Social Network Analysis and Mining (2022) 12:45
https://doi.org/10.1007/s13278-022-00873-0

ORIGINAL ARTICLE

Ridge count thresholding to uncover coordinated networks during onset of the Covid-19 pandemic

Spencer Lee Kirn[1] · **Mark K. Hinders**[1]

Received: 17 October 2021 / Revised: 1 January 2022 / Accepted: 28 February 2022

Abstract
In order to combat information operations (IO) and disinformation campaigns, one must look at the behaviors of the accounts pushing specific narratives and stories through social media, not at the content itself. In this work, we present a new process for extracting tweet storms and uncovering networks of accounts that are working in a coordinated fashion using ridge count thresholding (RCT). To do this, we started with a dataset of 60 million individual tweets from the early weeks of the Covid-19 pandemic. Coherent topics are extracted from this data by testing three different preprocessing pipelines and applying Orthogonal Nonnegative Matrix Factorization (ONMF). The most effective preprocessing pipeline used hashtag preclustering to downselect the total dataset to the 7 million tweets that included the top hashtags. Each topic identified by ONMF is described by a topic-tweet signal, crafted using the time stamp included in each tweet's metadata. These signals were broken down into tweet storms using RCT, which is calculated from the Dynamic Wavelet Fingerprint transform of each topic-tweet signal. Each tweet storm described a time of increased activity around a topic. Tweet storms identified in this way each represent some behavior in the underlying network. In total, we identified 39,817 total tweet storms that included about 2 million unique tweets. These tweet storms were used to identify networks of accounts that commonly co-occur within tweet storms to isolate those communities most responsible for driving narratives and pushing stories through social media. Through this process, we were able to identify 22 unique networks of accounts that were densely connected based on RCT tweet storm identification. Many of the identified networks exhibit obvious inauthentic behaviors that are potentially a part of an IO campaign.

Keywords Twitter · Information operations · Covid-19 · Topic modeling · Machine learning · Network analysis

1 Introduction

Disinformation is not new to the social media era. For example, Thomas Jefferson was quite skeptical of the veracity of newspapers in the early 1800s (Jefferson 1807). Disinformation was also passed between the Soviet Union and United States during the Cold War (Rid 2020). What is unique to this new era is the speed and efficiency of dissemination that social media provides to bad actors and malicious content creators. They often employ tactics such as networks of automated accounts and content polluters to game recommendation and trending algorithms in order to drive traffic to their disinformation. These tactics have been used to alter online behaviors (Bastick 2020), affect elections (Woolley 2020), and attack public officials, most recently during the COVID-19 pandemic (Barnes and Sanger 2020; Liu and Huang 2020; Wang et al. 2019) and civil unrest in the wake of the murder of George Floyd (Bradshaw and Howard 2018; Ferrara 2017; Cresci et al. 2017; Howard and Kollanyi 2016; Mueller 2019; Bessi and Ferrara 2016; Alba and Frenkel 2020; Ferrara et al. 2016; Broniatowski et al. 2018; Ferrara 2020; Nguyen and Catalan 2020; Schild et al. 2020). Disinformation, and the threat of it, has contributed to the erosion of our political discourse both on and off-line (Schneier 2020; Warzel 2020; Barrett 2020; Coppins 2020).

When presented with conflicting explanations of world events, humans are most likely to believe the ones that most closely align with their existing world view, especially if that information is coming from someone they

✉ Spencer Lee Kirn
slkirn@wm.edu

Mark K. Hinders
hinders@wm.edu

[1] William and Mary Applied Science Department, Williamsburg, VA 23187-8795, USA

Published online: 25 March 2022

know (Abu-El-Rub and Mueen 2019). Disinformation campaigns take advantage of this confirmation bias by using social media and infiltrating user echo chambers so as to target specific groups with their disinformation (Del Vicario et al. 2016). They often use large networks of automated accounts—called bots—disguised as regular people (Woolley 2020; Bessi and Ferrara 2016). These accounts post, like, and retweet each other to amplify their disinformation, taking advantage of recommendation algorithms put in place by social media platforms to boost their content toward the top of their target's feeds (Bradshaw 2019). These vulnerable users then begin sharing the disinformation to increase its reach online (Woolley 2020; Keller et al. 2020; Pierri et al. 2020; Yao et al. 2017; Zannettou et al. 2019). This bottom–up life cycle is quite unique from traditional forms of news and narratives and propaganda, which generally begin from a top, central, source and work their way down.

One complicating factor in any research within the disinformation field is deciding what actually *is* disinformation. This is an exceptionally difficult issue as much of the news ecosystem is quite partisan, and each party commonly accuses the other of propagating disinformation. Any diligent academic researcher will try to take their personal political biases out of any analysis conducted in this space. However, there are still significant unconscious biases that exist which affect how we approach these issues. One of the most glaring examples of this was that much of Academia and the Liberal news institutions called the *Lab-Leak Theory* of Covid-19, the theory that the virus inadvertently leaked from the Wuhan Institute of Virology, disinformation for much of 2020 (Bernstein 2021; Sills et al. 2021; Kormann 2021). However, in early 2021, the Biden Administration announced an investigation into this theory as a real possibility (Hunnicutt and Bose 2021). Often the term 'disinformation' is simply used by one side or the other to dismiss news supporting their opponents or opposing their own world view. This is why in this work, we have not defined anything beyond the glaringly obvious as disinformation. Instead, our work is to uncover the networks of accounts and bots that work together to promote their own narratives. These *information operations* (IO) might make use of disinformation to support their work, they might also use malinformation—real news presented in a misleading manner (Baines et al. 2020)—or a litany of other tactics to achieve their end.

It is the underlying behavior of the network pushing a specific story or narrative that will signal an IO in action, not characteristics of that story or narrative. Analyzing individual articles will never be sufficient because IO campaigns, and those running them can always adjust their tactics to get around detection software. To combat IO campaigns, one must look at the behaviors of the accounts pushing such information. The Dynamic Wavelet Fingerprint (DWFP) has been used in a wide range of applications to uncover signals buried deeply in noise (Hou and Hinders 2002; Hou et al. 2004; Bingham et al. 2009; Bertoncini and Hinders 2010; Bertoncini et al. 2012; Miller and Hinders 2014; Skinner et al. 2019; Hinders 2020; Rooney 2021). In previous work, we have shown the effectiveness of the DWFP for identifying common behavior within very large tweet storms (Kirn and Hinders 2020) and for identifying bot accounts (Kirn and Hinders 2021; Kirn 2021). In this work, we introduce ridge count thresholding (RCT), which isolates specific tweet storms within large collections of tweets to construct *tweet storm networks*. Tweet storm networks have users as nodes, which are connected if both users posted within the same tweet storms some set number of times. This creates dense networks of accounts that highlight both major and minor communities within the data. For this work, we use a dataset of more than 60 million tweets collected from the early weeks of the Covid-19 pandemic, March 14–28, 2020. This data were reduced to about 7 million tweets in the data preprocessing by filtering by the top hashtags. From this data, we identified 39,817 tweet storms, which consisted of just over 2 million unique tweets using RCT. This provided us with 22 specific communities that routinely post similar information during this time. The main contributions of this work are:

- Use of tweet preclustering to identify coherent sets of topics from large dataset of more than 60 million tweets
- Exhibition of RCT as an effective tool to identify potential IO campaigns
- 22 communities of densely connected accounts based on identified tweet storms

2 Methodology

Figure 1 shows a schematic diagram of how we approach this work. We begin with a set of over 60 million tweets over the 2 weeks from March 14–28, 2020. This is a large and computationally unwieldy dataset, thus it must be broken into subsets of tweets that represent broadly similar content, we refer to this step as *preclustering*. Preclustering has the added benefit of reducing the exceptional amount of noise within the dataset by dropping the tweets that are not related to any of the major clusters. Once we have tweets preclustered, each cluster is broken down into topics using Orthogonal Nonnegative Matrix Factorization (ONMF). The ONMF algorithm is optimized to identify the most coherent set of topics for each cluster. We test three different types of preclustering on a subset of the overall tweet volume to identify which provides the most robust set of topics. Using the extracted topics, we are able to construct topic-tweet signals, which describe how each topic propagated through the Twitterverse. From these signals, we isolate the areas of

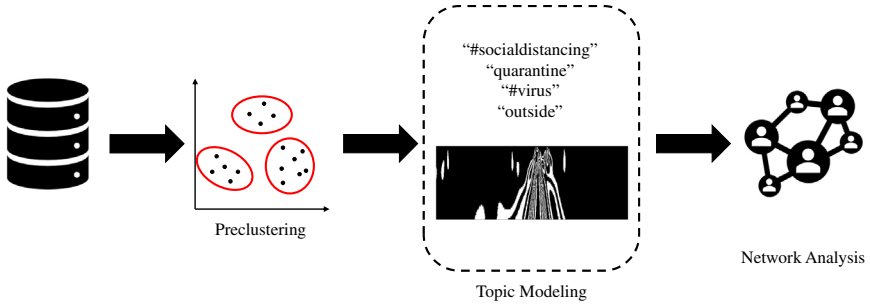

Fig. 1 Schematic diagram of the analysis method presented in this work. Tweets are pulled from a database and clustered into similar subsets of tweets. Each set of clustered tweets is then processed using an ONMF topic model to identify common threads of conversation and the related accounts within each. Topics are then passed through the DWFP to isolate tweet storms, which are used to create community networks of accounts appearing in the same set of tweet storms

interest—or *tweet storms*—using RCT. We use these tweet storms to create network representations of the overall Twitter activity during the time of data collection and identify communities within the overall traffic. We then down select to just the most highly connected accounts within these communities to identify those accounts that are driving narratives and are especially influential in order to understand how they react to the rapidly evolving Covid-19 news cycle.

2.1 Data

Our dataset was collected by Panacea Labs (Banda and Tekumalla 2020). It contains well over 100 million tweets beginning on January 1, 2020 through March 28, 2020 all referencing the Covid-19 pandemic. Panacea Labs continued to update this dataset as the pandemic evolved. At the time of this, writing the most recent data posted was on March 14, 2021. There is a significant increase in volume from a few thousand tweets a day to over four million tweets a day beginning on March 14th, 2020, when the group began specifically searching for the terms: 'COVD19', 'Coronavirus-Pandemic', 'COVID-19', '2019nCoV', 'CoronaOutbreak', 'coronavirus', 'WuhanVirus', 'covid19', 'coronaviruspandemic', 'covid-19', '2019ncov', 'coronaoutbreak', 'whuanvirus', among others. For this work, we select the time period from March 14–28, 2020, when the full tweet stream was running, to analyze the evolving conversations and narratives during this time. For this time period, we were able to gather 60,052,384 total tweets posted by 14,832,993 unique accounts. To access the data, we had to *hydrate* the tweets by querying Twitter's API using the provided tweet IDs to get the full tweet object. Tweets were gathered through the Python package Tweepy (Tweepy 2017) and saved as JSON files. Because we had to gather tweets through the Twitter API, with Tweet IDs provided by Panacea Labs, all previously deleted tweets and suspended accounts were inaccessible.

2.2 Topic modeling

Figure 2 shows a schematic illustration of the topic modeling process. First, tweets are taken from the Twitter API, and the text is preprocessed into a standardized form. The text for each tweet is then vectorized into individual tweet vectors, which are collected and formed into a term-tweet matrix, called $\boldsymbol{D}$. This matrix is then decomposed into two smaller matrices using ONMF. These matrices share a common dimension—topics, which are latent structures hidden within the overall $\boldsymbol{D}$ matrix and describe the underlying narratives occurring throughout the data. The term-topic matrix, $\boldsymbol{U}$ is used as a conversion matrix, it informs what each topic describes, this manifests itself through the topic terms, which are the highest value terms in the corresponding column of the $\boldsymbol{U}$ matrix. This work primarily focus on the topic-tweet signal created from the topic-tweet matrix, $\boldsymbol{V}$. This is constructed by isolating all tweets related to a given topic, then using the time stamps to plot the tweets out in time. Each signal is constructed of 20,820 data points, each representing the total number of relevant tweets posted in the corresponding minute between the first tweet posted in the dataset and the final. These topic-tweet signals are what we use to analyze the underlying behaviors and extract the networks of accounts pushing these narratives.

2.2.1 Tweet preprocessing

To ensure clean and understandable topics, tweets must be preprocessed before being passed to the ONMF engine. Take the tweet: "The United States has banned all travel from Europe #Covid-19". This is not a real tweet, though it is representative of tweets in the data. This short tweet introduces a significant amount of noise. Computational models identify common terms through string matching, which are case sensitive, meaning the terms '#Covid-19' and '#covid-19' are counted as two different terms. Thus, the first step in this process is to set all letters to lower case, and the tweet is tokenized, each

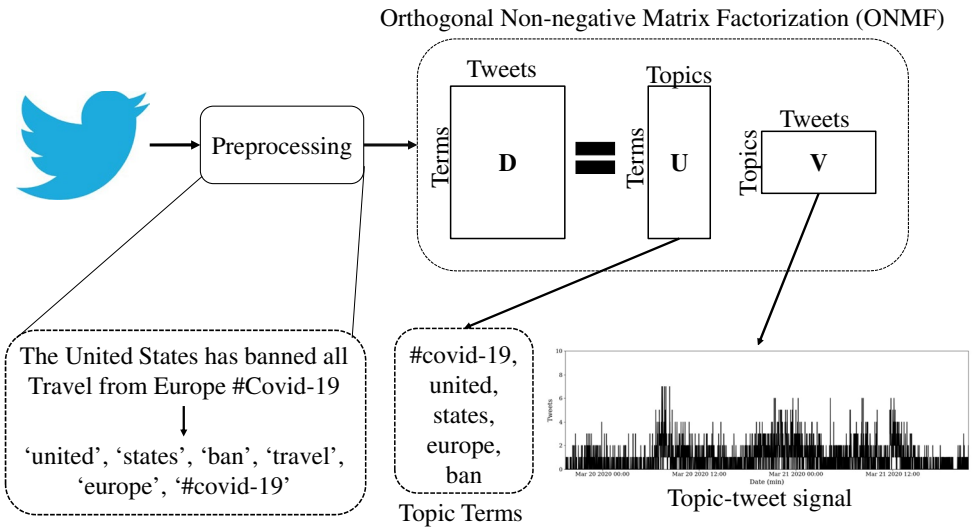

Fig. 2 Schematic illustration of the full topic modeling process. First tweets are taken from the Twitter API and preprocessed to clean the text. The cleaned text is then vectorized and transformed into a term-tweet matrix $\boldsymbol{D}$. The $\boldsymbol{D}$ matrix is decomposed into two matrices, the term-topic matrix $\boldsymbol{U}$ and the topic-document matrix $\boldsymbol{V}$ using ONMF. These two matrices describe the latent topics hidden within the data, $\boldsymbol{U}$ describes the terms that make up each topic, $\boldsymbol{V}$ gives the tweets related to each topic, which can be used to build topic-tweet signals

word is separated into its own entity: 'the', 'united', 'states', 'has', 'banned', 'all', travel', 'from', 'europe', '#covid-19'. All punctuation, not including hashtags '#', user mentions '@', and punctuation that is part of a word, i.e., 'u.s.', is removed. Once terms are tokenized and stripped of punctuation all stop words—i.e., 'and', 'to', etc.—are stripped out, and the remaining terms are lemmatized, or changed to their base form. This process is done using the Python library NLTK, which has a part-of-speech tagger and lemmatizer built in to its functionalities (Bird et al. 2009). At the end of the process, the above tweet is left as: 'united', 'states', 'ban', 'travel', 'europe', '#covid-19'.

