Understanding the Use of Images to Spread COVID-19 Misinformation on Twitter

YUPING WANG, Boston University, USA
CHEN LING, Boston University, USA
GIANLUCA STRINGHINI, Boston University, USA

While COVID-19 text misinformation has already been investigated by various scholars, fewer research efforts have been devoted to characterizing and understanding COVID-19 misinformation that is carried out through visuals like photographs and memes. In this paper, we present a mixed-method analysis of image-based COVID-19 misinformation in 2020 on Twitter. We deploy a computational pipeline to identify COVID-19 related tweets, download the images contained in them, and group together visually similar images. We then develop a codebook to characterize COVID-19 misinformation and manually label images as misinformation or not. Finally, we perform a quantitative analysis of tweets containing COVID-19 misinformation images. We identify five types of COVID-19 misinformation, from a wrong understanding of the threat severity of COVID-19 to the promotion of fake cures and conspiracy theories. We also find that tweets containing COVID-19 misinformation images do not receive more interactions than baseline tweets with random images posted by the same set of users. As for temporal properties, COVID-19 misinformation images are shared for longer periods of time than non-misinformation ones, as well as have longer burst times. When looking at the users sharing COVID-19 misinformation images on Twitter from the perspective of their political leanings, we find that pro-Democrat and pro-Republican users share a similar amount of tweets containing misleading or false COVID-19 images. However, the types of images that they share are different: while pro-Democrat users focus on misleading claims about the Trump administration's response to the pandemic, as well as often sharing manipulated images intended as satire, pro-Republican users often promote hydroxychloroquine, an ineffective medicine against COVID-19, as well as conspiracy theories about the origin of the virus. Our analysis sets a basis for better understanding COVID-19 misinformation images on social media and the nuances in effectively moderate them.

1 INTRODUCTION

People who spend time online are constantly bombarded by a deluge of information, consisting not only of text, but also of visuals like images, GIFs, and memes. With limited time, expertise, and investigative means, people usually have to take this information at face value and cannot reliably determine if it is true or not. The COVID-19 pandemic has exacerbated this problem, with a lack of knowledge about the virus allowing misinformation to spread in the early stage of the pandemic [52, 64].

A wealth of research has been conducted in the past two years to better understand the dynamics of COVID-19 related misinformation and its effect on our society and on public health measures [18, 26, 33, 51, 75, 76]. Most of this research has focused on textual content shared on social media; misinformation, however, is not solely composed of text but also of visuals. Images are more immediate than text, and can convey more complex messages than what can be contained in short social media posts (e.g., tweets) [47]. As a result, COVID-19 related image misinformation is particularly dangerous, because it can become viral and severely impact our society, for example by encouraging people not to protect themselves properly or promoting false cures. Despite the dangers posed by image-based COVID-19 misinformation, the research community has spent limited efforts to understand the problem, by either only analyzing images that appeared in news articles and were fact-checked [13] or by focusing on false information spread within a single country [34]. In [44],

Authors' addresses: Yuping Wang, Boston University, Boston, MA, USA, yupingw@bu.edu; Chen Ling, Boston University, Boston, MA, USA, ccling@bu.edu; Gianluca Stringhini, Boston University, Boston, MA, USA, gian@bu.edu.

2023. XXXX-XXXX/2023/3-ART $15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

researchers discuss how COVID-19 case illustrations are used by COVID-19 skeptics to support their own beliefs.

In this paper, we aim to shed light on how images are used to spread COVID-19 misinformation on Twitter. To this end, we collect 2.3M COVID-19 related tweets posted between March 1, 2020 to June 16, 2020. We then download 340K images included in those tweets. To facilitate manual analysis of these messages, we build a computational pipeline based on perceptual hashing techniques and clustering algorithms to group visually similar images together. We then develop a codebook to characterize COVID-19 misinformation images, identify five different types of COVID-19 misinformation images, and build a dataset of over 2.8K COVID-19 misinformation images posted on Twitter. We then perform a quantitative analysis on the tweets that contain COVID-19 misinformation images to answer the following research questions:

- RQ1: Do COVID-19 misinformation images generate more user engagement?
- RQ2: What are the temporal properties of COVID-19 misinformation images? Do COVID-19 misinformation images have a longer lifespan and longer burst times than non-misinformation images?
- RQ3: What are the characteristics of users who post COVID-19 misinformation images?

For RQ1, we compare the reactions (retweets and likes) to tweets containing COVID-19 misinformation images with baseline tweets, as well as baseline tweets containing random images posted by the same set of users. We find that tweets containing COVID-19 misinformation images do not receive significantly more engagement on Twitter. For RQ2, we compare the lifespan of COVID-19 misinformation images on Twitter with that of non-misinformation images, finding that tweets containing COVID-19 misinformation images are shared for longer periods of time, and they also tend to have longer burst times.

For RQ3, we apply a mixed approach to characterize the users who post COVID-19 misinformation images. We find that these users are quite diverse and from all over the world. Additionally, we find that a large portion of the US users in our dataset supports either the Republican Party or Democratic Party, and we find that users who support the Democratic and the Republican parties post a similar amount of tweets with misleading or false COVID-19 images. At a first glance, this is in contrast with previous work. For example, Lazer et al. [42] shows that registered Republicans are far more likely to share COVID-19 misinformation by citing URLs from fake news outlets than registered Democrats during the pandemic. Our analysis does however find that the type of COVID-19 misinformation images shared by supporters of the two parties is different. While pro-Republican users often promote COVID-19 conspiracy theories about the origin of the virus and advocate for the use of hydroxychloroquine to treat COVID-19, pro-Democrat users share false or misleading claims surrounding the response to the pandemic adopted by the Trump administration, as well as manipulated or forged images intended as satire.

Our results shed light on how images are used to spread COVID-19 misinformation on Twitter. Most interestingly RQ1 contradicts what was found by previous research on misinformation, which found that tweets containing false information receive more engagement [80, 82]. A potential reason is that past research followed a top-down approach, only looking for false stories that had been fact-checked, while our approach is bottom-up, identifying groups of misinformation images as they are posted online. We argue that more discussion is needed within the misinformation research community to better understand the advantages and disadvantages of different data approaches, and the biases that these choices might introduce in research results. We release our dataset of labeled COVID-19 image-based misinformation tweets at the following link[1] and we hope that it will spark

[1]https://doi.org/10.5281/zenodo.7581800

more research in this space by the computer-supported cooperative work and social computing community.

2 RELATED WORK

In this section, we first present the definition of misinformation that we follow in this paper. Next, we review previous work that studied text-based misinformation on social media. We then focus on research that looked at image-based misinformation, and finally discuss work that focused on false information spread specifically in the context of COVID-19. For a complete review of misinformation work, readers can refer to [6, 40, 97].

Definition of misinformation. Research on false information typically distinguishes between information that is spread with malicious intent (i.e., *disinformation*) and incorrect claims that are genuinely believed by whoever is posting them (i.e., *misinformation*) [43, 83, 84, 88]. While this distinction is important to understand the goal of people posting false information online and to design appropriate defenses, we argue that it is very challenging to infer the intent with which a piece of false information is posted online. For this reason, in this paper, we adopt the definition of misinformation proposed by Wu et al., which defines misinformation as "informative content that contains incorrect or inaccurate information" [88], regardless of the purpose with which it was posted.

Text misinformation. Various misinformation research projects focus on text misinformation posted on social media. One research direction is to develop automated approaches to detect false information, which typically are based on machine learning and natural language processing techniques [16, 72, 81, 87]. These approaches are however not a silver bullet, as identifying false information is a nuanced problem that is difficult to automate. Bozarth and Budak [10] found that the performance of trained models for false information detection varies significantly when trained and tested on different ground truth datasets, highlighting the challenges in generalizing research results in this space and developing general purpose approaches.

Another research direction is to investigate the propagation of misinformation by using qualitative and quantitative approaches. Vosoughi et al. [80] explored how textual information spreads on Twitter. The authors collected a set of tweets containing links to news articles that were fact-checked by organizations like Snopes and were either debunked or found to be true. Their analysis found that tweets containing links to debunked articles get shared more than those pointing to truthful articles.

Since fact-checking all news articles that appear on social media is a daunting task, and it is unfeasible for fact-checking organizations to cover them all, another line of research considers the trustworthiness of entire news outlets, instead of focusing on single articles [15, 43]. For example, researchers investigated the spread of articles from untrustworthy news outlets during the 2016 US presidential election [15, 30], and both concluded that although untrustworthy news outlets had a large influence on social media users, news articles written by trustworthy news outlets were still more widely shared than those from untrustworthy news outlets [15, 30]. By inspecting narratives around two distinct political themes, authors of [73, 84] showed that untrustworthy news outlets often coordinated with each other when shaping discourses around specific topics.

Zannettou et al. [93] conducted a large-scale measurement analysis investigating state-sponsored troll accounts active on Twitter. They focused on multiple aspects, including the temporal characteristics and content of tweets posted by Twitter trolls. Their results showed that state-sponsored trolls were influential and efficient in spreading URLs, and that trolls sponsored by different countries pushed distinct and often opposite political agendas.

Image-based misinformation. Another line of work investigates how image misinformation spreads over social media. Several approaches focus on developing automated detection methods to identify misinformation images, either manipulated images [1, 9, 95], or images that are taken out of context or

misinterpreted on social media [2, 7, 23, 36, 98]. The effectiveness and adoption of these approaches are impaired by the difficulty in building comprehensive ground truth of misinformation images, which usually have to be performed manually. Researchers have mitigated this problem by relying on images that have been fact-checked by organizations like Snopes [82]. In this paper, we develop what is, to the best of our knowledge, the first annotated dataset of COVID-19 misinformation images shared on Twitter. By making this dataset available to the public, we hope to foster additional research in automatically identifying misinformation images.

Other research used computational approaches to study the spread of image-based misinformation on social media. Previous work showed that images are commonly used on social media to spread misinformation [24, 66], as well as hateful content [38, 91]. Additionally, misinformation images are commonly used in political settings. Previous research found that these images were prevalent in public WhatsApp groups during election campaigns in India and Brazil [28, 67, 68], and that state-sponsored influence campaigns made wide use of images too [57, 92].

Wang et al. [82] analyzed the spread of *Fauxography* images on social media, which are images that are presented in an incorrect or untruthful fashion. They collected fact-checked images from Snopes, and identify 67k instances of those images that were posted on Twitter, Reddit, and 4chan, by looking at their visual similarity. They found that social media posts containing debunked images are more likely to be re-shared or liked than those that contain random images by the same set of users.

Zannettou et al. used visual similarity (i.e., perceptual hashing) and images annotated by the website KnowYourMeme to identify and study image memes posted on social media. As follow up work, Ling et al. [47] performed a mixed-method analysis of popular memes, looking for which indicators contribute to their virality.

In the spirit of this past research, in this paper, we study how images containing misinformation on COVID-19 are shared on Twitter. Unlike previous work, which relied on a top-down approach of looking for images that have been fact-checked or labeled by external organizations, we follow a bottom-up approach, grouping together images that look similar, developing a codebook to characterize image-based misinformation, annotating, and analyzing them.

COVID-19 misinformation. Given the severe impact that misinformation had on the societal response to the COVID-19 pandemic, researchers have been focusing on understanding this type of misinformation as well. To facilitate research in this space, scholars have released several COVID-19 related datasets, including tweets [18] and news articles [96]. Researchers used these datasets to investigate how social media users react to COVID-19 related health misinformation, related for example to the use of masks [89] or fake cures [51, 55]. Other work focused on investigating conspiracy theories related to COVID-19 [4, 56], often highlighting that these conspiracy theories often lead to serious consequences, like a raise in anti-Asian hatred [33, 71, 75].

The aforementioned work focuses on text when studying COVID-19 misinformation, and limited work has been conducted looking at images shared in this context. Javed et al. [35], in the journal extension of [34], studied COVID-19 textual and image misinformation shared in public WhatsApp groups and on Twitter in Pakistan, finding that the spread of this content appears to be organic and not carried out by bots on Twitter. Lee et al. [44] looked at manipulated and misleading data visualizations used to push false narratives surrounding COVID-19. Compared to these previous work, our study is more general as we look at any type of image used to carry out COVID-19 misinformation, and we do not focus on a single country but look at the entirety of Twitter (although we zoom into US-based users for RQ3).

