



Varieties of corona news: a cross-national study on the foundations of online misinformation production during the COVID-19 pandemic

Cantay Caliskan¹ · Alaz Kilicaslan²

Received: 30 April 2022 / Accepted: 11 November 2022 / Published online: 13 December 2022
© The Author(s), under exclusive licence to Springer Nature Singapore Pte Ltd. 2022

Abstract

Misinformation in the media is produced by hard-to-gauge thought mechanisms employed by individuals or collectivities. In this paper, we shed light on what the country-specific factors of falsehood production in the context of COVID-19 Pandemic might be. Collecting our evidence from the largest misinformation dataset used in the COVID-19 misinformation literature with close to 11,000 pieces of falsehood, we explore patterns of misinformation production by employing a variety of methodological tools including algorithms for text similarity, clustering, network distances, and other statistical tools. Covering news produced in a span of more than 14 months, our paper also differentiates itself by its use of carefully controlled hand-labeling of topics of falsehood. Findings suggest that country-level factors do not provide the strongest support for predicting outcomes of falsehood, except for one phenomenon: in countries with serious press freedom problems and low human development, the mostly unknown authors of misinformation tend to focus on similar content. In addition, the intensity of discussion on animals, predictions and symptoms as part of fake news is the biggest differentiator between nations; whereas news on conspiracies, medical equipment and risk factors offer the least explanation to differentiate. Based on those findings, we discuss some distinct public health and communication strategies to dispel misinformation in countries with particular characteristics. We also emphasize that a global action plan against misinformation is needed given the highly globalized nature of the online media environment.

Keywords Misinformation · Fake news · Social media · Computational social science · Cross-country analysis · Natural language processing

✉ Cantay Caliskan
cantay.caliskan@rochester.edu

¹ Goergen Institute for Data Science, University of Rochester, Rochester, USA

² Department of Sociology, Criminology and Anthropology, University of Wisconsin-Whitewater, Whitewater, USA

Introduction

“Anybody that wants a test can get a test. That’s what the bottom line is.” [1] was a sentence uttered by the president of the United States at a time when the daily testing capacity was about 75,000 nationally. Misinformation ranging from harmless rumors to extremely complex and dangerous conspiracy theories has been one of the most defining characteristics of social media use in recent years. The intensification of the use of social media as an information-sharing tool, coupled with a directly and indirectly forced lockdown during the COVID-19 pandemic, has multiplied the production of misinformation.

As provided in the example above, people of different walks of life including world leaders, Instagram celebrities, troll factories, and many others have been intentionally or unintentionally spreading misinformation. Assuming that these entities act rationally, people and institutions that use their own resources to produce those falsehoods must have a purpose. Inspired from the national-level analysis and empirical approach of the book *Varieties of Capitalism: The Institutional Foundations of Comparative Advantage* [2] and other articles and books that use ‘Varieties of...’ in their titles,¹ this article studies the relationship between country-specific variables, and the topics and content of misinformation. The study specifically looks at the variation in topic ratios, and the creativity in misinformation content over time by providing extensive data analysis and using the most extensive datasets on misleading and false news created during the global pandemic.²

Findings indicate that the ‘Varieties of...’ literature can be applied only to certain cases in this context. Specifically, countries struggling in a range of areas including press freedom and human development tend to produce news content similar to each other. In addition, news on animals, predictions and symptoms are the three biggest differentiators between countries, whereas news on conspiracies, medical equipment and risk factors offer the least explanation to differentiate. Based on those findings, we discuss some distinct public health and communication strategies to dispel misinformation in countries with particular characteristics, such as efforts to repair Western governments and pharmaceutical companies’ tarnished reputation in low human development countries. We also emphasize that a global action plan against misinformation is needed given the highly globalized nature of the online media environment and the ubiquitousness of the major conspiracy theories around COVID-19. The dataset used in this study is the largest dataset of online misinformation about COVID-19 that can be found in the literature as it comprises over 10 thousand falsehoods from 129 countries. Our study is also unique in the number of variables used: the countries are compared based on 14 variables aiming to measure

¹ Some well-known recent works from political science literature that use the ‘Varieties of...’ in their titles are [3–8].

² On this note, we *wholeheartedly* thank our attentive research assistant *Che Hoon Jeong* from Denison University in Granville, Ohio for his extensive work on compiling the topics for the misinformation dataset.

economic, informational, political, and socio-cultural environments in each, such as their level of democracy, trust in science, income inequality, and healthcare strength.

Misinformation during a pandemic

First, a clarification on the terms we use in this article. In parallel to the rise of social media and other online platforms over the past 15 years, there has been a proliferation of studies looking at the emergence, spread, consumption, and effects of misleading and false information by analyzing the phenomenon in various terms: how rumors spread on Twitter [9], real-world impacts of hoaxes at Wikipedia [10], how individuals consumed fake news prior to 2016 US presidential election [11], or how regular people, and not just state-supported media, actively participate in generating disinformation in Russia [12]. In this paper, we chose to use the term misinformation over the aforementioned alternatives. As defined in Merriam-Webster's dictionary [13], misinformation refers to “incorrect or misleading information”. According to this definition, any piece of information that is partly or fully false can be labeled as misinformation, irrespective of an intent to deceive on the part of those who produce or diffuse the information. Thus, misinformation is a broader term, which comprises disinformation, i.e., false information that is “deliberately” spread “in order to influence public opinion or obscure the truth” [13], as well as those shared inadvertently or unintentionally. As discussed in the “Data” section below, this term befits the dataset we use. In line with some other scholars [14, 15], we also avoid the popular term fake news both because it is polarizing and politically charged and also it is rather limited to forms of misinformation that are deliberately designed to mimic news from established and mainstream news organizations [16].

Misinformation during the COVID-19 pandemic has reached such high levels that World Health Organization (WHO) and other United Nations bodies recognized the need to fight against false and misleading information as a critical part of the global pandemic strategy [17]. Many of the falsehoods regarding COVID-19 have been inspired by and interacting with conspiracy theories that predated the pandemic, including those involving “Big Pharma”, GMOs, Bill Gates, “deep state”, or a ring of satanic pedophiles [18–20]. The outcomes have been tangible: misinformation has significantly contributed to the spread of the illness and preventable deaths by promoting ineffective and harmful treatments and discouraging people from basic prevention such as wearing a mask or maintaining social distance. For instance, a study shows that over just a few months in the first of half 2020, approximately 800 people died and many more were hospitalized after drinking methanol as a cure for coronavirus [21]. More recently, vaccine-related misinformation has curtailed the vaccination efforts of many countries around the world, including the US where as of September 2021 the rate of vaccinated individuals fell short of the Biden administration's original ambitions [22]. This was particularly worrying as hospitalization and death rates were much higher among the unvaccinated [23].

Unsurprisingly, there is a growing literature around misinformation regarding COVID-19, which can be arguably grouped into three main streams of research. A first stream strives to understand different types of misinformation in terms of

their sources, spread patterns, and influence as well as analyzing how misinformation differs from more accurate and factual information in those respects. The main findings indicate that most news shared online are accurate; however, they are less likely to be shared than inaccurate ones [24, 25], which are produced and spread by denser and more organized communities [26, 27]. In addition, we learn that more misinformation circulates on social media platforms than on traditional news media [28], and misinformation coming from public figures such as celebrities or politicians, although making up a relatively small part of online misinformation about COVID-19, generates higher levels of engagement and support than other types of falsehoods [29, 30].

A second group of research looks at what specific factors push individuals to believe in or share misinformation. Accordingly, they show that certain psychological predispositions, political leanings, and daily habits such as tendency to reject expert information [31, 32], political conservatism [33], and right-leaning media consumption [34] are positively correlated with expressing and propagating misinformed views about COVID-19; whereas others such as greater science knowledge and cognitive reflection [35, 36], and trust in science and scientists [37] are negatively associated with those. Furthermore, it is interesting that an individual's worry for personal health doesn't seem to have an effect on their propensity to share COVID-19 related misinformation [38].

A third category of research articles focuses on how exposure to misinformation affects health behavior during the COVID-19 pandemic. The findings are clear as follows: consuming and/or believing in misinformation has a significant negative effect on willingness to take preventive measures such as wearing a face mask [39], maintaining physical distancing [40, 41], or getting vaccinated [42, 43]. In addition, belief in COVID-19 related falsehoods, including conspiracies, predicts greater use of pseudoscientific practices such as consuming garlic [44] or hydroxychloroquine [45]. Taken together with the research streams discussed previously, these studies shed light on the level of threat that misinformation poses on public health.

In parallel with these three streams, a relatively small but emerging group of studies analyze how economic, political, and socio-cultural differences among countries impact misinformation during the pandemic. This body of work illustrates the significant effect of a number of county-level variables on sources and types of as well as exposure and susceptibility to misinformation. Hence, countries' levels of uncertainty avoidance [46]; political and media freedom, and mobile connectivity [47]; human development [48]; political conservatism and political control over media [49]; Gross Domestic Product [50]; media fragmentation and partisanship [51] have been found to shape the misinformation environment beyond the individual-level factors.

This study contributes to the existing literature by using the largest dataset of online misinformation to our knowledge: over 10 thousand falsehoods produced and shared across 129 countries.³ It is worth noting that, unlike most other studies

³ It is worth indicating that geolocation and translation into English has been done by the sources providing data to Poynter Institute [1].

focusing on online misinformation [14, 25, 52, 53], our study does not solely rely on Twitter, but uses false and misleading information from various sources, including social media platforms such as Facebook, WhatsApp, Instagram, and Telegram. In fact, Facebook leads the list by making up 4347 of those pieces of misinformation. We find this important, because Facebook is not only the biggest social media platform globally with its 2.8 billion users [54], it is also the most popular one for COVID-19-related misinformation [24]. For comparison, Twitter had slightly less than 400 million active users as of July 2021 [54]. The prominent role of Facebook in the spread of Covid-19-related misinformation has been documented in a number of settings including the USA (2020), and Egypt (2021) [55–57].

