

Measuring Community Resilience During the COVID-19 based on Community Wellbeing and Resource Distribution

Jaber Valinejad^{1*}, Zhen Guo², Jin-Hee Cho², Ing-Ray Chen²,

1 Harvard Medical School, Data and System Science Lab, Harvard University, Cambridge, MA, USA

2 The Department of Computer Science, Virginia Tech, National Capital Region Campus, Falls Church, VA, USA

* corresponding jvalinejad@mgh.harvard.edu

Abstract

The COVID-19 pandemic has severely harmed every aspect of our daily lives, resulting in a slew of social problems. Therefore, it is critical to accurately assess the current state of community functionality and resilience under this pandemic for successful recovery. To this end, various types of social sensing tools, such as tweeting and publicly released news, have been employed to understand individuals' and communities' thoughts, behaviors, and attitudes during the COVID-19 pandemic. However, some portions of the released news are fake and can easily mislead the community to respond improperly to disasters like COVID-19. This paper aims to assess the correlation between various news and tweets collected during the COVID-19 pandemic on community functionality and resilience. We use fact-checking organizations to classify news as real, mixed, or fake, and machine learning algorithms to classify tweets as real or fake to measure and compare community resilience (CR). Based on the news articles and tweets collected, we quantify CR based on two key factors, *community wellbeing* and *resource distribution*, where resource distribution is assessed by the level of *economic resilience, and community capital*. Based on the estimates of these two factors, we quantify CR from both news articles and tweets and analyze the extent to which CR measured from the news articles can reflect the actual state of CR measured from tweets. To improve the operationalization and sociological significance of this work, we use dimension reduction techniques to integrate the dimensions.

Introduction

Motivation

The recent outbreak of COVID-19 has disrupted every aspect of our daily lives. To absorb and adapt against COVID-19 in an agile manner and quickly recover from it, maintaining a healthy, socially connected, and prepared community is critical [1]. Community wellbeing is an essential asset to build a resilient community [2]. In addition, how resources are distributed in a community can present the community's resilience against a disaster like COVID-19. High accessibility to resources and their fair distribution are the keys to community resilience [3, 4]. Numerous sensing tools are available to assess community resilience, including online websites, social media, surveys, and infrastructure sensing. Among these sensing tools, while social media is an essential social sensing tool for revealing community behavior and thought, it has received little

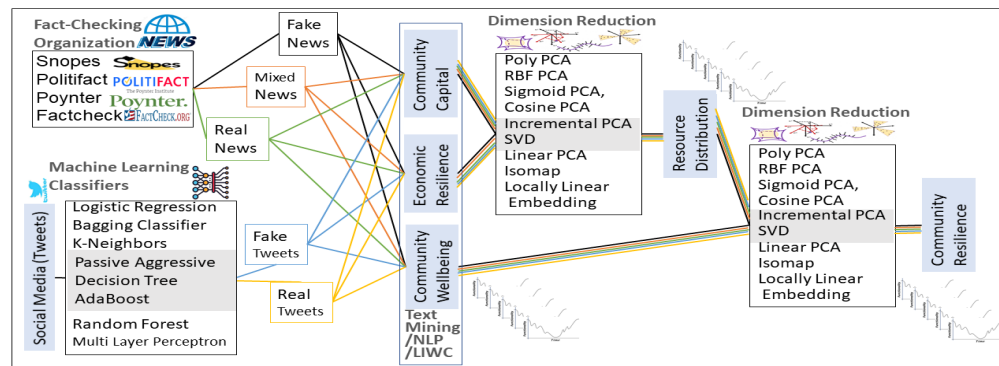


Fig 1. The proposed framework for assessing community resilience of various types of news/tweets via machine learning, natural language processing, and dimension reduction techniques.

attention previously. Seven out of ten Americans use social media to exchange personal information, interact with content, and connect with others [5]. According to a recent research [6], the psychological states of a whole population can be revealed through social media. Social media provides a platform for billions of users to communicate, express sentiments, and provide real-time updates about human interaction on a large scale [7]. Twitter is one of the major community social media platforms. In this regard, numerous studies have employed tweeter to evaluate population behavior [6, 7, 8, 9, 10, 11]. Unfortunately, fake news may negatively impact maintaining community wellbeing and equitable resource distribution during COVID-19. The Internet, social media, and mass media platforms have generated a large volume of information flow during the COVID-19. Part of the information volume spreads false information (e.g., misinformation or disinformation), rumors, fake news, or hoaxes [12]. Fake news is usually observed as more novel than real news; in addition, it flows on social/mass media noticeably faster, farther, and more broadly than real news [13]. Fake news has been commonly used to manipulate and propagate false information by appealing to users' ideological perspectives, emotions, and desires to spread their views to other people [14]. Thus, the dissemination of fake news via social/mass media may have an effect on people's social behavior. Social behavior changes can affect people's well-being and resource distribution, resulting in changes in community resilience. However, prior studies have rarely assessed community resilience via social media and have rarely investigated the correlation between various types of news and tweets from the community resilience's point of view.

Research Goal, Contributions, and Questions

In this work, we aim to quantify community resilience (CR) in terms of community wellbeing (CW) and resource distribution (RD). These two factors are quantified by natural language processing (NLP) tools on news articles that include real, mixed (i.e., half fake and half real), and fake news as well as tweets including real and fake tweets. We also examine the correlation between the measured CR from news articles and the actual state of CR captured from tweets on Twitter.

In Fig. 1, we illustrate our proposed framework for measuring community resilience of various types of news/tweets using machine learning, natural language processing, and dimension reduction techniques.

The **key contributions** of this work are as follows:

1. We develop novel community resilience metrics inspired by the system resilience

metric in the cybersecurity domain [15]. We define community resilience in terms of a community's absorption (or fault tolerance) to, adaptability to, and recoverability from attacks or failures (e.g., disasters). Specifically, we measure community resilience based on two attributes, namely, community wellbeing and resource distribution. We measure community wellbeing based on mental and physical wellbeing. We estimate resource distribution based on economic resilience, and community capital. To the best of our knowledge, measuring CR based on social media information has been rarely studied.

2. This work is the first to use news articles and Twitter to assess community resilience during the COVID-19. We use fact-checking to collect 4,952 full-text news articles and categorize them as real, mixed, or fake news. In addition, we retrieve tweets from 42,877,312 tweets IDs from Jan. 2020 to Jun. 2021. We use the top three machine learning (ML) algorithms, i.e., Passive-Aggressive Classifier, Decision Tree Classifier, and AdaBoost Classifier, to identify if a tweet is real or fake.
3. To boost the sociological significance of this work, we use dimension reduction techniques, including linear transformations, nonlinear transformations, and manifold learning to integrate various dimensions of community resilience. We will show that while the incremental principal component analysis (PCA) keeps temporal dependency information, it has a greater level of variance information ratio.
4. We analyze the correlation between measurements of CR attributes by each type of news (i.e., real, mixed, or fake) and tweets (i.e., real or fake). From this analysis, fake news is shown to influence people's behaviors towards undesirable states, undermining CR in reality. Moreover, the CR measured based on real or mixed news articles can reflect actual states of the CR measured from tweets.
5. We conduct a resilience analysis of various types of news (i.e., real, mixed, or fake) and tweets (i.e., real and fake) via an output-oriented analysis to show the values of each CR attribute over time, as well as a capacity-based analysis to demonstrate the time-averaged CR measurements. We also conduct statistical analyses to examine the correlation of CR attributes measured from news and tweets.

Our study will answer the following **research questions**:

1. *What are the main trends observed in community resilience and its key attributes, i.e., community wellbeing and resource distribution?*
2. *What are the key differences and correlations between the community resilience measured on various types of news and tweets?*
3. *What are the levels of the community resilience metrics, e.g., absorption and recovery during COVID-19 on various types of news and tweets?*

Research Assumptions and Limitations

We conduct our study by assuming the following intuitions. First, real tweets/news can represent community resilience better than mixed/fake tweets/news. Second, knowing a current situation with accurate information can lead people to make more rational decisions to handle a faced disaster, which is COVID-19 in this work. Although the scope of this work is limited to measuring and analyzing community resilience using tweets and news, further investigation to prove the above as the hypothesis will be conducted in our future work. As no research work cannot be faultless, our work also has a number of limitations:

1) While we gather all real and fake news propagated by the media, we only use Twitter to investigate the population's behavior. To analyze population behavior, surveys can provide high-quality data, albeit at a cost. While a national survey is beneficial, it is an expensive and time-consuming endeavor. Due to the fact that we wish to track multiple metrics of community resilience over an extended period of time, the availability of datasets is critical. Note that there is a trade-off between the quality, sample size, period, availability, and cost of datasets. Further research can compare the correlations between fake/real news and surveys.

2) We use anxiety, anger, and sadness to determine the level of community wellbeing. Additional wellbeing metrics can be added. This necessitates the development of new techniques for assessing other possible community well-being indicators.

3) While Twitter may not be representative of the US population, it can provide insight into how people live their lives. Nonetheless, considering additional social media platforms may be beneficial for future research.

Related Work

Community resilience (CR) refers to the ability of a social system to absorb the impact of the stress and cope with threats and adapt to post-event situations by reorganizing, changing, or learning to cope with the threat from the disasters [16, 17]. This definition is well aligned with the general concept of system resilience in terms of its fault tolerance (i.e., functioning under threats or errors), adaptability (i.e., adapting to disruptions), and recoverability (i.e., recovering quickly from the disrupted situations) [15]. Community resilience has been measured based on various types of metrics [18, 19, 20]. CR can be defined differently depending on different disasters faced in the past [21]. However, it has been commonly considered with a measure of resilience whether a society functions in terms of social, economic, institutional, infrastructure, community capital, and ecological aspects [22, 23].

Work [24] proposed the *wellbeing theory* discussing a measure of community wellbeing in terms of positive emotions, engagement, relationships, meaning, and accomplishment. [1] discussed 'health' in terms of behavioral, physical, social, and environmental wellbeing. Higher psychological wellbeing can introduce higher sustainability, equality, resilience, and inclusion [1, 24]. The key factors impacting people's resilience to disasters were also studied, such as family distress, available support systems, disruption of school/job programs, or loss of loved ones/property [25].

The distribution state of physical and social resources is another indicator of community resilience. Physical resources consist of critical infrastructures, electricity, water, food, medicine, emergency services capacity, transit capacity, grocery, pharmacy, or workplaces. Social resources include community capital and institutional resources [26], which allow people to interact with other people for their social activities. During the COVID-19, we observed aggressive panic buying behaviors of food, toilet papers, and sanitary products across countries or regions such as Singapore [27], Hong Kong [28], and Chinese mainland [29]. This is known to reduce community resilience due to a lack of balanced resource distribution.