Remarks. To the best of our knowledge, this work is the first that measures the impact of images that contain COVID-19 misinformation shared by global Twitter users, and the temporal characteristics

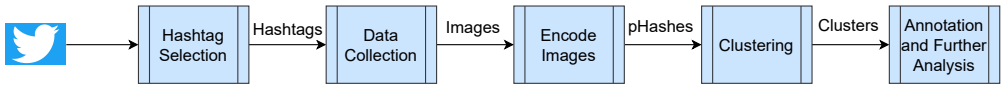


Fig. 1. Overview of our computational analysis pipeline.

of these images. We also analyze the political leanings of US-based users who share these misinformation images. This work is an additional step forward in the dialogue that the CSCW community has been having in the last years on understanding and tackling misinformation [29, 32, 74, 84], including image-based one [47]. Additionally, our dataset will be a useful asset for CSCW researchers wanting to study COVID-19 image-based misinformation, facilitating for example work in content moderation and misinformation detection.

3 METHODS & DATASET

To collect and annotate data for this paper we follow a mixed-method approach. At first, we perform a snowball sampling starting from COVID-19 related hashtags to identify a large number of related tweets from the Twitter Streaming API. We then download all images shared in tweets and apply perceptual hashing and clustering to group together images that look similar. Next, to identify images that contain misinformation we develop a codebook to facilitate the labeling of COVID-19 misinformation images and their categorization. Finally, we analyze the identified COVID-19 misinformation images to answer our research questions. A summary of our research pipeline is shown in Figure 1. In the following, we discuss each phase of our analysis pipeline in detail.

3.1 Dataset construction

As a first step, we need to collect Twitter data related to the COVID-19 pandemic and download the images contained in those tweets. To this end, we follow the same snowball sampling approach conducted by past research on hateful tweets [17]. For our analysis, we leverage data from the public 1% Twitter Streaming API [19, 41, 65].

Hashtag selection. First, we collect all public tweets returned by the API during the month of March 2020. We then aim to identify popular hashtags that are included in COVID-19 related tweets. To this end, we start by extracting all tweets that contain three hashtags: "COVID," "coronavirus," and "COVID19." We then proceed by identifying other hashtags that co-occur with these three, similarly to what was done by [17]. We finally select the 100 most popular co-occurring hashtags; adding these to our initial set, we have 103 total hashtags. The full list of hashtags has been shared anonymously at the following link.[2]

Data collection. By using the Twitter streaming API (which provides a 1% sample of all public tweets), we obtain a total of 505,346,347 tweets between March 1, 2020 and June 16, 2020. Then by using the 103 identified popular COVID-19 related hashtags in the hashtag selection step, we extract all tweets containing any of these hashtags from these 505M tweets. This gives us a total of 2,335,412 COVID-19 related tweets. Of these, 370,465 tweets contain image URLs, of which we are able to successfully download 339,952 images, which are shared by 339,891 tweets in June 2020. The tweets not included in the COVID-19 related dataset will also be used in RQ1 to establish baselines.

Note that we collect our own data instead of using existing datasets like the one compiled by Chen et al. [18] because for our analysis we need to compare tweets with images containing COVID-19 misinformation with baseline tweets posted by the same users sharing COVID-19 misinformation image tweets, to avoid bias generated by considering tweets from users with varying numbers of

[2]https://bit.ly/3XHPkb6

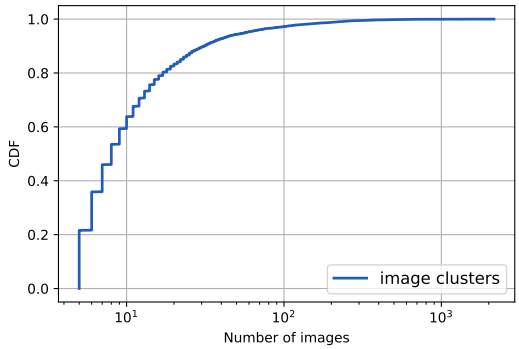


Fig. 2. The cumulative distribution function (CDF) of image cluster sizes.

followers [82, 90]. The baseline obtained from the Twitter Streaming API is more general since it contains tweets related to COVID-19, as well as tweets unrelated to COVID-19, which are missing in existing COVID-19 Twitter datasets [18]. Therefore, we rely on data collected from the Twitter Streaming API instead. Please see Section 4.1 for more details about how we build the baselines.

Ethics. In this work, we only use publicly available data posted on Twitter, and we do not analyze any personally identifiable information or interact with Twitter users in any way. As such, this research is not considered as a human subject by the IRB at Boston University. Nonetheless, we apply standard best practices when dealing with our data. For example, since there are human faces in our dataset, which create privacy concerns, in this paper we blur the face of people portrayed in images unless a) they are public figures, b) the image is a drawing, or c) the image comes from stock photography.

3.2 Grouping visually similar images

Images on social media do not appear in isolation, but rather are often re-shared and modified, becoming memes [24]. Therefore, to analyze how images are used in online discourse, we need techniques to group together not only occurrences of the same identical image, but also of visually similar images which might be variations of the same meme. This also reduces the workload required for human annotators when labeling images as misinformation, as we will discuss in Section 3.3. To identify and group together visually similar images we use the method used by [28] and [91]. We first calculate perceptual hashes (pHashes) for the images in our dataset, and then use clustering to group similar images together. In the rest of this section, we describe these two steps in detail.

Encode images. To find images that are visually similar we apply perceptual hashing to encode them. This technique returns binary vectors (pHashes) of 64 bits that are numerically close if the images are visually similar [53]. This allows us to find not only images that are identical, but also minor variations or meme derivatives of an image. We use the version of perceptual hashing provided by the ImageHash library [14], which previous work showed is robust to image transformations (e.g., slight rotation, skew) [91].

Image clustering. After obtaining the pHash values for all images in our dataset, we apply clustering to group together similar images. To this end, we use the DBSCAN clustering technique [25], which was used for similar purposes in previous work [91]. We learn from the experience of parameter selection from [91], setting the Hamming distance difference threshold for pHash values of two images to be considered visually similar to 6 and the minimum number of elements in a cluster to 5. The detailed process followed to select these thresholds is described in Appendix A. Our implementation of DBSCAN is based on [63].

Fig. 3. Little or indistinguishable visual dissimilarity of two images with distinct pHash values in our dataset.

As we described above, we obtain 339,952 images in total. Of these, 78,348 are obtained from original tweets, 261,604 are from retweets. Note that at this stage of our analysis we do not distinguish between images posted as original tweets and retweets. All images are fed into the DBSCAN clustering algorithm.

After clustering images, we group 148,987 images into 7,773 clusters. The cumulative distribution function (CDF) of the size of these clusters is shown in Figure 2. The median size of these clusters is 8, but there is a long tail of clusters that are much larger, with 10% of the clusters containing 31 or more images. In the subsequent annotation, we will focus on the images contained in clusters only. The reason is that we are interested in understanding how misinformation images are re-shared on Twitter, and if images did not cluster together it is safe to assume that they did not become popular on Twitter. In fact, considering that our data is a uniform 1% sample of all public tweets and that our minimum cluster size is 5, we can assume that images that do not form clusters in our dataset were likely shared publicly on Twitter less than 500 times at the time when the data was collected. In Appendix B we perform an additional analysis showing that tweets containing images that did not get clustered by our approach attracted significantly fewer retweets and likes than those that formed clusters, confirming our intuition. Other work in this area also focused on subsets of all images when performing annotation, by either only annotating the most shared images [68] or by taking a sample of images for annotation [28, 67]. Therefore, we believe that the selection criteria that we use to filter our data is appropriate for our purposes.

Evaluating image similarity within the same clusters. Before continuing with our annotation and analysis, it is of paramount importance to understand whether the clusters produced by our approach are of good quality. In other words, we want to assess whether images that are clustered together are either instances of the same original image or minor variations of it. To this end, we first look at how many clusters only contain identical images. We find that 6,128 out of 7,773 clusters only contain images with identical pHash values, which indicates all the images within the same cluster are visually identical [68, 92].

We then manually inspect the remaining 1,645 clusters that contain images with different pHash values. We find that these clusters fall within three categories:

Fig. 4. Minor variations or meme derivatives of two images with distinct pHash values in our dataset.

- Clusters containing images that although having distinct pHash values, they appear identical to the human eye. This is due to very small differences in the images. One such example is shown in Figure 3.
- Clusters containing images that are derivatives (i.e., minor variations) of other images in the cluster. An example of two images falling in this category is shown in Figure 4. As it can be seen, the meaning of the images is similar (adopt simple precautions to protect yourself from COVID-19 and do not listen to politicians), but while one image is targeted at a US audience, the second one is targeted at a Mexican one.
- Clusters containing images that do not appear visually similar to the human eye, despite having close pHash values. This is due to limitations in the pHash algorithm. For example, images where the background of a certain color dominates might be mistakenly grouped together. We consider these as false positives of our clustering approach.

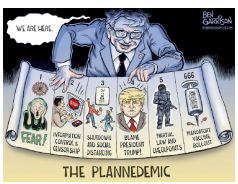
After inspecting all clusters, we find that 105 image clusters contain false positives. This translates into our approach having an accuracy of 98.6%, giving us confidence that our approach is reliable in grouping together similar images. Note that since the false positive clusters appear visually different to a human annotator, they are ignored in the later stages of our analyses to avoid biases in our results. In total, we have 7,668 clusters left, containing 146,192 images.

3.3 Identifying COVID-19 misinformation images

Because whether an image contains misinformation or not often depends on context, it is challenging to automatically label images as misinformation. To overcome this limitation, we manually annotate every image cluster in our dataset with the goal of building a credible ground-truth dataset [5, 21, 61].

In this section, we develop a codebook to guide the thematic annotation process for COVID-19 images on Twitter [11]. As mentioned previously, we use the definition of misinformation proposed by [88], which defines misinformation as "informative content that contains incorrect or inaccurate information." We divide the development of this codebook into two phases based on this definition. First, we use binary labeling to evaluate whether or not images are related to COVID-19. If so, then we call these images "*informative*." As a further step, we characterize the images that contain misinformation.

We use the following three steps to create our codebook and perform annotation: 1) Two researchers separately evaluate our dataset and provide preliminary codes based on thematic coding [11]. 2) We then discuss these preliminary codes and go through multiple rounds, utilizing a subset of the data to create a complete codebook. The procedure is repeated until the codebook reaches a point where



(a) Conspiracy theories on Bill Gates.

(b) Fauxtography of a lion wandering the streets of Russia.

(c) Misinformation on using a malaria drug to treat COVID-19.

(d) Wrong understanding of the threat severity of COVID-19.

(e) Other false claims.

Fig. 5. Types of COVID-19 misinformation images identified by our codebook.

future iterations would not improve it anymore. 3) The same two researchers classify the remainder of our dataset and discuss differences until a satisfactory consensus is obtained.

We next describe our process and our codebook in more detail.

Phase I: Labeling informative images. As previously stated, the initial stage of our annotation process is concerned with selecting informative images, which are images related to COVID-19. We begin by selecting 1,000 clusters at random from our dataset of 7,668 clusters. Two of the authors of this paper review and discuss every image in these clusters to develop a shared understanding of what an informative image looks like. The authors agree on the following criteria for an informative image based on this initial dataset:

- The image has no words or contain words in English or Chinese. We focus on these two languages because these are the two languages spoken by the researchers that annotated the dataset.
- The image must contain sufficient visual cues or words connected to the COVID-19 pandemic, such as RNA virus, public figures during the pandemic, and medical elements such as physicians, hospitals, face masks, and so on.

As long as one image in a cluster is informative, then we label this image cluster as informative. This is reasonable, because as we showed in Section 3.2, the accuracy of our clustering approach is high, and those clusters that did not produce good results were manually removed before proceeding to this annotation. Note the only goal of determining "informative" images is to filter out a smaller set of image candidates for us to manually identify COVID-19 misinformation images.