Another unique aspect of this study is the wide range of variables it uses to compare COVID-19 misinformation across countries. As the “Data” section below discusses, the countries are compared based on 14 variables aiming to measure political, economic, informational, and socio-cultural environments in each, such as the levels of corruption perception, human development, press freedom, and healthcare strength. The Health Belief Model posits that when facing a health threat like the COVID-19 pandemic, individuals’ likelihood of engaging in preventive health behaviors depend on the perceived level of threat and perceptions regarding the potential benefits of and barriers of participating in such behaviors [58]. The literature laid out above shows how misinformation affects those perceptions by downplaying the threat, distorting the facts around the origin and spread of the pandemic, and offering ineffective or harmful treatments. Therefore, we believe that our study can help public health authorities better leverage their country-level resources while fighting against online misinformation such as by strengthening media freedom or improving scientific literacy.

Data

The data for this study include 10,131 falsehoods about the COVID-19 Pandemic in varying intensities of misinformation.⁴ Observations have been collected from the Poynter CoronaVirus Facts/DatosCoronaVirus Alliance Database provided by the Poynter Institute [1]. Poynter Institute is a non-profit NGO focused on journalism and research and based in St. Petersburg, Florida (<https://www.poynter.org/>). The

⁴ The types of misinformation and their count as provided in the original (raw) dataset is provided here. False: 8615, Misleading: 655, MISLEADING: 383, Partly false: 132, NO EVIDENCE: 128, Mostly false: 104, misleading: 64, No evidence: 58, No Evidence: 49, PARTLY FALSE: 46, Explanatory: 38, Mostly False: 31, partly false: 21, Partially false: 18, MOSTLY FALSE: 14, Partly False: 13, no evidence: 12, missing context: 9, mostly false: 9, MOSTLY TRUE: 8, Misleading: 7, Missing context: 6, mainly false: 6, HALF TRUE: 6, false context: 4, Mostly True: 4, MISSING CONTEXT: 3, Partially False: 3, Partially true: 3, Two Pinocchios: 3, Fake: 3, Half True: 3, Inaccurate: 2, Partly FALSE: 2, misLEADING: 2, half true: 2, PARTLY TRUE: 2, Misleading/False: 2, Unproven: 2, "(Org. doesnt apply rating)": 2, Correct: 2, Missing Context: 1, partially false: 1, MISLEADING/FALSE: 1, EXPLANATORY: 1, mainly correct: 1, UNPROVEN: 1, True but: 1, Partly true: 1, Partially correct: 1, IN DISPUTE: 1, Mostly true: 1, false and misleading: 1, Mixed: 1, HALF TRUTH: 1, MiSLEADING: 1, Unlikely: 1, Misinformation / Conspiracy theory: 1, Fake news: 1, Unverified: 1.

dataset is updated daily and provides a comprehensive understanding on the evolution of misinformation during the progression of the pandemic. The dataset covers the period between January 2020 and February 2021. For each observation, Poynter Institute provides the fact-checker that has provided information on the intensity of misinformation to the institute, the date on which the story was published, origin country/countries/continents, the intensity of misinformation, a brief summary of the story, the original text, and the link to the story. Most of the misinformation has initially been published on social media outlets, such as Facebook, Instagram or YouTube. In most cases, it is hard to know the true origin of the misinformation, since they are posted on different social media outlets on the same day. We should note that our study is not the only one that uses Poynter Institute's dataset. A few other articles [14, 29, 47] also rely on the Poynter CoronaVirus Facts/DatosCoronaVirus Alliance Database. However, our study uses the highest number of falsehoods—over 10 000—not just among the studies that use this particular dataset, but among all published research articles that focus on misinformation around COVID-19.

This study stands apart from other work in that it uses a carefully controlled hand-labeling of topics of falsehood. More specifically, to extract more information from the dataset, we manually labeled 28 different topic categories that are mentioned by the Poynter Institute in association with COVID-19. (The topics that have been identified by the Poynter Institute are aid, animals, conspiracies, crime, cures, detection, food, governments, hospitals, individuals, insurance, laws, lockdown, medical equipment, medicine, origins, other diseases, predictions, prevention, religion, risk factors, spread, symptoms, travel, vaccines, videos, technology, and NGOs.) Manual labeling is a common practice in NLP-research and has been used in other studies, as well.⁵ In contrast to the suggestion by the Poynter Institute that each news belongs to exactly one topic, using excellent research assistance, we identified the news belonging to more than one topic and marked those accordingly. We used stratified sampling to check for the quality and consistency of the data collection process. The original dataset has later been enriched by using NLP by removing the stop words, using contraction mapping, removing links, emojis and hashtags, POS-tagging the words, and lemmatizing the news content. Brief and long summaries have then been merged to provide more information. The descriptive table below provides a few interesting facts about the dataset (Table 1).

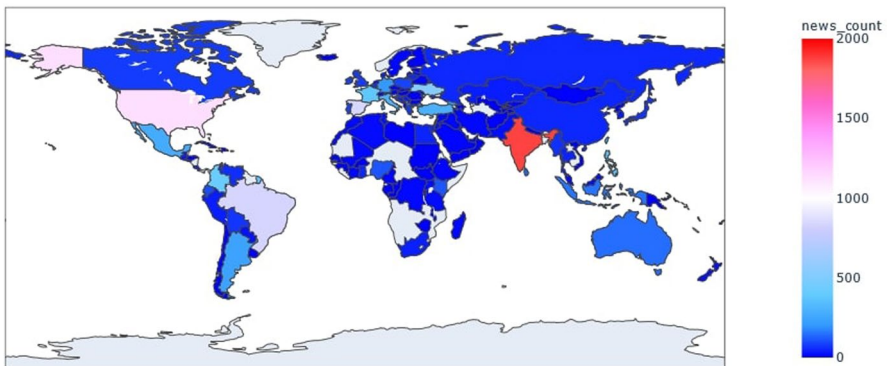
The map below shows the aggregate count of misinforming and misleading stories coming from each country. Most of the falsehoods were produced in India, the USA, Spain, Brazil and a few other European countries. As indicated by the map in Fig. 1, there is some correlation between the intensity of the pandemic and the amount of misinformation production. For example, as of September 2021, the USA, India, and Brazil were the top three countries in the world in terms of both the number of confirmed COVID-19 cases and deaths [65]. Also, as shown below, the period between March 2020 and May 2020 was the peak of misinformation

⁵ Hand-labeling of textual data is a technique that has been implemented in other related studies, as well [59–64].

Table 1 Descriptive table for the falsehood dataset

Description	Minimum	Average	Maximum
News count	Vanuatu (1)	85.7 per country	India (1892)
Brief summary length	3 words	17.1 words	113 words
Long summary length	2 words	27.3 words	198 words
Brief summary length (lemmatized)	1 word	10 words	63 words
Long summary length (lemmatized)	1 word	14.1 words	112 words
Merged summary length (lemmatized)	4 words	24.1 words	120 words
	1st	2nd	3rd
Social media source	Facebook (4347)	Twitter (1088)	WhatsApp (811)
Topic count	Individuals (4769)	Governments (2003)	Videos (1308)
Productive dates	03/19/20 (170)	03/23/2020 (154)	03/17/2020 (148)

Average Fake News Count by Country is 85.7

**Fig. 1** Geographic distribution of the falsehood counts

production worldwide, and during that time, Spain had a higher number of cases than other European nations such as Italy, France, or Germany, despite having a significantly smaller population size than them [17].

The stacked bar graph below shows the distribution of topics for each day, and the red line plot shows the number of stories published per day. As expected, there is a strong correlation between the two; nevertheless, there are some news that could not be assigned to any topic and some others with more than one topic. As one can observe below, the dataset shows signs of seasonality and trend. Specifically, the number of falsehoods reached its peak in the period March 2020–May 2020, in a time when there were significant uncertainties about the definition and implications of the Sars-Cov2 virus. It is worth noting that starting from June 2020, significantly fewer falsehoods have been published worldwide, arguably thanks to clarifications and new information regarding the origin and spread of the virus as well as methods

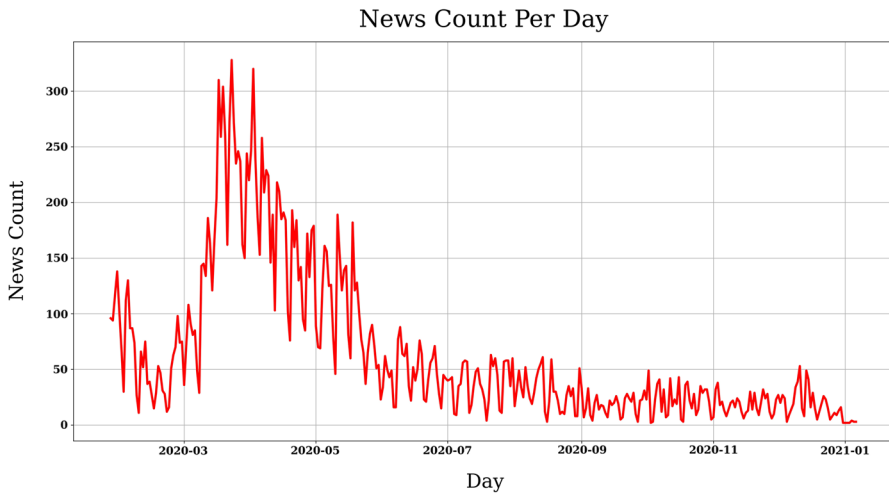


Fig. 2 News count per day

of treatment and prevention. Three key moments in the pandemic may have played important roles in reducing the amount of misinformation regarding COVID-19:

1. On May 27, 2020, Dr. Anthony Fauci, the director of the National Institute of Allergy and Infectious Diseases, announced that a vaccine would be ready by December 2020 [66], which offered a promise of an end to the pandemic and lockdowns.
2. On June 17, 2020, the WHO announced that it was stopping its trial of the hyped anti-malaria drug hydroxychloroquine after new data suggested that the drug was not effective for COVID-19. This helped dispel the myths regarding its benefits that were spread by individuals and organizations, including the influential French physician Didier Raoult, Donald Trump, and Russian state-owned media [67].
3. On July 7, 2020, the WHO announced that COVID-19 may be an airborne disease mainly transmitted through respiratory droplets [68]. This not only reinforced the message around the importance of mask-wearing, but also helped limit the spread of misinformation concerning transmission through food or packaging (Figs. 2, 3).