2.2.2 Tweet pooling

Even with the reduction of the vocabulary allowed by preprocessing phase, there is still a lack of term co-occurrence. To remedy, this Mehrotra et al. proposed tweet pooling (Mehrotra et al. 2013). Tweet pooling is the process by which tweets with similar hashtags are all pooled into one document to increase term co-occurrence. Here we test tweet pooling to see changes in the results of the ONMF topic model. For our application, we pooled hashtags as well as burst terms, which are terms whose burst-score raises above a predefined threshold,

$$b_s(m,t) = \frac{|\mathcal{M}(m,t) - \mu(m)|}{\sigma(m)} \tag{1}$$

where $\mathcal{M}(m,t)$ is the number of occurrences of term m at time t, $\mu(m)$ is the mean count for term m over the time of the topic model, and $\sigma(m)$ is the standard deviation for term m over the life of the topic model. If $b_s \geq \tau$ where τ is predefined, then all tweets with term m are combined into one document. We used $\tau = 5$ similar to Mehrotra (Mehrotra et al. 2013). A tweet that has multiple pooling terms will appear in multiple pooled documents.

2.2.3 Tweet preclustering

To truly understand what is going on in a dataset as large as the one we are using here, we must find a way to remove noise from the data, so we can analyze the signals underneath. To do this, we introduce preclustering, which subdivides a large dataset into smaller sections, on which we can run ONMF to pull out topics. This process drops much of the useless and uninformative tweets that obscure the latent topics. Furthermore, by subdividing large datasets commonly used in social media analysis, we are able to greatly increase computational speed. In this work, we test two different types of preclustering: hashtag and GMM and compare these to the results without preclustering.

Hashtag preclustering clusters tweets based on the hashtags used within a tweet. In this process, the top 1000 hashtags occurring in the data were identified and used to subdivide the full dataset. Since we are primarily interested in English tweets that we went through these 1000 hashtags and removed those that were not English. We also dropped a few of the top hashtags, including '#covid19' because they were part of the search terms used by Panacea Labs and thus occurred disproportionately often in the data. This left us with 595 total hashtags. All tweets that did not include one of these hashtags were dropped from analysis, which left us with a dataset of just over 7 million tweets to analyze.

The down side of hashtag preclustering is that we are certainly dropping too much data, since many tweets did not include hashtags. Another approach then is to form tweet vectors and cluster them using a clustering algorithm such as a Gaussian Mixture Model (GMM). We do form tweet vectors when we create the $\boldsymbol{D}$ matrix for the ONMF model,

however, these are exceptionally large vectors. Each tweet is described in M dimensions, where M is the size of the full vocabulary of terms in the data, usually on the order of 10^4. Clustering in high dimensions is difficult as the vector distances used to optimize clusters break down in such high dimensions. Thus, we trained a Word2Vec term embedding space to create term and tweet vectors using the Gensim library on Python (Řehůřek and Sojka 2010). This type of mode allows us to cast these tweet vectors into a much smaller dimension and still describe the same underlying behavior. Training term vectors like this also allows us to incorporate information about the semantic meaning of each term in the vocabulary, information that is not available to us in a term-frequency style embedding system.

The Word2Vec model defines a term matrix $\boldsymbol{W} \in \mathbb{R}^{M\times\kappa}$, where κ is the size of the embedding space, here κ is set to 100, as this value is common in Word2Vec applications in the literature. The tweet vector for tweet $n \in N$, $\vec{p}_n$, is defined as a linear combination of the terms, M_n contained within tweet n

$$\vec{p}_n = \sum_{m=1}^{M_n} \text{tf-idf}_{m,n}\vec{w}_m \tag{2}$$

where $M_n \in M$ is the subset of terms occurring in post n, and $\text{tf-idf}_{m,n}$ is the tf-idf score for the term m_i, and $\vec{w}_m$ is the Word2Vec embedding vector for term m. The tf-idf value for each term is calculated as

$$\text{tf-idf} = \text{tf}_{m,n} \times \log\left(\frac{N}{\text{df}_m + 1}\right) \tag{3}$$

where $\text{tf}_{m,n}$ is the term frequency which gives the number of occurrences of term m in tweet n, and df_m is the document frequency which gives the total number of tweets in which term m appears. Tweet vectors are clustered using a GMM.

2.2.4 Orthogonal NMF

NMF is a dimensionality reduction technique that reduces a given tweet from a high, term-based vector representation, to a much lower, topic-based vector representation. Each entry in this new representation describes a document's affiliation with a given topic. NMF is unique among topic models in that it restricts the resulting matrices to non-negative values, thus making the model highly interpretable. Furthermore, it uses matrix multiplication to update the model, making it easily portable to a GPU for efficient computation. It's soft clustering architecture and computational efficiency have allowed NMF to be applied to a wide array of applications both in text analysis and beyond (Berry et al. 2007).

To train an NMF model, we use three matrices: the term-tweet matrix $\boldsymbol{D} \in \mathbb{R}^{M\times N}$, the term-topic matrix $\boldsymbol{U} \in \mathbb{R}^{M\times K}$, and the topic-tweet matrix $\boldsymbol{V} \in \mathbb{R}^{K\times N}$ where M is the total number of terms in the overall vocabulary, and N is the total number of tweets that make up the data. Traditional NMF is expanded to include an orthogonality constraint on the term-topic matrix, $\boldsymbol{U}$ (Ding et al. 2006). Orthogonality constraints on $\boldsymbol{U}$ ensure that the topics extracted are diverse. This orthogonality constraint is given as

$$\begin{aligned} &\mathscr{L} = \frac{1}{2}||\boldsymbol{D} - \boldsymbol{UV}||_F^2 \\ &\text{s.t.} \quad \boldsymbol{U}, \boldsymbol{V} \geq 0 \quad \text{and} \quad \boldsymbol{U}^T\boldsymbol{U} \approx \boldsymbol{I}_K \end{aligned} \tag{4}$$

where $\boldsymbol{I}_K$ is the $K \times K$ identity matrix. The $\boldsymbol{D}$ matrix is a sparse, $\mathbb{R}^{M\times N}$ matrix where each entry $\boldsymbol{D}_{m,n}$ represents the term-frequency inverse document frequency (tf-idf) value for term $m \in M$ in tweet $n \in N$, calculated using (3). The orthogonality constraint can be rewritten using a Lagrangian penalty term

$$\mathscr{L} = \frac{1}{2}||\boldsymbol{D} - \boldsymbol{UV}||_F^2 + \frac{1}{2}tr[\boldsymbol{\Lambda}(\boldsymbol{U}^T\boldsymbol{U} - \boldsymbol{I}_K)] \tag{5}$$

where $\boldsymbol{\Lambda}$ is a $K \times K$ symmetric Lagrangian matrix.This loss equation results in the ONMF multiplicative update equations of

$$\begin{aligned} &\boldsymbol{U} \leftarrow \boldsymbol{U} \circ \sqrt{\frac{\boldsymbol{DV}^T}{\boldsymbol{UU}^T\boldsymbol{DV}^T}} \\ &\boldsymbol{V} \leftarrow \boldsymbol{V} \circ \frac{\boldsymbol{U}^T\boldsymbol{D}}{\boldsymbol{U}^T\boldsymbol{UV}}. \end{aligned} \tag{6}$$

where the square root is added to the $\boldsymbol{U}$ term in (6) by Ding (Ding et al. 2006) to keep entries in $\boldsymbol{U}$ from ballooning. The multiplicative update equations provided in (6) allow for ONMF to be easily ported onto a GPU for efficient calculations.

2.2.5 Topic scoring

To measure the best topic embeddings, we use the coherence score (Mimno et al. 2011). Coherence is calculated based on common term co-occurrences within topic tweets and through the dataset as a whole

$$C(t, U(t)) = \sum_{m=2}^{M^{(t)}} \sum_{l=1}^{m-1} \log \frac{Z(u_m^{(t)}, u_l^{(t)}) + 1}{Z(u_l^{(t)})} \tag{7}$$

where C is the coherence score, $M^{(t)}$ defines the set of topic terms for the given topic, t, $Z(u_l^{(t)})$ is the frequency of term $u_l^{(t)}$ in the entire corpus, and $Z(u_m^{(t)}, u_l^{(t)})$ is the number of documents in which the terms $u_l^{(t)}$ and $u_m^{(t)}$ co-occurred. A perfect coherence score between two terms, u_m, u_l will occur when all documents containing u_l also have u_m and has a

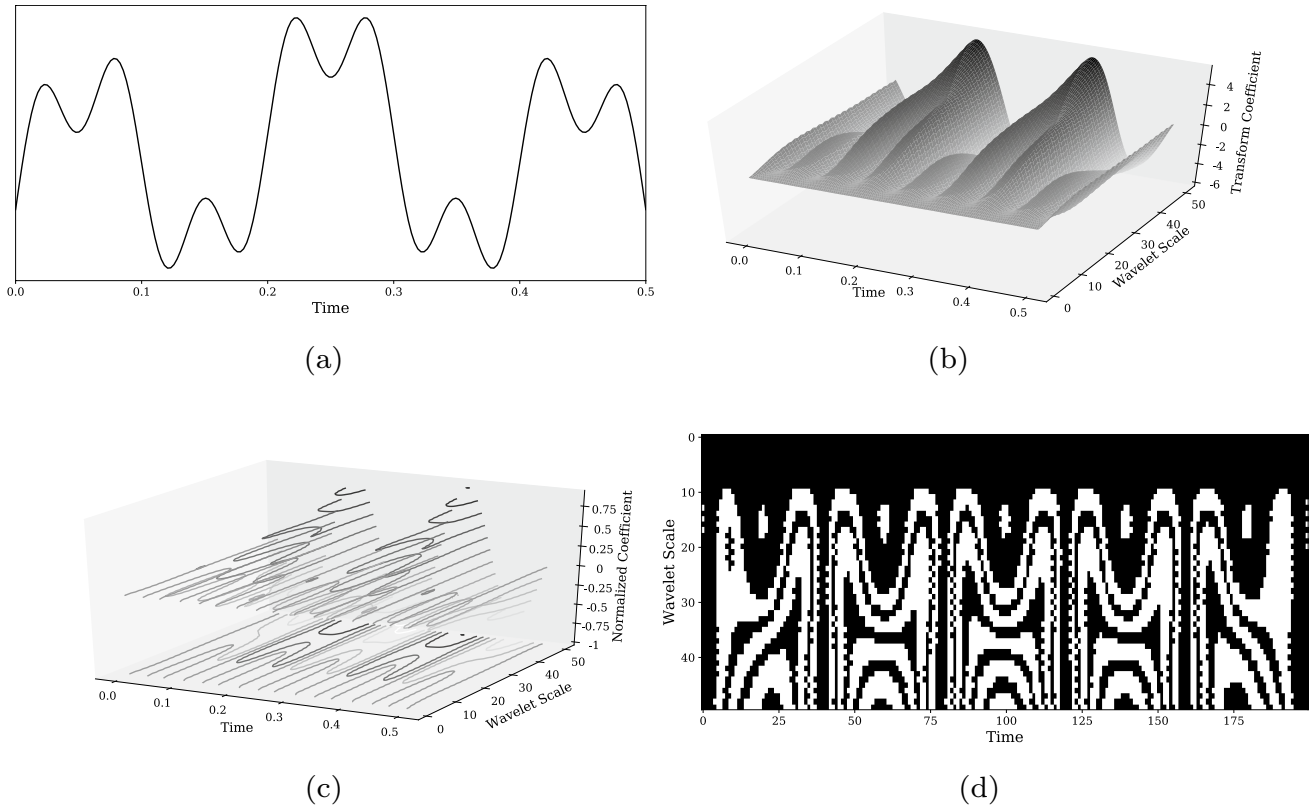

Fig. 3 Schematic illustration the DWFP process. A time-domain signal $f(t)$, (**a**), is passed through a continuous wavelet transform (8). This produces a three-dimensional surface describing the data (**b**). Thie surface is normalized between −1 and 1, and a thick contour slice operation is performed (**c**). This is cast down to two dimensions to create a binary image called a wavelet fingerprint (**d**)

value of just above 0, resulting from the smoothing term in the numerator of (7).

2.3 Dynamic wavelet fingerprint technique

Each topic extracted from the ONMF model sill contains significant noise. For example, one topic in our data is focused on the Covid-19 outbreak in Italy, with top terms: '#coronavirusitalia', 'italy', '#estadodealarma', 'meanwhile', 'phone', and 'happen'. The first tweet storm within this topic centers around one viral tweet which read: 'meanwhile, in Italy.. #Covid_19, #estadodealarma #coronavirusitalia #CoronaOutbreak' and linked to a video of Italian citizens playing music on their porches. This viral tweet caused an increase in traffic around this topic. However, if we scan out into areas of outside of this increased traffic we find many tweets that are not relevant, such as: '#Covid_19 is a glaring reason for Medicare for All. Meanwhile, Biden has 0 plans of backing M4A.' A tweet about U.S. politics that has no relevance to Italy's Covid-19 outbreak outside of sharing a few terms. So, we need a way to isolate these areas of increased activity—what we call tweet storms—and drop the rest of the noise. This could be done using a tweet thresholding mechanic, where if the total number of tweets is above some value, we window around those until it falls below that threshold. However, this quickly falls apart when we try to analyze topics with vastly different volumes. For example, in this work it was not uncommon to have some topics peak at several hundred tweets, while others may have only peaked at 10 or fewer. This could be remedied by looking at the derivatives of topics to isolate periods of rising volume, however this approach is limited in scope. Instead, we look to use the DWFP, which uses wavelets to transform a signal into a binary image. This method has the benefit of being normalized, so we look solely at the underlying behaviors without the overall volume of tweets distorting the results (Kirn and Hinders 2020, 2021; Kirn 2021).

Figure 3 shows the four steps to the DWFP. We first take a signal, shown in Fig. 3a, filter it and pass it through a continuous wavelet transform (CWT),

$$\boldsymbol{C}(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t)\psi_{a,b}(t)\mathrm{d}t \tag{8}$$

where a represents the wavelet scale, b is the time, and $\psi_{a,b}$ is the specific wavelet defined by a and b parameters. The

CWT creates a three-dimensional surface plot describing the signal's behavior in wavelet space, shown in Fig. 3b. There are three axes to the surface plot: time, wavelet scale, and transform coefficient. Time is the same as it is in the original signal, this is possible because of the compact support of the wavelets, each point in time describes the point on which the wavelet was initiated. Wavelet scale is the length of the wavelet, higher scales correspond to longer wavelets. This is the wavelet version of frequency in a Fourier paradigm. Transform coefficient is the value of the corresponding wavelet scale at the given time.