After the two authors independently label the 1,000 image clusters as either informative or non-informative by checking every image in these image clusters, we calculate the Cohen's Kappa between the annotators and find perfect agreement ($\kappa = 0.991$) [49]. This shows that two annotators

Type	#Cluster	#Images
Conspiracies on COVID-19.	91	1,403
Wrong understanding of the threat severity of COVID-19.	10	294
Fauxtography.	27	455
Wrong medical advice.	41	605
Other false claims.	23	478

Table 1. Overview of cluster numbers and number of images for each of the misinformation types.

strongly agree with each other, verifying the codebook's validity. After establishing that the codebook is mutually agreed by the annotators, the rest of the images in our collection are labeled by the first author by checking every image in these image clusters. Finally, out of 7,668 clusters, we identify 2,316 informative clusters containing 39,367 images.

Note that if the text in a tweet is related to COVID-19 this does not necessarily imply that the images included in this tweet are also related to COVID-19. For example, a tweet discussing the pandemic in the text might include a generic image like a reaction GIF.

Phase II: Characterizing COVID-19 misinformation images. While identifying images as informative is important for comprehending the problem, not all informative images contain misinformation. In fact, a wealth of good information is posted on social media to urge individuals to take responsibility and take measures for halting the pandemic. In this phase, we want to identify the features of COVID-19 misinformation images, by analyzing image themes. We start by having both annotators look over the labeled informative images from Phase I. We follow a loose approach with the objective of getting a basic idea of the themes present in the dataset. Annotators then convene to discuss their findings.

Eventually, the annotators identify five types of misinformation, the specifics of which are presented below.

(1) *Conspiracies on COVID-19.* Conspiracy theories, particularly those involving science, medicine, and health-related issues, are common since long before the COVID-19 pandemic [58]. A large body of research has demonstrated that conspiracy theories can cause individuals to reject information from competent authorities, raising worries about the potential for popular conspiracy theories to diminish people's willingness to follow public health advice [27, 62, 78]. In this work we define conspiracy images as visuals that provide a theory that explains an occurrence or set of circumstances as the product of a hidden scheme by generally strong conspirators [79]. Figure 5(a) shows an example of this type of misinformation.

(2) *Fauxtography.* Previous research indicates that a minority of misinformation is created from scratch [12]. A prominent example of information that is repurposed with the goal of misleading is fauxtography, where a photo is altered or miscaptioned to convey false information [82]. An example of this type of misinformation is shown in Figure 5(b), where a news screenshot presents a lion wandering on the street. The caption says that "Russia unleashed more than 500 lions on its streets to ensure that people are staying indoors during the pandemic outbreak." In reality, the picture is depicting a lion that was roaming the streets of a South African city after being released by a local film company in 2016, and it has nothing to do with COVID-19. [3]

(3) *Wrong medical advice.* Our dataset contains a number of instances of incorrect medical advice, in accordance with previous work studying health misinformation during the COVID-19 pandemic [50]. Examples include not wearing masks or fake cures against COVID-19. According to past research, verified Twitter handles (including organizations/celebrities) are

[3]https://www.snopes.com/fact-check/russia-release-lions-coronavirus/

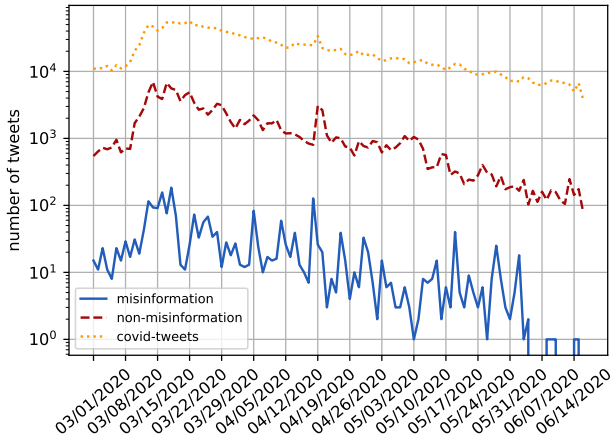


Fig. 6. Number of tweets containing COVID-19 misinformation images, non-misinformation images, and overall COVID-19 tweets in our dataset appearing every day during our observation period.

also active in either generating (new tweets) or distributing misinformation (retweets) [70]. In Figure 5(c), the screenshot of a news broadcast claims "Malaria drug can treat coronavirus." [4]

(4) *Wrong understanding of the threat severity of COVID-19.* There are so-called COVID-19 skeptics [44], who do not believe COVID-19 poses a serious threat [54, 86], putting anyone who believes this message in peril. Examples include claiming COVID-19 is a fabrication or that COVID-19 is a common flu. Figure 5(d) shows an example of this type of misinformation.

(5) *Other false claims.* The text on images may also contain other false claims. One example is the image shown in Figure 5(e), claiming that the Trump family refused to donate money for COVID relief. In reality, former US President Donald Trump donated his salary from the last quarter of 2019 to combat COVID-19.[5]

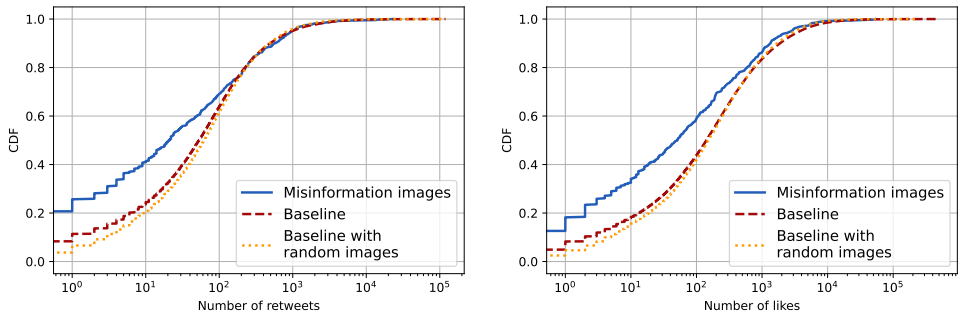
We have limited resources to determine the goal of each cluster of pictures in disseminating misinformation, as well as the accuracy of the misinformation. Therefore, to label images as misinformation we use information gathered from trustworthy fact-checking websites like AP News and WHO Mythbusters.[6]

We label an image as misinformation if it falls into at least one of the five categories. As a result, a single image might belong to two or more categories [34]. All the 2,316 informative clusters generated during the first phase are annotated by two researchers. Similar to Phase I, the two annotators inspect all the images in these image clusters. As long as one image is labeled as misinformation, we label the image cluster as misinformation, and we check every image in the corresponding cluster to see if this image is an image with misinformation. After some deliberation, the two annotators agree on 165 image clusters to be COVID-19 misinformation image clusters, containing 2,418 images

[4]Note that as the medical consensus evolves, so does the common knowledge of what is wrong medical advice. For example, the US Food and Drug Administration (FDA) issued an emergency use authorization (EUA) authorizing the use of hydroxychloroquine to treat certain COVID-19 patients between March 28, 2020 and June 15 (https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-revokes-emergency-use-authorization-chloroquine-and). Since our data collection period partially overlaps with this EUA, the consensus around hydroxychloroquine changed during our study. Nonetheless, following our definition of misinformation, we still consider this type of content as misinformation, as the use of this drug to cure COVID-19 was later debunked.

[5]https://www.cnbc.com/2020/03/03/trump-donates-his-2019-q4-salary-to-help-combat-coronavirus.html

[6]https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters



(a) Retweets (b) Likes

Fig. 7. CDFs of retweets and likes for tweets with COVID-19 misinformation images vs. baseline tweets.

posted by 2,404 users. Further inspection reveals only one image among these images clusters is not a COVID-19 misinformation image, which is also ignored. At last, we have 165 image clusters identified as COVID-19 misinformation image clusters, containing 2,417 images posted by 2,403 users. Table 1 provides an overview of the number of clusters identified for each category and of the number of images in each of the categories.

4 RESULTS

The 165 COVID-19 misinformation image clusters contain 2,417 images included in 2,417 tweets in total while the remaining 7,503 non-misinformation image clusters include 143,774 images shared in 143,755 tweets. Figure 6 shows the time occurrences of tweets containing both types of images in our data. In Figure 6, we also plot the time occurrence of the 2.3M COVID-19 tweets that are selected from the general tweets, as explained in Section 3.1. As it can be seen, the occurrence of three types of tweets increased at the beginning of the observation period and then gradually declined. The highest number of tweets containing non-misinformation images appeared on March 14, 2020 with a total of 6,956 occurrences, while the day with the highest number of tweets containing COVID-19 misinformation images was March 18, 2020 when the total number of occurrences was 184. As for COVID-19 tweets, the highest number of instances appeared on March 22, 2020, with 55,951 occurrences in total. After these dates, the number of tweets gradually decreased.

In the following, we present our analysis to answer our three research questions.

4.1 RQ1: Do COVID-19 misinformation images generate more user engagement?

We use the number of retweets and likes that Twitter posts receive to characterize user engagement. The raw Twitter data collected using the Twitter streaming API contains real-time activity, i.e., the streaming API gathers tweets as soon as they are posted. However, tweets get re-shared and liked over time, and only looking at this early snapshot is not enough to evaluate the engagement collected by tweets. To comprehensively assess long-term engagement, we re-download the tweets in our dataset based on their tweet IDs, the process of which is called *hydration*.[7] This enables us to know the actual number of retweets and likes of a tweet at the time of hydration.

Twitter posts can be classified as original tweets, retweets, and quote tweets. The difference between quote tweets and retweets is that quote tweets contain some extra text from the users who

[7]https://developer.twitter.com/en/docs/twitter-api/v1/tweets/post-and-engage/api-reference/get-statuses-lookup

quote. After hydration, we find that due to limitations in the Twitter API, we cannot retrieve the actual number of retweets and likes of normal retweets [82].[8] For this reason, the assessment of engagement retweets and likes of a normal retweet post is conducted by hydrating the original tweet that produced the retweet post [82].

To investigate whether tweets containing COVID-19 misinformation images receive more engagement, we extract a set of random tweets posted by the same set of users who share the tweets with COVID-19 misinformation images. This is to eliminate potential bias introduced by comparing users with a different number of followers [82, 90]. Our goal is to compare the engagement distribution of baseline tweets to tweets containing COVID-19 misinformation tweets.

To assemble the data for this analysis, we first take the tweet IDs for all tweets in our dataset that contain COVID-19 misinformation images. To avoid duplication potentially introduced by retweets, for those tweets that are retweeted we take the tweet IDs of the original tweets instead. We then deduplicate this set to ensure that each tweet ID is only considered once in this experiment. After this process, we obtain 635 unique tweets containing COVID-19 misinformation images for hydration shared by 565 users. Finally, we hydrate these tweet IDs in April 2021. This is done to ensure that our analysis considers the latest number of retweets and likes. After this process, we obtain 483 unique tweets posted by 429 users, and 152 tweets are not available.

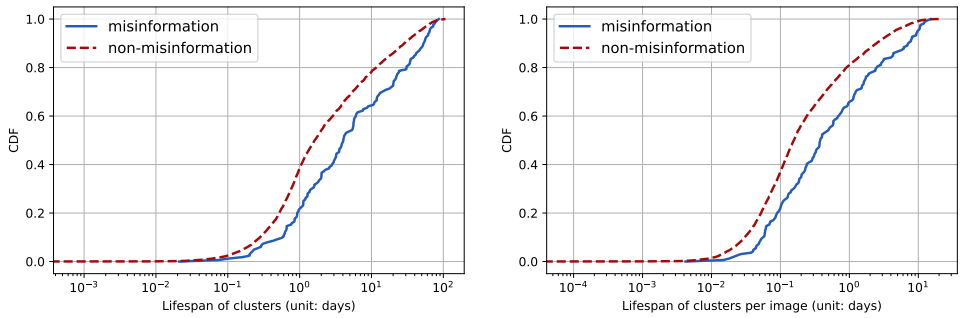
To build the set of baseline tweets for comparison, we obtain all tweets that are contained in the 505M tweets obtained by using Twitter streaming API (See Section 3.1) and are posted by the 429 users who shared COVID-19 misinformation images and follow the same process described above, while in addition removing any tweet that is already considered as part of the COVID-19 image misinformation tweets. The amount of these deduplicated tweets is 63,283. In total, we obtain 59,644 unique baseline tweets after hydration, and 3,639 unique tweets are not available. The hydration is conducted at the same time described above, in April 2021.