As noted above, a significant contribution of this study is the assignment of the falsehoods to topics. Similarly, the exploration of co-occurrences between topics is of great importance to understand how groups of countries with varying characteristics produce misinformation. The graph below shows the aggregated results for topic co-occurrences. In the graph, the diagonal values show the counts for topics in the dataset, and the non-diagonal values indicate the number of co-occurrences between topics. Most falsehoods refer to an individual creating a story, and therefore have been classified as ‘Individuals’. Other topics with a significant amount

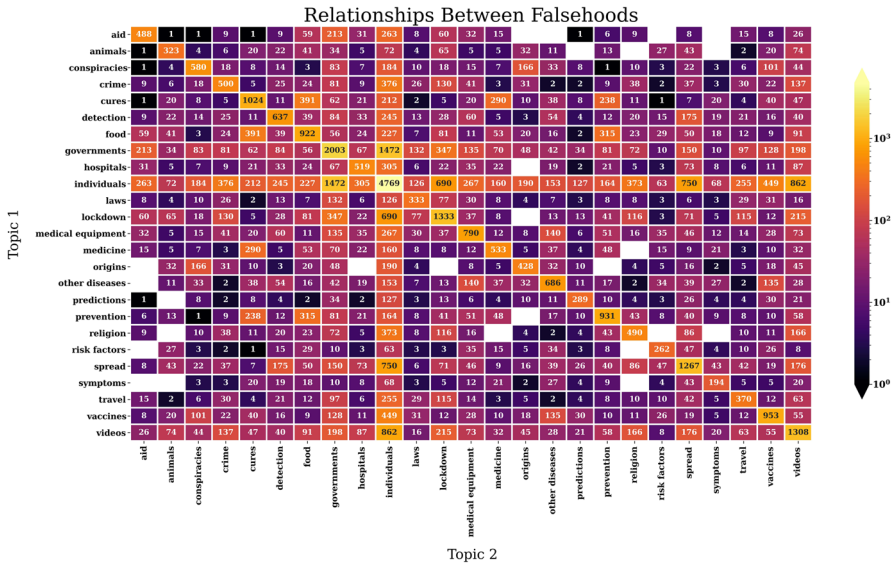


Fig. 3 Topics and co-occurrences

of co-occurrence are ‘Governments’, ‘Conspiracies’, ‘Cures’, ‘Food’, ‘Prevention’, ‘Spread’, and ‘Medicine’. We explain these co-occurrences as follows: a great percentage of falsehoods were created between January 2020 and May 2020, a period in which there were still lots of unknowns about the origin, spread, cure, and prevention of the disease. This created a fertile ground for fully or partially misleading stories (e.g., benefits of eating a particular food as a protective measure) as well as outright conspiracy theories (e.g., Sars-Cov2 virus being a biological weapon). Starting from June 2020, the aforementioned announcements by leading figures and organizations seem to have helped reduce the misinformation around those subjects. However, despite a significant reduction in its amount, misinformation continued to be produced, and started to focus on other issues such as vaccines, politicians, and the impact of COVID-19 on healthcare systems. The count values in the graph have been colorized using a logarithmic scale, and there are a few instances where no co-occurrence has occurred (those observations have been colored as gray). On average, each falsehood has 2.165 co-occurring topics, and the maximum number of topics found in a single observation is 8.

During the collection and processing of the dataset, we were pleasantly surprised to see news with a varying degree of credibility and fantasy. There were stories with complete divergence from reality, nevertheless, they sounded credible; others gave the impression that they were produced in distant corners of a world of fantasy. Overall, we can divide the stories into two main categories. On the one hand, there were those that had some ground in reality but still distorted the facts and misled the reader, whether intentionally or not. One example was a story published in Brazil that claimed the FDA had warned the public that COVID-19 vaccines cause stroke. In reality, the FDA had prepared a table listing all possible side effects that

must be looked after. That did not mean those side effects were known to happen after immunization with vaccines. On the other hand, some stories did not have any basis in facts and deserved to be labeled conspiracy theories. For instance, a story published in Georgia used a fake quote attributed to an American virologist and claimed that 5G high-frequency towers were installed to control humans implanted with microchips through vaccines, and the Rockefeller and Mason families were the financiers of this project. A range of examples showing this variation and the diversity of topics are in the Table 2.

As mentioned above, we compared the 129 countries in the dataset based on 14 variables aiming to measure economic (e.g. income inequality), political (e.g. trust in government), informational (e.g. press freedom), and socio-cultural (e.g. trust in science) environments in each. While selecting those variables, our goal was to draw a thorough picture of countries to understand what factors impact the production of COVID-19-related misinformation in different settings. A detailed discussion on the social, economic, and political variables we have used can be found in “Part I” of the Appendix.

Research questions

The rich dataset on misinformation provides opportunities to make statistical comparisons between countries (e.g., their social and political characteristics), topics of the news, and the content of their text. In addition, since the evolution of the pandemic was a socially dynamic phenomenon, the fourth aspect is time. The examination of the dataset shows that there is considerably high variation between individual observations over time, and this study aims to find if these micro-variations can lead to meaningful macro-level comparisons. The richness of detail and the opportunities offered by the variables guided us to construct a methodological framework to find the larger patterns in the data.

By attempting to cover a large ground, this study aims to conceptually and empirically contribute to the literature by grouping the countries according to the predominant types of falsehoods they produce. In order to cover the dynamic evolution of misinformation over the course of 13 months during the pandemic, the paper looks at four pillars of analysis that can be grouped under topic analysis and content analysis. With this background in mind, the paper aims to be one of the pioneering contributions to literature. Despite the early optimism stemming from vaccine rollouts, as of September 2021, COVID-19 was still a major health threat around the world due to factors including new mutations and many countries missing their vaccination targets. Particularly, we can predict that a sizable number of low- and middle-income countries will be fighting against it for a long time due to unacceptably low vaccination rates—e.g., only around 2% received at least one dose of a vaccine in low-income countries [83]. In addition, socially and biologically, the field is still characterized by known unknowns and unknown unknowns; therefore, conceptual contributions may provide helpful guidance to researchers and policy-makers. Theoretically, the paper also aims to give comparative politics and sociology scholars opportunities to look deeper into the reasons why different countries

Table 2 Examples of falsehoods

Brief description of the story	Origin	Publication medium	Topics assigned	Fact checking	Nature of the misinformation
The WHO stated that wearing a mask while playing sports increases the levels of carbon dioxide we absorb	Spain	Various media outlets	Medical equipment	The WHO recommends not wearing a mask when carrying out intense sports activity, but it never said it is due to an increase in carbon dioxide	Distortion of reality
Studies show the flu shot increases a person's susceptibility to COVID-19	Australia	Facebook	Vaccines, risk factors, other diseases	In fact, the study referred to in the post shows that the flu shot does not make people more susceptible to any other respiratory infection	Complete fabrication
There is a sign outside the Bill & Melinda Gates Foundation that reads as "Center for Global Human Population Reduction."	United States	Facebook	Individuals, conspiracies	There is no such center run by the Gates, nor a sign as described	Complete fabrication
A diet rich in alkaline foods can eliminate the coronavirus	Brazil	Facebook	Cures, food	There is no food capable of preventing or curing someone from Covid-19	Complete fabrication
Bing Liu, a researcher who was about to discover a vaccine for COVID-19, was murdered in the US	Brazil	Instagram	Crime, individuals, vaccines	Although a researcher named Bing Liu was killed in the US on May, he was not about to discover a vaccine for COVID-19	Distortion of reality
President Donald Trump attended a recital of Quran to fight coronavirus	India	Facebook	Governments, individuals, religion	The video is real, but it is from 2017, when Trump attended the inaugural prayer service after taking the oath as the 45th president of the United States	Distortion of reality

produce different types of falsehoods and to analyze which socio-cultural, economic, and political variables affect the misinformation environment more than others. Methodologically, the paper takes advantage of a variety of statistical techniques, including a selection of network similarity algorithms. The use of network similarity algorithms to compare texts has largely been neglected in the computational social science literature.

It is also important to mention that before conducting this study, we considered a different strategy as well. In fact, initially, we extracted hundreds of millions of tweets from more than ten countries around the world and calculated the similarity between those tweets and the dataset on falsehoods. However, this approach resulted in no findings, and using different text similarity algorithms, we were not able to identify any matches. This led us to believe that a targeted approach to analyze misinformation could be more effective than trying to discover patterns in large but random samples.

A closer elaboration of the research questions has been provided below.

Topic analysis

The first two sets of research questions analyze the causal mechanism of topic selection by different groups of countries. The countries have been grouped by using economic, informational, political, and socio-cultural variables that we have introduced in the Data section. The set of questions below help to understand the macro patterns in the dataset by minimizing the errors associated with labeling through manual classification.

RQ1) Divisive and connective topics

RQ1a) In terms of topic creation, what are the topics that two groups of countries utilize in the most comparable amount vis-a-vis each other? In other words, what are the topics that are the most connective?

RQ1b) What are the topics that are the most divisive?

RQ2) Topic co-occurrences

RQ2a) What are the topics that co-occur the most?

RQ2b) Are some groups of countries statistically significantly different from others in terms of topic co-occurrence?

RQ2c) Is there a time frame in the evolution of COVID-19 in which topics were more similar to each other?

Content analysis

The second group of questions look more deeply at the content of the news by calculating the similarity between and across news associated with different topics and also analyze the news from the perspective of ‘unusualness’. (This will be explained

in greater detail in the Methods section.) By doing this, we aim to understand if there is any association between topic correlation or content similarity across different groups of countries. The goal is to find out if countries have been inspired from each other in terms of content creation and how this relates to variables collected.

RQ3) Content similarity

RQ3a) Are there groups of countries that produce news that are significantly more similar to each other?

RQ3b) How does the similarity between news change over time?

RQ4) Misinformation unusualness/creativity

RQ4a) Are there groups of countries that are more creative than others in content formation?