Using the surface plot from Fig. 3b, we normalize the surface along the transform coefficient axis from −1 to 1 and then conduct a thick contour slice operation over the transform coefficient. This slice operation is shown in Fig. 3c. This contour slice operation is cast down into a two dimensional binary image, shown in Fig. 3d.

2.3.1 Ridge count thresholding

Figure 4 shows an example of how RCT is used to extract localized tweet storms within a topic. This is necessary because, even with effective topic modeling there are many tweets that are irrelevant to the overall topic. In general, these are tweets that have a term in common with the top terms, but nothing else. However, since tweets are so short, having just a top term in a topic present within the tweet will cause it to be included.

To suppress the noise that is inherent within these topics, we need to isolate areas of peak activity, and to do this, we use RCT. For this process, we start with a topic-tweet signal, shown in Fig. 4a. This signal is then passed through a wavelet denoising function to smooth the signal and remove excess noise. The smoothed signal is passed through the DWFP to get the binary image shown in Fig. 4b.

The DWFP image provides us with a visual representation of the three-dimensional surface calculated by the wavelet transform. Areas of dense ridges show higher volume in this surface. These are the areas with the most interesting activity that we need to isolate to understand the underlying behavior within a broader topic. This can be calculated using the ridge count (RC) of the DWFP image. RC is calculated by counting the number of ridges—instances where a column changes from 0 to 1 or 1 to 0—in each column of a DWFP image. Once we have the full RC signal for the DWFP image that we then deploy a 5-min rolling average function to smooth the RC signal.

To extract tweet storms within the RC signal, we set two thresholds. The lower threshold is set to 3, and the upper threshold is set to 8, shown with the horizontal red lines in Fig. 4c. Tweet storms are identified by first selecting all points above the upper threshold. For each of these points, we scan left and right to find the points when the RC drops below the lower threshold for at least 5 steps. This process sets the bounds for tweet storms, which are shown using the gray boxes in Fig. 4a and c. The values of 3 and 8 used here were heuristic values identified to work well for windowing groups of tweets that were very focused on a specific subject. The value 8 was chosen because it was large enough that only very dense areas of wavelet activity would reach that threshold, but also low enough that we could see multiple tweet storms for topics. As this value moves up, we see fewer, though more focused, tweet storms. While lower values would lead to more, noisier tweet storms. The value 3 was chosen as a way to cut off tweet storms before the overall volume dropped too low. Raising this value would lead to more focused tweet storms, though we would lose many relevant tweets on either tail of the time window. While lowering the value would result in more noise within the data. Future work to optimize these values will be a necessary follow on to the work presented here.

3 Results and analysis

3.1 Data subset testing

Before running topic modeling on the entire dataset, we tested two different methods of preclustering tweets to optimize the output topics: GMM, and hashtag. We also ran ONMF without any preclustering as a control, and we tested the results of pooling tweets for each preclustering method. This resulted in six different trials. From each trial, we measured the average coherence of the extracted topics to identify the most effective topic spaces. For computational simplicity, we kept this trial to just the first 2 days of data. Results are shown in Table 1.

3.1.1 No preclustering

Running ONMF on the full dataset at once produced unsatisfactory results because there was too much noise in the data. Many topics were generic, identifying very broad topics in the data and not extracting the specific narratives or viral tweets that we are interested in. Overall, the coherence of the extracted topics was the worst of all the trials, with tweet pooling improving the results over no pooling. Contributing to the exceedingly generic topics identified here, is that there were only 280 topics identified using tweet pooling, and 270 without pooling. That few topics is not nearly sufficient to extract specific narratives in such a large dataset.

Not using any form of preclustering also caused significant computational issues. Running ONMF on just these 2 days of data was computationally very difficult and took significantly longer—days, not hours—than the GMM and hashtag clustered tweet sets. This is not an option as we look

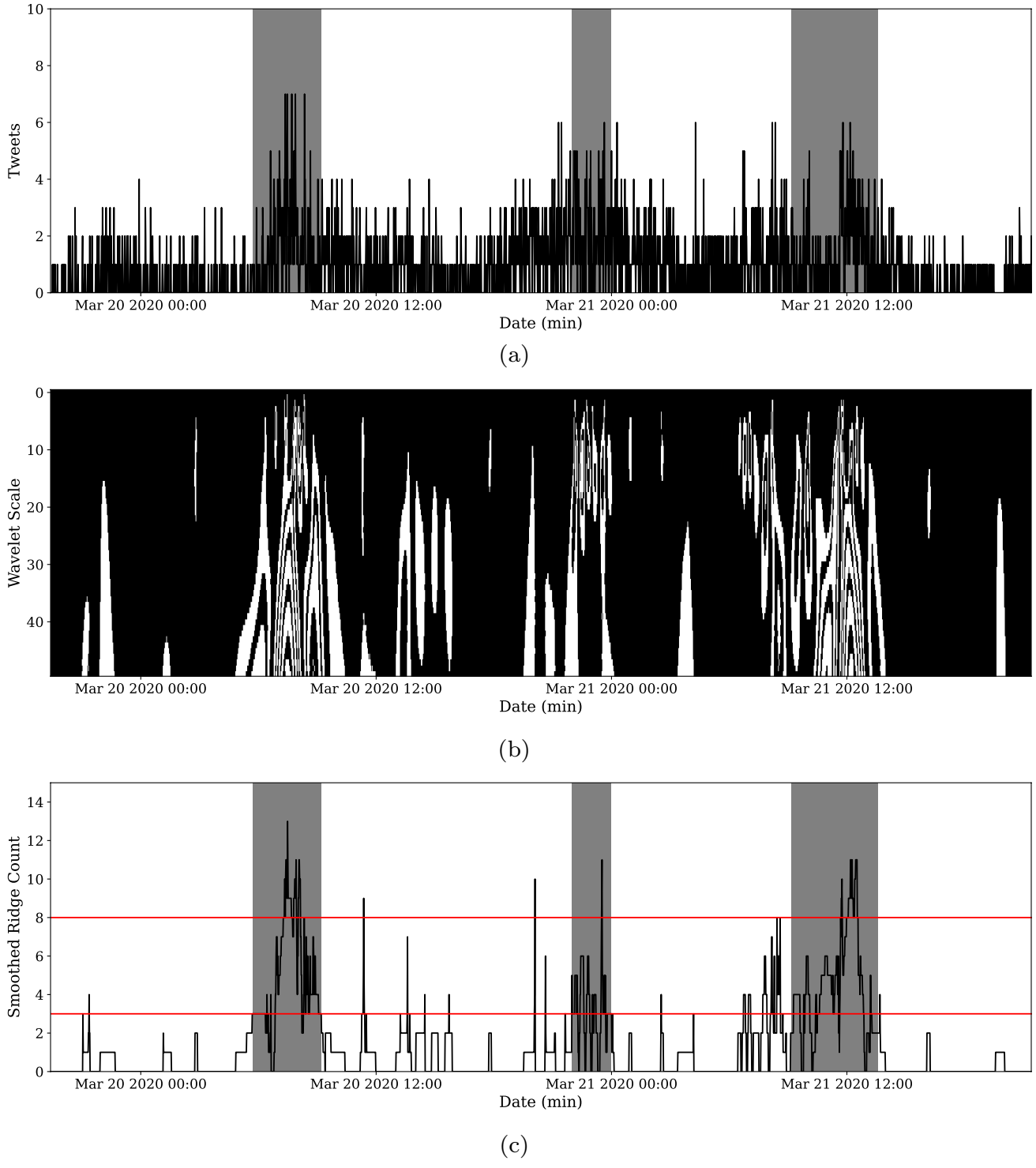

Fig. 4 Example of how RCT is used to extract localized tweet storms within a topic. **a** We start with a tweet signal from a topic calculated using ONMF. This signal is then passed through a wavelet denoising function to smooth the signal and remove excess noise. **b** The smoothed signal is passed through the DWFP to create a binary image. **c** We take the RC of the DWFP image and use upper and lower thresholds to isolate points of increased activity. Upper and lower thresholds are shown by the horizontal red lines. Extracted tweet storms are shown by the gray boxes in **a** and **c**

Table 1 Results from the ONMF trials on a 2 day sample of tweets

	No preclus. No pool	No preclus. Pool	GMM No pool	GMM Pool	Hashtag No pool	Hashtag Pool
Average topic coherence	− 3.11	− 2.92	− 2.63	− 2.53	− 1.16	− 1.14
Total topics	270	280	2135	1995	7860	8525

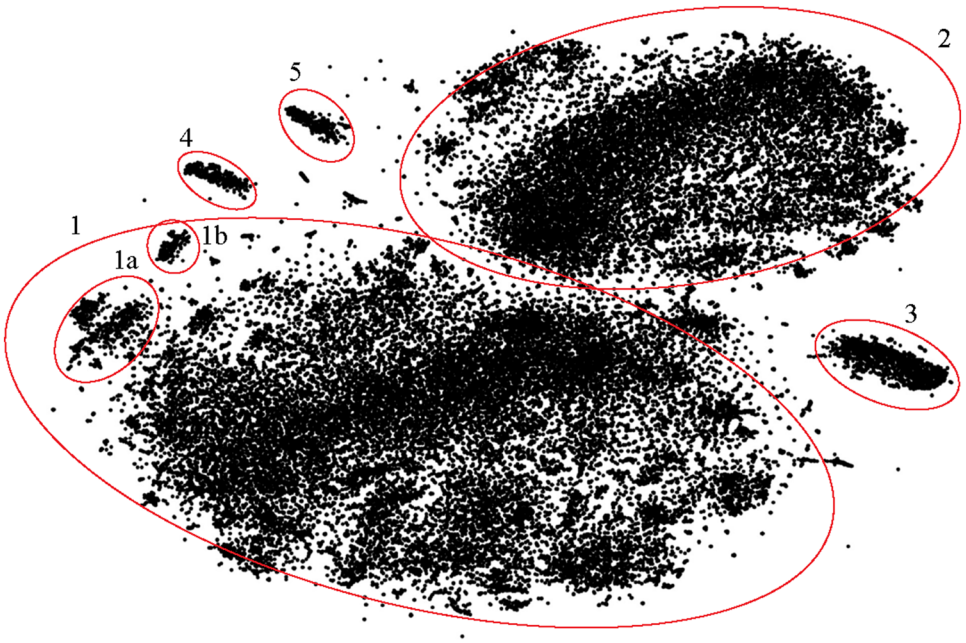

Fig. 5 t-SNE plot of 1 h of tweets using Word2Vec tweet embedding. Languages are effectively separated with English tweets labeled 1, Spanish tweets labeled 2, French tweets labeled 3, Indonesian tweets labeled 4, and Italian tweets labeled 5. 1a and 1b show two distinct topics within the English tweets, however most other structure is lost in the noise of the embedding space

to analyze much larger datasets, so some method of preclustering tweets before ONMF is required to make the process computationally feasible.

3.1.2 GMM preclustering

GMM preclustering relies on Word2Vec embeddings for each tweet in the dataset. We trained a Word2Vec model using the Gensim Library (Řehůřek and Sojka 2010) in Python and converted each tweet to a 100-degree vector by calculating embeddings as the tf-idf value of a term in a tweet multiplied by that term's embedding vector calculated using (2). We opted to train our own Word2Vec model because no open source embedding space would be able to handle the new terms critical to understanding this dataset such as 'covid', not to mention hashtags and user mentions. In total, we identified 32 different clusters within the trial data. This resulted in a total of 2135 topics identified without tweet pooling and 1995 topics identified with tweet pooling.

Figure 5 shows Word2Vec embedding space for 1 h of tweets using t-distributed stochastic neighbor embedding (t-SNE) (Van der Maaten and Hinton 2008). The t-SNE plot was created using SciKit-Learn in Python (Pedregosa et al. 2011). What we found was that the GMM model was able to effectively cluster tweets by language, but we did not see the partisan tweet clustering that we were expecting to see. In Fig. 5, each ellipse labeled 1-5 shows a cluster of tweets almost entirely in the same language English, Spanish, French, Indonesian, and Italian, respectively. This is exactly what we expected to see from this kind of embedding space. Furthermore, 1a and 1b show two different clusters of specific topics that occurred over this time period: the passing of the Families First Act by House Democrats and reporting on a Google Covid-19 tracking site, respectively. This provides promising results, however when expanding to a larger dataset over all 48 h, many of these topics become buried in the noise. Thus, when running GMM to extract the specific clusters we find that beyond dividing by language there is not much benefit of this kind of embedding.

Some adjustments could be made to make this kind of tweet clustering more effective, such as dropping short—one and two word—tweets, which do not provide much information, but account for a large portion of the data. We could also run clustering hour-by-hour to pick up on temporally local topics instead of trying to analyze the full dataset in one attempt. This would also be more computationally feasible as we expand to larger datasets and would offer a possibility of real-time application.

3.1.3 Hashtag preclustering

From Table 1, it can be seen that the most effective topic sets were extracted using hashtag preclustering. In total, we were able to identify 7860 topics without using tweet pooling and 8525 topics using tweet pooling. Pooling tweets by keywords provided more topics and slightly more coherent topics, however, this can be computationally expensive as it requires the module to iterate through every input tweet without much payoff in topic coherence. This expense grows linearly with the total tweet input, so as we work with larger datasets that this will become an even greater issue in the processing pipeline. Thus, when analyzing the full 2 weeks of data we opt to use hashtag preclustering without pooling before running ONMF.

3.2 Full dataset analysis

For the full dataset, we ran ONMF on hashtag preclustered data with no tweet pooling. In total, we found 18,190 topics with an average coherence score of −1.54. From each topic created via the ONMF model, RCT was used to isolate tweet storms. In total, we identified 99,570 total tweet storms. All tweet storms with less than 25 tweets and those that were not in English were dropped leaving 39,817 total tweet storms, with 2,058,255 unique tweets.

Graphical methods have been shown to be quite effective for social media analysis (Boshmaf et al. 2015) and particularly for automated bot detection (Hurtado et al. 2019; Beskow and Carley 2020; Abu-El-Rub and Mueen 2019). Thus, we leverage this type of analysis to identify connected accounts. For each tweet storm, we took the corresponding accounts and created a sparse, binary, user-tweet storm matrix, $\boldsymbol{X} \in \mathbb{R}^{U \times S}$, where U is the total number of users, and S is the total number of tweet storms and $\boldsymbol{X}_{u,s} = 1$ if user u posted in tweet storm s. The goal here is not necessarily to classify the specific tweet storms, but instead how users overlap with one another within those tweet storms. So, we multiplied $\boldsymbol{X}$ with its transpose to get the matrix $\boldsymbol{G} \in \mathbb{R}^{U \times U}$. Each entry in the matrix, $\boldsymbol{G}_{i,j}$ represents the number of tweet storms in which users u_i and u_j co-occurred. For this application, the values on the diagonal are unimportant, so they are simply zeroed out.