The CDFs of the retweets and likes produced by tweets containing COVID-19 misinformation images (labeled as "Misinformation images") and by baseline tweets (labeled as "Baseline") are shown in Figure 7(a) and 7(b), respectively. Our observation is that baseline tweets are more likely to produce more engagement than tweets containing COVID-19 misinformation images: Baseline tweets receive a median of 53 retweets while COVID-19 misinformation images receive a median of 21 retweets. Similarly, baseline tweets receive a median of 142 likes while COVID-19 misinformation images receive a median of 51 likes.

To evaluate the difference between these distributions, we use two-sample Kolmogorov-Smirnov tests (K-S test) [46]. We compare the tweets containing COVID-19 misinformation with baseline tweets, and the results show that the difference between these two categories is statistically significant at the $p < 0.01$ level with $D = 0.181$ and $D = 0.178$ for retweets and likes, respectively. Thus, we reject the null hypothesis that tweets containing COVID-19 misinformation images receive the same level of engagement as baseline tweets.

Previous research showed that tweets containing images are more likely to receive engagements on social media [45, 82]. To reduce this bias, similar to previous work [68, 82], we further add one baseline which is composed of 28,390 tweets that contain random images posted by the same 429 users. This set of images is drawn from all baseline tweets that contain images, and these images do not include COVID-19 misinformation images that we identified by our approach [68, 82, 90]. Again, we plot the CDFs of the retweets and likes produced by the added baseline, which is labeled as "Baseline with random images." We observe that tweets containing COVID-19 misinformation

[8]In the Twitter JSON files, the field "retweet_count" of a retweet is equal to the field "retweet_count" of the corresponding original tweet, and the field "favorite_count" of a retweet, which shows the number of likes the tweet receives is always 0, even if the retweet receives a like.



(a) Raw lifespan of COVID-19 misinformation images vs. non-misinformation images

(b) Normalized lifespan of COVID-19 misinformation images vs. non-misinformation images

Fig. 8. CDF of the lifespan of COVID-19 misinformation images and non-misinformation images in our dataset.

images also tend to produce less engagement than those with other images: Baseline tweets with random images receive a median of 62 retweets while COVID-19 misinformation images receive a median of 21 retweets. Similarly, baseline tweets with random images receive a median of 151 likes while COVID-19 misinformation images receive a median of 51 likes.

We evaluate the difference between the two distributions by using a two-sample K-S test. The comparison is conducted between tweets with COVID-19 misinformation images and tweets with random images, and the results show that the difference between these two categories is statistically significant at the $p < 0.01$ level with $D = 0.216$ and $D = 0.205$ for retweets and likes, respectively.

Note that we do not directly compare tweets that contain COVID-19 misinformation images and tweets that contain non-misinformation images that are included in the 7,503 clusters, because the overlap of users sharing them is low: of the 429 users who share COVID-19 misinformation images, only 130 share a total of 420 non-misinformation images. Instead, we use random images, which include all the images other than the misinformation images shared by the same set of 429 users.

This result shows that the tweets containing COVID-19 misinformation images are not more popular than baseline tweets, as well as baseline tweets with random images.

Takeaways of RQ1. From RQ1, we find that COVID-19 misinformation images do not produce as many engagements as the two baselines. In Section 6 we discuss the implications that this finding has for the field of misinformation studies.

4.2 RQ2: What are the temporal properties of COVID-19 misinformation images? Do COVID-19 misinformation images have a longer lifespan and longer burst times than non-misinformation images?

Another interesting research question beyond the number of likes or retweets that a COVID-19 misinformation image receives is understanding the temporal properties of COVID-19 misinformation images on Twitter. In RQ2 we aim to answer this question. In particular, we investigate the lifespan and burst time of COVID-19 misinformation images compared to those of non-misinformation images, respectively.

We define the time between the first tweet containing an image in a cluster and the last tweet containing an image from the same cluster posted as the *lifespan* of an image. The lifespan comparison

is conducted only between tweets with images to eliminate the effect caused by tweets without images.

Figure 8(a) shows the CDFs of the raw lifespan of COVID-19 misinformation images and non-misinformation images, which corresponds to 165 COVID-19 misinformation image clusters and 7,503 non-misinformation image clusters, respectively. We can see that COVID-19 misinformation images tend to linger on Twitter longer than non-misinformation images: Non-misinformation images have a median raw lifespan of 1.62 days while COVID-19 misinformation images have a median raw lifespan of 4.05 days.

We further use a two-sample K-S test to verify the difference between the two distributions. The result shows that the difference is statistically significant between the two distributions at the $p < 0.01$ level with D=0.202. Therefore, we reject the null hypothesis that COVID-19 misinformation images and non-misinformation images have the same level of lifespan.

Noticing that the size of clusters may influence the lifespan of clusters, we normalize the raw lifespan of clusters by the number of images for each cluster and present the CDFs of the normalized lifespan of COVID-19 misinformation images and non-misinformation images in Figure 8(b). Still, we find that the normalized lifespan of COVID-19 misinformation images is more likely to last longer than non-misinformation images: Non-misinformation images have a median normalized lifespan of 0.16 days while COVID-19 misinformation images have a median normalized lifespan of 0.38 days. Similarly, we use a two-sample K-S test to check the differences between the two distributions. The result shows that the difference is statistically significant between the two distributions at the $p < 0.01$ level with D=0.220. We conclude both the raw and normalized lifespan of COVID-19 misinformation images are longer than that of non-misinformation ones.

We further analyze the burst time of images [68]. We define burst time as the time between two consecutive shares of two images from one cluster, similar to [68]. Figure 9 shows the CDFs of burst times of misinformation images and non-misinformation images. The burst time of misinformation images tends to be longer than that of non-misinformation images: The median burst time of the misinformation image is 0.705 hours while the median burst time of the non-misinformation image is 0.276 hours. Again, we inspect the difference between the two CDFs by using a two-sample K-S test, the result of which indicates that the difference is statistically significant between the two distributions at the $p < 0.01$ level with D=0.165. This allows us to reject the null hypothesis that COVID-19 misinformation images and non-misinformation images have the same level of burstiness.

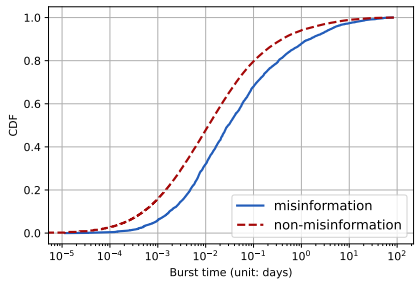


Fig. 9. Burst time of COVID-19 misinformation and non-misinformation images appearing on Twitter.

	Users	Tweets
United States	1,282 (45.8%)	1,354 (46.9 %)
India	261 (9.3%)	263 (9.1%)
United Kingdom	142 (5.1%)	142 (4.9%)
Malaysia	99 (3.5%)	99 (3.4%)
South Africa	90 (3.2%)	91 (3.2%)

Table 2. Top 5 countries, dependencies, and areas of special sovereignty of users sharing COVID-19 misinformation images.

Note that in RQ1, we use COVID-19 misinformation images compared with random images while in RQ2 we use COVID-19 misinformation images compared with non-misinformation images.

However in RQ2, we do not compare COVID-19 misinformation images with random images for temporal properties because as we described in Section 3.1, we only download images that are included in COVID-19 tweets, and we do not download random images in June 2020. When we analyze the temporal properties of COVID-19 misinformation images in 2021, various random images are not available for us to download. The quality of random image clusters obtained in this way is affected by the unavailable images, and it introduces bias. In addition, compared with the size of general tweets obtained by Twitter Streaming API and the size of COVID-19 tweets (505M vs. 2.3M), we may have 200 times more clusters for manual inspection to ensure the quality of these clusters, which makes it infeasible for us to remove clusters that are incorrectly clustered. Therefore, we use the existing non-misinformation image clusters to do the comparison, which takes into account the effect that images may bring, and it is still a good indicator to characterize that COVID-19 misinformation images may have longer lifespans and burst times.

Takeaways of RQ2. As for temporal properties of COVID-19 misinformation images, we find that they tend to have longer lifespans, regardless of raw or normalized, and burst times compared with non-misinformation images. This suggests that COVID-19 misinformation images may linger longer on Twitter and have a longer-term negative effect on Twitter users.

4.3 RQ3: What are the characteristics of users who post COVID-19 misinformation images?

In the first two research questions, we analyzed the characteristics of COVID-19 misinformation images and of the tweets discussing them. In this section, we switch the attention to the users posting COVID-19 misinformation images, looking for their characteristics and for patterns. We first look at geographical information to better understand the geographic makeup of the users in our dataset. Next, we look at the profile description of users, looking for popular hashtags. Finally, we look at the political leanings of the users in the US posting COVID-19 misinformation images on Twitter. From the 165 image clusters that contain COVID-19 misinformation images, we have 2,417 tweets that contain COVID-19 misinformation images, which include 165 original tweets, and 2,252 retweets. The metadata of these retweets contains 475 unique original tweets. After removing the overlapping original tweets, in total, we have 2,887 tweets containing COVID-19 misinformation images posted by 2,801 users.

Analysis of user locations. We look at the users who post COVID-19 misinformation images based on their location (see previous studies like [3, 94]). To do so, the first author of this paper manually checks all the 2,801 users to determine their home locations by using location features and indicators from users' profiles in the collected Twitter user metadata, which include location fields, self-descriptions in bios, and flag icons [3, 94]. This information is at the granularity level of countries, dependencies, and areas of special sovereignty, e.g., Puerto Rico and Hong Kong. [9] If the profiles do not provide enough information, the author then infers the home locations of these users by using the content [31] of their tweets posted in 2020, as well as interaction information (e.g., mentions, retweets, or replies) generated in 2020 between these users and users whose home locations are explicit [3, 8, 37, 48]. If such information is still not enough to infer their location, we classify them as "unknown."

After the manual inspection, we find that the location for 2,656 users is known, where for 1,845 users the location is obtained from their metadata, and the other 811 users are inferred from their tweets. For the remaining 145 users, the location is unknown. The home locations of users who post COVID-19 misinformation images are more than 10 countries, dependencies, and areas of special sovereignty, including English-speaking countries, e.g., US, UK, and Canada, as well as non-English

[9] https://www.state.gov/dependencies-and-areas-of-special-sovereignty/

speaking countries, e.g., China, Indonesia, Mexico. We present the Top 5 countries, dependencies, and areas of special sovereignty that the home locations of users who post COVID-19 misinformation images are in Table 2.

From Table 2, we can see that users from the top 5 countries, dependencies, and areas of special sovereignty account for 66.9% of all users who share tweets with COVID-19 misinformation images in our dataset. Users from the US are the most, and their portion is almost half of all the users. US users also share nearly half of all the tweets that contain COVID-19 misinformation images. We also observe that in the top 5 countries from Table 2, English is a *de facto* official language,[10] which is partly because the hashtags we select are mostly from English and we exclude images that contain text other than Chinese and English from being COVID-19 misinformation image candidates.

Note that location estimation methods described above may not work for the estimation of the city or finer level of location granularity, however, we still argue that this location estimation approach with the full available information can be used as the best guess for the country level of location granularity, i.e., the most likely one, which is useful to help us understand the constitution of user locations.

Analysis of user bios. We start by measuring the most common hashtags used in their user biographies among all the 2,801 users who post COVID-19 misinformation images, as previous work shows that hashtags in bios can reveal user characteristics, for example, their political leaning [20]. Only 324 users in our dataset (11.6%) include hashtags in their bios. The top 20 hashtags in user bios are shown in Table 3. As we can see several hashtags suggest the user's support of Republican presidential candidate Trump, e.g., #MAGA ("Make America Great Again"), #KAG ("Keep America Great"), #Trump2020, #BuildTheWall, etc. as well as hashtags that are commonly associated to conservatism, e.g., #Conservative and #NRA ("National Rifle Association"). We can also find hashtags that are associated with the Democratic party or Anti-Trump movements, e.g., #Biden2020, #Resist [77], and #FBR (often stands for following back resistance). Interestingly, there are also hashtags related to QAnon, which is a US far-right movement promoting various conspiracy theories [59, 60]. Such hashtags include #Q, #QAnon, #WWG1WGA ("When we go one, we go all," which is a popular slogan among QAnon adherents [60]).