RQ4b) How does creativity evolve over time?

As the research questions suggest, the paper aims to offer a descriptive perspective into the creation of a framework on misinformation production. The statistical tools used in the paper are elaborated closely in the next section.

Methods

The methodological tools used in the paper have been chosen to find similarities and differences between individual and groups of observations. To answer the four sets of research questions indicated above, we employed a variety of tools, including t-test, calculation of entropy and GINI index as a measure of information gain, k-means++ clustering, network similarity algorithms, and content comparison algorithms (NLP). (For the analysis, Python programming language and associated libraries were used.)

The data on falsehoods were collected from the Poynter Institute using webscraping techniques. (Poynter Institute allows the use of their data for research purposes.) The data was later manually processed to associate each observation with at least one topic from a collection of 28 different topics (topics were identified through the examples provided on Poynter Institute's website). Around 500 observations could not be associated with any topic and therefore discarded. For the classification of topics, a few unsupervised clustering options have been tested, such as latent Dirichlet allocation [84] and non-negative matrix factorization [85]; nevertheless, the most coherent results have been obtained through manual labeling.

As previously mentioned, the starting point of this paper is the assumption that the nature of misinformation production is highly dependent on the personalities of countries that can be associated with certain socio-cultural, political, informational, and economic characteristics. In that sense, our study follows previous work such as Sauvy's "Three Worlds, One Planet" (1952) [86] that coined the term *Third World*; Huntington's "The Clash of Civilizations?" (2000) [87]; Hall and Soskice's *Varieties*

of *Capitalism: The Institutional Foundations of Comparative Advantage* (2001) [2]; and Wallerstein's *The Capitalist World Economy* (1979) [88] with its core versus periphery distinction. However, as different from them, we do not prioritize a specific dimension (e.g., economic systems or “culture”) as the primary distinguishing variable; instead, we aim at drawing a more comprehensive picture of countries by using 14 variables ranging from income inequality to trust in science and scientists to colonization history. In order to simplify and automate the classification of countries, a generally accepted and useful clustering algorithm, k-Means++ [89] was used. To pre-process the data for clustering, categorical variables have been converted into a 5-point Likert scale, and 0–1 normalization has been applied to all variables. The optimal number of clusters was determined using the cluster variation (SSE) and the ‘elbow method’. Two clusters came out as the optimal number, and six as the second optimal choice; to better represent the variation among countries, we chose six.

To handle the missing data in the datasets, different imputation methods were used. For the missing social, economic and political observations, a technique called “multivariate feature imputation” was implemented [90]. This technique uses a two-dimensional matrix as the input and models each feature with missing values as a function of other features using an iterated round-robin fashion. This is suitable for our case, since the variables at hand possibly have causal connections. To fill in the missing values in the time series datasets (for similarity and unusualness), K-nearest neighbors (KNN) [91–94] algorithm was used. The assumption behind this algorithm is that missing observations can be approximated by the values of the closest points, most frequently by taking the average of ‘k’ many points around the missing observation. As argued in a multitude of works using KNN for imputation (This approach is believed to work well in time series data with missing observations for which the best predictor of the missing points are the values temporally closest to them.

In addition, the same set of social, political, cultural, and economic variables were used to reduce dimensionality using principal components analysis. For the PCA, two dimensions came out to be the optimal choice based on the scree plot. Two other dimension reduction techniques, namely t-SNE [95] and spectral embedding [96], were considered; however, PCA was preferred as a traditional method to obtain two variables that are not correlated. The correlation map between the variables used and the reduced dimensions can be seen in the plot below (significant correlation values are marked with a black box). As Fig. 4 shows, most of the social, cultural, and political variation can be explained by four variables: corruption perceptions index and health coverage (PCA—Dimension 1), GINI Index, and trust in government (PCA—Dimension 2). Figure 5 is a representation of the variables after dimension reduction (PCA). As evidenced by it, countries in the dataset can be successfully grouped into six clusters (by using k-means++) with the help of the variables listed below. Findings have been consistent and robust after running the clustering algorithm ten times.

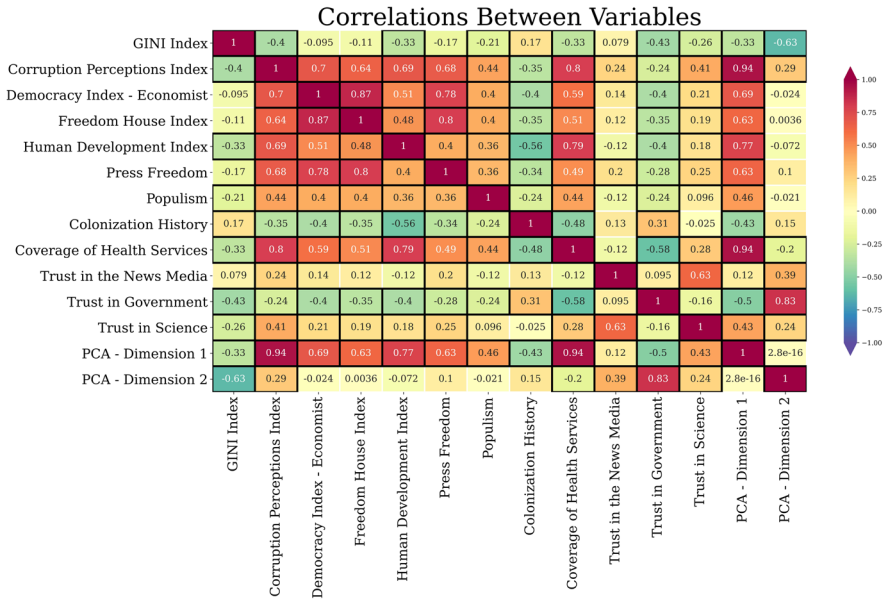


Fig. 4 Correlations between variables

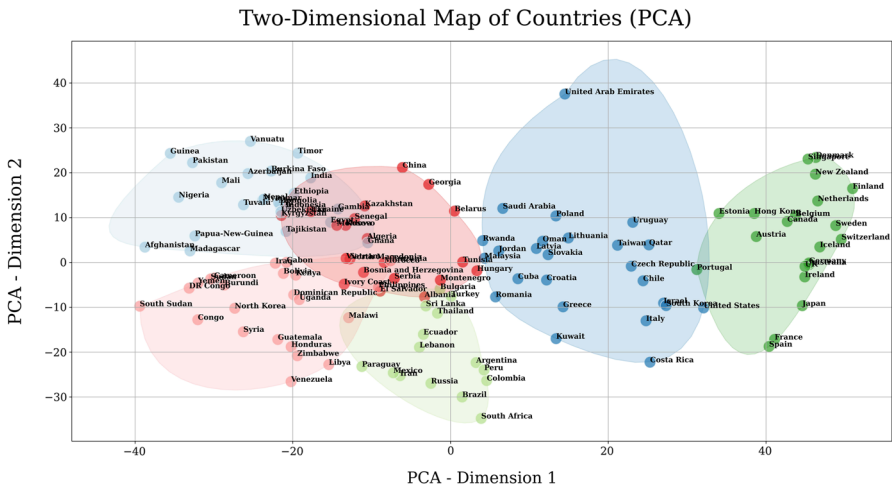


Fig. 5 Countries in different social, economic and political clusters

In order to compare the use of topics across groups of countries, an idea employed by decision trees was used. When decision trees are applied for classification goals, entropy and GINI index are the two most frequently used cost functions to calculate information gain/purity of the classes obtained by splitting the data. Thus, we wanted

to find, given two groups of countries that are different by a single feature (for example, two groups of countries with different levels of democracy), which topic (in terms of its frequency) is the most different and which topic is the most similar among the two. Across the social, cultural, economic, and political variables, hundreds of comparisons between groups of countries were made, and the count values for most divisive (most different) and the most connective (most similar) topics were identified. Finally, these count values were inversely weighted by the count of the associated topic in the dataset to obtain a ranking for the most divisive and connective topics.

The paper assumes that the diversity of word usage in news reflects creativity; thus, more ‘unusually’ worded news are more creative. To measure the unusualness of the observations, 3-g and 4-g for the cleaned and lemmatized news were found. These n-grams were then used to calculate the TF-IDF score of each observation, which corresponds to the sum of TF-IDF scores for each n-gram associated with that observation. Observations with higher TF-IDF scores are believed to be more important and more creative; those with lower scores are considered as less unusual. This information was then used to compute how the unusualness changes over time and across groups of countries.

To calculate the document similarity between different observations, several considerations and attempts have been made, ranging from more generally accepted and earlier algorithms to more advanced techniques. Specifically, Word Mover Distance [97], Universal Sentence Encoder provided by Google [98], BERT embeddings [99], and Knowledge-based Measures [100] have been explored in the earlier phases of the analysis. All these models have turned out to be computationally too expensive to find the cross-similarities between over 10,000 documents. To solve this problem, TF-IDF (term frequency-inverse document frequency) scores have been calculated for all documents following the cleaning and lemmatization process [101]. Eventually, cosine similarities have been found in over 50 million cross-comparisons. These similarities have then been aggregated to make cross-country comparisons using *t* tests.

Finally, network similarity algorithms were applied to compare adjacency matrices composed of bi-weekly aggregated topic correlations between documents. Topic similarities can be represented as graph data since one document can only have a limited number of topics, and more than one topic presented in a single document can change the impact of the misinformation dramatically (holistic assumption). To calculate the similarities between topic correlation matrices, the following two advanced graph similarity algorithms were used: Frobenius distance [102] and quantum-JSD distance [103]. The aggregated relative similarity matrix between topic ratios has been provided as an example in Fig. 6. Topics with yellow-to-red colored cross-similarities are more closely related, and topics with yellow-to-blue colored cross-similarities are rarely mentioned together. A closer elaboration on this relationship will be provided in the “Results” section. For more information on how network similarities have been calculated, please refer to the “Appendix” section. In a similar approach, PERMANOVA [104] and Anosim [105] techniques that allow the comparison of $n \times n$ -dimensional topic correlation matrices were put to test; however, ultimately, the tests were not reported because of the impact of data size on the results.