The matrix $\boldsymbol{G}$ is a user network graph, where each node represents an individual user, and edges are drawn between users if they post in the same tweet storm. To reduce the noise within the data, we drop all accounts that tweeted less than 10 times. Furthermore, we drop all edges that have a weight of less than 10. The minimum edge weight of 10 bakes in the redundancy occurring in the ONMF topics. Since we preclustered by hashtag, it was not uncommon for viral tweets and prominent discussions to appear in multiple hashtag clusters, thus we would often have very similar tweet storms extracted from different topics. Setting these thresholds dropped the network to 12,057 total nodes, with 377,843 edges. The full network is shown in Fig. 6, all network graphics in this paper were created using Gephi Bastian et al. (2009). Nodes are shown in black with size relative to the node's overall degree.

Figure 6 shows the overall network created using RCT. It can be seen that the network is constructed from several densely connected clusters of nodes. We use the Louvain Method (Blondel et al. 2008) and the Python-Louvain library[1] to isolate these individual communities. Running this over the full graph results in 65 total communities, of which 22 are made of more than 10 accounts. Table 2 describes these 22 communities. For each community, we report the total number of nodes, as well as the edge density, i.e., total number of edges out the maximum possible and the average weight of the edges. We also used the bot identification API, Botometer to get the bot probabilities for every account according to their algorithm (Yang et al. 2019b). While Botometer is not effective for identifying if a specific tweeter is a bot (Rauchfleisch and Kaiser 2020), it does do a sufficient job at giving a general sense of the overall bot presence within a community of accounts. The Average Botometer Score reported in Table 2 gives the average English Complete Automation Probability (CAP) value for all active accounts within each community. The Deleted/Suspended Percentage column reports the total percentage of accounts within that community that have been either deleted or suspended by Twitter between the time the tweets were hydrated and the time of this writing. We should note that we are unable to discern between an account that has been deleted by the user and an account that has been suspended by Twitter for their behavior.

Figure 7 shows a comparison of Botometer CAP English scores (Yang et al. 2019a) among a sample of all accounts within the data, Fig. 7a, and for all accounts within the RCT network shown in Figs. 6, 7b. For the sample of all accounts within the full dataset, we find a slightly bimodal distribution with a low peak around 0.25 and then a much larger peak at 0.75. Overall, the average Botometer CAP English score for the accounts in Fig. 7a is 0.60, with 2194 accounts deleted, which equates to about 15% of the sample. Figure 7b shows a distribution with a mode of 0.75, similar to before, but it is no longer bimodal. Based on these distributions, it seems as if we were able to drop most of the human accounts from the data through the RCT mechanism, and we are left with accounts that—according to Botometer—are suspicious. Furthermore, we found that of the accounts within this network 3120 had been deleted or suspended between the time the data were collected, and these accounts

[1] https://python-louvain.readthedocs.io/en/latest/.

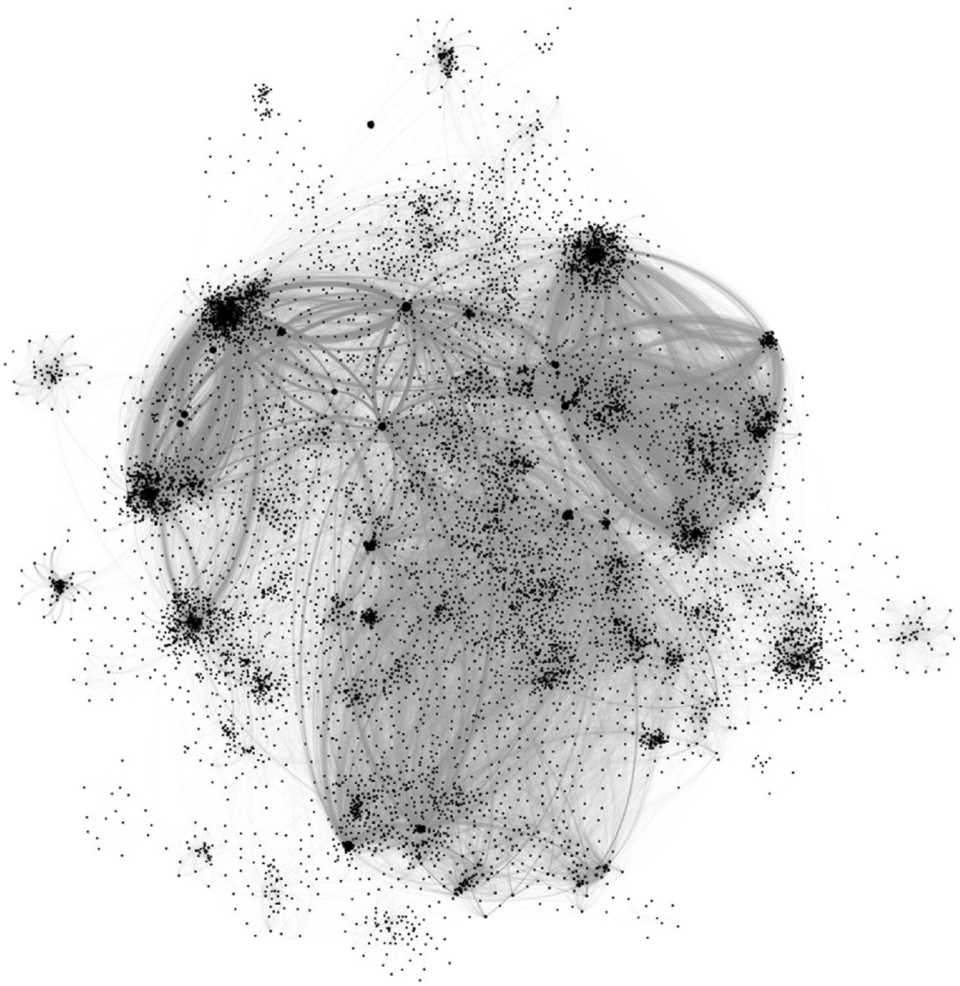

Fig. 6 Overall user network after dropping all users who tweeted less than 10 times and all edges of less than 10. Nodes tend to cluster into communities of densely connected nodes, which we identify using the Louvain Method for community detection

were checked through the Botometer API. This represents more than 25% of the accounts shown in Fig. 6. This is particularly suspicious because during much of the latter part of 2020 and early 2021 Twitter became much more strict about the type of content allowed on their platforms as criticisms of rampant misinformation increased (Guynn 2020; Conger 2021). Thus, it is likely much of these deleted accounts were deleted by Twitter for promoting some form of misinformation.

We identified two major types of communities among the 22 reported in Table 2. The first and most common type is the Constant Spam Community (CSC). Within the 22 total communities we identified, 16 could be classified as a CSCs. These are called CSCs because they continuously posted throughout the 2 weeks our data covered. CSCs tend to have more total accounts that are sparsely connected, relative to the other type of community, a Targeted Spam Community (TSC). TSCs generally have fewer total nodes, but are much more densely connected. These communities are named as such because they showed very little activity, except for a few isolated instances where their overall tweet volume spiked significantly. Below we discuss some of the most interesting communities detected through this method.

3.2.1 Community 0

Figure 8 shows the basic information about Community 0. In this figure, and the similar figures for each community below, (a) shows a zoomed in image of the specific network with the most central nodes highlighted in red, (b) shows the specific posting patterns for this community through the span of the dataset, and (c) shows the DWFP image of two individual days of posting activity.

Community 0 is a CSC. This community exhibits a strong diurnal pattern in how they post. It was also almost entirely retweeting other information, nearly 94% of their tweets were retweets—17,556 out of 18,679 total tweets.

The network's traffic seems random and unorganized, however looking at a few specific points in their overall traffic shows potentially coordinated behavior. Two of these are highlighted in Fig. 8b. The point labeled A in Fig. 8b shows a significant spike in localized activity, from

Table 2 Communities of 10 or more users as extracted from the network shown in Fig. 6 using the Louvain method

Com.	Com. type	Total nodes	Total tweets	Edge density	Avg. edge weight	Avg. botometer score	Deleted/suspended percentage (%)	Total tweet storms
0	CSC	1258	18,669	0.06	24.11	0.76	66	12,114
1	CSC	190	3209	0.20	30.21	0.72	29	5816
3	CSC	735	14,857	0.09	16.71	0.77	17	6934
4	CSC	926	19,226	0.05	16.34	0.78	72	10,090
5	CSC	3191	54,389	0.01	13.65	0.75	14	20,157
7	CSC	1453	26,704	0.08	13.50	0.75	20	9848
8	CSC	324	5937	0.45	17.01	0.72	20	8667
9	CSC	198	3210	0.56	32.44	0.74	15	6157
11	CSC	475	7879	0.05	17.16	0.69	12	8634
14	CSC	171	2369	0.08	15.12	0.64	16	3047
15	CSC	364	6365	0.30	13.53	0.75	30	4813
16	CSC	1426	35,497	0.01	13.41	0.75	17	15,608
17	TSC	51	1088	0.82	15.01	0.80	20	209
20	CSC	762	21,378	0.02	14.21	0.73	10	8699
21	TSC	90	1449	0.76	14.06	0.72	10	275
23	CSC	82	2571	0.09	12.77	0.81	21	764
24	TSC	47	878	0.75	12.44	0.78	11	261
29	TSC	13	230	0.25	11.05	0.78	77	115
32	CSC	12	239	0.45	14.06	0.75	8	210
35	TSC	118	1624	0.99	49.66	0.83	26	162
37	CSC	49	973	0.13	13.70	0.79	24	565
51	CSC	16	177	0.23	11.11	0.83	13	79

a mean below 5 tweets/min to almost 15. This shows the nodes in Community 0 posting about the ongoing Democratic Debate occurring between Senator Bernie Sanders and Vice President Joe Biden. The main spike here specifically refers to Vice President Biden accidentally referring to Covid-19 as Ebola. There were over 100 tweets around Biden's faux pas, every one of them referencing a mental decline in the former Vice President. The point labeled B in Fig. 8b shows this network celebrating President Trump signing Covid-19 economic relief package—the CARES Act—into law.

Table 3 shows the top 10 topics posted by this network. From these topics, it is clear that this community is an American, Conservative community. These topics show a strong focus on China—Topics 1, 3, and 9—as well as other general American politics—Topics 5, 6, and 7.

Of the 8 highlighted nodes in Fig. 8a, half have been deleted or suspended by Twitter. Of the four remaining, three remain actively posting, while the fourth fully stopped any Twitter activity on June 20, 2020. Only one of these four accounts has any personal details within their profile, and this is a only a blurry selfie for a profile picture.

3.2.2 Community 1

Figure 9 shows the basic information about Community 1. The structure of this network is shown in Fig. 9a. There are two main clusters of accounts that are densely connected. From these, there are three central nodes, highlighted in red in Fig. 9a. Of these three, two have been suspended, the other account posted heavily every day from their creation—July 2019—until November 9, 2020, when they stopped cold. This date is relevant because it is exactly a week after the U.S. Presidential Election, suggesting this account was only created to support President Trump's reelection campaign. This account has no personal details, with no bio nor profile picture and seems to retweet American Conservative news and respond to liberal politicians.

Figure 9b and c show the signal and corresponding wavelet fingerprint for how this community posts. This network seems to consistently post without much obvious coordination. No singular event appeared to cause a significant spike in activity like the ones that we observed in Community 0. Unlike Community 0, Community 1's fingerprint in Fig. 9c shows constant activity, confirming what is observed within the signal.

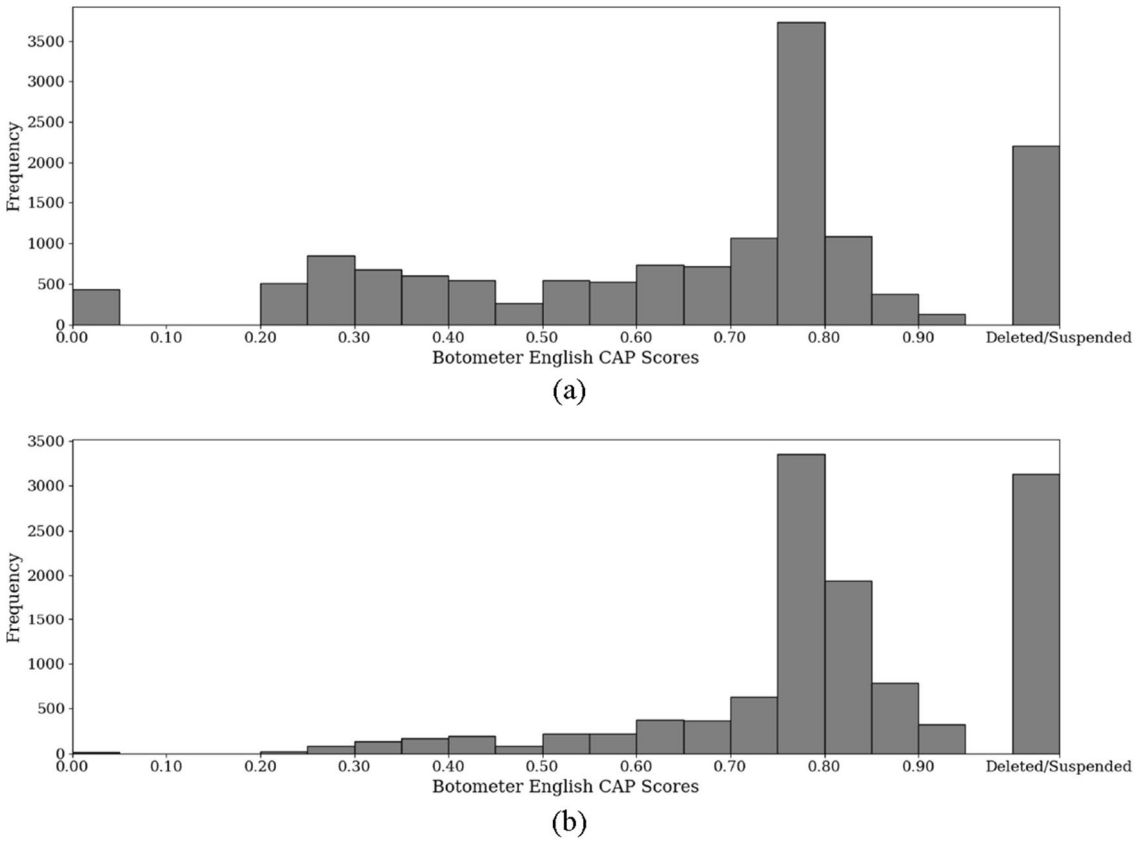

Fig. 7 Comparison of Botometer CAP English scores among: **a** random sample of all accounts from the full dataset ($n = 14,996$), **b** all accounts in Fig. 6