This result indicates that many accounts posting COVID-19 misinformation images are focused on US politics. This is in line with the fact that the majority of the users in our dataset are from the US (1,282, as shown in Table 2).

Analysis of US users' political leanings. To better understand their political leanings, we further annotate the US-based users with their political leanings [20, 39].

We determine to manually check the US-based users who post COVID-19 misinformation images based on their profiles in our collected Twitter user metadata and their tweets posted in 2020 to annotate their political leanings [20, 39, 90]. To do so, we have developed a codebook which is shown below:

- *"Pro-Republican."* We use "Pro-Republican" to represent the political leanings of users who identify themselves as Trump supporters or Republican supporters who are not against Trump. The pro-Republican users often use keywords like "MAGA,""KAG," and "Trump2020" to show their support for Trump or keywords like "Republican" and "Conservative" to show their support for the Republican Party in their profiles and tweets.

[10]https://en.wikipedia.org/wiki/List_of_countries_and_territories_where_English_is_an_official_language

Hashtag	Count
#MAGA	82
#KAG	43
#WWG1WGA	38
#Resist	32
#2A	26
#Trump2020	18
#TheResistance	18
#FBR	17
#1A	13
#Patriot	12
#Resistance	12
#Q	10
#TRUMP2020	9
#NRA	9
#QAnon	8
#followbackhongkong	8
#Biden2020	7
#Conservative	7
#resist	7
#BuildTheWall	7

Table 3. Top 20 hashtags in the bios of user profiles

	Users	Tweets
pro-Republicans	508 (39.6%)	538 (39.7 %)
QAnon	111 (8.7%)	117 (8.6%)
pro-Democrats	546 (42.6%)	586 (43.2%)
Neutral	118 (9.2%)	118 (8.7%)
N/A	110 (8.6%)	112 (8.3%)

Table 4. Political leanings of users post misinformation images tweets among users from the United States

- *"QAnon."* Among the pro-Republicans users[11], we code those who support the QAnon movement as "QAnon." QAnon adherents often describe themselves with keywords like "QAnon" and "WWG1WGA" in their profiles and tweets.
- *"Pro-Democrat."* We code users who identify themselves as Democratic supporters or who are against Trump as users whose political leaning are "Pro-Democrat." The pro-Democrat users often use keywords like "Democrat" and "Biden2020" to show their support for the Democratic Party and keywords like "Resist" and "FBR" to show they are against Trump in their profiles and tweets.
- *"Neutral."* If a user supports neither Republicans nor Democrats in their profile and tweets, we classify their political leaning as "neutral."
- *"N/A."* If there is no clue in the profile to show the political leaning of a user, and their profile is suspended or their account is protected, then we classify their political leaning as "N/A."

Following the approach of similar work [20, 39, 90], two authors of this paper randomly select 300 users from the 1,282 US-based users and independently code these users based on their profiles in our collected Twitter user metadata and their tweets posted in 2020 to determine their political leanings. To evaluate the agreement between the two annotators, we need to calculate Cohen's Kappa between the annotation results of the two authors. Since QAnon users are a subset of Pro-Republican users and Cohen's Kappa does not apply to the case in which one item is given two labels, we split the annotation process into two parts:

- In the first part, we only focus on the four categories "Pro-Republican," "Pro-Democrat,""Neutral," and "N/A." Then we calculate the Cohen's Kappa between the results of the two authors and

[11]One of the core beliefs among QAnon adherents is to support Trump. See https://en.wikipedia.org/wiki/QAnon for further explanations.

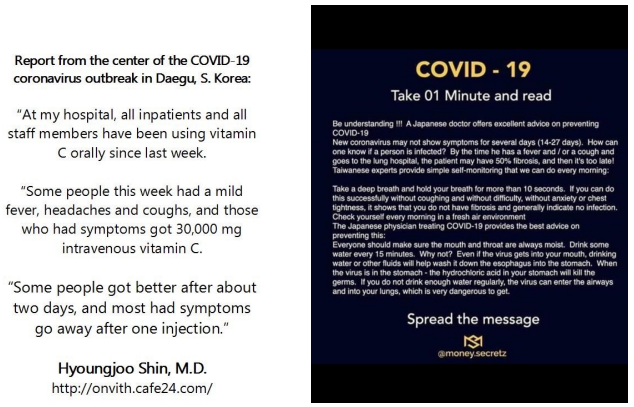


Fig. 10. Images spreading COVID-19 treatment rumors in our dataset.

find a very high agreement ($\kappa = 0.890$), which suggests the two annotators strongly agree with each other, verifying the validity of this codebook about the four categories "Pro-Republican," "Pro-Democrat,""Neutral," and "N/A."

- After that, the two authors discuss and determine 152 users out of these 300 users as Pro-Republican. Then in the second part, the two authors code these 152 users as "QAnon" or not independently. We again calculate the Cohen's Kappa between the QAnon coding results of the two authors and we find the two authors highly agree with each other ($\kappa = 0.941$), confirming the validity of this codebook about the category "QAnon."

After establishing that the codebook is reliable, the remaining users are annotated by the first author. The annotation result is shown in Table 4, where we can see that more than 80% of users who post COVID-19 misinformation images in the US present indicators that they support a political party. Also, we find that the number of tweets with COVID-19 misinformation images shared by pro-Republican and pro-Democrat users is close. In Section 5 we further investigate what type of COVID-19 misinformation images is shared by the supporters of the two political parties, finding that supporters of the two political parties share different types of COVID-19 misinformation images. We also find that although QAnon supporters are not mainstream pro-Republican, they are still a non-negligible part of the users who support the Republican party: QAnon theory adherents are around 21.9% of all users who are pro-Republicans.

Takeaways of RQ3. Overall, we find that users who post COVID-19 misinformation images are from numerous countries and regions, and nearly half of them are from the US. Among the users from the US, we find most of them show explicit political leanings, and the numbers of tweets posted by users who support Republicans and Democrats are close to each other. In Section 5 we further analyze the type of misinformation images shared by users with different political leanings.

5 CASE STUDIES

In this section, we present three case studies to better illustrate the types of COVID-19 misinformation narratives that we observe in our dataset. First, we examine some narratives on false or unconfirmed treatments for COVID-19. We next proceed to look at viral conspiracy theories about Bill Gates, who is blamed for creating the virus. Finally, we show examples of COVID-19 misinformation images shared by pro-Republican and pro-Democrat users, showing that users with different political leanings share different types of COVID-19 misinformation images.

Fig. 11. Images promoting conspiracies on Bill Gates in our dataset.

False and unconfirmed treatments for COVID-19. Throughout the COVID-19 pandemic, several rumors about cheap and accessible treatments against COVID-19 have emerged on the Web. However, these treatments are not effective against the virus, or have never been confirmed by rigorous clinical trials. We see 33 such image clusters in our dataset. The image on the left of Figure 10 encourages people to take vitamin C to treat COVID-19. This treatment was debunked by institutes like the NIH.[12] Another alternative treatment example is shown on the right of Figure 10, claiming that drinking water can prevent people from getting COVID-19. Likewise, this treatment is also ineffective as explained by BBC.[13]

Conspiracies on Bill Gates. Bill Gates has been the target of conspiracy theories since the beginning of the COVID-19 pandemic. According to The New York Times and Zignal Labs, he was named 1.2 million times in the two months leading up to the initial global pandemic in 2020 [22]. These alternative narratives are directed toward Bill Gates himself and the Bill and Melinda Gates Foundation.

Bill Gates conspiracies are aimed at him and his philanthropic work in advancing global health issues. In our dataset, we find 14 clusters related to Bill Gates conspiracy theories. Two examples of such images are shown in Figure 11. In the left image, Bill Gates is accused of testing the COVID-19 vaccine on Indian children without their consent. In the right one, a fauxtography image is miscaptioned to claim to portray a rally demanding the arrest of Bill Gates for crimes against humanity.

Misinformation images shared by pro-Republican and pro-Democrat users. The COVID-19 pandemic has become a polarizing issue in the US and has been at the center of competing narratives during the 2020 General Election. As shown in Section 4.3, in our analysis we found that pro-Democrat and pro-Republican users share a similar amount of tweets with COVID-19 misinformation images. In our dataset, we see pro-Democrat and pro-Republican users involve in sharing 49 and 92 COVID-19 misinformation image clusters, respectively, for a total of 586 tweets shared by pro-Democrat users and 538 tweets shared by pro-Republican users. This is somewhat surprising since previous work [42] shows that the amount of COVID-19 misinformation articles shared by pro-Republican users exceeds tremendously than those shared by pro-Democrat users, and one would expect that this ratio might represent the general trend for COVID-19 misinformation image sharing between Republicans and Democratic supporters on Twitter. However, in our case, we find that with respect to sharing COVID-19 misinformation images, the amounts of tweets shared by

[12] https://ods.od.nih.gov/factsheets/DietarySupplementsInTheTimeOfCOVID19-Consumer/#:~:text=Research%20hasn't%20clearly%20shown,and%20minerals%20to%20work%20properly.

[13] https://www.bbc.com/future/article/20200319-covid-19-will-drinking-water-keep-you-safe-from-coronavirus

(a) (b) (c)

Fig. 12. COVID-19 misinformation images shared by users supporting the Democratic party.

pro-Republican and pro-Democrat users are much closer. In this section we aim to shed light on this finding, analyzing what type of COVID-19 misinformation images is shared by users with different political affiliations.

We find that it is rare for Republican and Democrat users to share the same COVID-19 misinformation images on Twitter: only 17 image clusters are shared by both pro-Democrat and pro-Republican users, and there are 32 out of 49 image clusters and 75 out of 92 image clusters that are only shared by pro-Democrat and pro-Republican users respectively. We present six examples that are shared mutually exclusive between the pro-Democrat and pro-Republican users in Figure 12 and Figure 13.

Among the 32 image clusters shared by pro-Democrat users, the most shared ones are false or misleading claims surrounding the Trump administration's response to the pandemic, where 9 image clusters belong to this category. Figure 12(a) shows an image shared by a pro-Democrat user. The caption criticizes President Trump for refusing to accept WHO-supplied test kits, a claim that has been debunked as not true.[14] The image comes with a tweet stating "This all day" and three hashtags "#COVID19US," "#TrumpRefusedTestKits," and "#TrumpIsTheWORSTPresidentEVER." This indicates a tendency of pro-Democrat users to share false claims for political gain, for example by making the Trump Administration's response to the pandemic appear worse than it actually was. Another common COVID-19 misinformation image type is related to various false claims targeting issues other than Trump, which has 5 clusters. One such image example in Figure 12(b) falsely claims pandemics happen every 100 years, which is also proved to be incorrect.[15]

Additionally, we find several examples of fauxtography images shared by pro-Democrat users. These images are also shared with the goal of criticizing the Trump administration's response to COVID-19 but often have a satirical angle. 7 out of 32 image clusters in this group belong to this category. One example is shown in Figure 12(c), which is a meme showing former President Trump screaming at dead bodies around him to draw attention to his approval rate. This image is a composition of a photo documenting the tragic scene of dead bodies inside a truck posted by a New York nurse[16] and a popular meme showing Trump yelling at a boy mowing the White House lawn.[17]

[14]Please see https://www.factcheck.org/2020/03/biden-trump-wrong-about-who-coronavirus-tests/ for fact-checking
[15]Please see https://www.statesman.com/story/news/politics/elections/2020/04/10/fact-check-has-pandemic-occurred-every-100-years/984128007/ for fact-checking
[16] https://www.dailymail.co.uk/news/article-8167283/Horrifying-moment-dead-bodies-loaded-refrigerated-truck-forklift.html
[17]https://knowyourmeme.com/memes/trump-yelling-at-lawn-mowing-boy

(a) (b) (c)

Fig. 13. COVID-19 misinformation images shared by users supporting the Republican party.

The purpose of this fauxtography is likely to satirize the former president's alleged obsession with approval ratings during a time of great tragedy.