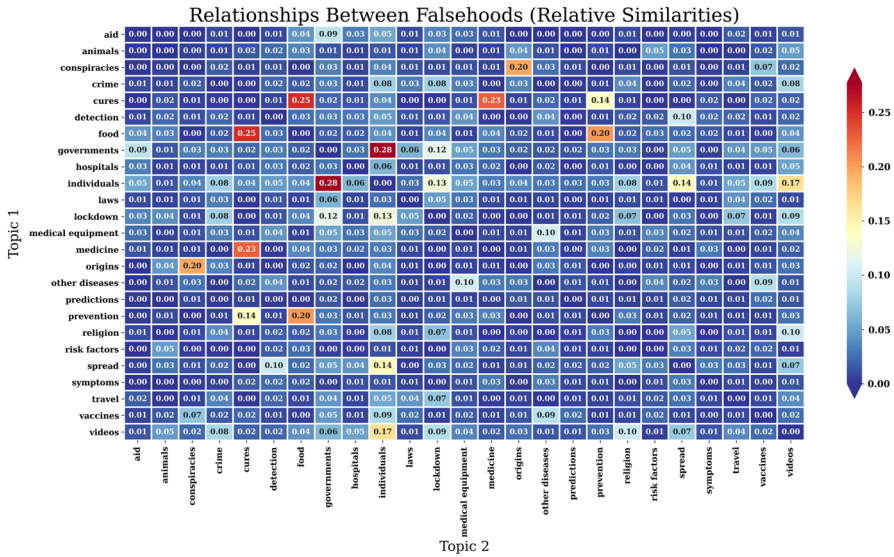


Fig. 6 Relative similarities between falsehoods of all topics

Empirical results

Topic analysis

The topic analysis focuses on two questions as previously mentioned: divisive and connective topics (i) and topic co-occurrences (ii). The results obtained for the first case indicate significant variance in the power of topics to differentiate groups of countries from each other. Thus, clusters of countries can be strongly associated with topics and vice versa. Topics were used to separate countries into clusters and these clusters were compared with the groups generated through the use of social, economic, cultural, and political variables. The results show that some topics lead to a much greater amount of cluster purity when used for generating groups. The analysis to calculate purity has been repeated by using two cost functions, entropy, and GINI Index, and the results are the same. The table below provides a ranking for the most connective and divisive topics. In each comparison the name for the most connective and divisive topic has been obtained and the number of times a topic appears as the most connective or the most divisive has been recorded. Finally, the count values have been inversely weighted by the falsehood count associated with that particular topic. The table below shows the ranking for most connective and divisive topics and these values. The inversely weighted values explain how strongly connective or divisive a topic is compared to others (Table 3).

As seen above, “conspiracies” is the most shared topic category across groups of countries. We explain this finding as follows: the major conspiracy theories,

Table 3 Ranking of most connective and most divisive topics

	Minimum entropy/GINI index (most connective)	Maximum entropy/GINI index (most divisive)
1	Conspiracies (0.071)	Animals (0.387)
2	Medical equipment (0.058)	Predictions (0.135)
3	Risk factors (0.053)	Symptoms (0.067)
4	Other diseases (0.048)	Laws (0.063)
5	Vaccines (0.047)	Travel (0.046)
6	Travel (0.046)	Risk factors (0.046)
7	Aid (0.029)	Origins (0.030)
8	Religion (0.029)	Prevention (0.029)
9	Animals (0.028)	Spread (0.017)
10	Symptoms (0.026)	Hospitals (0.015)
11	Laws (0.024)	Aid (0.012)
12	Origins (0.023)	Conspiracies (0.012)
13	Predictions (0.021)	Other diseases (0.010)
14	Medicine (0.017)	Medicine (0.009)
15	Hospitals (0.015)	Detection (0.008)
16	Crime (0.014)	Medical equipment (0.008)
17	Cures (0.012)	Religion (0.006)
18	Detection (0.009)	Food (0.005)
19	Videos (0.008)	Crime (0.004)
20	Food (0.005)	Videos (0.002)
21	Lockdown (0.004)	Vaccines (0.001)
22	Governments (0.003)	Governments (0.000)
23	Spread (0.003)	Cures (0.000)
24	Prevention (0.002)	Individuals (0.000)
25	Individuals (0.002)	Lockdown (0.000)

including those pointing to Bill Gates-led plots to implant digital microchips to control people, marking the virus as a biological weapon created by Chinese or American scientists, and those demonizing pharmaceutical companies as agents that worsen the pandemic and conceal the effective treatments, are produced by a small number of individuals and organizations with political and financial goals. Then, these are shared globally in the form of news stories occasionally through media outlets, but primarily via social media posts. In fact, a recent investigation conducted by the Associated Press and the Atlantic Council’s Digital Forensic Research Lab found that a few “superspreaders”—people and organizations such as Kevin Barrett, an anti-Semitic former lecturer on Islam, and the Montreal based “Centre for Research on Globalization”—were responsible for a great percentage of conspiracies on the origin of COVID-19 circulating online [106]. Similarly, a study published right before the COVID-19 pandemic found that 54% of all anti-vaccine ads on Facebook were funded by two organizations, even though most of the ads appeared to be grass-root discussions by concerned parents and neighborhood

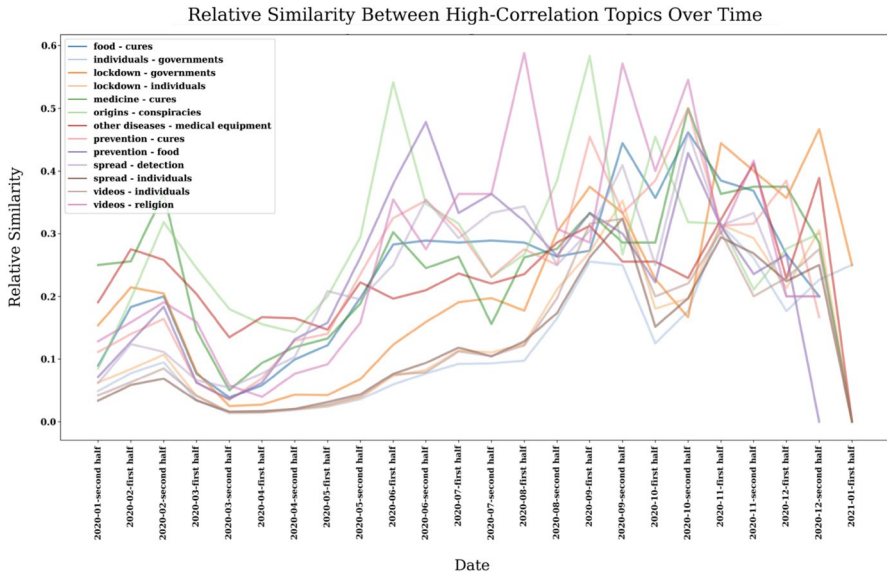


Fig. 7 Relative similarity between high-correlation topics over time

groups [107]. Thus, in addition to raising concerns around the use of Facebook and similar platforms to spread misinformation, this finding indicates that conspiracy theories regarding COVID-19 have a global appeal cutting across socio-cultural, economic, informational, and political variables that divide the countries.

Secondly, we looked at the co-occurrence dynamics of the topics in the topic analysis section. A one-to-one match between each topic gives close to 400 possibilities for topic pairs. Among those, co-occurring topics with an aggregated relative similarity of more than 0.1 have been selected and their number of co-occurrences have been inversely weighted by the total count of both topics (comparable to Jaccard similarity) in periods of two weeks. In other words, a matrix similar to the one in Fig. 5 has been produced for every two weeks topic pairs with high relative similarity have been observed. This gave us Fig. 6 below. The high similarity co-occurring topics are food-cures, individuals-governments, lockdown-governments, lockdown-individuals, medicine-cures, origins-conspiracies, other diseases-medical equipment, prevention-cures, prevention-food, spread-detection, spread-individuals, videos-individuals, and videos-religion. The figure below suggests that there is a pattern in the co-occurrence of the topics and the time series dataset can be clustered into the following two groups: before April 2020 and after. Higher values correspond to greater weighted topic co-occurrence and lower values indicate that co-occurrence has become weaker (Fig. 7).

Lastly, bi-weekly relative similarity matrices have been treated as networks and compared to each other using network similarity algorithms. This provided a systematic way to compare the dynamics of topic ratios in the misinformation dataset over time. To validate the results, two different algorithms (Frobenius distance and quantum-JSD distance) have been used and the results have been evaluated against