Social media activities influence community resilience [30] in terms of social wellbeing and community capital. Official and informal sources use social media to spread information to handle a disaster for public safety, such as social distance, sanitation, food or transportation availability, or business hours. In addition, social media provide good networking tools to engage people with a community or government guidance [31]. However, false information has often been propagated through social media, such as fake news or rumors, which can easily amplify fear, anxiety [27, 32], outright racism, disgust, and mistrust [28]. These unnecessary misperceptions have been

the key to triggering irrational, undesirable responses to disasters. In the literature, people's responses and behaviors to the COVID-19 have been measured by analyzing social media information. The examples include emotions and psychological states extracted from the datasets of Weibo users using the linguistic inquiry, word count (LIWC) framework [33, 34], risk perception, negative emotions (e.g., sadness, anger, anxiety), and behavioral responses (e.g., panic buying) to COVID-19 from the dataset of Sina Weibo, Baidu search engine, and Ali e-commerce marketplace using LIWC [35]. Aggressive panic buying behaviors were more prominently observed when more misinformation or rumors on the COVID-19 were disseminated [35]. Emotions (e.g., surprise, disgust, fear, anger, sadness, anticipation, joy, and trust) in replies were also captured from real and false tweets using the National Research Council Canada (NRC) [36] and LIWC [13]. Mingxuan *et al.* [37] measured people's mental health based on emotions extracted from social media data, which was analyzed using machine learning (ML) or NLP techniques [38, 39].

However, to our knowledge, no prior work has estimated community resilience based on community wellbeing and resource distribution using both social media news articles (i.e., real, mixed, and fake) and tweets (i.e., real and fake) to compare their measurements and investigate their correlations.

Measurement of Community Resilience Using Social Media Information

In this section, we discuss how community resilience is measured using social media information, including both news articles and tweets.

Community Resilience Metrics

We measure the community functionality in terms of community wellbeing and resource distribution. Fig. 2 represents the community functionality, $CF(t)$, with time t . We define community resilience based on the concept of system resilience [15], consisting of absorption (i.e., fault tolerance), adaptability, and recoverability. We interpret the time until a community does not function as the time period for absorption, namely TFA (i.e., time from t_0 to t_1). Absorption (ABS) refers to the community's capacity to absorb the shock and adverse effects caused by COVID-19. High TFA implies that the community tolerates hardships introduced by a disaster so that the community can still function by providing at least critical, minimum services, such as food, employment, schools, or health services. Note that a higher absorption is more desirable. Community Non-Functioning (CNF) is a term that refers to situations in which the community's functionality falls below a critical threshold. We denote the deadlock functionality threshold by b . We call the time from t_1 to t_3 the *time under community non-functioning* (TNF). A shorter TNF is considered more desirable, representing fast failure and fast recovery. By following the conventional concept of system reliability, the *mean time to recovery* (MTTR), we defined the time to recovery (TTR) estimated from the time the community reaches a critical functionality point (t_1) to the time it fully recovers from the disaster and reaches at the initial normal state (t_4). Recovery (REF) refers to the community's capacity to recover from COVID-19. The recoverability effectiveness (RE) refers to how much the community has recovered from the minimum functionality point, t_2 , to the current point at t_4 . Note that a higher level of recovery is more desirable. We consider the whole period from the outbreak of a disaster (e.g., COVID-19) to the time a community is fully recovered, t_4 , as the time period for adaptability (TA). Depending on how the community handles the disaster, TA may not face TNF but directly recover from a less functionality state to a full functionality state.

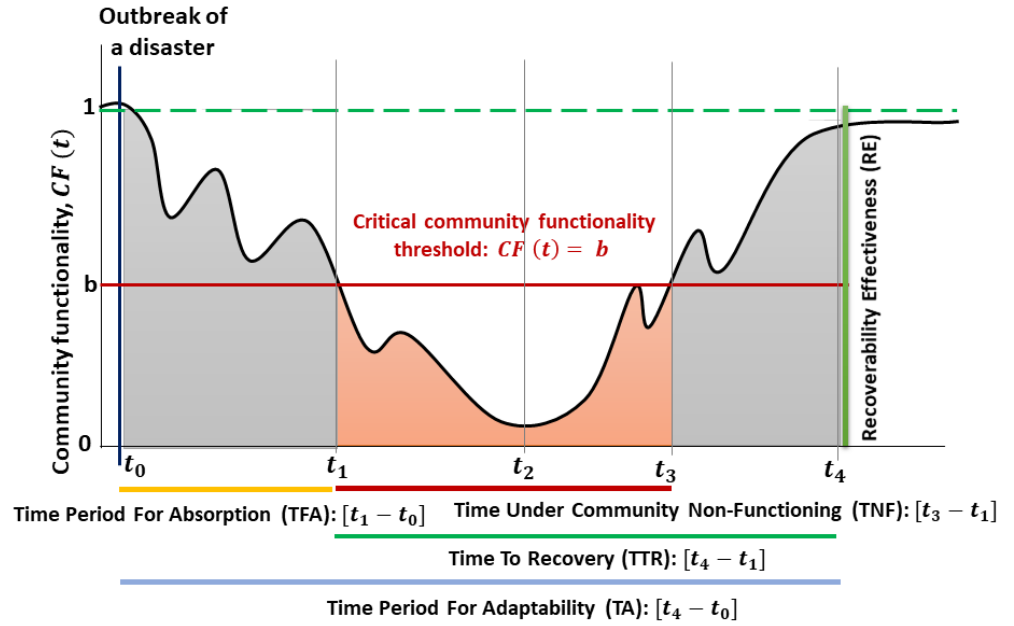


Fig 2. The evolution of community functionality ($CF(t)$) from the outbreak of a disaster (e.g., COVID-19) to the full recovery of a community.

Higher absorption, recovery, and adaptability are more desirable, which means the more area under the curve a community has, the more resilient it is.

We estimate $CF(t)$ based on the levels of community wellbeing ($CW(t)$) and resource distribution ($RD(t)$) at time t . Here, CR is measured by:

$$CR_{[a,b]} = \int_{t=a}^b CF(t) dt = \int_{t=a}^b f(CW(t), RD(t)) dt, \quad (1)$$

where $[a, b]$ denotes the time period used to calculate CR . Note that CW and RD are treated equally in this work. For fair consideration of each component, we use a normalized value of CW and RD as a real number ranging in $[0, 1]$ using *min-max scaling* [40]. Function f in its simple form can be the average of CW and RD . However, depending on the relative relevance of CR in a given domain, CW and RD can be weighted differently. In order to improve the operability of this work, we will explore the appropriate f function. We will demonstrate that the incremental PCA function is the best f function.

To determine the average CF during the period of the COVID-19, we measure ABS , CNF , and REF as follows:

- ABS is the average CF during the time period for absorption, which is given by:

$$ABS = \frac{\int_{t_0}^{t_1} CF(t) dt}{t_1 - t_0}. \quad (2)$$

- CNF is the average CF over the time under the critical area of CF , which is measured by:

$$CNF = \frac{\int_{t_1}^{t_3} CF(t) dt}{t_3 - t_1}. \quad (3)$$

We assume that a community is entirely dysfunctional when its CR is below the threshold b .

- *REF* refers to the average *CF* during the period of recovery, which is obtained by: 209

$$REF = \frac{\int_{t_1}^{t_4} CF(t)}{t_4 - t_1}. \quad (4)$$

Integrating community resilience components 210

Since community resilience encompasses a variety of dimensions, the manner in which these characteristics are interwoven is critical. One strategy is to use a weighted average. To improve the operationalization and sociological significance of this work, we use dimension reduction techniques, including linear transformations, nonlinear transformations, and manifold learning to combine two main dimensions into one dimension, i.e., resource distribution or community resilience. We use multiple dimension techniques to determine which one performs better. Thus, the polynomial (Poly) Kernel PCA, the Gaussian radial basis function (RBF) Kernel PCA, the sigmoid Kernel PCA, the cosine Kernel PCA, the incremental PCA, the linear PCA, the SVD, the isomap, and the Locally Linear Embedding are the methods used to calculate resource distribution and community resilience. The variance information ratio (derived using the eigenvalues' values), the reconstruction error, and the time-related correlations (time corr) are shown in Table 1. It is preferable to have a higher level of variance information ratio and a lower level of reconstruction error. While we are concerned with minimizing error, we also want to retain time-series information. In other words, this type of data is intrinsically associated with temporal dependency. As a result, we can determine the time-related correlation, or the correlation between the integrated result and each of its components. To be more precise, we calculate the correlation between resource distribution and each of community capital and economic resilience. Plus, we calculate the correlation between community resilience and each of community wellbeing and resource distribution. To maintain the temporal dependency information, at least one correlation should be positive. If two dimensions are raised, the integrated results should also increase. In Table 1, two techniques stand out, namely incremental PCA and SVD. Furthermore, based on the results of other techniques, there are scenarios in which there are two negative correlations. Because incremental PCA has a greater level of variance information ratio, we select it as the ultimate dimension reduction strategy. 211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236

Now we describe how to estimate CW and RD as below. 237

Measuring Community Wellbeing 238

A lack of community wellbeing (CW) under disasters can either increase people's vulnerability to early deaths or injuries, or trigger irrational behavior, such as panic buying [41]. Wellbeing is measured by the extent of people's moods, such as anxiety, depression, and anger, which have long been recognized as typical symptoms of wellbeing illness [42, 43, 44]. Therefore, we obtain the extent of community wellbeing from the features of *anxiety*, *sadness*, and *anger*, extracted from linguistic inquiry and word count (LIWC) categories. 239
240
241
242
243
244
245

Measuring Resource Distribution 246

Resource distribution (RD) also measures part of CR [3, 4, 26] where the high functioning in RD refers to the high ability that a community can provide services to its inhabitants related to economic, infrastructure, institutional, and community capital resources. We assume that sufficient and well-distributed resources can contribute to the community that can better resist, recover, and/or overcome a disaster. We measure RD in terms of how well each service is provided. RD is measured by: 247
248
249
250
251
252

$$RD = f(EF(t), CCF(t)), \quad (5)$$

Table 1. The variance information ratio (Var), the reconstruction error (error), and the time-related correlations (Time corr) of Resource distribution and community resilience by using the polynomial (Poly) Kernel PCA, the Gaussian radial basis function (RBF) Kernel PCA, the sigmoid Kernel PCA, the cosine Kernel PCA, the incremental PCA, the linear PCA, the SVD, the isomap, and the Locally Linear Embedding.