On the other hand, we find 27 out of 75 clusters indicate that pro-Republican users post COVID-19 misinformation images advocating for conspiracy theories and false claims about China or Democrats. For example, Harvard University Professor Charles Lieber was charged for hiding his links with a Chinese University just after the outbreak of the COVID-19 pandemic in 2020,[18] which caused a conspiracy theory that the coronavirus might be a bioweapon of China, as shown in Figure 13(a). The associated text in the tweet reads "#CoronaVirus arrests· · ·," which indicts the user connected the charge with COVID-19. Another example, shown in Figure 13(b), is a conspiracy theory about the origin of COVID-19, claiming that House Speaker Nancy Pelosi took part in releasing the coronavirus to help President Biden get elected. The text associated with this image has the hashtag "#Coronavirus, " and asserts that the exaggeration of COVID-19 is the biggest political fraud in history. Another common narrative shared by pro-Republican users promoted the use of hydroxychloroquine to treat COVID-19,[19] which was later proven to be ineffective, a narrative that was also embraced by former President Trump, making it a controversial political issue, unlike other COVID-19 treatment rumors. An example is an image in Figure 13(c), which falsely claims that hydroxychloroquine is effective against COVID-19 and conjecture that people do not want to use hydroxychloroquine because they are reluctant to admit Trump is correct. The user who posted the image tagged such response as "#TrumpDerangementSyndrome," in the associated Tweet text.

When looking at users who support QAnon, we find the tendency to support conspiracy theories surrounding COVID-19. We find 53 COVID-19 misinformation image clusters posted by QAnon adherent users, and 33 of them promote conspiracies, in line with the actions of the community that were observed by previous work [59, 85]. The image in Figure 14(a) attacks the role of the CDC, claiming that it is a vaccine company. The user posted this image with a passage of text with similar content as the text on the image. Another image example is shown in Figure 14(b) and the associated text they posted reads " the #Covid_19 is Bilderberg after dinner entertainment #Agenda21." Here Bilderberg meeting is a secret annual meeting participated by powerful people around the world[20], and Agenda 21 is a goal set by the United Nations for global development.[21] This suggests this

[18] https://en.wikipedia.org/wiki/Charles_M._Lieber
[19] https://www.who.int/news-room/questions-and-answers/item/coronavirus-disease-(covid-19)-hydroxychloroquine
[20] https://en.wikipedia.org/wiki/Bilderberg_meeting
[21] https://en.wikipedia.org/wiki/Agenda_21

(a) (b)

Fig. 14. COVID-19 misinformation images shared by users who are QAnon adherents.

QAnon adherent user believes that the outbreak of the COVID-19 pandemic is planned by the United Nations and other powerful entities, aiming to construct the so-called "New World Order."

6 DISCUSSION

In this section, we first discuss the implications of our findings for the field of misinformation research. We then reason on what our results mean for platforms aiming to moderate and mitigate misinformation. Finally, we discuss the limitations of our study.

Implications for misinformation research. We present a codebook that aims to characterize the various types of COVID-19 misinformation images. As the pandemic progresses and new narratives emerge, this codebook can be used by other misinformation researchers to better characterize new narratives and how they propagate on social media. Additionally, we publicly release our dataset of tweets containing COVID-19 misinformation images, and we hope that this dataset will be used by social computing and computer vision researchers to develop better tools to automatically identify online misinformation.

Although our dataset is rather small, we make several findings that are of interest to the misinformation research community. We hope that this work will inspire researchers in the CSCW community and beyond to investigate these findings on alternative data, helping to build a better understanding of how misinformation is created and spreads.

Unlike previous works on textual [80] and visual [82] misinformation, we find that image-based COVID-19 misinformation does not receive more engagement than baseline tweets, as well as baseline tweets with random images on Twitter. A possible reason is that those previous works focused on news stories and images that had been analyzed and debunked by fact-checking organizations. While analyzing false news and images that are popular enough to be fact-checked is useful to understand the misinformation phenomenon, this can introduce a selection bias where smaller false narratives that do not meet the threshold for fact-checking are simply ignored.

From the perspective of political leanings, we find that users supporting the Republican and Democratic parties post a similar rate of COVID-19 misinformation images. This is interesting, because it may be different than what was found by previous work [42]. Upon further inspection, we find that the supporters of the two parties share different types of false or misleading images. For instance, we find that Democrats often share misleading and false claims about the Trump administration's response to the pandemic, as well as manipulated images (i.e., fauxtography) used as a satirical commentary to such response. Republican users, on the other hand, prominently

share conspiracy theories about the virus, claiming that it was manufactured by powerful entities (e.g., China, the Democratic Party, United Nations) to achieve various nefarious goals. Republican supporters also often share images promoting the use of hydroxychloroquine to treat COVID-19. These findings show that there are different types of COVID-19 misinformation images that become popular on social media, some more dangerous than others, and that social network companies should take this into account when choosing how to mitigate them [90]. These results also highlight the need to develop approaches able to identify the *intent* with which a misleading image is posted. For instance, posts that are advocating for the adoption of some unproved cure are potentially dangerous to the community and should be considered by platforms for moderation, while posts that satirize such cures are not a threat to public health. At the moment, the research community lacks automated methods to distinguish between the two.

Instead of following a top-down approach dictated by fact-checking websites, in this paper, we follow a bottom-up one, where our data analysis and clustering identify COVID-19 misinformation images that are worth being studied. What we find is that COVID-19 misinformation images are in general not more popular than tweets with random images, but they are shared for longer periods of time compared with non-misinformation images. Going forward, we argue that misinformation researchers should combine the different approaches, relying on solid ground truth provided by fact-checkers, but also relying on real-world data to identify stories that might not meet the threshold to be fact-checked, but nonetheless are discussed on social media for long periods of time.

Implications for platform moderation. This research highlights the types of image-based misinformation that are shared on Twitter during the COVID-19 pandemic. We find a set of challenges and opportunities that can be adopted by platforms like Twitter aiming to curb the misinformation problem.

First, we find that conspiracy narratives that target politicians and wrong medical advice are particularly common on Twitter. These narratives make it challenging for the public to be properly informed, and can hamper public measures like mask mandates and vaccination requirements. In this setting, a promising avenue is using soft moderation measures [90], where misinformation tweets are not taken down but labels providing additional information and resources are provided to the user.

Second, we find that COVID-19 misinformation images on Twitter have a longer lifespan. In particular, the same misinformation image can be used for long periods of time and even resurface to promote new false narratives. For example, images of Bill Gates promoting vaccinations have been popular ground for conspiracy theorists well before the COVID-19 pandemic, and have risen back to popularity as efforts to develop the COVID-19 vaccine mounted. This gives an opportunity for platforms to develop image-based mitigation able to identify emerging misinformation narratives that use new versions of old images.

Our approach can allow platforms to identify and moderate misinformation images that are pervasive but are never shared by popular accounts, and therefore risk remaining under the radar with current approaches. We argue that by looking at images that are shared hundreds of times on Twitter (i.e., form large clusters following our approach) moderators could identify and curb emerging misinformation narratives, potentially before a popular account re-shares that information and makes it visible to even more users.

Finally, we find that memes are often used to promote false narratives on Twitter. This opens up a number of challenges for platforms because the line between satire and harmful information is often blurry and very context-dependent. We argue that Twitter and other social media platforms should invest in technology that can identify the purpose and the context in which an image is used to improve automated misinformation detection systems.

Limitations of our study. In this paper, we provide a comprehensive characterization of COVID-19-related misinformation images. Nonetheless, this work still comes with limitations. First, the set of

hashtags that we use is English only. Since the pandemic is global, this unavoidably missed tweets posted by people speaking other languages. Future work might replicate our analysis of data from other languages. Second, COVID-19 tweets may not use the hashtags that we selected, causing our analysis to miss relevant images. Third, our selection criteria for human annotation is that images need to cluster together and the cluster size needs to be five or higher. We have to do this to keep the analysis manageable for the two annotators and to ensure the quality of clusters. This means that we may miss less popular images; given that our data source is the Twitter 1% sample, we expect these images in smaller clusters to be shared less than 500 times on Twitter at the time when we collect data. Also, since some Twitter users have already been suspended from Twitter and their profiles do not provide enough information, it prevents us to determine the characterization of some users. Our study also inherits the limitations of using the public Twitter Streaming API, which only provides us with a view of 1% of all public tweets. While this dataset is partial, we believe that it still helps us to infer general trends in the share of COVID-19 misinformation images on Twitter during the observation period.

In addition to that, our work annotates images that are grouped into clusters. These images are generally more popular, which makes our results biased toward images that go viral. Future research may find efficient ways to annotate images that are less popular and are likely to be pervasive on social networks. Another issue is that the number of COVID-19 misinformation images identified by our study is small, and therefore further research is needed to investigate if these results generalize to larger datasets and settings. Though the research has the above limitations, we believe our research still helps people better understand what impact the misinformation images have on Twitter, and provides insight on further Twitter moderation.

7 CONCLUSION

In this paper, we build a large dataset of misinformation images related to COVID-19 posted on Twitter between March and June 2020. We develop a codebook to characterize the various types of COVID-19 misinformation images related to the virus, from false medical advice to conspiracy theories. We then use this dataset to understand how COVID-19 misinformation images are used on social media. We find that these images do not receive more retweets and likes than tweets with random images. On the other hand, COVID-19 misinformation images are shared for longer periods of time than non-misinformation images on Twitter. We also find that COVID-19 misinformation images are shared by users who support the Democratic and the Republican party in similar numbers, but there is a difference in the type of images that the two groups share. While Democratic users often share misleading facts on the Trump administration's response to the pandemic, together with manipulated satirical images to critique this response, Republican users often share conspiracy theories about the origin of the virus, as well as images advocating for false treatment of hydroxychloroquine against COVID-19. Our findings help researchers gain a better understanding of image-based misinformation on social media, and identify a number of challenges and opportunities for further research in this space.

Acknowledgments. This work was supported by the National Science Foundation under grants CNS-2114407 and CNS-1942610, by a grant from the Media Ecosystems Analysis Group (MEAG), by the Boston University Hariri Institute for Computing, and by the Boston University Institute for Health System Innovation & Policy. We would also like to thank Savvas Zannettou for his early help with setting up the data analysis infrastructure.

REFERENCES

[1] Sara Abdali, Rutuja Gurav, Siddharth Menon, Daniel Fonseca, Negin Entezari, Neil Shah, and Evangelos E Papalexakis. 2021. Identifying Misinformation from Website Screenshots. In *Proceedings of the International AAAI Conference on*

Web and Social Media, Vol. 15. 2–13.

[2] Sahar Abdelnabi, Rakibul Hasan, and Mario Fritz. 2022. Open-Domain, Content-based, Multi-modal Fact-checking of Out-of-Context Images via Online Resources. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 14940–14949.

[3] Oluwaseun Ajao, Jun Hong, and Weiru Liu. 2015. A survey of location inference techniques on Twitter. *Journal of Information Science* 41, 6 (2015), 855–864.

[4] Syeda Zainab Akbar, Anmol Panda, Divyanshu Kukreti, Azhagu Meena, and Joyojeet Pal. 2021. Misinformation as a Window into Prejudice: COVID-19 and the Information Environment in India. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (2021), 1–28.

[5] Afra Feyza Akyürek, Lei Guo, Randa Elanwar, Prakash Ishwar, Margrit Betke, and Derry Tanti Wijaya. 2020. Multi-label and multilingual news framing analysis. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 8614–8624.

[6] Firoj Alam, Stefano Cresci, Tanmoy Chakraborty, Fabrizio Silvestri, Dimiter Dimitrov, Giovanni Da San Martino, Shaden Shaar, Hamed Firooz, and Preslav Nakov. 2022. A Survey on Multimodal Disinformation Detection. In *Proceedings of the 29th International Conference on Computational Linguistics*. International Committee on Computational Linguistics, 6625–6643.

[7] Shivangi Aneja, Chris Bregler, and Matthias Niener. 2021. COSMOS: Catching Out-of-Context Misinformation with Self-Supervised Learning. arXiv:2101.06278 [cs.CV]

[8] Lars Backstrom, Eric Sun, and Cameron Marlow. 2010. Find me if you can: improving geographical prediction with social and spatial proximity. In *Proceedings of the 19th international conference on World wide web*. 61–70.