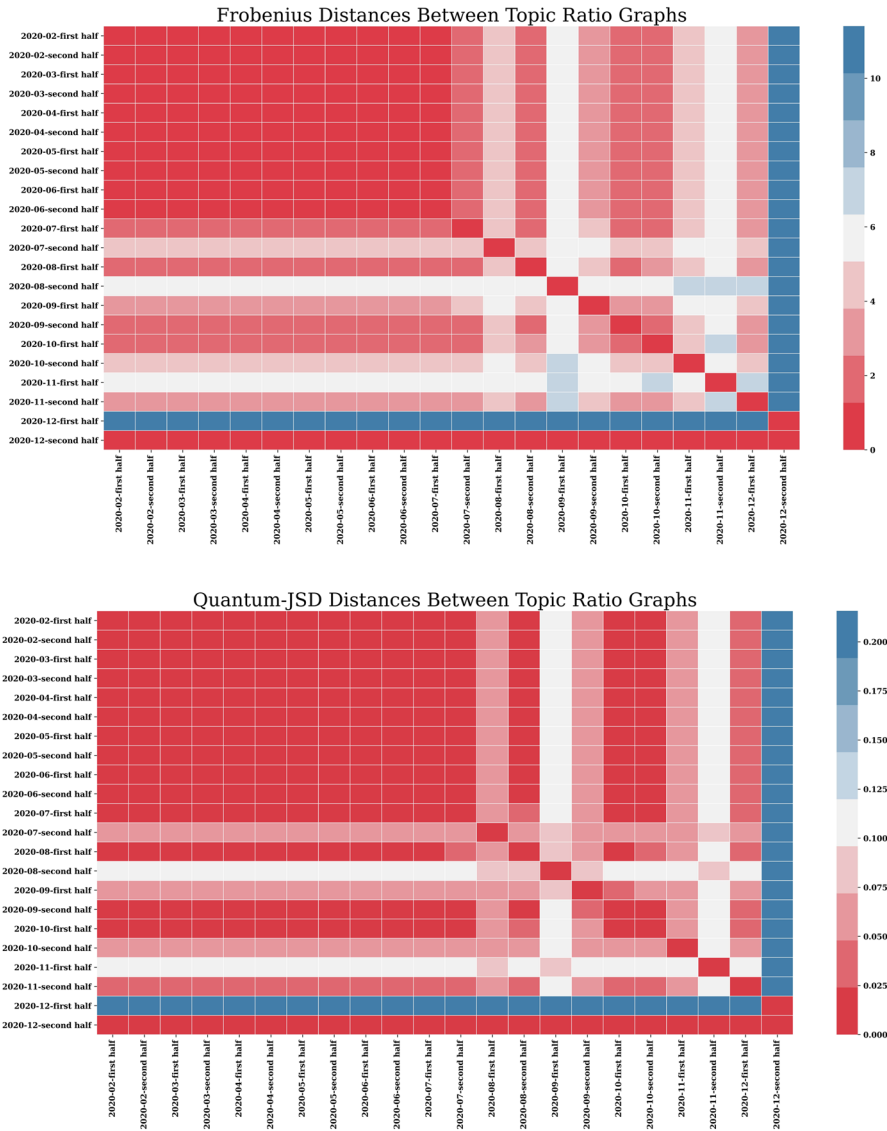


Fig. 8 Frobenius and quantum-JSD distances between the bi-weekly relative similarity graphs of topic ratios

each other. The results indicate that in the first few months of the pandemic, topic ratios have been comparable to each other; specifically, starting from February 2020 until the end of June 2020, results suggest that there has not been much variation. This finding is also reinforced by the results provided in Fig. 6 relatively more conservatively: highly correlated topics formed a pattern until the end of May 2020. This suggests intense cross-country exchanges and learning from each other in the

first few months of the pandemic. The graphs below show the similarities (or distances) between the bi-weekly relative similarity graphs. The rectangles in the intersection of two time points show the distance between two topic ratio graphs. Red values are associated with greater similarity and blue values correspond to lower similarity scores. In addition, to have complete data for the bi-weekly periods, the first and the last time series observations have been truncated (Fig. 8).

We interpret those results in line with the discussion in the “Data” section above. As we also mentioned there, the period until May/June 2020—the first few months of the pandemic— was characterized by uncertainties about the definition and implications of the Sars-Cov2 virus and the highest intensity of misinformation production. More specifically, there were still lots of unknowns about the origin, spread, cure, and prevention of the disease; each among the topics with a significant amount of co-occurrence. Starting from June 2020, with key announcements by leading figures and organizations regarding the origin and spread of the virus as well as methods of treatment and prevention—e.g. the WHO’ announcement that COVID-19 may be an airborne disease mainly transmitted through respiratory droplets—we saw a decline in the number of falsehoods related to those popular and highly co-occurring topics, while misinformation started to focus on other issues such as vaccines, politicians, and the impact of COVID-19 on healthcare systems.

Content analysis

Content similarity

In this section, we tried to understand if countries of similar social, economic, and political backgrounds produce news with similar content, or if there is a statistically significant difference between countries with different social, economic and political endowment in terms of content creation. The assumption was that the behavior of people is strongly related to national variables [108, 109] and this ultimately translates into writing. In fact, there is an extensive literature showing that individual’s everyday behaviors such as financial decisions [110], consumption habits [111], and health behaviors [112] are associated with national variables, including cultural values, human development levels, or business systems. Similarly, scholars point to how national-level factors such as political systems, economic indicators, or press freedom have determining impacts on the “journalistic cultures” [113], which, in turn, shape how different topics, including climate change [114] and international migration [115] are covered.

We broke the countries down into groups and compared the aggregated mean of pairwise similarity between the news for bi-weekly periods and for the whole dataset. The instances in which a comparison results in statistically significantly higher similarity results than the other sets of comparisons have been identified. These instances can be observed from the graphs below. The remainder of the comparisons can be found in the “Appendix”. On the whole, countries in the fourth cluster, countries in West and South-Asia, socialist/Arab-oil-based/advanced city economies, countries with low HDI (human development index), countries with very

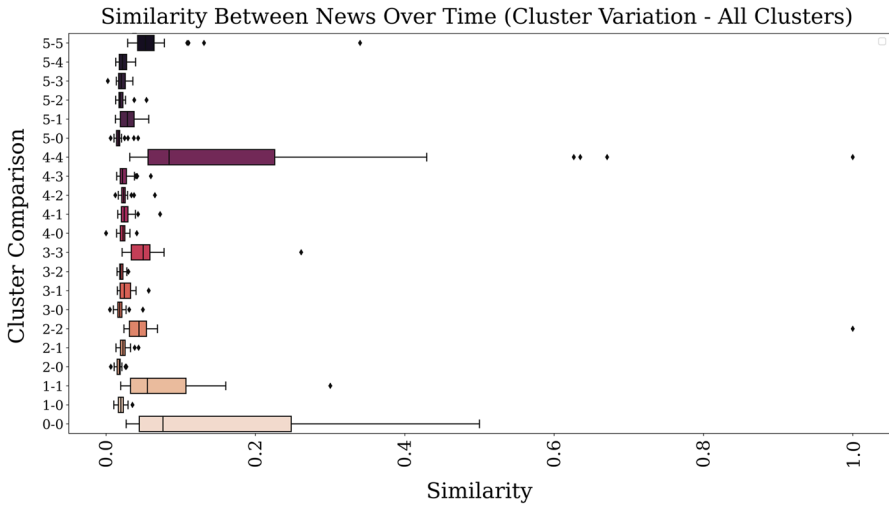


Fig. 9 Similarity between news over time, cluster comparison

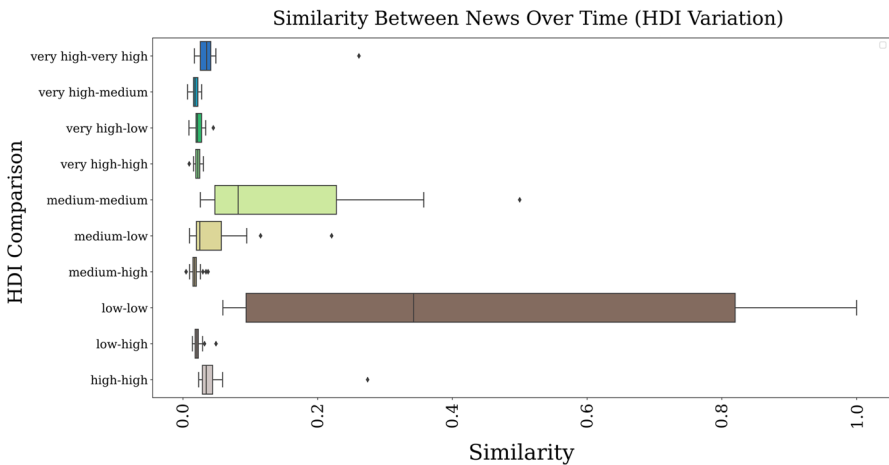


Fig. 10 Similarity between news over time, HDI comparison

serious press freedom problems produce news that are more similar to each other (the remainder of the comparisons have been reported in the “Appendix”) (Figs. 9, 10, 11, 12, 13).

We believe some of those findings are particularly worth discussing here. First, it should be noted that an overwhelming percentage of countries classified as

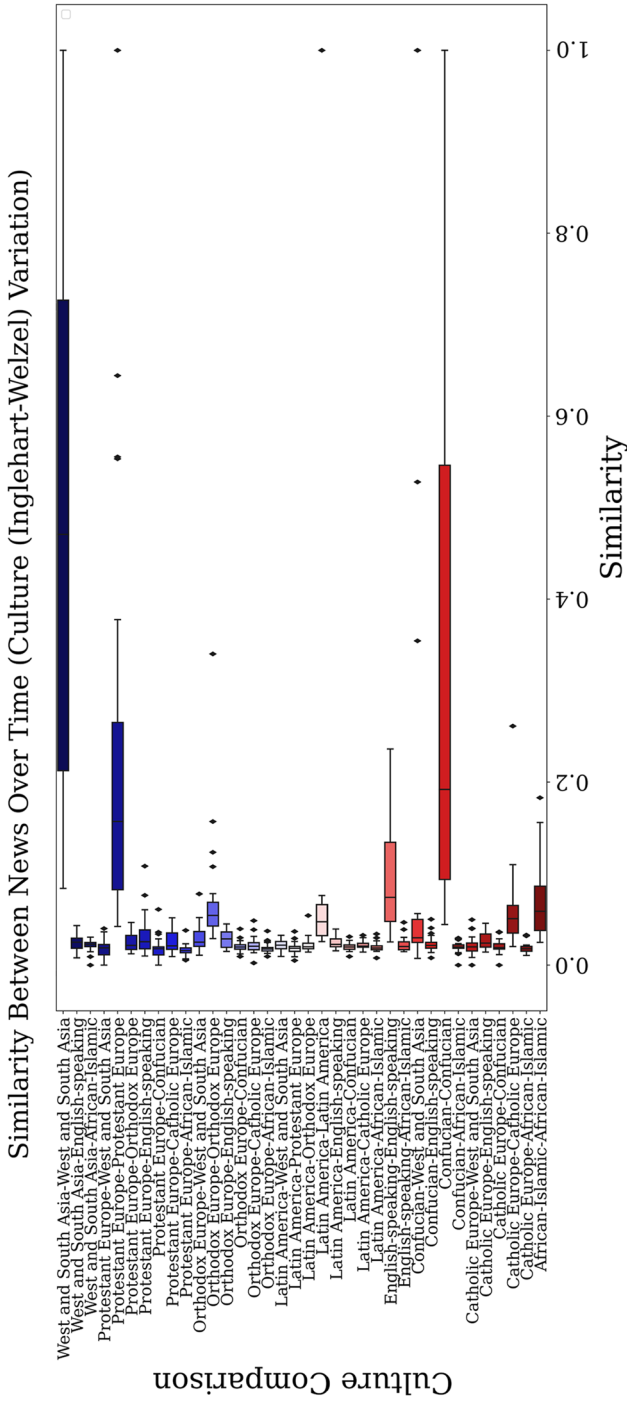


Fig. 11 Similarity between news over time, culture comparison

Similarity Between News Over Time (Business Systems (Witt) Variation)