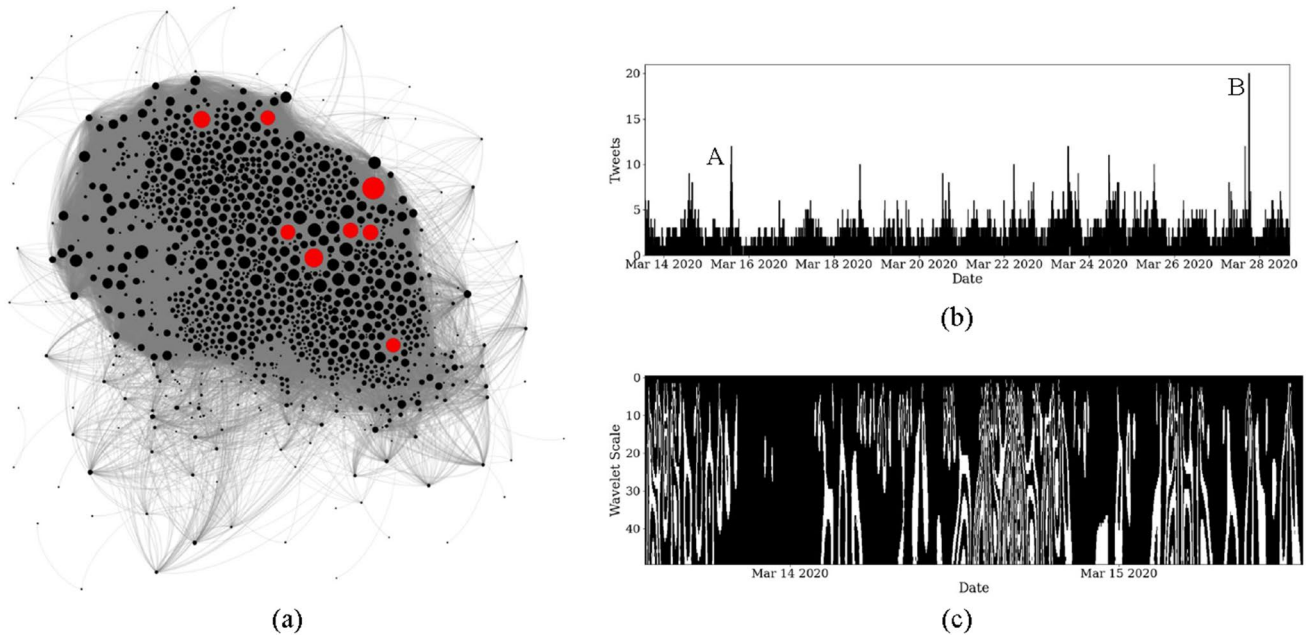

Fig. 8 Results from Community 0. **a** Shows the network diagram of all nodes, those in red are the most central nodes driving traffic. **b** Shows the overall tweet signal from the network. **c** Shows the DWFP transform of the network traffic for the first 2 days

Table 3 Top 10 topics from Community 0

Topic	Topic terms
1	#china, chinese, communist, party, #wuhan
2	#usnsmercy, response, hospital, patients, covid
3	China, virus, call, #chinaliedpeopledied, world
4	#coronavirus, #covid-19, #covid_19, #qanon, #covid2019
5	Americans, #democratshateamerica, america, think, stimulus
6	Coronavirus, #foxnews, #americafirst, bill, #dobbs
7	@realdonaldtrump, trump, president, #kag, #trump2020
8	Will, people, stop, help, country
9	#wuhanvirus, #chinesevirus, #chinavirus, #wuhancoronavirus, crisis
10	#covid19, #wwg1wga, #qanon, #deepstate, need

#covid19Canada #HealthCare #frontlines". This tweet is one of the most viral tweets within our dataset, so it appears in multiple communities. It is also like one of the reasons that these two clusters are connected. Because it has four different hashtags that it is included within each of those hashtag's preclusters from which topics were calculated. This kind of redundancy is a weakness of the hashtag preclustering method used in this work.

3.2.3 Community 4

Community 4 is a community of American Conservatives with a total of 926 accounts, which posted 19,226 total tweets over the 2 weeks of data. Much of the activity within this network centered around three accounts, highlighted

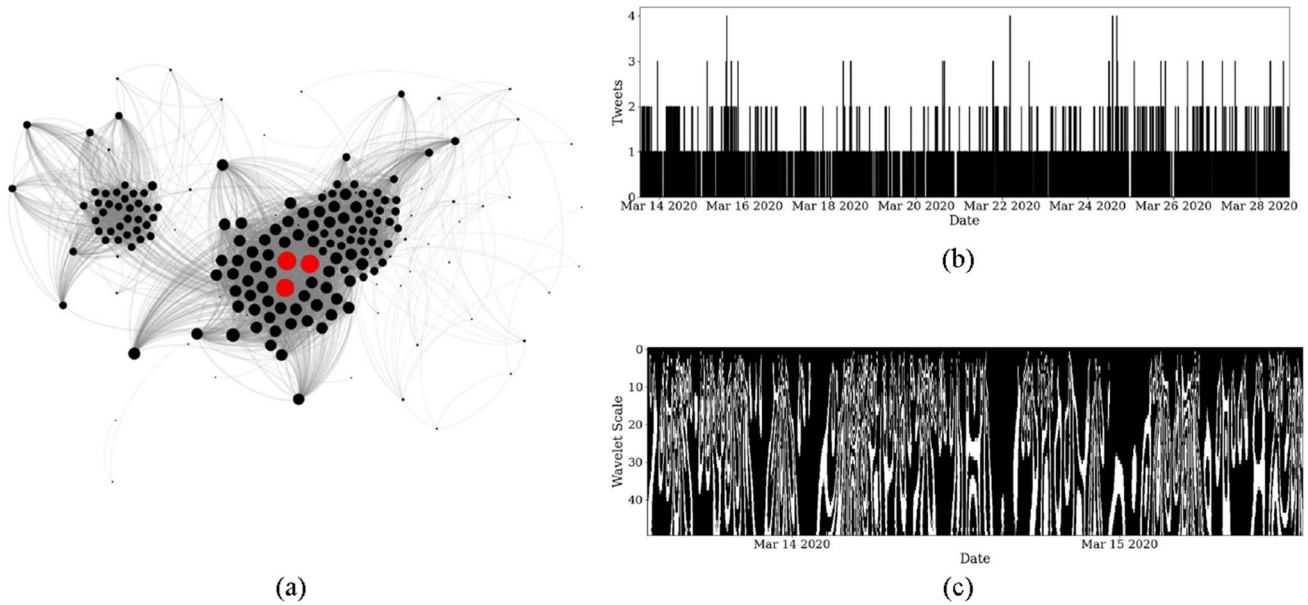

Fig. 9 Results from Community 1. **a** Shows the network diagram of all nodes, those in red are the most central nodes driving traffic. **b** Shows the overall tweet signal from the network. **c** Shows the DWFP transform of the network traffic for the first 2 days

Table 4 shows the top 10 topics posted by Community 1. The separation of clusters of nodes in Fig. 9a can be seen through these topics. The smaller cluster of nodes was predominantly Canadian Twitter accounts who posted about the pandemic and issues within Canada. These are seen in Topics 1, 2, and 7. The larger cluster of nodes, which included the three most central nodes, wasAmerican Conservative accounts posting about China and President Trump. Much of the overlap within these clusters appear to be from tweets discussing the pandemic in general or from viral tweets such as one that read "Dropped my wife off at the hospital this morning for her 12hr shift. While 95% of the world is distancing from the #coronavirus, health professionals are putting their armor on and attacking it. My wife is a hero.

in red in Fig. 10. Each of these accounts has since been deleted or suspended by Twitter, similar to much of the rest of this community. In total, 72% of the accounts—669 out of the 926 accounts—have been suspended or deleted, which is a good indication that we have identified a network that was pushing significant disinformation or violating Twitter's terms of service in some way. For this community specifically, their disinformation seems to be centered around the QAnon conspiracy theory. In fact, the account with the highest degree within the whole network was a common promoter of QAnon (LaFrance 2020). Table 5 shows the top 10 topics from all tweets posted from this community. Of these three—3, 6, and 10—show at least one common

Table 4 Top 10 topics from Community 1

Topic	Topic terms
1	#covid19, #cdnpoli, canada, test, canadians
2	Wife, hero, #covid19canada, #frontlines, armor
3	Coronavirus, trump, president, test, @realdonaldtrump
4	#china, #chinesevirus, #wuhanvirus, #chinavirus, communist
5	#coronavirus, cases, #coronavirusoutbreak, #covid2019, #covid-19
6	#wuflu, #coronaviruspandemic, #wuhancoronavius, #chinapneumonia, malicious
7	#ableg, #cdnpoli, #abpoli, alberta, @jkenney
8	Will, people, just, many, covid
9	China, world, virus, global, pandemic
10	Need, time, make, help, work

QAnon hashtag. In total almost 25% of the tweets posted by users in this network, 4462 tweets, included either #wwg-1wga (where we go one we go all) or #qanon, two very common QAnon hashtags. This was about 12% of the total tweet volume that included either of these two hashtags throughout the full dataset. These hashtags also appeared in the topics from Community 0, though there was not the same overall volume for these as was seen from Community 4.

Like most CSCs, this community shows strong diurnal patterns in how they post and react to each other. Each day this generally tops out at around 6 tweets a minute, shown in Fig. 10b. The corresponding fingerprint for the first 2 days is shown in Fig. 10c. We can see in this fingerprint the repetitive pattern that we would expect from this kind

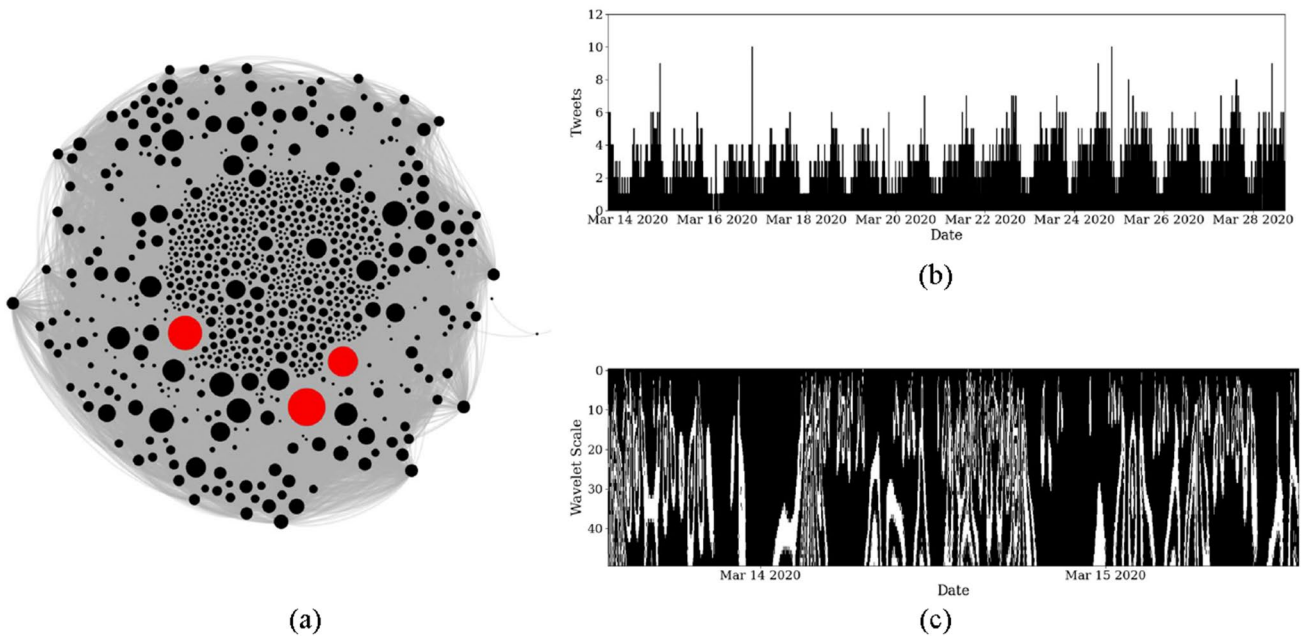

Fig. 10 Results from Community 4. **a** Shows the network diagram of all nodes, those in red are the most central nodes driving traffic. **b** Shows the overall tweet signal from the network. **c** Shows the DWFP transform of the network traffic for the first 2 days

Table 5 Top 10 topics from Community 4

Topic	Top terms
1	#coronavirus, #covid_19, #trump2020nowmorethanever, #china, #coronavirusoutbreak
2	China, chinese, virus, #chinesevirus, #china
3	@realdonaldtrump, @potus, america, americans, #americafirst
4	#trump2020, #kag, #wwg1wga, #paintourcountryred, #2a4life
5	#maga, #trump, #twgrp, #mighty200, #qanon
6	#tcot, #ccot, #covid-19, #coronaviruspandemic, #foxandfriends
7	#qanon2018, #qanon2020, #democratshateamerica, #chinesecoronavirus, #qanon
8	Trump, president, will, people, news
9	Coronavirus, #foxnews, bill, pelosi, senate
10	#covid19, #coronavirusoutbreak, need, like, people

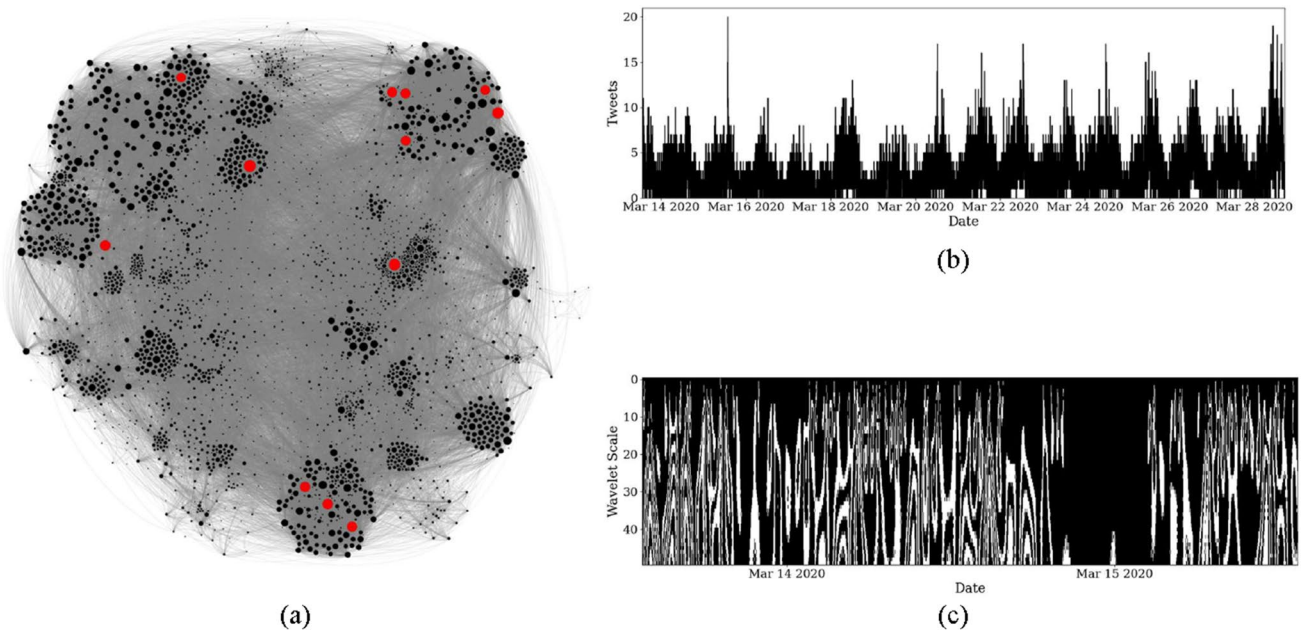

Fig. 11 Results from Community 5. **a** Shows the network diagram of all nodes, those in red are the most central nodes driving traffic. **b** Shows the overall tweet signal from the network. **c** Shows the DWFP transform of the network traffic for the first 2 days

Table 6 Top 10 topics from Community 5

Topic	Topic terms
1	Trump, #trumpgenocide, donald, call, watch
2	#covid19, #stayhome, #flattenthecurve, spread, #socialdistancing
3	Test, positive, tests, kits, covid
4	People, suffer, give, must, spread
5	Coronavirus, #mog, #maga, pandemic, #trump
6	Need, stay, tell, home, please
7	#covid-19, #coronaoutbreak, #trumpliedpeopledied, #dumptrump2020, #worst-presidentinhistory
8	Will, #onevoice1, bill, health, take
9	Cases, deaths, total, number, confirm
10	#coronavirus, #coronaviruspandemic, #coronavirusoutbreak, #pandemic, #covid_19

of community. This pattern repeats day-after-day from this network.