Integrated Metrics	News/Tweets	Info	Nonlinear Transformation				Linear Transformation			Manifold Learning	
			Poly Kernel PCA	RBF Kernel PCA	Sigmoid Kernel PCA	Cosine Kernel PCA	Incremental PCA	PCA	SVD	Isomap	Locally Linear Embedding
Resource distribution	Real News	Var/Error	0.953	0.936	0.996	0.989	0.983	0.983	0.437	1.92e-3	3.23e-8
		Time Corr	(-,+)	(-,+)	(-,+)	(-,+)	(-,+)	(-,+)	(+,-)	(-,+)	(+,-)
	Mixed News	Var/Error	0.915	0.903	0.981	0.986	0.944	0.950	0.411	6.31e-3	2.07e-7
		Time corr	(-,+)	(-,+)	(-,+)	(-,+)	(+,-)	(-,+)	(+,-)	(-,+)	(-,+)
	Fake News	Var/Error	0.698	0.578	0.858	0.891	0.563	0.563	0.518	3.72e-2	1.33e-6
		Time corr	(-,+)	(-,+)	(-,+)	(-,+)	(+,-)	(-,+)	(+,-)	(-,+)	(-,+)
	Real Tweets	Var/Error	0.749	0.729	0.724	0.714	0.739	0.740	0.579	3.74e-2	3.02e-8
		Time corr	(-,+)	(-,+)	(-,+)	(-,+)	(+,-)	(-,+)	(+,-)	(-,+)	(-,+)
Fake Tweets	Var/Error	0.679	0.634	0.623	0.944	0.521	0.625	0.508	4.57e-2	3.42e-6	
	Time corr	(+,+)	(+,+)	(-,+)	(+,-)	(+,+)	(+,+)	(+,+)	(+,+)	(-,+)	
Community resilience	Real News	Var/Error	0.757	0.594	0.654	0.953	0.640	0.642	0.420	4.1e-2	7.63e-7
		Time corr	(+,+)	(+,+)	(-,+)	(-,+)	(+,+)	(+,+)	(+,+)	(+,+)	(+,-)
	Mixed News	Var/Error	0.663	0.694	0.706	0.665	0.684	0.684	0.697	4.9e-2	4.9e-7
		Var corr	(-,+)	(-,+)	(-,+)	(-,+)	(+,+)	(-,+)	(+,+)	(-,+)	(+,+)
	Fake News	Var/Error	0.837	0.826	0.721	0.693	0.743	0.760	0.920	1.89e-2	4.11e-6
		Time corr	(-,+)	(-,+)	(-,+)	(-,+)	(+,+)	(-,+)	(+,+)	(-,+)	(-,+)
	Real Tweets	Var/Error	0.788	0.862	0.962	0.949	0.865	0.870	0.918	1.76e-2	7.15e-6
		Time corr	(-,+)	(-,+)	(-,+)	(-,+)	(+,+)	(-,+)	(+,+)	(-,+)	(+,-)
Fake Tweets	Var/Error	0.654	0.646	0.525	0.932	0.489	0.595	0.665	5.76e-2	8.3e-7	
	Time corr	(-,+)	(+,-)	(-,+)	(-,+)	(-,+)	(-,+)	(+,+)	(+,+)	(+,+)	
Average Var/ Error			0.775	0.738	0.785	0.909	0.706	0.743	0.519	3.13e-2	1.84e-6

where $EF(t)$ and $CCF(t)$ refer to the level of states related to economic, and community capital functioning, respectively, with an equal weight considered. Again, depending on the domain requirement, its weight can be differently considered. As discussed before, function f can be as simple as the average of $EF(t)$, and $CCF(t)$. However, in order to improve the operationalization of this work, We will demonstrate that the incremental PCA function is the best f function.

Each component of RD, including $EF(t)$, and $CCF(t)$, is measured by LIWC categories as follows:

- *Economic functionality* (EF) is the economic capacity of a given community before and after a disaster. The examples include housing capital, employment, income, signal sector employment dependence, or business sizes. Economic functioning is captured by extracting the amount of words related to money or work, such as the increased use of work-related (e.g., ‘job,’ ‘majors,’ ‘xerox’), money-related (e.g., ‘Audit,’ ‘cash,’ or ‘owe’) terms in the LIWC categories.
- *Community capital* indicates a community’s ability to provide social activity services to its inhabitants and build trust among them. We assess community capital in terms of the language patterns representing community cooperation using the LIWC categories as follows:
 - *Communication Efficiency*: The increased use of complex words and words with more than six letters has been identified as being inefficient for communication and cooperation [45]. To measure this, we calculate the opposite degree of ‘Words > 6 letters.’
 - *Group-Oriented Communications*: The frequent use of first-person pronouns, such as ‘we,’ ‘us,’ ‘our,’ indicates group interaction [46]. In psychological linguistics, it is known that assent-related languages (e.g., ‘agree,’ ‘OK,’ ‘yes’) point to group

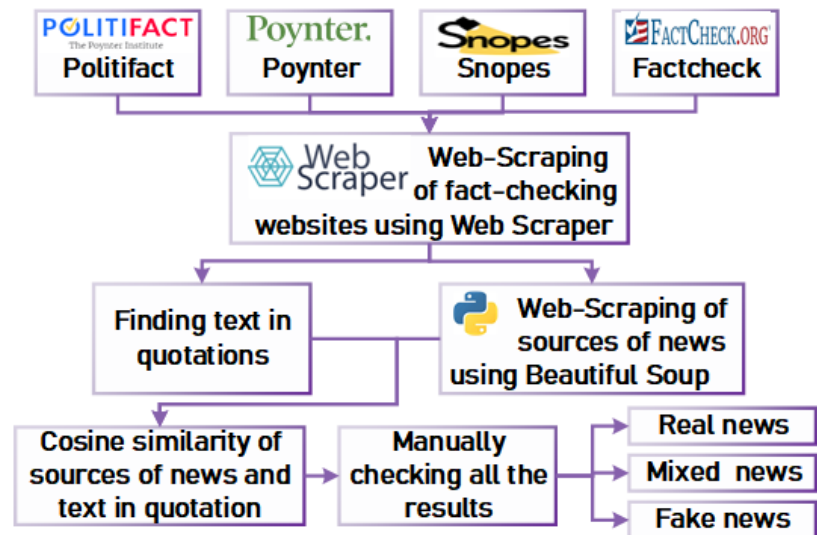


Fig 3. Collecting news based on web-scraping and manual cleaning.

consensus and cooperation [47]. Hence, we measure the frequency of words using the ‘first-person plural’ pronouns and ‘assent’ in the LIWC categories. 278 279

- *Social Process-Related Communications*: We measure increased social engagement and cooperation [48, 49] based on the frequency of social process languages obtained by ‘friend’ and ‘family.’ 280 281 282

The presence of more words within a category indicates a higher value. For fair comparison, we normalize the value of each attribute in CR by dividing the accumulated degree by the number of words, representing the extent of each attribute ranging in [0, 1] as a real number. Note that we can define community wellbeing, community capital, economic resilience, resource distribution, and community resilience in terms of absorption, adaptability, and recoverability components. 283 284 285 286 287 288

Procedures of Measuring CR via Social Media Information 289

Collecting News Using Web-Scraping 290

We describe the process of finalizing information associated with news in Fig. 3. The information includes the text of news articles, issues, subjects, misconceptions, and the title of news articles for all the articles published over time. We use a two-stage web-scraping method to collect these contents. The web crawling process begins with the Google Chrome Extension ‘Web Scraper – Free Web-Scraping’ [50]. This tool allows interaction with the website from which we scrape data to identify the HTML tags required to extract data from fact-checking websites. We can export the results as a CSV file containing external links to the original articles. Then, we use the Python library BeautifulSoup [51] to analyze external links and scrap the original articles and additional tags that were difficult to web-scrape with the first tool. Additionally, we extract the quotation’s text from news scraped from fact-checking organizations. Then, we compare the *cosine similarity* [52] of this quoted text to the news obtained via external links to choose the most appropriate news text automatically and double-check them manually. Note that we filter the so-called ‘most appropriate news’ by capturing the original news text. The original news text is filtered out by excluding text quoted from other sources. We leverage the automatic web-scraping techniques to capture only the original news text solely written by the author of the given news article. 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307

Table 2. News types based on the classifications of three fact-checking organizations

Type of news	Snopes	Politifact	Poynter	Factcheck
Real	true, mostly true	true, mostly true	-	-
Fake	mostly false, false	mostly false, false, pants on fire	fake	fake
Mixed	mixture	half true	-	-
Number	2413	927	1308	304

Table 3. The numbers of various types of news and tweets per month considered in this study.

Year		2020												2021						All
Month		Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.	Jan.	Feb.	Mar.	Apr.	May	Jun.	All
News	Real	0	3	35	47	27	18	28	16	12	15	8	32	27	13	23	25	21	10	360
	Mixed	0	2	29	35	24	6	10	5	7	8	13	11	16	13	10	7	9	2	207
	Fake	49	120	485	468	267	108	131	87	77	116	60	122	61	86	129	115	126	61	2668
	All	49	125	549	550	318	132	169	108	96	139	81	165	104	112	162	147	156	73	3235
tweets	Real	124	81	316	208	138	91	46	40	29	35	20	196	114	88	87	93	83	76	1865
	Fake	1993	1378	8463	5644	3391	1989	1364	904	720	894	559	3304	2716	2018	1762	1852	1854	1595	42400
	All	2117	1459	8779	5852	3529	2080	1410	944	749	929	579	3500	2830	2106	1849	1945	1937	1671	44265

Classification of News Articles

We extract 4,952 real, mixed, and fake news articles talking about COVID-19 based on the results of four fact-checking organizations, including Snopes [53], Politifact [54], Poynter [55], and Factcheck [56]. We gather 2413, 927, 1308, and 304 news articles talking about COVID-19 for Jan. 2020 - Jun. 2021 from these four organizations, respectively. It is not uncommon for fake news to be examined by several facts checking organizations. According to our datasets, no disagreement is found between these fact-checking outcomes across organizations. The categories of Snopes of interest include true, mostly true, mixture, mostly false, false news. Similarly, Politifact uses tag news with true, mostly true, half true, mostly false, false, and pants on fire news. We categorize news articles into real, mixed, or fake, as described in Table 2. Using these classifications, we collect all news articles from the archived news regarding COVID-19 from these organizations for Jan. 2020 - Jun. 2021.