[9] Belhassen Bayar and Matthew C Stamm. 2016. A deep learning approach to universal image manipulation detection using a new convolutional layer. In *Proceedings of the 4th ACM workshop on information hiding and multimedia security*. 5–10.

[10] Lia Bozarth and Ceren Budak. 2020. Toward a better performance evaluation framework for fake news classification. In *Proceedings of the international AAAI conference on web and social media*, Vol. 14. 60–71.

[11] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[12] J Scott Brennen, Felix Simon, Philip N Howard, and Rasmus Kleis Nielsen. 2020. Types, sources, and claims of COVID-19 misinformation. *Reuters Institute* 7, 3 (2020), 1.

[13] J Scott Brennen, Felix M Simon, and Rasmus Kleis Nielsen. 2021. Beyond (mis) representation: visuals in COVID-19 misinformation. *The International Journal of Press/Politics* 26, 1 (2021), 277–299.

[14] Johannes Buchner. 2020. A Python Perceptual Image Hashing Module: ImageHash. https://github.com/JohannesBuchner/imagehash.

[15] Ceren Budak. 2019. What happened? the spread of fake news publisher content during the 2016 us presidential election. In *The World Wide Web Conference*. 139–150.

[16] Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. 2011. Information credibility on twitter. In *Proceedings of the 20th international conference on World wide web*. 675–684.

[17] Despoina Chatzakou, Nicolas Kourtellis, Jeremy Blackburn, Emiliano De Cristofaro, Gianluca Stringhini, and Athena Vakali. 2017. Mean birds: Detecting aggression and bullying on twitter. In *Proceedings of the 2017 ACM on web science conference*. 13–22.

[18] Emily Chen, Kristina Lerman, and Emilio Ferrara. 2020. Tracking Social Media Discourse About the COVID-19 Pandemic: Development of a Public Coronavirus Twitter Data Set. *JMIR Public Health and Surveillance* 6, 2 (2020), e19273.

[19] Kaiping Chen, Zening Duan, and Sijia Yang. 2022. Twitter as research data: Tools, costs, skill sets, and lessons learned. *Politics and the Life Sciences* 41, 1 (2022), 114–130.

[20] Michael D Conover, Bruno Gonçalves, Jacob Ratkiewicz, Alessandro Flammini, and Filippo Menczer. 2011. Predicting the political alignment of twitter users. In *2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing*. IEEE, 192–199.

[21] Keith Cortis and Brian Davis. 2021. A Dataset of Multidimensional and Multilingual Social Opinions for Maltas Annual Government Budget. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 15. 971–981.

[22] Daisuke Wakabayashi, Davey Alba, and Marc Tracy. 2020. Bill Gates, at Odds With Trump on Virus, Becomes a Right-Wing Target. https://www.nytimes.com/2020/04/17/technology/bill-gates-virus-conspiracy-theories.html.

[23] Prateek Dewan, Anshuman Suri, Varun Bharadhwaj, Aditi Mithal, and Ponnurangam Kumaraguru. 2017. Towards understanding crisis events on online social networks through pictures. In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. 439–446.

[24] Yuhao Du, Muhammad Aamir Masood, and Kenneth Joseph. 2020. Understanding visual memes: An empirical analysis of text superimposed on memes shared on twitter. In *Proceedings of the International AAAI Conference on Web and*

Social Media, Vol. 14. 153–164.
[25] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise.. In *KDD*, Vol. 96. 226–231.
[26] Emilio Ferrara. 2020. What types of COVID-19 conspiracies are populated by Twitter bots? *First Monday* 25 (2020).
[27] Daniel Freeman, Felicity Waite, Laina Rosebrock, Ariane Petit, Chiara Causier, Anna East, Lucy Jenner, Ashley-Louise Teale, Lydia Carr, Sophie Mulhall, et al. 2020. Coronavirus conspiracy beliefs, mistrust, and compliance with government guidelines in England. *Psychological medicine* (2020), 1–13.
[28] Kiran Garimella and Dean Eckles. 2020. Images and misinformation in political groups: Evidence from WhatsApp in India. *Harvard Kennedy School Misinformation Review* 1 (2020).
[29] Amira Ghenai and Yelena Mejova. 2018. Fake cures: user-centric modeling of health misinformation in social media. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–20.
[30] Nir Grinberg, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer. 2019. Fake news on Twitter during the 2016 US presidential election. *Science* 363, 6425 (2019).
[31] Bo Han, Paul Cook, and Timothy Baldwin. 2014. Text-based twitter user geolocation prediction. *Journal of Artificial Intelligence Research* 49 (2014), 451–500.
[32] Md Mahfuzul Haque, Mohammad Yousuf, Ahmed Shatil Alam, Pratyasha Saha, Syed Ishtiaque Ahmed, and Naeemul Hassan. 2020. Combating Misinformation in Bangladesh: roles and responsibilities as perceived by journalists, fact-checkers, and users. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020), 1–32.
[33] Bing He, Caleb Ziems, Sandeep Soni, Naren Ramakrishnan, Diyi Yang, and Srijan Kumar. 2021. Racism is a virus: Anti-Asian hate and counterspeech in social media during the COVID-19 crisis. In *Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. 90–94.
[34] R Tallal Javed, Mirza Elaaf Shuja, Muhammad Usama, Junaid Qadir, Waleed Iqbal, Gareth Tyson, Ignacio Castro, and Kiran Garimella. 2020. A First Look at COVID-19 Messages on WhatsApp in Pakistan. In *2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. 118–125.
[35] R Tallal Javed, Muhammad Usama, Waleed Iqbal, Junaid Qadir, Gareth Tyson, Ignacio Castro, and Kiran Garimella. 2022. A deep dive into COVID-19-related messages on WhatsApp in Pakistan. *Social Network Analysis and Mining* 12, 1 (2022), 1–16.
[36] Zhiwei Jin, Juan Cao, Yongdong Zhang, Jianshe Zhou, and Qi Tian. 2016. Novel visual and statistical image features for microblogs news verification. *IEEE transactions on multimedia* 19, 3 (2016), 598–608.
[37] David Jurgens. 2013. That's what friends are for: Inferring location in online social media platforms based on social relationships. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 7. 273–282.
[38] Douwe Kiela, Hamed Firooz, Aravind Mohan, Vedanuj Goswami, Amanpreet Singh, Casey A Fitzpatrick, Peter Bull, Greg Lipstein, Tony Nelli, Ron Zhu, et al. 2021. The hateful memes challenge: competition report. In *NeurIPS 2020 Competition and Demonstration Track*. PMLR, 344–360.
[39] Juhi Kulshrestha, Motahhare Eslami, Johnnatan Messias, Muhammad Bilal Zafar, Saptarshi Ghosh, Krishna P Gummadi, and Karrie Karahalios. 2019. Search bias quantification: investigating political bias in social media and web search. *Information Retrieval Journal* 22, 1 (2019), 188–227.
[40] Srijan Kumar and Neil Shah. 2018. False information on web and social media: A survey. In *arXiv:1804.08559*.
[41] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. 2010. What is Twitter, a social network or a news media?. In *Proceedings of the 19th international conference on World wide web*. 591–600.
[42] David Lazer, Damian J Ruck, Alexi Quintana, Sarah Shugars, Kenneth Joseph, Nir Grinberg, Ryan J Gallagher, Luke Horgan, Adina Gitomer, Aleszu Bajak, et al. 2021. The COVID States Project# 18: Fake News on Twitter. (2021).
[43] David MJ Lazer, Matthew A Baum, Yochai Benkler, Adam J Berinsky, Kelly M Greenhill, Filippo Menczer, Miriam J Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, et al. 2018. The science of fake news. *Science* 359, 6380 (2018).
[44] Crystal Lee, Tanya Yang, Gabrielle D Inchoco, Graham M Jones, and Arvind Satyanarayan. 2021. Viral Visualizations: How Coronavirus Skeptics Use Orthodox Data Practices to Promote Unorthodox Science Online. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–18.
[45] Yiyi Li and Ying Xie. 2020. Is a picture worth a thousand words? An empirical study of image content and social media engagement. *Journal of Marketing Research* 57, 1 (2020), 1–19.
[46] Bernard Lindgren. 1993. *Statistical Theory*. Vol. 22.
[47] Chen Ling, Ihab AbuHilal, Jeremy Blackburn, Emiliano De Cristofaro, Savvas Zannettou, and Gianluca Stringhini. 2021. Dissecting the Meme Magic: Understanding Indicators of Virality in Image Memes. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–24.
[48] Jeffrey McGee, James Caverlee, and Zhiyuan Cheng. 2013. Location prediction in social media based on tie strength. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*. 459–468.
[49] Mary L McHugh. 2012. Interrater reliability: the kappa statistic. *Biochemia medica* 22, 3 (2012), 276–282.

[50] Shahan Ali Memon and Kathleen M Carley. 2020. Characterizing covid-19 misinformation communities using a novel twitter dataset. *arXiv preprint arXiv:2008.00791* (2020).
[51] Nicholas Micallef, Bing He, Srijan Kumar, Mustaque Ahamad, and Nasir Memon. 2020. The role of the crowd in countering misinformation: A case study of the COVID-19 infodemic. In *2020 IEEE International Conference on Big Data (Big Data)*. 748–757.
[52] Joanne M Miller. 2020. Do COVID-19 conspiracy theory beliefs form a monological belief system? *Canadian Journal of Political Science/Revue canadienne de science politique* 53, 2 (2020), 319–326.
[53] Vishal Monga and Brian L Evans. 2006. Perceptual image hashing via feature points: performance evaluation and tradeoffs. *IEEE transactions on Image Processing* 15, 11 (2006).
[54] Matt Motta, Dominik Stecula, and Christina Farhart. 2020. How right-leaning media coverage of COVID-19 facilitated the spread of misinformation in the early stages of the pandemic in the US. *Canadian Journal of Political Science/Revue canadienne de science politique* 53, 2 (2020), 335–342.
[55] Ece C Mutlu, Toktam Oghaz, Jasser Jasser, Ege Tutunculer, Amirarsalan Rajabi, Aida Tayebi, Ozlem Ozmen, and Ivan Garibay. 2020. A stance data set on polarized conversations on Twitter about the efficacy of hydroxychloroquine as a treatment for COVID-19. *Data in brief* 33 (2020), 106401.
[56] Lynnette Hui Xian Ng and Kathleen M Carley. 2021. The coronavirus is a bioweapon: classifying coronavirus stories on fact-checking sites. *Computational and Mathematical Organization Theory* 27, 2 (2021), 179–194.
[57] Lynnette Hui Xian Ng, JD Moffitt, and Kathleen M Carley. 2022. Coordinated through a Web of Images: Analysis of Image-based Influence Operations from China, Iran, Russia, and Venezuela. *arXiv preprint arXiv:2206.03576* (2022).
[58] J Eric Oliver and Thomas Wood. 2014. Medical conspiracy theories and health behaviors in the United States. *JAMA internal medicine* 174, 5 (2014), 817–818.
[59] Antonis Papasavva, Max Aliapoulios, Cameron Ballard, Emiliano De Cristofaro, Gianluca Stringhini, Savvas Zannettou, and Jeremy Blackburn. 2022. The gospel according to Q: Understanding the QAnon conspiracy from the perspective of canonical information. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 16. 735–746.
[60] Antonis Papasavva, Jeremy Blackburn, Gianluca Stringhini, Savvas Zannettou, and Emiliano De Cristofaro. 2021. Is it a Qoincidence?: An Exploratory Study of QAnon on Voat. In *Proceedings of the Web Conference 2021*. 460–471.
[61] Chan Young Park, Xinru Yan, Anjalie Field, and Yulia Tsvetkov. 2021. Multilingual contextual affective analysis of LGBT people portrayals in Wikipedia. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 15. 479–490.
[62] Irena Pavela Banai, Benjamin Banai, and Igor Mikloušić. 2022. Beliefs in COVID-19 conspiracy theories, compliance with the preventive measures, and trust in government medical officials. *Current Psychology* 41, 10 (2022), 7448–7458.
[63] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011).
[64] Gordon Pennycook and David G Rand. 2021. The psychology of fake news. *Trends in cognitive sciences* (2021).
[65] Jürgen Pfeffer, Katja Mayer, and Fred Morstatter. 2018. Tampering with Twitters sample API. *EPJ Data Science* 7, 1 (2018), 50.
[66] Jingnong Qu, Liunian Harold Li, Jieyu Zhao, Sunipa Dev, and Kai-Wei Chang. 2022. DisinfoMeme: A Multimodal Dataset for Detecting Meme Intentionally Spreading Out Disinformation. *arXiv preprint arXiv:2205.12617* (2022).
[67] Julio CS Reis, Philipe Melo, Kiran Garimella, Jussara M Almeida, Dean Eckles, and Fabrício Benevenuto. 2020. A dataset of fact-checked images shared on whatsapp during the brazilian and indian elections. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 14. 903–908.
[68] Gustavo Resende, Philipe Melo, Hugo Sousa, Johnnatan Messias, Marisa Vasconcelos, Jussara Almeida, and Fabrício Benevenuto. 2019. (Mis) information dissemination in WhatsApp: Gathering, analyzing and countermeasures. In *The World Wide Web Conference*. 818–828.
[69] Erich Schubert, Jörg Sander, Martin Ester, Hans Peter Kriegel, and Xiaowei Xu. 2017. DBSCAN revisited, revisited: why and how you should (still) use DBSCAN. *ACM Transactions on Database Systems (TODS)* 42, 3 (2017), 1–21.
[70] Gautam Kishore Shahi, Anne Dirkson, and Tim A Majchrzak. 2021. An exploratory study of covid-19 misinformation on twitter. *Online social networks and media* 22 (2021), 100104.
[71] Xinyue Shen, Xinlei He, Michael Backes, Jeremy Blackburn, Savvas Zannettou, and Yang Zhang. 2022. On Xing Tian and the Perseverance of Anti-China Sentiment Online. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 16. 944–955.
[72] 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).
[73] Kate Starbird. 2017. Examining the alternative media ecosystem through the production of alternative narratives of mass shooting events on Twitter. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 11.