3.2.4 Community 5

Community 5 is the largest community extracted using the RCT method. It has more than double the number of nodes as the next largest community. These accounts posted over 50,000 tweets during the 2 weeks of data collection. Figure 11a shows the network structure from this community. Community 5 represents a large network of American Liberal accounts based on a review of their tweets and topics extracted using ONMF, shown in Table 6.

Figure 11b shows the overall posting patterns from Community 5. This network exhibits a diurnal pattern in how they post, though individual days seem to have more variability. The largest uptick in volume from this community came during the Democratic presidential debate, where there was significant support for both Vice President Biden and Senator Sanders within the network.

Community 5 does not show one set of central nodes, instead there are several, smaller clusters of accounts. Each of the most central nodes appears to be some form of a retweet spam account that consistently retweets liberal news articles or prominent liberal politicians. We ran the Louvain Method over this community to isolate subcommunities of accounts. Most of these subcommunities have the same basic

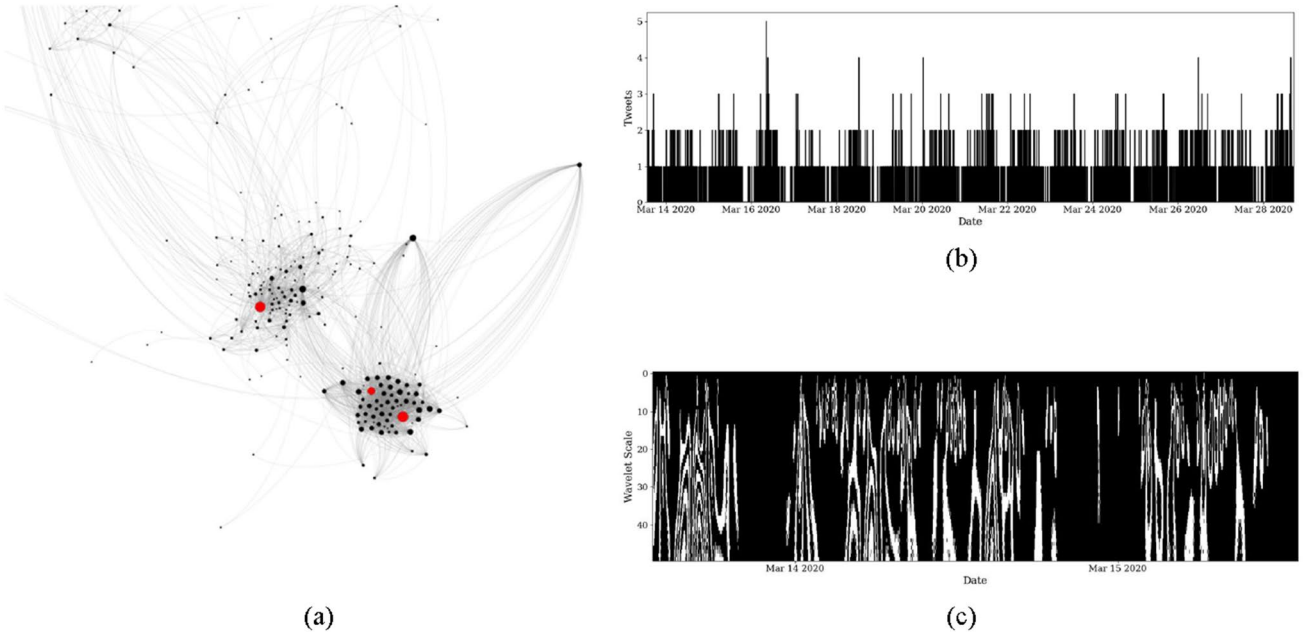

Fig. 12 Subcommunity 10 identified from Community 5. **a** Shows the network diagram of all nodes, those in red are the most central nodes driving traffic. **b** Shows the overall tweet signal from the network. **c** Shows the DWFP transform of the network traffic for the first 2 days

shape as community 4, with a few central nodes surrounded by smaller ones who ping off of the central one. We identified 17 subcommunities within this network. On average, these subcommunities were significantly more dense than the network as a whole, with an average density of 0.27.

Figure 12 shows a zoomed-in look at one of these subcommunities: subcommunity 10. This subcommunity includes 228 accounts and is focused around three main accounts who consistently put out Liberal messaging that is immediately retweeted or parroted in some way by the rest of the network. This subcommunity also heavily pushed two hashtags: '#mog' and '#onevoice1'. These two hashtags refer to two different liberal Twitter news aggregators. In the case of '#mog', there is one central account that posts news articles with that hashtag, which are heavily retweeted. However, '#onevoice1' seems to be more of a collection of accounts that post news and videos with that hashtag. There seems to be significant overlap in the accounts that are pushing both of these hashtags. In total almost 30% of that 3848 tweets posted by this subcommunity included one of these two hashtags.

3.2.5 Community 8

Figure 13a shows the network structure of Community 8 from Table 2. This community include 324 total accounts that are densely connected, with an overall density of 0.45. If we drop the nodes with the lowest degree, those with degree of less than 30, then we find a fully connected network of 215 accounts. The construction of this community is exceptionally similar to that of Community 9, which shows the same high density among the top accounts.

Figure 13b and c show specific posting patterns from this community. The overall community here shows a constant stream of tweets through all time. Figure 13c shows this very clearly in the DWFP. Unlike previous communities, where there are clear breaks shown in the DWFP image, Community 8 shows constant high scale behavior all through the evening, meaning that there is not much of a diurnal pattern. Around March 24–March 27, there is an increase in the overall volume put out by this network.

Table 7 shows the top 10 topics calculated for this community of accounts. There does not appear to be a clear localization for the accounts within this network. Topic 3 references India's Covid response, while Topic 4 references Australia's, and Topic 5 references the UK's. This dispersion of account locations is why there is less of a diurnal pattern in the overall tweet volume shown in Fig. 13b. The main focus that seems to unite these accounts despite their location, is the focus on China. Topics 6 and 8 both reference China. Overall, 20% of the tweets posted by this community include the term 'China'. The dense network connections despite perceived regional differences is an indication of a potential IO campaign.

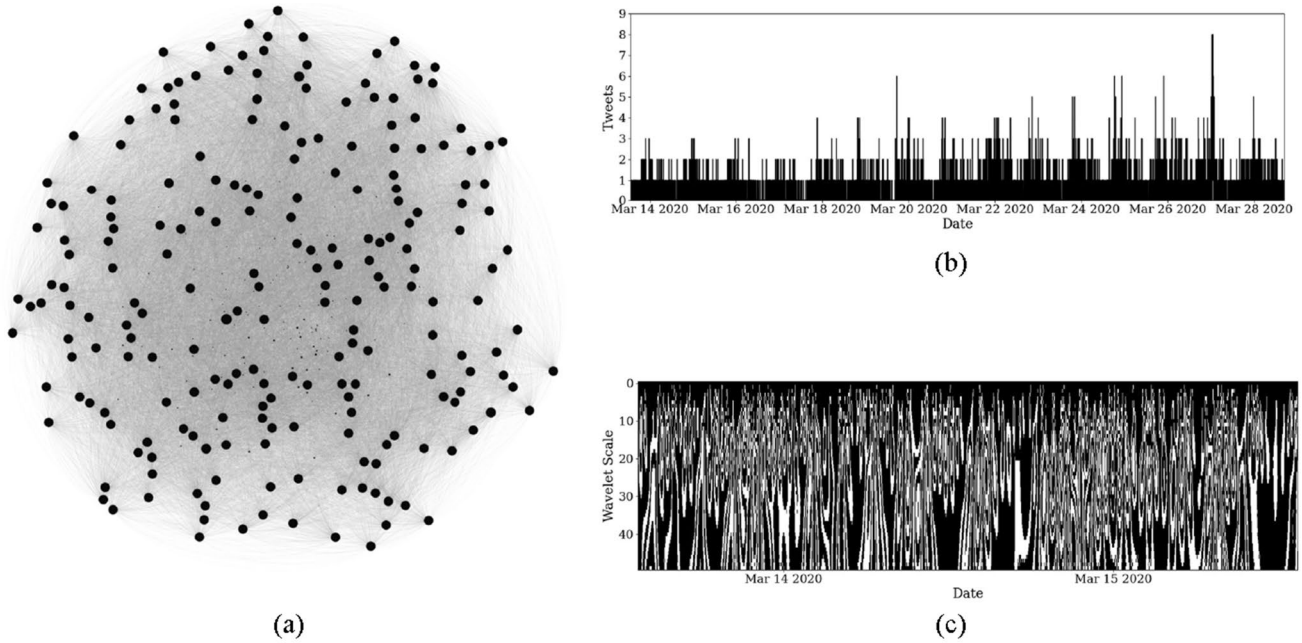

Fig. 13 Results from Community 8. **a** Shows the network diagram of all nodes. **b** Shows the overall tweet signal from the network. **c** Shows the DWFP transform of the network traffic for the first 2 days

Table 7 Top 10 topics from Community 8

Topic	Topic terms
1	Lead, #stayhomesavelives, conference, mild, video
2	Cases, total, deaths, report, confirm
3	Coronavirus, people, #indiafightscorona, covid, home
4	#covid19, australia, 2020, #covid19australia, march
5	#coronaviruspandemic, #covid-19, #coronavirusupdates, #coronapocalypse, #covid-19uk
6	China, #wuhanvirus, #chinesevirus, #chinaliedpeopledied, #chinavirus
7	#coronavirus, #covid2019, #virus, #corona, #breaking
8	#china, world, chinese, #wuhan, #ccp
9	Will, today, measures, essential, close
10	#coronavirusoutbreak, #coronavirusupdate, #covid_19, #coronacrisis, #covid2019

3.2.6 Community 15

Figure 14a shows the network diagram for Community 15. This community included 364, densely connected accounts. The four most central nodes within this community are highlighted in red in Fig. 14a. Of these four most central accounts, three have since been deleted.

Figure 14b and c show the overall tweet traffic driven by this community. These show a vague diurnal pattern, though there is still a constant stream of activity through all hours of the day and night. As the weeks progress the overall volume of tweets put out by these accounts increases.

The top 10 topics extracted from this community's tweets are shown in Table 8. Every single topic in this table, save for Topic 8, directly mentions China and the Chinese Communist Party (CCP). Most accounts within the community claim to be located in Hong Kong and post a constant stream of anti-Chinese sentiments. Overall, 67% of the tweets posted by this community included the term 'China'. Similar to Community 8, this is the kind of sharp focus on a singular subject that is unlikely to be observed from a network of human Twitter accounts, even if those accounts are quite passionate about some subject.

3.2.7 Community 17

Figure 15a shows the network diagram for Community 17. This community only has 51 total accounts, but they are very densely connected with the second highest density of all communities identified.

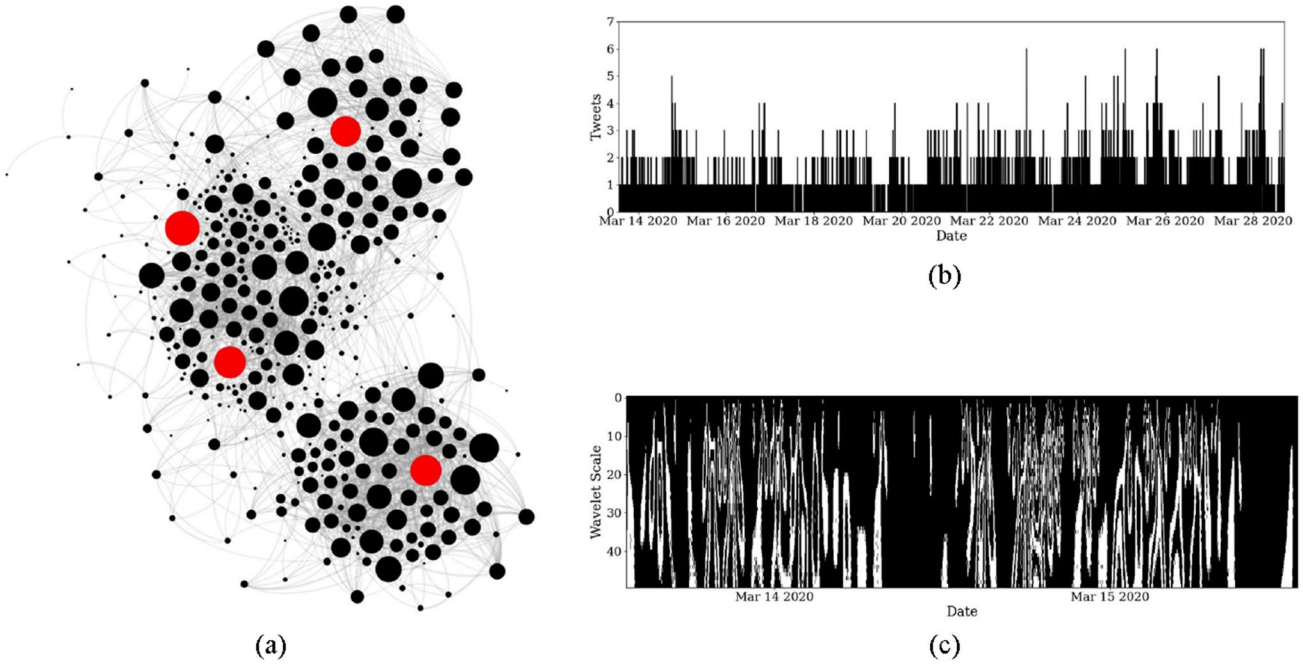

Fig. 14 Results from Community 15. **a** Shows the network diagram of all nodes, those in red are the most central nodes driving traffic. **b** Shows the overall tweet signal from the network. **c** Shows the DWFP transform of the network traffic for the first 2 days

Table 8 Top 10 topics from Community 15

Topic	Topic terms
1	#coronavirus, #china, #who, #taiwan, #covid-19
2	#covid19, #hongkong, hong, #coronavirusoutbreak, kong
3	China, coronavirus, pandemic, make, global
4	People, #ccpchina, like, #boycottchina, many
5	#ccp, hold, accountable, petition, sign
6	Chinese, call, #chinavirus, spread, communist
7	#chinazi, #chinesecoronavirus, #chinesevirus, #communistvirus, #chinaliedpeopledie
8	Cases, virus, corona, total, deaths
9	#wuhanvirus, #chinaliedpeopledied, #chinesevirus, #chinavirus, #chinaisasshoe
10	#ccpvirus, #covid2019, #coronaviruspandemic, #coronavirusoutbreak, #coronaviruschina

Since this community is so densely connected, similar to Community 8, there are no central nodes. Instead, this community seems to work as a collective. Figure 9b and c show how this community posts in both the time domain as well as the DWFP. Unlike all other communities discussed above, this community only seems to post at specific times with targeted messaging, a hallmark of TSCs. Unlike CSCs, which show constant activity throughout the data, TSCs show very little activity, until all at once they begin posting about some specific event or cause.