Processing of News Articles for Analysis

We extract 3,437 news articles tagged with COVID-19 and coronavirus. After processing the initial cleaning, such as checking news with a correct tag, we come up with 3,235 news, consisting of 360 real news, 207 mixed news, and 2,668 fake news. After eliminating repetitive or irrelevant news, we select 207 news at random out of each pool of different types of news for fair consideration. Table 3 provides the distribution of published news and tweets considered across months. As in Table 3, we observe a significant amount of news articles published in Mar./Apr. 2020 and prominently there is a higher amount of fake news and tweets compared to those of real counterparts.

The news sources are mainly newspaper interviews, TV interviews, viral images, Journals, Press releases, digital ads, campaign ads, meeting in white houses, Story, TV segments, social media, or press conferences. The news is in the format of photos, infographics, videos, text, or interviews. As photos, infographics, videos, or interviews are not in the format of text, there is a challenge to analyze them. The fact-checking organizations put text and explanations related to each of them. Hence, we use the text generated by the fact-checking organizations to analyze them. We also use the converted format of the photo, infographic, or video for our analysis. We use the release date of the news to determine when a news article is published. The fact-checking organizations (i.e., Snopes, Politifact, Poynter, and Factcheck) categorize news into various classes based on Table 2.

Table 4. Prediction Performance of Various Machine Learning Classifiers

ML Classifier	Accuracy	Precision	Recall	F-score
Passive Aggressive	0.995	0.995	0.995	0.995
Logistic regression	0.984	0.984	0.984	0.984
Bagging Classifier	0.618	0.779	0.598	0.532
K-Neighbors	0.671	0.782	0.655	0.622
Decision Tree	0.994	0.994	0.994	0.994
Random Forest	0.519	0.623	0.5	0.346
AdaBoost	0.995	0.995	0.995	0.995
Multi Layer Perceptron	0.966	0.967	0.966	0.966



(a) Real news (b) Mixed news (c) Fake news (d) All news

Fig 4. Word cloud for real, mixed, fake, and all news for Jan. 2020 – Jun. 2021.

Collecting COVID-19-Related Tweets

Twitter, one of the most famous platforms, has above 313 million active users who generate 500 million tweets per day [57, 58]. Hence, we investigated 42,877,312 tweet IDs for Jan. 2020 – Jun. 2021. Note that we limited tweets to the US and we ended up with 44,265 tweets. Furthermore, we ordered these tweets chronologically, as in Table 3, showing a significant amount of tweets generated during Mar./Apr. 2020.

Classifying All Tweets as Real or Fake Based on Three machine learning (ML) Classifiers

We first classify tweets as real or fake. We first train eight existing ML classifiers on the datasets described in [59], which contain 23,481 fake tweets and 21,417 real news articles. We then select the top three ML classifiers, i.e., Passive-Aggressive, Decision Tree, and AdaBoost based on their prediction performance, as shown in Table 4. Finally, we predict the truthfulness of each tweet using these three ML algorithms and determine the final prediction for each tweet based on the majority rule of the three ML classifiers (i.e., at least two ML classifiers should give the same prediction result).

Identifying Physical-Psycho-Social States and Behavioral Patterns using LIWC

We use the LIWC as our text-mining tool for the analyses of COVID-19 related news and tweets because it contains a wealth of physical-psychosocial characteristics and behavioral patterns. Prior to analyzing them with the LIWC, all tweets are sorted by month and cleaned using various NLP tools (i.e., nltk, string, stopwords, RegexpTokenizer, and regexp) for each type of news (i.e., real, mixed, or fake) and tweet (i.e., real or fake). We begin text cleaning by removing HTML, punctuation, stop words, and stammering words. Following that, we extract all LIWC features relevant to CR assessment.

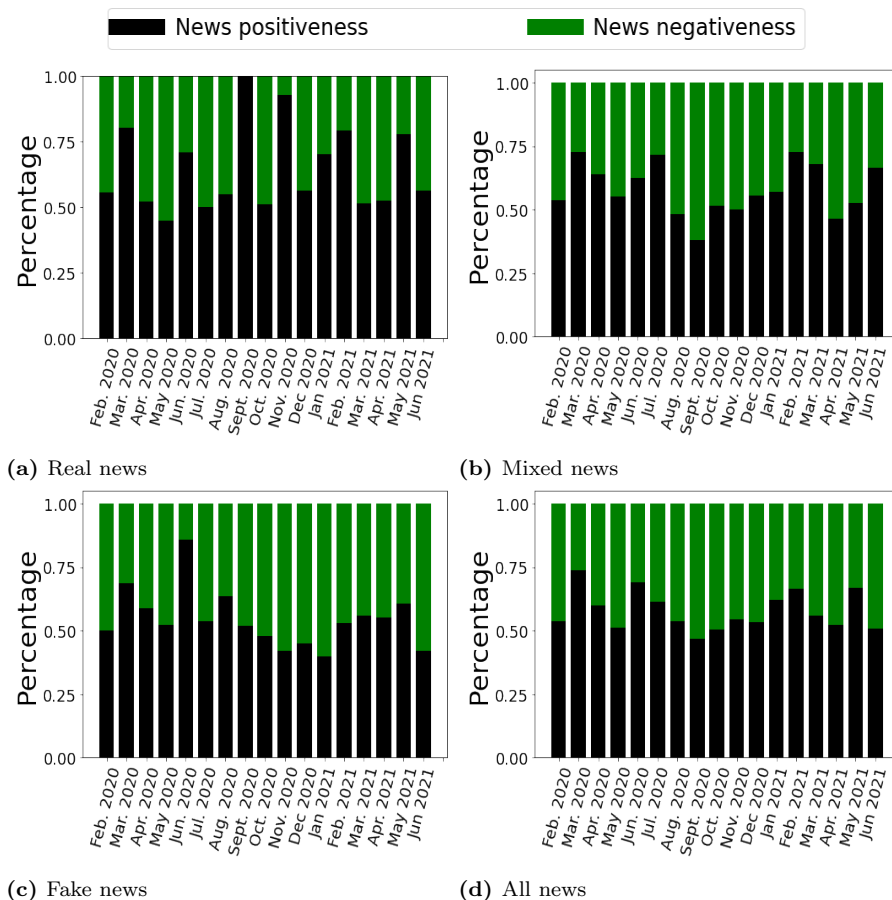


Fig 5. The positiveness and negativeness of news about the COVID-19 for Jan. 2020 – Jun. 2021.

Experimental Results & Analysis

News Analyses

Fig. 4 illustrates the word cloud associated with real, mixed, fake, and all news. Fig. 5 plots the positive and negative sentiments associated with various types of news over time. Table 5 shows the frequency of various topics under different types of news. Politics is the most popular subject. Medical and health, entertainment, and business are also popular topics affecting community resilience. In May and Sep. 2020, real news has the least positive and negative sentiment. In Sep. 2020 and Mar. 2020, mixed news has the least positive and negative content. In Jan. 2021 and Jun. 2020, fake news has the least positive and negative sentiment. In Sep. and Mar. 2020, all news is at its least positive and least negative, respectively. The subject of each news item is determined by fact-checking organizations, such as Snopes and Politifact.

Community Wellbeing Assessment

The output-oriented analysis measurements provide accurate information about the trend and dynamic change of functionality in a given community [60]. From Feb. 2020 to Jun. 2021, Fig. 6 depicts the normalized degree of output-oriented community wellbeing (CW) as measured by real, mixed, and fake news as well as real and fake tweets. Fake news and fake tweets demonstrate similar CW patterns. The peak of CW

Table 5. Frequency of Different Types of News Collected Under Various Topics for Jan. 2020 – Jun. 2021.

Source	Subject (an amount of news)
Real news	politics (87), medical (29), fauxtography (17), entertainment (13), business (12), viral (5), phenomena (5), crime (5), history (5), health (5)
Mixed news	coronavirus (67), politics (48), health (32), facebook (29), public (19), medical (17), fact (16), checks (16), posts (10), budget (8)
Fake news	politics (97), medical (39), fauxtography (13), entertainment (9), junk (9), news (9), viral (7), phenomena (7), technology (6), business (5)
All news	politics (229), medical (84), coronavirus (67), health (38), fauxtography (31), facebook (29), entertainment (23), business (22), public (17), fact (16)

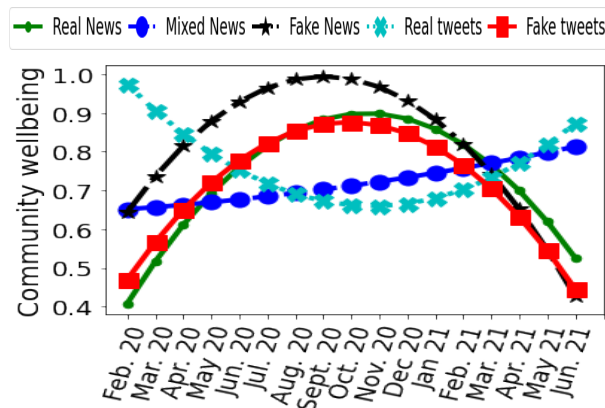


Fig 6. Community wellbeing measured by different types of news (i.e., real, mixed, and fake) and tweets (i.e., real and fake).

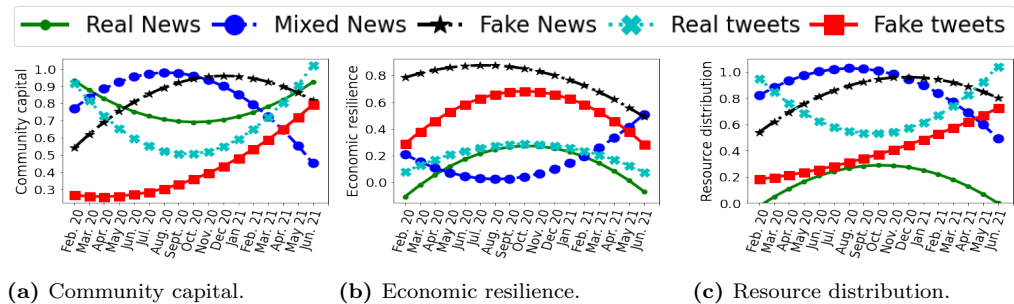
in fake tweets and real/fake news occurs in Sep. 2020. On the other hand, the peaks of CW in real tweets and mixed news occur in Feb. 2020 and Jun. 2021, respectively. We also observe that CW reaches its lowest point by the end of 2020 under real tweets. This result aligns well with the trends reported by the US Census Bureau [61] that since the COVID-19 outbreak in Feb. 2020, people’s wellbeing had deteriorated by the end of 2020.