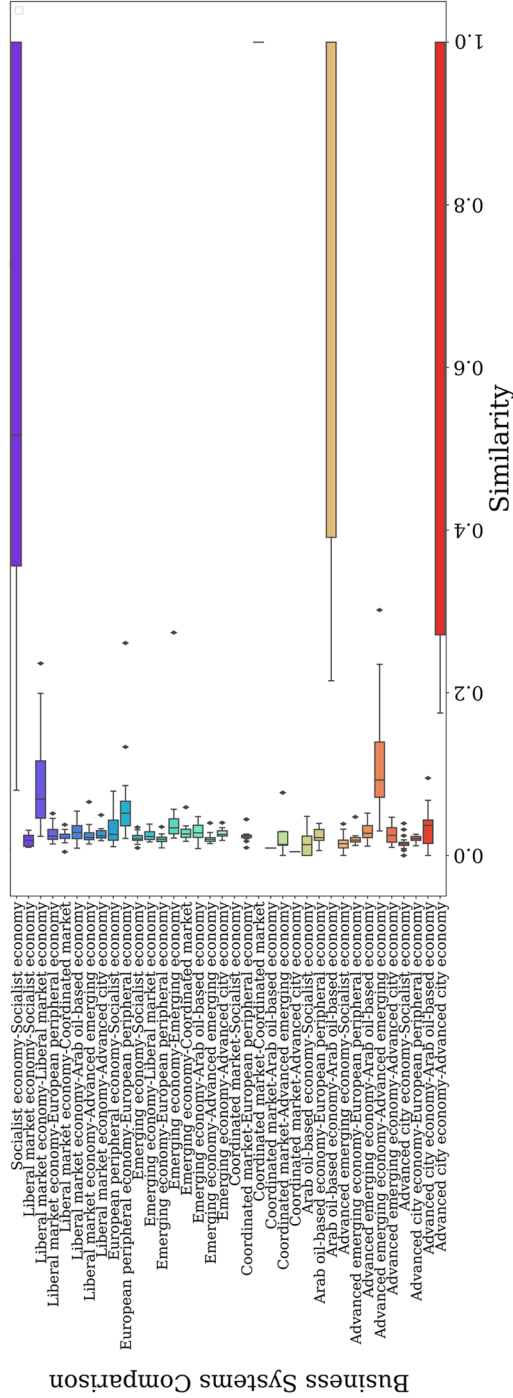


Fig. 12 Similarity between news over time, business systems comparison

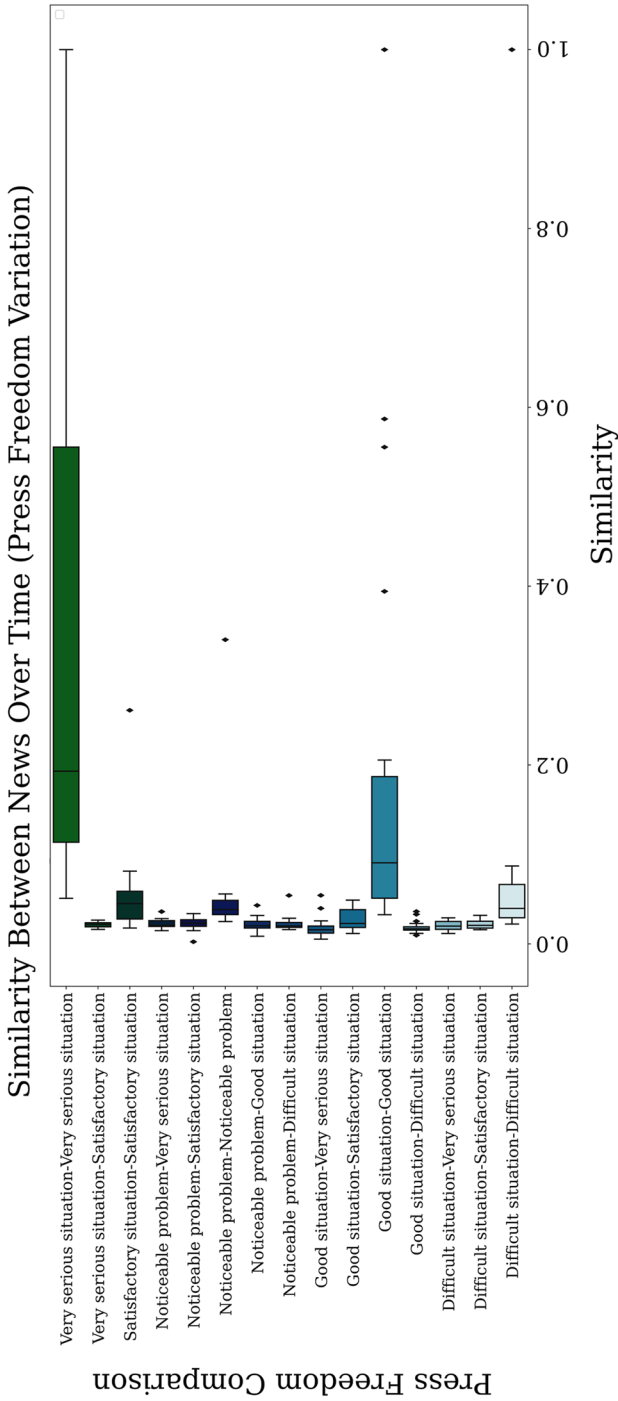


Fig. 13 Similarity between news over time, press freedom comparison

having low HDI are located in sub-Saharan Africa. A number of studies indicate that the COVID-19-related misinformation in the African continent has a few distinct characteristics, and we speculate that this might explain why the news are more similar to each other. Specifically, falsehoods related to unproven local remedies [116] and those stemming from religious beliefs [117] are found to be particularly common in Africa. In addition, distrust towards international bodies [118] and the history of unethical Western medical practices in the continent [119] are some of the other factors fuelling misinformation. In fact, our dataset offers some interesting examples. In Ivory Coast, stories claiming that neem leaf works against COVID-19 were posted thousands of times on social media despite no evidence. Similarly, falsehoods claiming that the Rwandan president Paul Kagame censured the WHO for rejecting a herbal tonic were widely shared on Facebook and Twitter across African nations, including Nigeria.

Countries classified as having “very serious” press freedom issues by the reporters without borders also produced more similar news. In line with our discussion on the “journalistic cultures” above, we believe that this might reflect the effects of government control and censorship of the media, which largely shape both the content and tone of the coverage with regard to the pandemic. In fact, in this group of countries— including Egypt, Iran, Saudi Arabia, China, and Vietnam, among others—not just the traditional media sources, but also the social media networks are subject to heavy government control [72]. For instance, a report by the social media exchange—an NGO working to advance digital rights in the Arabic-speaking region—shows how Egyptian authorities prosecuted a number of journalists, doctors, and activists who circulated news on the social media on the COVID-19 outbreak—e.g., the number of infections or deaths—that did not match the official discourse and numbers [120].

A third interesting finding is that the countries in the fourth cluster—the light pink colored cluster in Fig. 3 above—generated more similar news in terms of the content. Some of the countries in this group are Iraq, the Democratic Republic of Congo, Venezuela, Honduras, Kenya, Bolivia, Uganda, and Yemen. A few of the members of this cluster have also low HDI and/or very serious press freedom issues; therefore, the explanations above can partially apply to those countries. However, four distinct characteristics identify the countries in this cluster: a high perception of corruption of the public sector, a high degree of mistrust towards the government, a high level of economic inequality, and largely ineffective health service provision. Taken together, these factors point to an environment of weak state capacity and a low level of trust in public institutions. In fact, studies show that there are remarkable correlations among those variables. For example, while a study finds a strong relationship between high levels of economic inequality and

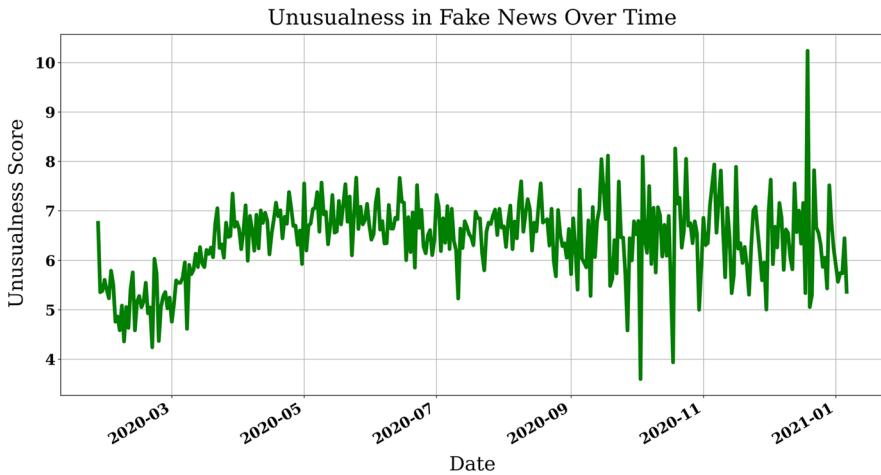


Fig. 14 Aggregated average unusualness scores over time

low levels of trust in national institutions across the EU member countries [121], another one conducted in post-Soviet nations shows that there is a negative association between perception of corruption and trust in public institutions such as police, national and regional governments, and courts [122]. Given that the state capacity and trust in public institutions are integral to an effective pandemic strategy—affecting people’s compliance with restrictions and willingness to get vaccinated, governments’ success in enforcing lockdowns and other isolation practices, etc.—it is not surprising that those countries produced more similar news. Our dataset includes several fascinating falsehoods particularly common to the countries in this cluster. Reflecting the mistrust towards the government and its capacity to supervise the pandemic efforts, a news story in Zimbabwe alleged that a medical laboratory conducted clinical trials for a possible vaccine and led to the death of 68 out of 80 volunteers in total. Similarly, mistrust towards the government and a high level of political polarization undoubtedly fostered misinforming news such as the one in Bolivia that was published in July 2020 and claimed that President Maduro was extending the full lockdown until January 2021.