Table 9 shows the top 10 topics posted by Community 17. All of these topics focus on keeping oneself clean and protected from Covid-19 based on the teachings of an Indian religious leader Dr. Gurmeet Ram Rahim (@gurmeetramrahim). The spike occurring on March 24, 2020 in Fig. 15b represents a community posting about Dr. Gurmeet Ram Rahim's most recent advice to increase one's immunity to Covid-19, which was to eat paneer (an Indian cheese) and pistachios and practice yoga and meditation. Topics 3, 9, and 10 all show topics relevant to this spike. This shows clear inauthentic behavior in how these accounts are promoting this Doctor's advice, which also does not seem to be supported by any medical science—at least that they provide within their messages.

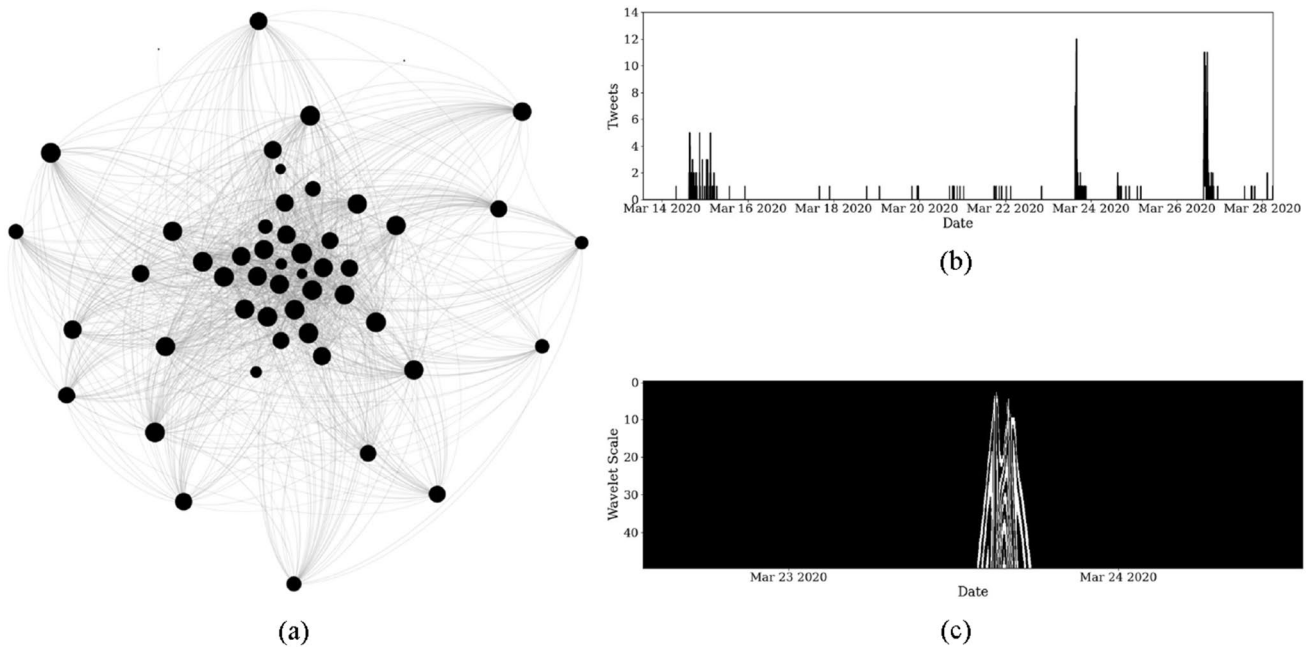

Fig. 15 Results from Community 17. **a** Shows the network diagram of all nodes. **b** Shows the overall tweet signal from the network. **c** Shows the DWFP transform of the network traffic for the first 2 days

Table 9 Top 10 topics from Community 17

Topic	Topic terms
1	Important, understand, spreads, distancing, lockdown
2	Stay, home, away, community, ones
3	Food, physical, mental, panner, recommend
4	Safe, urge, everyone, give, government
5	Wash, hand, importance, prevent, hands
6	India, days, chain, break, lockdown
7	#coronavirusupdates, #covid-19, clean, traditional, greet
8	Help, will, maintain, patients, cleanliness
9	@gurmeetramrahim, @derasachasauda, spread, stop, aware
10	Daily, immunity, #stayhomestaysafe, paneer, minutes

3.2.8 Community 35

Community 35 shows the most obvious bot-like behavior of any community within the data. The network is shown in Fig. 16a and is almost completely connected. Over 99% of possible edges exist. There is no central node, but instead all nodes act as a collective. Furthermore, this network is isolated from all other networks within this dataset as there is no edge between any one of these accounts and any other account within the network in Fig. 6. This was common among TSCs.

Figure 16b and c show the posting patterns from this community in both the time domain and the DWFP. These illustrate a community which rarely posted for the first 12 days of data. In total, this network posted 24 tweets before March 27th. However, seemingly unprompted, on March 27th around 7:30 UTC (3:30 EDT) this network sent out a flurry of about 1600 tweets over a 4 h span, which topped out at almost 50 tweets in a minute. Table 10 shows the top 10 topics identified from this community. All the included tweets were targeting the United States and many accused the U.S. of using Covid-19 as a bioweapon. Of these tweets, most were from a set of copy and pasted text including: “The one who spread H1N1, diphtheria, chikungunya, and cholera in #Yemen for 5 years of aggression is the one who spread #Coronavirus aroung (sic) the world. #US_is_the_enemy_of_humanity #5YearsOfWarOnYemen”. Every instance of this tweet even included the typo “aroung” and none of these were retweets, instead individual tweets posted by different, seemingly unrelated, accounts. Table 11 shows the top 10 most common tweets posted within this network of accounts. For each tweet, we also report the number of times that tweet appeared. Very few of these were retweets. For example, the first entry in Table 11 was posted 37 times and of those, only 4 were retweets. Every other instance was an account posting it independently. This is a very clear IO campaign using inauthentic behaviors to attempt to alter public discourse at this time. Their primary goal here was likely to get the hashtag ‘#5YearsOfWarOnYemen’ to trend and when outside user clicked on it they would see some of these tweets and further propagate this message.

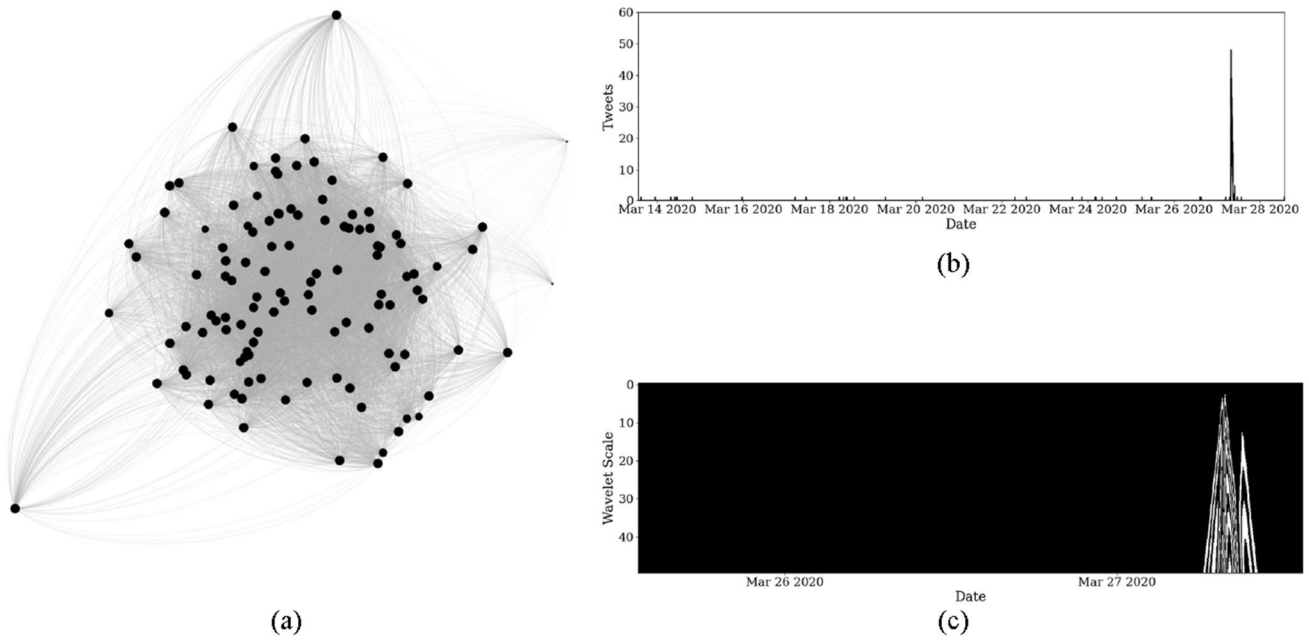

Fig. 16 Results from Community 35. **a** Shows the network diagram of all nodes. **b** Shows the overall tweet signal from the network. **c** Shows the DWFP transform of the network traffic during the spike in activity

Table 10 Top 10 topics from Community 35

Topic	Topic terms
1	Many, kill, direct, service, facilities
2	Import, equipment, drugs, medical, need
3	Suffer, blockade, continue, time, world
4	Siege, humanity, lift, impose, country
5	Spread, aggression, crimes, years, cruel
6	Biological, warfare, weapon, native, americans
7	Yemenis, greenlight, enter, hold, accountable
8	Reach, humanitarian, provide, carry, current
9	People, countries, yemeni, coronavirus, threaten
10	Health, destroy, system, sector, outbreak

4 Discussion

IO are campaigns operated by covert actors—i.e., political parties or large corporations—to influence public discourse through targeted media. These can, though do not always, include misinformation. In order to identify IO in the wild, we must be able to identify disparate accounts that are commonly posting and promoting similar information. In this work, we have shown the ability to identify densely connected communities of Twitter accounts using topic modeling with ONMF and isolating tweet storms within topics using RCT. The method presented here is novel because it does not rely on follower-following relationships, which can be noisy and easily gamed. Furthermore, follower-following relationships are difficult and time consuming to construct due to rate limits put in place by the Twitter API. Instead, we focus on the content being posted by accounts to identify topics, and from those, we use the specific post timings to isolate networks of accounts that are commonly posting the same topics concurrently.

Two different types of communities were identified. The first, and most common, was the CSC. These communities were commonly much larger and less densely connected than the other type—TSCs. CSCs generally center around several large retweet spam accounts. These accounts commonly have a follower ratio—ratio of users that follow them to those they follow—close to one. This is a common way of building out one's footprint on Twitter, by following all those who follow you.

CSCs could be IO networks pushing specific agendas or they could be groups of like-minded accounts that share similar information. For example, Community 4 is a CSC that promotes far-right American Conservative narratives, including the QAnon conspiracy theory. It is unclear if this was an IO that was driven by some covert actor(s) looking to push this conspiracy, or if we had discovered a network of QAnon accounts. Either way, it is useful to be able to localize this information online so fact checkers can combat the dangerous misinformation they are spewing, and misinformation can be traced backward to the source. Not all these communities put out harmful misinformation, of course.

Table 11 Top 10 most commonly copied tweets from Community 35

Tweet text	Posts
60% of the health facilities in #Yemen went out of service due to direct targeting by US-Saudi airstrikes, which killed many patients. You can imagine the catastrophe if Saudi was able to enter #Coronavirus to #Yemen! #5YearsOfWarOnYemen	37
UAE & Saudi aim at entering #Coronavirus to #Yemen with a greenlight from US to kill more Yemenis. Yemenis hold the US accountable #5YearsOfWarOnYemen	34
Listen to the voice of humanity and lift the siege imposed on #Yemen in order for us to be able to combat #Coronavirus if it reaches our country! #5YearsOfWarOnYemen	31
The US internationally-banned bombs killed too many innocents in #Yemen, & caused chronic diseases to many. However, US and its agents in the area aim to enter #Coronavirus to kill more #Yemenis #5YearsOfWarOnYemen	31
We call upon WHO to carry out responsibility in its humanitarian mission to provide medications and medical equipment to combat #Coronavirus if it reached #Yemen. In the current health crisis, WHO is held accountable for its role #5YearsOfWarOnYemen	31
A #Coronavirus outbreak in #Yemen will cause a serious disaster after the US-Saudi coalition destroyed the health system. #5YearsOfWarOnYemen	28
Although the whole world has ramped up their efforts to face the #Coronavirus pandimic, #WHO has not provided anything to help the battered health system in #Yemen to contribute to #Coronavirus response in case it sneaked in to a besieged country! #5YearsOfWarOnYemen	28
Battering the health sector in #Yemen by airstrikes and blocking medicine and medical equiment from entering, and aiming at entering #Coronavirus to Yemen are the goals of the US-Saudi coalition #5YearsOfWarOnYemen	27
Biological warfare has been used by US and its allies in their aggression on #Yemen for 5 years killing too many children and women by deadly viruses and diseases. The world shall put an end to the American Tyranny in the situation of #Coronavirus #5YearsOfWarOnYemen	27
If the US and countries of aggression coalition wouldn't lift the blockade on #Yemen under the #Coronavirus pandemic that threatens all humanity, when will they?! #5YearsOfWarOnYemen	27

Communities 3 and 7 are networks of Indian accounts that are promoting and spamming Covid-19 information and protocols from the Indian Government to ensure as many people understand what is happening as possible.

TSCs often exhibit obvious automated and coordinated behaviors occurring within the Twittersphere. The best example of this from the data was Community 35. This was a network of accounts that claimed to be located in Yemen and all at once posted the exact same tweets—even down to the typos—spamming a hashtag and pushing their disinformation about Covid-19 being a U.S. derived bioweapon. Not all of these types of communities are obviously nefarious in nature. For example, Community 21 showed the same behavior, spamming promotions for meditation and yoga videos for people to do while under lockdown. Though this still highlights a coordinated effort to distort and alter public discourse.