230–239.

[74] Kate Starbird, Ahmer Arif, and Tom Wilson. 2019. Disinformation as collaborative work: Surfacing the participatory nature of strategic information operations. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–26.

[75] Fatemeh Tahmasbi, Leonard Schild, Chen Ling, Jeremy Blackburn, Gianluca Stringhini, Yang Zhang, and Savvas Zannettou. 2021. Go eat a bat, Chang!: On the Emergence of Sinophobic Behavior on Web Communities in the Face of COVID-19. In *Proceedings of the Web Conference 2021*. 1122–1133.

[76] Samia Tasnim, Md Mahbub Hossain, and Hoimonty Mazumder. 2020. Impact of rumors and misinformation on COVID-19 in social media. *Journal of preventive medicine and public health* 53, 3 (2020), 171–174.

[77] New York Times. 2017. "Resist" Is a Battle Cry, but What Does It Mean? https://www.nytimes.com/2017/02/14/us/politics/resist-anti-trump-protest.html.

[78] Joseph E Uscinski, Adam M Enders, Casey Klofstad, Michelle Seelig, John Funchion, Caleb Everett, Stefan Wuchty, Kamal Premaratne, and Manohar Murthi. 2020. Why do people believe COVID-19 conspiracy theories? *Harvard Kennedy School Misinformation Review* 1 (2020).

[79] Jan-Willem van Prooijen and Karen M Douglas. 2018. Belief in conspiracy theories: Basic principles of an emerging research domain. *European journal of social psychology* 48, 7 (2018), 897–908.

[80] Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *Science* 359, 6380 (2018).

[81] William Yang Wang. 2017. "Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. 422–426.

[82] Yuping Wang, Fatemeh Tahmasbi, Jeremy Blackburn, Barry Bradlyn, Emiliano De Cristofaro, David Magerman, Savvas Zannettou, and Gianluca Stringhini. 2021. Understanding the Use of Fauxtography on Social Media. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 15. 776–786.

[83] Tom Wilson and Kate Starbird. 2020. Cross-platform disinformation campaigns: lessons learned and next steps. *Harvard Kennedy School Misinformation Review* 1 (2020).

[84] Tom Wilson, Kaitlyn Zhou, and Kate Starbird. 2018. Assembling strategic narratives: Information operations as collaborative work within an online community. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–26.

[85] Julia Carrie Wong. 2018. What is QAnon? Explaining the bizarre rightwing conspiracy theory. *The Guardian* (2018).

[86] Meaghan Wray. 2020. Corona challenge: TikTok star films herself licking airplane toilet seat. https://globalnews.ca/news/6718358/tiktok-toilet-seat-lick-coronavirus/.

[87] Liang Wu and Huan Liu. 2018. Tracing fake-news footprints: Characterizing social media messages by how they propagate. In *Proceedings of the eleventh ACM international conference on Web Search and Data Mining*. 637–645.

[88] Liang Wu, Fred Morstatter, Kathleen M Carley, and Huan Liu. 2019. Misinformation in social media: definition, manipulation, and detection. *ACM SIGKDD Explorations Newsletter* 21, 2 (2019), 80–90.

[89] Neil Yeung, Jonathan Lai, and Jiebo Luo. 2020. Face off: Polarized public opinions on personal face mask usage during the COVID-19 pandemic. In *2020 IEEE International Conference on Big Data (Big Data)*. 4802–4810.

[90] Savvas Zannettou. 2021. " I Won the Election!": An Empirical Analysis of Soft Moderation Interventions on Twitter. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 15. 865–876.

[91] Savvas Zannettou, Tristan Caulfield, Jeremy Blackburn, Emiliano De Cristofaro, Michael Sirivianos, Gianluca Stringhini, and Guillermo Suarez-Tangil. 2018. On the Origins of Memes by Means of Fringe Web Communities. In *Proceedings of the Internet Measurement Conference 2018*. 188–202.

[92] Savvas Zannettou, Tristan Caulfield, Barry Bradlyn, Emiliano De Cristofaro, Gianluca Stringhini, and Jeremy Blackburn. 2020. Characterizing the Use of Images in State-Sponsored Information Warfare Operations by Russian Trolls on Twitter. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 14. 774–785.

[93] Savvas Zannettou, Tristan Caulfield, William Setzer, Michael Sirivianos, Gianluca Stringhini, and Jeremy Blackburn. 2019. Who Let The Trolls Out?: Towards Understanding State-Sponsored Trolls. In *Proceedings of the 10th ACM Conference on Web Science*. 353–362.

[94] Xin Zheng, Jialong Han, and Aixin Sun. 2018. A survey of location prediction on twitter. *IEEE Transactions on Knowledge and Data Engineering* 30, 9 (2018), 1652–1671.

[95] Peng Zhou, Xintong Han, Vlad I Morariu, and Larry S Davis. 2018. Learning rich features for image manipulation detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 1053–1061.

[96] Xinyi Zhou, Apurva Mulay, Emilio Ferrara, and Reza Zafarani. 2020. ReCOVery: A Multimodal Repository for COVID-19 News Credibility Research. In *Proceedings of the 29th ACM International Conference on Information and Knowledge Management*. 3205–3212.

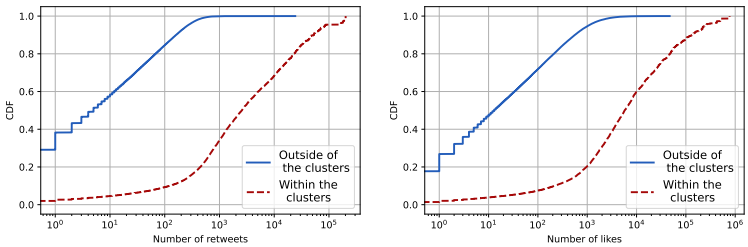
[97] Xinyi Zhou and Reza Zafarani. 2020. A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. *ACM Computing Surveys* 53 (2020), 1–40.

Distance	#Cluster	#Images clustered	%Noise
2	7,590	144,398	57.5%
4	7,674	146,612	56.5%
6	7,773	148,987	56.2%
8	7,854	152,368	55.2%
10	7,827	161,522	52.5%

Table 5. Overview of cluster parameter performance I.

Distance	%Correctly grouped clusters
2	99.5%
4	99.5%
6	99.5%
8	97%
10	86%

Table 6. Overview of cluster parameter performance II.



(a) Retweets (b) Likes

Fig. 15. Engagement of tweets that contain images within the clusters and tweets that contain images outside of the clusters, respectively.

[98] Dimitrina Zlatkova, Preslav Nakov, and Ivan Koychev. 2019. Fact-Checking Meets Fauxtography: Verifying Claims About Images. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. 2099–2108.

A PARAMETER SELECTION FOR CLUSTERING

There are two parameters in DBSCAN algorithm to be determined: maximum distance ε and the minimum size of a cluster minPts [69]. The maximum distance ε determines the maximum distance (with an arbitrary distance measure) between two points that are considered neighbors for each other in one cluster. The minimum size of a cluster minPts specifies the minimum number of elements in a cluster. The members in Group with fewer elements than minPts are treated as noise points.

In practice, previous work applies a heuristic method to determine the parameter minPts and the minPts are set to default value 4 for two-dimensional data (See Section 4.2 of paper [25] and Section 4.1 of paper [69]). In Section 4.1 of paper [69], it is suggested that for datasets with high dimensions, it could improve results by increasing minPts. Since the vectors in our dataset are 64-bit binary vectors (i.e., vectors with sixty-four dimensions), and we determine the distance between two vectors by using Hamming distance (i.e., the number of bits that are different between two vectors), and we refer to the parameters used in [91], where minPts are set heuristically as 5, we decide to set the minPts as 5. Then we turn to determine the parameter ε.

Following a similar parameter selection method described in appendix A of paper [91], we find that when clustering by varying the parameter ε, the percentage of noise does not change a lot. The result is shown in Table 5.

Next, we randomly select 200 clusters for each threshold and manually check the clusters. We find that among candidates for distance 2, 4, and 6, 199 clusters are totally correctly clustered among 200 clusters for each threshold, while for distance 8 and 10, 194 and 172 clusters are totally correctly clustered, respectively. The proportion of correctly grouped clusters for each threshold is shown in Table 6.

First, between the thresholds 8 and 10, we select 8 because it has a higher percentage of correctly grouped clusters. Then among thresholds 2, 4, and 6, we select 6 because threshold 6 has a lower percentage of noise while maintaining the same percentage of correctly grouped clusters. Finally, Between the final two threshold candidates 6 and 8, we select 6 because it has a higher percentage of correctly grouped clusters while the percentage of noise differ little between the two thresholds.

B EVALUATE POPULARITY OF IMAGES IN THE CLUSTERS VS IMAGES NOT IN THE CLUSTERS

In total, we have 339,891 tweets that contain images within 2.3M COVID-19-related tweets. As described in Section 4.1, we hydrate the original tweets, quoted tweets, and the original tweet part of the retweets in October 2020. In total, we have 1,424,481 unique tweets for hydration and after hydration, we obtain 1,272,929 unique tweets, among which we have 122,679 unique tweets containing images that are grouped into clusters, while we obtain 165,859 unique tweets containing images that are outside of the clusters. Note that compared with the hydration conducted in April 2021, which is used for investigating RQ1, the hydration conducted in October 2020 does not take general tweets into consideration. Therefore, we cannot use the hydrated tweets in October 2020 to answer RQ1.

We plot the CDFs of the two types of images in Figure 15. The median for the retweets of tweets that contain images within the clusters and tweets that contain images outside of the clusters are 2,576 and 5, respectively while the median for the likes of tweets that contain images within the clusters and tweets that contain images outside of the clusters are 5,733 and 13, respectively. The further two-sample K-S tests show that differences between these two categories are statistically significant at the $p < 0.01$ level with $D = 0.75$ and $D = 0.81$ for retweets and likes, respectively. These results reject the null hypothesis that tweets containing images within the clusters receive the same level of engagement as tweets containing images outside of the clusters. In the cases of both likes and retweets, tweets that contain images within the clusters tend to have more engagement than tweets that contain images outside of the clusters. We conclude that tweets containing images within the clusters are more likely to generate more engagement than tweets containing images outside of the clusters.