM20] Kai-Cheng Yang, Christopher Torres-Lugo, and Filippo Menczer. Prevalence of low-credibility information on twitter during the covid-19 outbreak. *arXiv preprint arXiv:2004.14484*, 2020.

[ZMFZ20] Xinyi Zhou, Apurva Mulay, Emilio Ferrara, and Reza Zafarani. Recovery: A multimodal repository for covid-19 news credibility research. *arXiv preprint arXiv:2006.05557*, 2020.