Community Capital, Economic Resilience, and Resource Distribution Assessment

Fig. 7 illustrates the output-oriented degree of community capital, economic resilience, and resource distribution measured from the news (i.e., real, mixed, and fake) and tweets (i.e., real and fake) collected for Feb. 2020 – Jun. 2021. From this figure, we observe that real tweets and real news typically follow similar trends for Community capital, and economic resilience. Similarly, fake tweets and fake news also exhibit approximately similar trends.

For real news and tweets, economic functionalities are at their peak in Sep. 2020, while community capital is at the lowest level. Community capital shows its trend in the opposite direction of economic functionality for real/mixed news and real/fake tweets. This is because when a community is threatened due to the impact introduced by a disaster, people are more likely to cooperate for survival.

The incremental PCA method calculates the resource distribution based on community capital and economic resilience. The findings indicate that the trends in mixed/fake news and real/fake tweets are comparable to those in community capital. On the other hand, the trend in real news about resource distribution tracks the economic



(a) Community capital. (b) Economic resilience. (c) Resource distribution.
Fig 7. Community capital, economic resilience, and resource distribution measured based on different types of news and tweets for Feb. 2020–Jun. 2021.

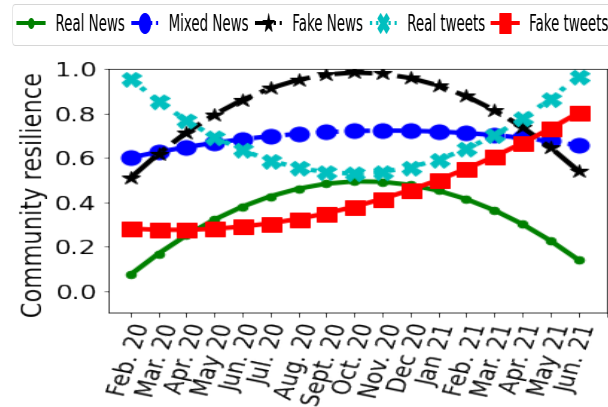


Fig 8. Output-oriented analysis of community resilience measured based on different types of news and tweets for Feb.2020–Jun. 2021.

functionality trend. Fake tweets and fake news both exhibit the same pattern in terms of resource distribution. Simultaneously, real and mixed news follow similar trends.

Community Resilience Assessment

Output-Oriented Resilience Assessment

We first measure CR over time (i.e., Feb. 2020 to Jun. 2021) for the output-oriented resilience assessment, as shown in Fig. 8. Although real news shows that community resilience begins to improve by the end of 2020, it also begins to deteriorate in 2021. In 2021, people’s wellbeing has been worsened. This is probably because people become tired of long-term restrictions in their daily lives, such as social distancing and online schooling/working, especially with the emergence of COVID-19 variants. These factors may drive people to become more pessimistic about the full recovery from the pandemic.

Note that the incremental PCA method calculates the community resilience based on community wellbeing and resource distribution. The findings indicate that the trends in real/fake news and real/fake tweets are comparable to those in resource distribution. On the other hand, the trends in real, mixed, and fake news about resource distribution track the community wellbeing trends. Note that resource distribution and community wellbeing follow the same pattern for real and fake news.

Capacity-based Resilience Assessment

Capacity-based measurements are time-averaged community resilience (CR) measurements of a given community, indicating the degree of functionality of the

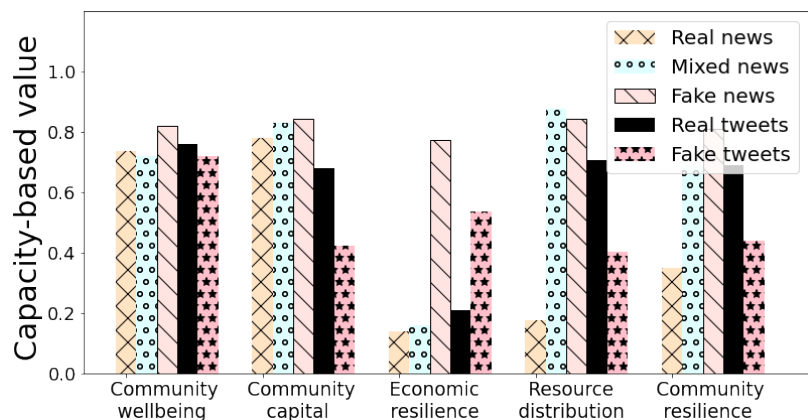


Fig 9. Capacity-based analysis of community wellbeing, community capital, economic resilience, resource distribution, and community resilience.

community [60]. Fig. 9 illustrates the capacity-based values of all resilience-related metrics, including community wellbeing, community capital, economic resilience, resource distribution, and finally, community resilience, measured using real, mixed, and fake news as well as real and fake tweets.

We observe from Fig. 9 that fake news is in a better state of community wellbeing (CW). In other words, released fake news implies that CW is adequate and likely underestimates the detrimental effect of the COVID-19. Additionally, people’s communication via fake tweets demonstrates a significant level of isolation, whereas real tweets show a higher level of community capital. Fig. 9 shows that while fake news presents a high degree of economic resilience, real news shows a low degree of economic resilience under the COVID-19. A possible reason is that fake news can trigger panic buying, thus eroding economic resilience. Similarly, fake news has a greater level of resource distribution than real news. Finally, fake news shows higher CR than real news. Fake news has the potential to mislead people into taking inappropriate actions in response to the COVID-19 by forming unrealistic optimism about the future. For instance, some fake news suggests that smoking, self-medicating with antibiotics, and wearing multiple surgical masks help combat COVID-19. This information is not only impractical, but also potentially jeopardizing community resilience.

Absorption, Community Non-Functioning, and Recovery

Table 6 shows the measurement values of community functionality (CF) metrics, including Absorption (ABS), Community Non-Functioning (CNF), Recovery (REC), Time for Absorption (TFA), Time under Community Non-Functioning (TNF), and Time To Recovery (TTR) (see Fig. 2) for news and tweets, with the critical CF threshold, b , varying in the range of 0.2 to 0.5 in increment of 0.1.

Fake news induces a higher level of absorption for all critical CF threshold values than real news. Additionally, fake news typically exhibits the greatest degree of recovery. Fake news fosters distrust among the public, despite the fact that trust is a critical component of transparent risk communication, collaboration, and the cooperation of individuals to overcome catastrophic events. The negative outputs of fake news create problems not only in handling COVID-19 but also in recovering from it.

Real news induces a 17-month recovery for all critical CF threshold values, while the absorption level is 0-1 month. This means that CR steadily increased from Feb. 2020 to Jun. 2021. In other words, with real news, the community can recover very quickly following the initial degradation of functionality. Additionally, the number of months

Table 6. Absorption (ABS), Community Non-Functioning (CNF), Recovery (REC), Time for Absorption (TFA), Time under Community Non-Functioning (TNF), and Time To Recovery (TTR) for News and tweets with the Critical Community Functionality Threshold, b , varying over the range of 0.2-0.5.

Adaptability Indexes	$b = 0.2$					$b = 0.3$					$b = 0.4$					$b = 0.5$				
	News			tweets		News			tweets		News			tweets		News			tweets	
	Real	Mixed	Fake	Real	Fake	Real	Mixed	Fake	Real	Fake	Real	Mixed	Fake	Real	Fake	Real	Mixed	Fake	Real	Fake
ABS	0	0.6	0.51	0.68	0.28	0	0.6	0.51	0.68	0	0	0.6	0.51	0.68	0	0	0.6	0.51	0.68	0
CNF	0.13	0	0	0	0	0.17	0	0	0	0.28	0.25	0	0	0	0.31	0.35	0	0	0	0.34
REC	0.35	0.69	0.81	0.68	0.46	0.35	0.69	0.81	0.68	0.44	0.35	0.69	0.81	0.68	0.44	0.35	0.69	0.81	0.68	0.44
TFA	0	1	1	9	3	0	1	1	9	0	0	1	1	9	0	0	1	1	9	0
TNF	3	0	0	0	0	5	0	0	0	5	9	0	0	0	9	17	0	0	0	12
TTR	17	17	17	9	15	17	17	17	9	17	17	17	17	9	17	17	17	17	9	17

Note that TFA, TNF, and TTR refer to the month-based average values.

during which the community is non-functioning ranges from 0 to 17 months, depending on the critical threshold level. For example, TNF is equal to 17 months when $b = 0.5$ for real news, which means that the community functionality from the perspective of real news is less than 0.5 for all 17 months. Understandably, as the critical threshold level increases, the time duration associated with community dysfunction and recovery increases, while that associated with absorption decreases. On the other hand, mixed news has a higher level of absorption than fake news. Both fake news and mixed news show a higher level of absorption than that of real news. This implies that the level of community functionality is initially high and gradually declines, whereas real news demonstrates a rapid decline in community functionality at the start. Therefore, we can conclude that mixed/fake news tends to underestimate the negative impact of COVID-19 on the community. Real tweets, on the other hand, exhibit a high absorption level when $b = 0.2 - 0.5$, indicating that individuals believe the community is highly functional.

Tables 7- 10 show the measurement values of community wellbeing, resource distribution, community capital, and economic functionality metrics, including Absorption (ABS), Community Non-Functioning (CNF), Recovery (REC), Time for Absorption (TFA), Time under Community Non-Functioning (TNF), and Time To Recovery (TTR), respectively. Based on the results:

- Community wellbeing's point of view: While fake news induces the highest level of absorption for all critical CF threshold values, real news typically exhibits the greatest degree of recovery.

- Resource distribution's point of view: While fake news induces a higher level of absorption for all critical CF threshold values compared to real news, mixed news has the highest level of absorption. Additionally, fake news typically exhibits the greatest degree of recovery.

- Economic functionality's point of view: Fake news induces the highest level of absorption and recovery for all critical CF threshold values,

- Community capital's point of view: It is similar to Resource distribution.

Statistical Analyses of News and Tweets

Table 11 shows the findings from our statistical analyses on the correlation between news and tweets. The statistical analyses include Pearson correlation (PC), Kendall tau correlation (KC), parametric statistical hypothesis tests (PT; Student's t-test), and non-parametric statistical hypothesis tests (NT; Mann-Whitney U Test). The Pearson correlation (PC) and Kendall tau correlation coefficients demonstrate the linear and monotonic relationships between two variables, x and y [62]. We choose Pearson's correlation coefficient to investigate if there is a linear statistical relationship or association between a resilience metric measured from real/mixed/fake news (x) vs. the same resilience metric measured from real/fake Tweets (y). The Pearson correlation

Table 7. Absorption (ABS), Community Non-Functioning (CNF), Recovery (REC), Time for Absorption (TFA), Time under Community Non-Functioning (TNF), and Time To Recovery (TTR) for News and tweets with the Critical wellbeing Functionality Threshold, b , varying over the range of 0.2-0.5.