Being able to identify these networks is the first step in identifying an IO campaign. Presumably, these accounts posting about Covid-19 being a bioweapon are part of a broader campaign, and if we are able to isolate this kind of behavior, then the obvious next step is to use the key terms and hashtags these accounts are pushing to identify more accounts involved within this campaign. These accounts need not be isolated to Twitter either. Tracing these kinds of narratives through various social media channels, including main stream platforms such as Reddit and Facebook or more niche ones such as 4Chan or Gab, will provide researchers with the ability to trace derived narratives back toward their online origins to identify the source(s). This is especially important during the onset of a global catastrophe such as the Covid-19 pandemic, because there is a vacuum of information. When this happens unscrupulous actors can fill the void with anything they please such as disinformation and conspiracy theories. We have shown the ability to highlight the networks pushing this kind of bunk using RCT in the work presented here. This will be necessary as the information ecosystem continues to evolve on social media, and these actors become more proficient in exploiting vacuums in collective knowledge.

4.1 Limitations

In the current state, this method identifies Coordinated Networks of accounts that all post within the same tweet storms. However, this method does not distinguish those networks that are part of IO campaigns from those that might simply be networks of friends or like-minded individuals posting together. In the above analysis, we looked at each CN and discussed their behaviors and the underlying accounts that make them up. However, to identify if these networks are truly part of an IO campaign we need to analyze each network in a sophisticated way. This includes looking at the specific posting patterns of the CN as well as the individual accounts that make it up.

Another limitation of the presented method is in the hashtag preclustering phase. While this step was vital to allow the ONMF model to extract the topics necessary to build out the tweet storm networks, it also dropped a significant amount of data from the analysis. In future work, this can be addressed by creating a preclustering step that not only clusters by hashtag, but also by keywords that can be calculated using the burst term measurement illustrated by Mehrotra et al. (2013).

The overall timing of the model itself also represents something of a limitation. Overall, the model takes about a few days to a week to completely run. This includes running the ONMF over the hashtag clusters on the high performance computing cluster provided by William & Mary, which took anywhere from 15 min to several hours per cluster depending on the total number of tweets included. Once all the topics were calculated from the hashtag clusters the tweet storm extraction and network construction phases took about a day. This is far too long for a real time application of this model where we want to identify networks in IO campaigns as they operate. Much of this can be improved by running both the ONMF phase and the tweet storm extraction phases in parallel across a distributed network. This would drastically cut the overall amount of time required.

5 Conclusion and future work

In this work, we presented a new process for extracting tweet storms and uncovering networks of accounts that are working in a coordinated fashion using ridge count thresholding (RCT). To do this, we started with a dataset of 60 million individual tweets from the early weeks of the Covid-19 pandemic. These tweets were preclustered into similar sets of tweets using shared hashtags. Each of these clusters was then passed through an ONMF model to extract topics from within the data. Each topic is described by a topic-tweet signal, crafted using the time stamp included in each tweet's metadata. These signals were broken down into tweet storms using RCT, which is calculated from the DWFP transform of each topic-tweet signal. Each tweet storm described a time of increased activity around a topic. Tweet storms identified in this way each represent some behavior in the underlying network. We use this to identify networks of accounts that commonly co-occur within these tweet storms to identify those communities most responsible for driving narratives and pushing stories through social media. Through this process, we were able to identify 22 total networks of accounts that were densely connected based on RCT tweet storm identification. These accounts were more likely to be bots based on the Botometer score than an average account in the overall data. We were also able to identify the specific narratives and stories that these networks were pushing, by isolating the network's tweets and running topic modeling on those subsets of tweets.

Many of the identified Communities discussed in the analysis are still quite large and appear to be quite noisy. Thus, future work will be necessary to optimize the parameters used in this work. This includes both DWFP parameters as well as the upper and lower thresholds necessary for RCT. These will need to be tuned with a dataset in which IO campaigns are known to exist or at least one with known IO narratives. Beyond just tuning the parameters, more research needs to be done on the actual structure and posting patterns of individual communities. In this work, we saw that most communities included a few central spam accounts with many other, smaller accounts surrounding them in support. Further research can be put in to identifying if this is a standard construction of an IO campaign and, if so, what specific posting patterns can we identify with the DWFP that sets these networks apart from other networks of regular users.

Acknowledgements The authors acknowledge William & Mary Research Computing for providing computational resources and/or technical support that have contributed to the results reported within this paper. URL: https://www.wm.edu/it/rc.

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

Declarations

Conflict of interest The authors have no relevant financial interests to disclose.

References

Abu-El-Rub N, Mueen A (2019) Botcamp: bot-driven interactions in social campaigns. In: The world wide web conference. ACM, pp 2529–2535

Alba D, Frenkel S (2020) Medical expert who corrects Trump now a target of the far right. The New York Times

Baines D, Elliott RJ et al (2020) Defining misinformation, disinformation and malinformation: an urgent need for clarity during the Covid-19 infodemic. Technical report

Banda JM, Tekumalla R (2020) A Twitter dataset of 40+ million tweets related to COVID-19. https://doi.org/10.5281/zenodo.3723940

Barnes JE, Sanger DE (2020) Russian intelligence agencies push disinformation on pandemic. The New York Times

Barrett B (2020) Russia doesn't want Bernie Sanders. It wants chaos. Wired

Bastian M, Heymann S, Jacomy M (2009) Gephi: an open source software for exploring and manipulating networks. http://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154

Bastick Z (2020) Would you notice if fake news changed your behavior? An experiment on the unconscious effects of disinformation. Comput Hum Behav 116:106633

Bernstein J (2021) Bad news: selling the story of disinformation. Harper's Magazine

Berry MW, Browne M, Langville AN, Pauca VP, Plemmons RJ (2007) Algorithms and applications for approximate nonnegative matrix factorization. Comput Stat Data Anal 52(1):155–173

Bertoncini CA, Hinders MK (2010) Fuzzy classification of roof fall predictors in microseismic monitoring. Measurement 43(10):1690–1701. https://doi.org/10.1016/j.measurement.2010.09.015

Bertoncini CA, Rudd K, Nousain B, Hinders M (2012) Wavelet fingerprinting of radio-frequency identification (RFID) tags. IEEE Trans Ind Electron 59(12):4843–4850. https://doi.org/10.1109/TIE.2011.2179276

Beskow DM, Carley KM (2020) You are known by your friends: leveraging network metrics for bot detection in Twitter. In: Open source intelligence and cyber crime. Springer, pp 53–88

Bessi A, Ferrara E (2016) Social bots distort the 2016 U.S. presidential election online discussion. First Monday 21(11-7)

Bingham J, Hinders M, Friedman A (2009) Lamb wave detection of limpet mines on ship hulls. Ultrasonics 49(8):706–722. https://doi.org/10.1016/j.ultras.2009.05.009

Bird S, Klein S, Loper E (2009) Natural language processing with Python: analyzing text with the natural language toolkit. O'Reilly Media, Inc, Sebastopol

Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. J Stat Mech Theory Exp 10:10008

Boshmaf Y, Logothetis D, Siganos G, Lería J, Lorenzo J, Ripeanu M, Beznosov K (2015) Integro: leveraging victim prediction for robust fake account detection in OSNS. NDSS 15:8–11

Bradshaw S (2019) Disinformation optimised: gaming search engine algorithms to amplify junk news. Internet Policy Rev 8(4):1–24

Bradshaw S, Howard PN (2018) The global organization of social media disinformation campaigns. J Int Affairs 71(1.5):23–32

Broniatowski DA, Jamison AM, Qi S, AlKulaib L, Chen T, Benton A, Quinn SC, Dredze M (2018) Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate. Am J Public Health 108(10):1378–1384

Conger K (2021) Twitter, in widening crackdown, removes over 70,000 QAnon accounts. New York Times

Coppins M (2020) The billion-dollar disinformation campaign to reelect the president: how new technologies and techniques pioneered by dictators will shape the 2020 election. The Atlantic

Cresci S, Di Pietro R, Petrocchi M, Spognardi A, Tesconi M (2017) The paradigm-shift of social spambots: evidence, theories, and tools for the arms race. In: Proceedings of the 26th international conference on world wide web companion. International World Wide Web Conferences Steering Committee, pp 963–972

Del Vicario M, Vivaldo G, Bessi A, Zollo F, Scala A, Caldarelli G, Quattrociocchi W (2016) Echo chambers: emotional contagion and group polarization on Facebook. Sci Rep 6:37825

Ding C, Li T, Peng W, Park H (2006) Orthogonal nonnegative matrix t-factorizations for clustering. In: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, pp 126–135

Ferrara E (2017) Disinformation and social bot operations in the run up to the 2017 French Presidential Election. First Monday 22(8)

Ferrara E (2020) # Covid-19 on Twitter: bots, conspiracies, and social media activism. arXiv preprint arXiv:200409531

Ferrara E, Varol O, Davis C, Menczer F, Flammini A (2016) The rise of social bots. Commun ACM 59(7):96–104

Guynn J (2020) 'Significant and growing public health challenge,' Twitter cracks down on COVID-19 vaccine misinformation. USA Today

Hinders MK (2020) Intelligent feature selection for machine learning using the dynamic wavelet fingerprint. Springer, Berlin. https://doi.org/10.1007/978-3-030-49395-0

Hou J, Hinders MK (2002) Dynamic wavelet fingerprint identification of ultrasound signals. Mater Eval 60(9):1089–1093

Hou J, Leonard KR, Hinders MK (2004) Automatic multi-mode lamb wave arrival time extraction for improved tomographic reconstruction. Inverse Probl 20(6):1873–1888. https://doi.org/10.1088/0266-5611/20/6/012

Howard PN, Kollanyi B (2016) Bots, #strongerin, and #Brexit: computational propaganda during the UK-EU referendum. Available at SSRN 2798311

Hunnicutt T, Bose N (2021) Biden orders review of COVID origins as lab leak theory debated. Reuters

Hurtado S, Ray P, Marculescu R (2019) Bot detection in reddit political discussion. In: Proceedings of the fourth international workshop on social sensing, pp 30–35

Jefferson T (1807) From Thomas Jefferson to John Norvell, 11 June 1807. https://founders.archives.gov/documents/Jefferson/99-01-02-5737

Keller FB, Schoch D, Stier S, Yang J (2020) Political astroturfing on Twitter: how to coordinate a disinformation campaign. Polit Commun 37(2):256–280

Kirn SL (2021) Uncovering information operations on Twitter using natural language processing and the dynamic wavelet fingerprint. Doctoral dissertation, The College of William and Mary

Kirn SL, Hinders MK (2020) Dynamic wavelet fingerprint for differentiation of tweet storm types. Soc Netw Anal Min 10(1):4

Kirn SL, Hinders MK (2021) Bayesian identification of bots using temporal analysis of tweet storms. Soc Netw Anal Min 11(1):1–17

Kormann C (2021) The mysterious case of the Covid-19 lab-leak theory. The New Yorker

LaFrance A (2020) The prophecies of Q: American conspiracy theories are entering a dangerous new phase. The Atlantic

Liu PL, Huang LV (2020) Digital disinformation about Covid-19 and the third-person effect: examining the channel differences and negative emotional outcomes. Cyberpsychol Behav Soc Netw 23(11):789–793

Mehrotra R, Sanner S, Buntine W, Xie L (2013) Improving LDA topic models for microblogs via tweet pooling and automatic labeling. In: Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval, pp 889–892

Miller CA, Hinders MK (2014) Classification of flaw severity using pattern recognition for guided wave-based structural health monitoring. Ultrasonics 54(1):247–258. https://doi.org/10.1016/j.ultras.2013.04.020

Mimno D, Wallach H, Talley E, Leenders M, McCallum A (2011) Optimizing semantic coherence in topic models. In: Proceedings of the 2011 conference on empirical methods in natural language processing, pp 262–272

Mueller RS (2019) Report on the investigation into Russian interference in the 2016 presidential election. US Department of Justice, Washington

Nguyen A, Catalan D (2020) Digital mis/disinformation and public engagement with health and science controversies: fresh perspectives from Covid-19. Media Commun 8(2):323–328

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011) Scikit-learn: machine learning in Python. J Mach Learn Res 12:2825–2830

Pierri F, Artoni A, Ceri S (2020) Investigating Italian disinformation spreading on Twitter in the context of 2019 European elections. PLoS One 15(1):e0227821

Rauchfleisch A, Kaiser J (2020) The false positive problem of automatic bot detection in social science research. Berkman Klein Center Research Publication, Cambridge

Řehůřek R, Sojka P (2010) Software framework for topic modelling with large corpora. In: Proceedings of the LREC 2010 workshop on new challenges for NLP frameworks. ELRA, Valletta, Malta, pp 45–50

Rid T (2020) Active measures: the secret history of disinformation and political warfare. Farrar, Straus and Giroux, New York

Rooney M (2021) Characterization of wireless communications networks using machine learning and 3D electromagnetic wave propagation simulations. Doctoral dissertation, The College of William and Mary

Schild L, Ling C, Blackburn J, Stringhini G, Zhang Y, Zannettou S (2020) " Go eat a bat, chang!": an early look on the emergence of sinophobic behavior on web communities in the face of Covid-19. arXiv preprint arXiv:200404046

Schneier B (2020) Bots are destroying political discourse as we know it. The Atlantic

Sills J, Bloom JD, Chan YA, Baric RS, Bjorkman PJ, Cobey S, Deverman BE, Fisman DN, Gupta R, Iwasaki A, Lipsitch M, Medzhitov R, Neher RA, Nielsen R, Patterson N, Stearns T, van Nimwegen E, Worobey M, Relman DA (2021) Investigate the origins of COVID-19. Science 372(6543):694

Skinner E, Kirn S, Hinders M (2019) Development of underwater beacon for Arctic through-ice communication via satellite. Cold Reg Sci Technol 160:58–79. https://doi.org/10.1016/j.coldregions.2019.01.010

Tweepy (2017) Streaming with tweepy–tweepy 3.5.0 documentation. http://tweepy.readthedocs.io/en/v3.5.0/streaming_how_to.html

Van der Maaten L, Hinton G (2008) Visualizing data using t-SNE. J Mach Learn Res 9(11):2579–2605

Wang Y, McKee M, Torbica A, Stuckler D (2019) Systematic literature review on the spread of health-related misinformation on social media. Soc Sci Med 240:112552

Warzel C (2020) Twitter is real life. The New York Times

Woolley S (2020) The reality game: how the next wave of technology will break the truth. PublicAffairs

Yang KC, Varol O, Davis CA, Ferrara E, Flammini A, Menczer F (2019) Arming the public with artificial intelligence to counter social bots. Hum Behav Emerg Technol 1(1):48–61

Yang KC, Varol O, Hui PM, Menczer F (2019) Scalable and generalizable social bot detection through data selection. arXiv preprint arXiv:191109179

Yao Y, Viswanath B, Cryan J, Zheng H, Zhao BY (2017) Automated crowdturfing attacks and defenses in online review systems. In: Proceedings of the 2017 ACM SIGSAC conference on computer and communications security, pp 1143–1158

Zannettou S, Caulfield T, De Cristofaro E, Sirivianos M, Stringhini G, Blackburn J (2019) Disinformation warfare: understanding state-sponsored trolls on Twitter and their influence on the web. In: Companion proceedings of the 2019 world wide web conference, pp 218–226