Adaptability Indexes	$b = 0.2$					$b = 0.3$					$b = 0.4$					$b = 0.5$				
	News			tweets		News			tweets		News			tweets		News			tweets	
	Real	Mixed	Fake	Real	Fake	Real	Mixed	Fake	Real	Fake	Real	Mixed	Fake	Real	Fake	Real	Mixed	Fake	Real	Fake
ABS	0.41	0.65	0.82	0.77	0.72	0.41	0.65	0.82	0.77	0.72	0.41	0.65	0.82	0.77	0.72	0	0.65	0.84	0.77	0
CNF	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.41	0	0.43	0	0.46
REC	0.74	0.72	0.43	0.74	0.44	0.74	0.72	0.43	0.74	0.44	0.74	0.72	0.43	0.74	0.44	0.74	0.72	0.43	0.74	0.72
TFA	1	1	17	10	17	1	1	17	10	17	1	1	17	10	17	0	1	16	10	0
TNF	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	2
TTR	17	17	1	8	1	17	17	1	8	1	17	17	1	8	1	17	17	1	8	17

Note that TFA, TNF, and TTR refer to the month-based average values.

Table 8. Absorption (ABS), Community Non-Functioning (CNF), Recovery (REC), Time for Absorption (TFA), Time under Community Non-Functioning (TNF), and Time To Recovery (TTR) for News and tweets with the Critical resource distribution Functionality Threshold, b , varying over the range of 0.2-0.5.

Adaptability Indexes	$b = 0.2$					$b = 0.3$					$b = 0.4$					$b = 0.5$				
	News			tweets		News			tweets		News			tweets		News			tweets	
	Real	Mixed	Fake	Real	Fake	Real	Mixed	Fake	Real	Fake	Real	Mixed	Fake	Real	Fake	Real	Mixed	Fake	Real	Fake
ABS	0	0.88	0.54	0.67	0	0	0.88	0.54	0.67	0	0	0.88	0.54	0.67	0	0	0.9	0.54	0.67	0
CNF	0.08	0	0	0	0.19	0.18	0	0	0.23	0.18	0	0	0	0.27	0.18	0.49	0	0	0.31	
REC	0.18	0.49	0.84	0.72	0.4	0.18	0.49	0.84	0.72	0.4	0.18	0.49	0.84	0.72	0.4	0.18	0.49	0.84	0.72	0.4
TFA	0	17	1	9	0	0	17	1	9	0	0	17	1	9	0	0	16	1	9	0
TNF	8	0	0	0	2	17	0	0	0	6	17	0	0	0	9	17	1	0	0	12
TTR	17	1	17	9	17	17	1	17	9	17	17	1	17	9	17	17	1	17	9	17

Note that TFA, TNF, and TTR refer to the month-based average values.

Table 9. Absorption (ABS), Community Non-Functioning (CNF), Recovery (REC), Time for Absorption (TFA), Time under Community Non-Functioning (TNF), and Time To Recovery (TTR) for News and tweets with the Critical economic Functionality Threshold, b , varying over the range of 0.2-0.5.

Adaptability Indexes	$b = 0.2$					$b = 0.3$					$b = 0.4$					$b = 0.5$				
	News			tweets		News			tweets		News			tweets		News			tweets	
	Real	Mixed	Fake	Real	Fake	Real	Mixed	Fake	Real	Fake	Real	Mixed	Fake	Real	Fake	Real	Mixed	Fake	Real	Fake
ABS	0	0.21	0.77	0	0.53	0	0	0.77	0	0	0	0	0.77	0	0	0	0	0.79	0	0
CNF	0.06	0.08	0	0.13	0	0.14	0.11	0	0.21	0.29	0.14	0.12	0	0.21	0.33	0.14	0.14	0.49	0.21	0.38
REC	0.14	0.16	0.49	0.21	0.29	0.14	0.16	0.49	0.21	0.53	0.14	0.16	0.49	0.21	0.53	0.14	0.16	0.49	0.21	0.53
TFA	0	1	17	0	17	0	0	17	0	0	0	0	17	0	0	0	0	16	0	0
TNF	10	12	0	6	0	17	14	0	17	2	17	15	0	17	4	17	16	1	17	6
TTR	17	16	1	17	1	17	17	1	17	17	17	17	1	17	17	17	17	1	17	17

Note that TFA, TNF, and TTR refer to the month-based average values.

Table 10. Absorption (ABS), Community Non-Functioning (CNF), Recovery (REC), Time for Absorption (TFA), Time under Community Non-Functioning (TNF), and Time To Recovery (TTR) for News and tweets with the Critical community capital Functionality Threshold, b , varying over the range of 0.2-0.5.

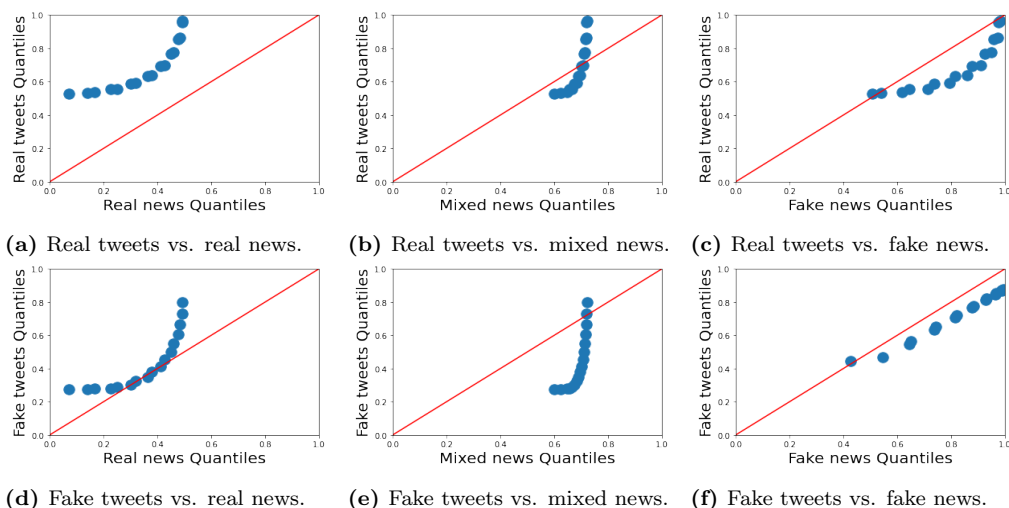
Adaptability Indexes	$b = 0.2$					$b = 0.3$					$b = 0.4$					$b = 0.5$				
	News			tweets		News			tweets		News			tweets		News			tweets	
	Real	Mixed	Fake	Real	Fake	Real	Mixed	Fake	Real	Fake	Real	Mixed	Fake	Real	Fake	Real	Mixed	Fake	Real	Fake
ABS	0	0.88	0.54	0.67	0	0	0.88	0.54	0.67	0	0	0.88	0.54	0.67	0	0	0.9	0.54	0.67	0
CNF	0.08	0	0	0	0.19	0.18	0	0	0.23	0.18	0	0	0	0.27	0.18	0.49	0	0	0.31	
REC	0.18	0.49	0.84	0.72	0.4	0.18	0.49	0.84	0.72	0.4	0.18	0.49	0.84	0.72	0.4	0.18	0.49	0.84	0.72	0.4
TFA	0	17	1	9	0	0	17	1	9	0	0	17	1	9	0	0	16	1	9	0
TNF	8	0	0	0	2	17	0	0	0	6	17	0	0	0	9	17	1	0	0	12
TTR	17	1	17	9	17	17	1	17	9	17	17	1	17	9	17	17	1	17	9	17

Note that TFA, TNF, and TTR refer to the month-based average values.

Table 11. The Statistical Analysis of Various Functionalities for Three News Compared To Two Types of tweets: Pearson Correlation (PC), Kendall Tau Correlation (KC), Parametric Statistical Hypothesis Tests (PT), and Non-Parametric Statistical Hypothesis Tests (NT)

Source		News															
Feature		Wellbeing			Community capital			Economic resilience			Resource distribution			Community resilience			
Type		Real	Mixed	Fake	Real	Mixed	Fake	Real	Mixed	Fake	Real	Mixed	Fake	Real	Mixed	Fake	
Tweets	PC	Real	-1	-0.2	-0.73	0.97	-0.88	-0.56	0.99	-0.76	0.63	-0.97	-0.86	-0.63	-0.98	-0.86	-1
		Fake	0.96	-0.19	0.94	0.3	-0.86	0.41	0.99	-0.76	0.62	-0.09	-0.76	0.53	-0.17	0.18	-0.26
	KC	Real	-1	-0.18	-0.65	0.88	-0.68	-0.47	0.88	-0.68	0.5	-0.88	-0.68	-0.56	-0.88	-0.66	-0.88
		Fake	0.76	-0.06	0.88	0	-0.44	0.41	0.88	-0.68	0.5	0.06	-0.38	0.38	0	0.22	0
	PT	Real	✓	✓	✓	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✓	✗
		Fake	✓	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗
	NT	Real	✓	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗
		Fake	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗

Note that ✓ and ✗ mean following or not following the same distribution, respectively.



(a) Real tweets vs. real news. (b) Real tweets vs. mixed news. (c) Real tweets vs. fake news. (d) Fake tweets vs. real news. (e) Fake tweets vs. mixed news. (f) Fake tweets vs. fake news.

Fig 10. The Quantile-Quantile (Q-Q)-plot of news and tweets used to measure community resilience where x-axis refers to the quantiles of real, mixed, or fake news and y-axis indicates the quantiles of real, fake, or all tweets.

coefficient assumes that both x and y are normally distributed. When this assumption does not hold, we rely on a non-parametric approach, such as Kendall tau correlation, which does not make any assumption about distribution. According to Table 11, fake tweets and news have a positive correlation for resilience-related features with a probability of 80 percent. Pearson and Kendall tau correlations (PC and KC) indicate that the correlations between fake news and real tweets are negative, with a probability of 80 percent. We also found that mixed news negatively correlates with real and fake tweets across all types of CR attributes with a probability of 95 percent. Parametric and non-parametric statistical hypothesis tests (PT and NT) demonstrate the distribution's similarity across multiple scenarios. Fig. 10 illustrates the Quantile-Quantile (Q-Q)-plot for community resilience in relation to various news types (i.e., real, mixed, or fake) and tweet types (i.e., real or fake). We observe that fake tweets and real tweets exhibit similarity in their distributions with the probability of 60 percent. This similarity implies that both tweets can properly reflect the actual states of community resilience (CR) regardless of their truthfulness. Furthermore, analyzing social media information and predicting CR can provide a useful indicator to measure how our community is functioning against a disaster such as COVID-19.

Summary of Resilience-Related Analysis

We summarize the findings obtained from the discussion above as follows:

- Based on fake news, the public may believe that the community is resilient, which is not the case. Additionally, the results indicate that fake news shares the same viewpoint. They underestimate COVID-19's adverse effects and demonstrate a higher level of resilience than that measured by real news. This perspective prolongs the time required for actual complete recovery. Further, based on this finding, we observe that fake news is not always pessimistic or negative.
- From community resilience point of view, mixed news is more optimistic than real news showing higher resilience. This may be because mixed news contains fake news, which underestimates the impact of COVID-19.
- Compared to propagated fake tweets, propagated fake news is more unrealistic from community resilience point of view. They demonstrate a greater capacity for community resilience. This finding is reasonable because the source of fake news frequently intends to cause harm, whereas fake news may be spread by people who may have no bad intent but mistakenly believe it or have no knowledge to judge the information credibility.
- From community resilience point of view, propagated real news is slightly more negative than original real tweets, showing a lower level of community resilience. This is because the original intent of fake news originators has been diluted through the process of propagation.

Conclusion

This section summarizes the key contributions made in this work and answers the research questions raised in Section "Research Goal, Contributions, and Questions". In addition, we suggest future research directions.

Summary of the Key Contributions

In this paper, we analyzed community resilience (CR) during the COVID-19 pandemic in the US from Feb. 2020 to Jun. 2021 based on both news articles and tweets on social media. We measured CR based on two main dimensions developed in this paper: community wellbeing (CW) and resource distribution (RD). We also developed two different dimensions to measure RD: economic resilience and community capital. We leveraged the information provided by fact-checking organizations such as Politifact, Poynter, Snopes, and Factcheck to collect 4,952 full-text news articles and categorize them as real, mixed, or fake news. On the other hand, to identify real and fake tweets, we used the top three machine learning (ML) algorithms among eight ML algorithms being evaluated, i.e., Passive-Aggressive Classifier, Decision Tree Classifier, and AdaBoost Classifier. The three ML algorithms showed at least 95% accuracy in classifying 42,877,312 tweet IDs from Jan. 2020 to Jun. 2021 into fake and real tweets based on the majority rule as our experimental tweets dataset. To improve the operationalization and sociological significance of this work, we used dimension reduction techniques, including linear transformations, nonlinear transformations, and manifold learning to integrate various dimensions of community resilience. We provided the output-oriented and capacity-based resilience analyses for various types of news and tweets and investigated their general trends and relationships. In addition, we evaluated community resilience in terms of the meantime to absorption, community

non-functioning, and recovery under various critical community functionality thresholds that determine the deadlock of community failure. 561
562

Answers to the Research Questions 563

RQ1. *What are the main trends observed in community resilience and its key attributes, i.e., community wellbeing and resource distribution?* 564
565

Answer. Among the PCAs with various kernel types, the SVD, the isomap, and the Locally Linear Embedding, we used the incremental PCA to integrate dimensions of resource distribution and community resilience due to the higher level of variance information ratio and the preservation of temporal dependency information. In September 2020, CW reached its peak in fake tweets and real/fake news. The peaks of CW in real tweets and mixed news, on the other hand, occur in February 2020 and June 2021, respectively. Additionally, we observe that CW reaches a low point by the end of 2020 when real tweets are used. Plus, the findings suggest that the resource distribution trends observed in mixed/fake news and real/fake tweets are comparable to those observed in community capital. On the other hand, the trend in real news about resource distribution corresponds to the trend in economic functionality. Take note that both real and fake news follow the same pattern in terms of resource distribution and community wellbeing. Community resilience trends in real/fake news and real/fake tweets are comparable to resource distribution trends. On the other hand, trends in real, mixed, and fake news regarding resource distribution are similar to the trends in community wellbeing. Fake news has a more even distribution of resources than real news. Finally, fake news has a higher community resilience than real news. By creating unrealistic optimism about the future, fake news has the potential to mislead people into taking inappropriate actions in response to the COVID-19. 566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584

RQ2. *What are the key differences and correlations between the community resilience measured on various types of news and tweets?* 585
586

Answer. According to the findings, fake tweet articles have an 80% probability of correlating positively with fake news for resilience-related characteristics. Additionally, Pearson and Kendall tau correlations indicate that the correlation between fake news and real tweets is negative, with an 80 percent probability. Additionally, we discovered that mixed news has a 95% probability of negatively correlating with real and fake tweets across all types of CR attributes. Statistical hypothesis tests, both parametric and non-parametric, demonstrate the distribution's similarity across multiple scenarios. We observe that fake and real tweets have a 60% probability of having similar distributions. This implies that fake tweets can accurately reflect the actual state of community resilience (CR), regardless of their veracity. 587
588
589
590
591
592
593
594
595
596

RQ3. *What are the level of the community resilience metrics, e.g., absorption and recovery during COVID-19 on various types of news and tweets?* 597
598

Answer. According to Tables 6- 10, both fake and mixed news exhibit a greater level of absorption than real news for all critical CF threshold values and resilience-related characteristics. Fake news has the highest level of absorption for all critical threshold values, both in terms of community well-being and economic functionality. The number of months that the community is unable to function (TNF) varies between 0 and 17 months, depending on the critical threshold value. For real news, TNF is equal to 17 months when $b = 0.5$. Fake news typically exhibits the greatest degree of recovery in terms of community functionality, economics, community capital, and resource distribution. As a result, we can conclude that mixed/fake news frequently underestimates COVID-19's negative impact on the community. The negative consequences of fake news complicate not only the handling of COVID-19, but also the recovery process. 599
600
601
602
603
604
605
606
607
608
609
610

Future Research Directions

We suggest the following future research directions.

First, in this work, we only used Twitter to gather all real and fake news to investigate the behavior of the population. One should be extremely careful in analyzing social media information. Surveys can provide high-quality data for analyzing population behavior, albeit at a cost. Additional research can be conducted to examine the correlations between fake/real news and survey responses. Nonetheless, considering additional social media platforms for future research may be beneficial.

Second, while we propose in this work to quantify community resilience using social media data (e.g., tweets), we are still in the first phase of this journey, namely enhancing community resilience. The literature does not include a thorough examination of the prediction models for community resilience. As a result, more sophisticated models are required to forecast how distinct communities will respond to a variety of events and epidemics. It specifically calls for developing a multi-agent model that accounts for the spread of fake news. The approach described in this work must be extended further to validate the model. The next step on this path is to predict output-oriented community resilience using machine and deep learning techniques.

Third, we consider community capital and economic resilience as resource distribution metrics in this work. Apart from these metrics, institutional and infrastructure resilience are also critical aspects of resource distribution that can have an effect on community resilience. Additionally, we use anxiety, anger, and sadness to ascertain the community's level of wellbeing. Additional metrics for wellbeing can be added. This necessitates the development of new techniques for assessing additional potential indicators of community wellbeing.

Finally, one can choose appropriate engagement strategies, such as collaborative adaptive management and joint fact-finding, based on a community's social characteristics and the perspectives of its stakeholders, in order to determine appropriate policies to enhance community resilience.

References

1. J. I. Ruzek Disaster response, mental health, and community resilience Available at <https://www.psychiatrictimes.com/view/disaster-response-mental-halth-and-community-resilience>, Accessed: 08-11-2021
2. Cantrill, S. and Australia, T. and Fernandes, H. Building resilient communities: The importance of integrating mental health and wellbeing in effective development thinking and practice. Micah Triennial, 2018 Sep.
3. Patel, S. S. and Rogers, M. B. and Amlôt, R. and Rubin, G. J. What do we mean by 'community resilience'? A systematic literature review of how it is defined in the literature PLoS Currents, 2017 Feb., 9.
4. Rego, A. and Mehta, S. Opportunities and challenges in risk resilient recovery. World Hospitals and Health Services: the Official Journal of the International Hospital Federation, 2005 Jan.; 41 (4): 33 –35.
5. Pew Research Center Social media facts sheet Available at <http://www.pewinternet.org/fact-sheet/social-media>, Accessed: 04-08-2022
6. Jaidka, Kokil and Giorgi, Salvatore and Schwartz, H Andrew and Kern, Margaret L and Ungar, Lyle H and Eichstaedt, Johannes C Estimating geographic

- subjective well-being from Twitter: A comparison of dictionary and data-driven language methods *Proceedings of the National Academy of Sciences*, 2020; 117 (19): 10165–10171.
7. McCormick, Tyler H and Lee, Hedwig and Cesare, Nina and Shojaie, Ali and Spiro, Emma S Using Twitter for demographic and social science research: Tools for data collection and processing *Sociological methods & research* 2017, 46 (3): 390 –421.
 8. DeJohn, Amber D and Schulz, Emily English and Pearson, Amber L and Lachmar, E Megan and Wittenborn, Andrea K Identifying and understanding communities using Twitter to connect about depression: cross-sectional study *JMIR mental health*, 2018; 5 (4): e9533.
 9. Costello, Cory and Srivastava, Sanjay and Rejaie, Reza and Zalewski, Maureen Predicting Mental Health From Followed Accounts on Twitter *Collabra: Psychology*, 2021;7 (1).
 10. McCormick, Tyler and Lee, Hedwig and Cesare, Nina and Shojaie, Ali Using twitter for demographic and social science research: Tools for data collection *WASHINGTON UNIV SEATTLE*: 2014.
 11. Komorowski, Marlen and Do Huu, Tien and Deligiannis, Nikos Twitter data analysis for studying communities of practice in the media industry *Telematics and Informatics*, 2018; 35 (1): 195—212.
 12. Tambuscio, M. and Ruffo, G. and Flammini, A. and Menczer, F. Fact-checking effect on viral hoaxes: A model of misinformation spread in social networks *Proceedings of the 24th International Conference on World Wide Web*, 2015 May.: 977–982
 13. Vosoughi, S. and Roy, D. and Aral, S. The spread of true and false news online *Science*, 2018 Mar.; 359 (6380): 1146—1151
 14. Sivek, S. C. Both Facts and Feelings: Emotion and News Literacy *Journal of Media Literacy Education*, 2018; 10 (2): 123–138
 15. Cho, J. H. and Xu, S. and Hurley, P. M. and Mackay, M. and Benjamin, T. and Beaumont, M. STRAM: Measuring the trustworthiness of computer-based systems *ACM Comput. Surv.*, 2019 Feb.; 51 (6)
 16. Valinejad, J. and Mili, L. and Triantis, K. and von Spakovsky, M. and van der Wal, N. Stochastic Multi-Agent-Based Model to Measure Community Resilience-Part 2: Simulation Results *arXiv preprint arXiv:2004.05185*
 17. Valinejad, Jaber and Mili, Lamine and van der Wal, Natalie and von Spakovsky, Michael and Xu, Yijun Multi-Dimensional Output-Oriented Power System Resilience based on Degraded Functionality 2021 IEEE Power & Energy Society General Meeting (PESGM), Washington, D.C., USA
 18. Links, J. M. and Schwartz, B. S. and Lin, S. and Kanarek, N. and Mitrani-Reiser, J. and Sell, T. K. and Watson, C. R. and Ward, D. and Slemp, C. and Burhans, R. and others =COPEWELL: a conceptual framework and system dynamics model for predicting community functioning and resilience after disasters *Disaster Medicine and Public Health Preparedness*, 2018 Feb.; 12 (1): 127 –137

19. Cutter, S. L. and Barnes, L. and Berry, M. and Burton, C. and Evans, E. and Tate, E. and Webb, J. A place-based model for understanding community resilience to natural disasters *Global Environmental Change*, 2008 Oct., 18 (4), 598–606
20. A. Ostadtaghizadeh and A. Ardalan , D. Paton and H. Jabbari and H. R. Khankeh Community disaster resilience: a systematic review on assessment models and tools *PLOS Currents Disasters*, 2015 Apr.; 7
21. Sherrieb, K. and Norris, F. H. and Galea, S. Measuring capacities for community resilience *Social Indicators Research*, 2010 Nov.; 99 (2); 227–247
22. H. Caia and N. S.N. Lama and Y. Qiangb and L. Zoua and R. M. Corrella and V. Mihunov A synthesis of disaster resilience measurement methods and indices *International Journal of Disaster Risk Reduction*, 2018 Oct.; 31: 844–855
23. S.L. Cutter and C.G. Burton and C.T. Emrich Disaster resilience indicators for bench-marking baseline conditions *J. Homel. Secur. Emerg. Manag.*, 2010 Aug., 7(1)
24. Springgate, B. F. and Wennerstrom, A. and Meyers, D. and Allen, C. E. and Vannoy, S. D. and Bentham, W. and Wells, K. B. Building community resilience through mental health infrastructure and training in post-Katrina New Orleans *Ethnicity & Disease*, 2011; 21 (3 Suppl 1): S1-20-9
25. Inter-Agency Standing Committee and others IASC guidelines on mental health and psychosocial support in emergency settings Geneva, Switzerland: IASC 2006
26. Varghese, J. and Krogman, N. T. and Beckley, T. M. and Nadeau, S. Critical analysis of the relationship between local ownership and community resiliency *Rural Sociology*, 2006 Sep.; 71 (3): 505–527
27. Ho, C. S. and Chee, C. Y. and Ho, R. C. Mental health strategies to combat the psychological impact of COVID-19 beyond paranoia and panic *Ann Acad Med Singapore*, 2020 Mar.; 49 (1): 1–3
28. Smith, G. D. and Ng, F. and Li, W. H. C. COVID-19: Emerging compassion, courage and resilience in the face of misinformation and adversity *Journal of Clinical Nursing*, 2020 May; 29 (9): 1425
29. Hua, J. and Shaw, R. Coronavirus (COVID-19) infodemic and emerging issues through a data lens: The case of China *International Journal of Environmental Research and Public Health*, 2020 Jan.; 17 (7): 2309
30. Van B, J. J. and Baicker, K. and Boggio, P. S. and Capraro, V. and Cichocka, A. and Cikara, M. and Crockett, M. J. and Crum, A. J. and Douglas, K. M. and Druckman, J. N. and others Using social and behavioural science to support COVID-19 pandemic response *Nature Human Behaviour*, 2020; 4 (5): 1–12
31. Taylor, M. and Wells, G. and Howell, G. and Raphael, B. and others The role of social media as psychological first aid as a support to community resilience building *The Australian Journal of Emergency Management*, 2012 Jan.; 27 (1)
32. Reddy, H. and Raj, N. and Gala, M. and Basava, A. Text-mining-based Fake News Detection Using Ensemble Methods *International Journal of Automation and Computing*, 2020 Apr.; 17 (2): 1–12

33. Li, S. and Wang, Y. and Xue, J. and Zhao, N. and Zhu, T. The impact of COVID-19 epidemic declaration on psychological consequences: a study on active WEIBO users *International Journal of Environmental Research and Public Health*, 2020 Jan.; 17 (6): 20–32
34. Tausczik, Y. R. and Pennebaker, J. W. The psychological meaning of words: LIWC and computerized text analysis methods *Journal of Language and Social Psychology*, 2010 Mar.; 29 (1): 24–54
35. Hou, Z. and Du, F. and Jiang, H. and Zhou, X. and Lin, L. Assessment of public attention, risk perception, emotional and behavioural responses to the COVID-19 outbreak: social media surveillance in China *medRxiv*, 2020
36. Sentiment and emotion Lexicons National Research Council Canada (NRC) Available at <https://nrc.canada.ca/en/research-development/products-services/technical-advisory-services/sentiment-emotion-lexicons>, Accessed: 08-11-2021
37. Ju, Mingxuan and Song, Wei and Sun, Shiyu and Ye, Yanfang and Fan, Yujie and Hou, Shifu and Loparo, Kenneth and Zhao, Liang Dr. Emotion: Disentangled Representation Learning for Emotion Analysis on Social Media to Improve Community Resilience in the COVID-19 Era and Beyond *Web Science*, 2021 Apr.: 518–528
38. Coppersmith, G. and Dredze, M. and Harman, C. Quantifying mental health signals in Twitter *Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, 2014 Jan.: 51–60
39. Molyneaux, H. and O'Donnell, S. and Kakekaspan, C. and Walmark, B. and Budka, P. and Gibson, K. Community resilience and social media: Remote and rural First Nations communities, social isolation and cultural preservation *International Rural Network Forum. Whyalla and Upper Spencer Gulf*, 2012 Sep.
40. Han, J. and Pei, J. and Kamber, M. *Data mining: Concepts and techniques* Elsevier, 2011 Jan.
41. Prince, M. and Patel, V. and Saxena, S. and Maj, M. and Maselko, J. and Phillips, M. R. and Rahman, A No health without mental health *The Lancet*, 2007 Sep.; 370 (9590): 859–877
42. Krendl, Anne C and Perry, Brea L The impact of sheltering in place during the COVID-19 pandemic on older adults' social and mental well-being *The Journals of Gerontology: Series B*, 2021 Feb.;76 (2): e53–e58
43. Faelens, Lien and Hoorelbeke, Kristof and Soenens, Bart and Van Gaeveren, Kyle and De Marez, Lieven and De Raedt, Rudi and Koster, Ernst HW Social media use and well-being: A prospective experience-sampling study *Computers in Human Behavior*, 114 , 2021 Jan.; 106510
44. Paredes, Mario R and Apaolaza, Vanessa and Fernandez-Robin, Cristóbal and Hartmann, Patrick and Yañez-Martinez, Diego The impact of the COVID-19 pandemic on subjective mental well-being: The interplay of perceived threat, future anxiety and resilience *Personality and Individual Differences*, 2021 Feb.; 170 (170): 110455

45. Mehl, Matthias R and Gosling, Samuel D and Pennebaker, James W Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *Journal of Personality and Social Psychology*, 2006 May. ; 90 (5): 862
46. Simmons, Rachel A and Gordon, Peter C and Chambless, Dianne L Pronouns in marital interaction: What do you and I say about marital health? *Psychological Science*, 2005 Dec.; 16 (12): 932–936
47. Sexton, J Bryan and Helmreich, Robert L Analyzing cockpit communications: The links between language, performance, error, and workload *Human Performance in Extreme Environments*, 2000 Oct.; 5 (1): 63–68
48. Lazega, Emmanuel and others The collegial phenomenon: The social mechanisms of cooperation among peers in a corporate law partnership Oxford University Press on Demand, 2001
49. Newman, Matthew L and Groom, Carla J and Handelman, Lori D and Pennebaker, James W Gender differences in language use: An analysis of 14,000 text samples *Discourse Processes*, 2008 May.; 45 (3): 211–236
50. Webscraper.io Web Scraper - Free Web Scraping Available at <https://chrome.google.com/webstore/detail/web-scraper-free-web-scra/jnhgnonknehpejjeheh11klipmbmhn?hl=en>, Accessed: 08-11-2021
51. Richardson, Leonard Beautiful soup documentation Available at <https://beautiful-soup-4.readthedocs.io/en/latest/>, Accessed: 08-11-2021
52. Li, Baoli and Han, Liping Distance weighted cosine similarity measure for text classification *International Conference on Intelligent Data Engineering and Automated Learning*, 2013 Oct.: 611–618
53. Snopes COVID-19 Available at <https://www.snopes.com/tag/\uppercase{covid-19}/>, Accessed: 08-11-2021
54. Politifact COVID-19 Available at <https://www.politifact.com/coronavirus/>, Accessed: 08-11-2021
55. Poynter COVID-19 Available at <https://www.poynter.org/ifcn-\uppercase{covid-19}-misinformation/>, Accessed: 08-11-2021
56. Factcheck COVID-19 Available at <https://www.factcheck.org/a-guide-to-our-coronavirus-coverage/>, Accessed: 08-11-2021
57. Internet Live Stats Twitter usage statistics, Available at <http://www.internetlivestats.com/twitter-statistics>, Accessed: 04-08-2022
58. Twitter Twitter usage/company facts, Available at <https://about.twitter.com/company>, Accessed: 04-08-2022
59. Ahmed, Hadeer and Traore, Issa and Saad, Sherif Detecting opinion spams and fake news using text classification *Security and Privacy*, 1 (1), 2018 Jan.: e9

60. Zobel, C. W. and Baghersad, M. Analytically comparing disaster resilience across multiple dimensions *Socio-Economic Planning Sciences*, 2020 Mar.; 69 (1): 100678
61. U.S. Census Bureau Household pulse survey Available at <https://www.census.gov/programs-surveys/household-pulse-survey/data.htm>, Accessed: 08-11-2021
62. Croux, Christophe and Dehon, Catherine Influence functions of the Spearman and Kendall correlation measures *Statistical Methods & Applications*, 2010; 19 (4): 497–515