Misinformation unusualness/creativity

In the last part of the paper, groups of countries have been compared against each other in terms of the creative word usage in the news they published. The average

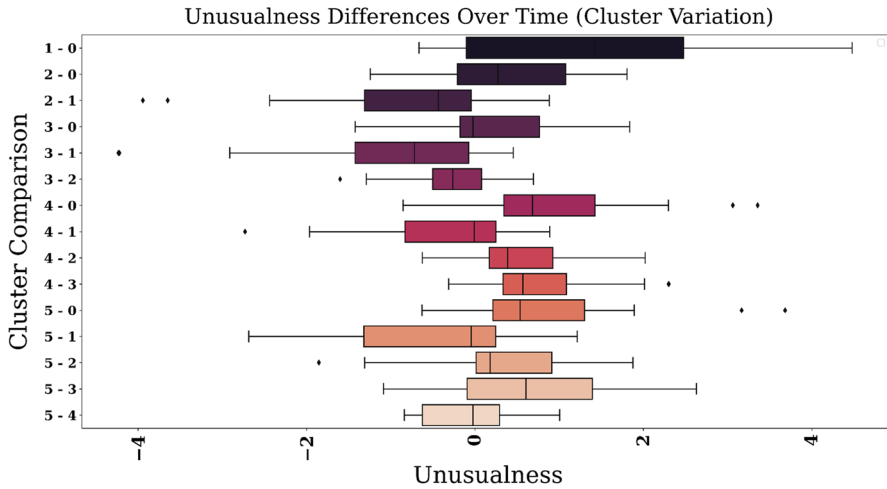


Fig. 15 Comparison of unusualness scores between clusters of countries

value of unusualness of one group was compared with the average unusualness extracted from the other group. Since the difference is taken into consideration, statistically significant results should be much higher than zero. We provided the average aggregated unusualness score in Fig. 14. The figure shows that an initial lack of creativity in the first few months of the pandemic was followed by an increase and relative stability throughout 2020.

Breaking down the countries into groups and comparing the levels of creativity between them did not provide any results that are statistically significantly different from zero. Thus, our expectation that countries of different backgrounds would choose word-groups according to their taste did not come true. A comparison between clusters of countries created using the social, cultural, economic, and political variables in the dataset has been provided below in Fig. 14. We believe that this lack of meaningful difference across clusters of countries in terms of creative word usage can be explained by three main factors. First, it points to a highly globalized media environment in the sense that media outlets across nations share vocabulary and discourses to a great extent. The digital media, and the Internet more broadly, have created “a new global language” [123] with specific neologisms and novel syntactic, orthographic, and lexical commonalities among world languages, such as heavy use of emojis and emoticons, abbreviations, and acronyms [124]. Second, research shows significant differences between truthful news and falsehoods regarding their linguistic characteristics as the latter use more words related to anxiety, more superlatives, sensationalistic writing, and overly emotional language [125, 126]. Third, as previously mentioned, a great percentage of falsehoods in our dataset are from Facebook and a few other social media outlets. Given the studies showing that misinformation spreads really fast on social media platforms [24, 127] and that most COVID-19-related falsehoods were produced by a very small number of

individuals [128], the lack of statistically significant difference among the groups of countries is not too surprising (Fig. 15).

Discussion and conclusion

The “Varieties of... “literature has influenced more than a generation of scholars and practitioners worldwide. There have been politicians to use the arguments first offered by Hall and Soskice to transform the institutional structure of their countries (such as the British Labour Party politician, Ed Miliband, when he was Leader of the Opposition) and many other scholars who expanded the typologies first proposed by Hall and Soskice. Academically, we hope that our paper will provide a strong comparative perspective to an emerging literature. We also agree that the “Varieties of...” conceptualization is deterministic in nature; however, as recent media-viewer experience suggests, local and global media, policymakers, and transnational institutions have also been looking at pandemic-related policy success and failure from a cross-national, and mostly deterministic perspective. Thus, many are wondering why some countries have been more successful than others in mitigating the human costs of the pandemic, while others have been less so in an environment where local political leaders are looking for the best non-local practices. From a practical sense, we believe that the arguments and facts laid out here may contribute to the public health efforts to fight against misinformation, which continues to take lives in a myriad of ways, such as by discouraging people from getting vaccinated or promoting fraudulent and dangerous products.

To conclude, we want to reiterate four of the key contributions that this paper provides to the literature, and particularly to the tools to be used by global public health circles. First, our study is truly unique in terms of its data and methodology—it comprises over 10 thousand falsehoods from 129 countries; its data come from a variety of sources, including the most widely used social media platform globally, i.e., Facebook; and it uses 14 different variables aiming to measure political, economic, informational, and socio-cultural environments in each country in order to compare COVID-19 misinformation across them. We believe that the resulting clustering of countries into groups offers avenues for developing distinct public health and communication strategies to dispel misinformation in countries with particular characteristics.

Second, and relatedly, the findings give clues on what those strategies should be. For instance, our analysis suggests that countries with low HDI (mainly located in sub-Saharan Africa) produce misinformation related to unproven local remedies and those stemming from certain religious beliefs as well as from distrust of international organizations and Western medical practices. This shows the importance of working with local religious leaders and healers and repairing Western governments and pharmaceutical companies’ tarnished reputation. Likewise, given that countries with severe press freedom issues (e.g., those implementing outright censorship of news and bans on social media platforms) generate similar news, global public health circles should design an anti-misinformation strategy specific to those

nations, which necessitates going beyond using online platforms that are at risk of being censored.

Third, our study indicates that there have been successful anti-misinformation efforts throughout the pandemic, but significant challenges persist. More specifically, we found that the types of falsehoods that were particularly common in the first few months of the pandemic and were widely shared across countries (mainly those about the origin, spread, cure, and prevention of the disease) got effectively addressed by announcements coming from leading figures and organizations such as WHO or Anthony Fauci, resulting in a decline in the number of falsehoods related to those topics. However, the findings also reveal two worrying trends, among others: (1) conspiracy theories are common among all groups of countries, which can be explained by the fact that they are originated by a small number of individuals and organizations (aka misinformation “superspreaders”), but are effectively disseminated across the globe; (2) in countries with weak state capacity and a low level of trust in public institutions, misinformation creates a particularly dangerous vicious circle—distrust of government fosters the production of falsehoods, which in turn further weakens governments’ ability to supervise the pandemic efforts. Accordingly, we argue that international organizations and leading figures in global health should strengthen their efforts to reach out to those populations and develop effective strategies against the dissemination of the major conspiracies.

Fourth, we found that while the most prominent misinformation topics vary across groups of countries, the word-groups used in misinforming news stories are remarkably similar. In line with the “glocalization” literature [129, 130], we interpret this as the coexistence of globalizing and localizing processes—on the one hand, socio-economic, cultural, political, and informational characteristics of countries clearly affect the types of falsehoods Internet users are exposed to; but, on the other hand, the tone and structure of falsehoods do not show much variance, which points to a highly globalized online media environment. Given those findings, we argue that even though implementing country-specific strategies (e.g., improving scientific literacy in a country where it is currently weak) is crucial, a global action plan against misinformation is also very much needed.

Appendix

Part I: Social, economic, and political variables

See Table 4.

Table 4 Country information data

Name of the index	Organization/individual that compiled the index	Year of data collection	Number of countries represented	Indicators	Assigned categories
Freedom in the World	Freedom House [69, 70]	2019	129	Electoral process; political pluralism and participation; functioning of government; freedom of expression and belief; associational and organizational rights; rule of law; personal autonomy and individual rights	Free; partly free; not free
Edelman Trust Barometer	Edelman Data and Intelligence [71]	2018	26	Public trust in government through an online survey	Scale (lowest = 21 highest 86)
Corruption Perceptions Index	Transparency International [72]	2019–2020	127	How corrupt the public sector is perceived to be (based on expert opinion and surveys of business people)	Scale (lowest = 12 highest = 88)
World Press Freedom Index	Reporters Without Borders [73]	2019	127	Online questionnaire filled by lawyers, sociologists, and media professionals evaluating: pluralism, media independence, media environment and self-censorship, legislative framework, abuses, transparency, and the quality of the infrastructure that supports the production of news and information	Good situation, satisfactory situation, problematic situation, difficult situation, very serious situation

Table 4 (continued)

Name of the index	Organization/individual that compiled the index	Year of data collection	Number of countries represented	Indicators	Assigned categories
Reuters Institute Digital News Report	Reuters Institute for Study of Journalism [74]	2019	37	Online questionnaire filled by a representative sample of people evaluating overall trust in the news	% of those who trust in the news (lowest = 22, highest = 59)
The Wellcome Global Monitor	Wellcome Trust [75]	2018	115	Questionnaire filled by a representative sample of people face-to-face or via telephone evaluating people's trust in scientists	% of people who have a high level of trust in scientists (lowest = 2, highest = 54)
Index of Populist Rhetoric	Luigi Curini (Università degli Studi di Milano) (2019) [76]	2012–2018	37	The populist discourse of the most recent presidents and prime ministers who came to power since 2012 (using The Global Populism Database by the Guardian and Team Populism)	Scale (1 = not populist, 8 = very populist)
World Cultural Map	World Values Survey [77]	2017–2020	92	Face-to-face interview with a representative sample of people	Confucian, Orthodox Europe, African-Islamic, Latin America, West and South Asia, Catholic Europe, English-speaking, Protestant Europe

Table 4 (continued)

Name of the index	Organization/individual that compiled the index	Year of data collection	Number of countries represented	Indicators	Assigned categories
Clusters of Business Systems	Michael A. Witt et al. (2018) [78]	2013	50	Education, employment relations, finance, interfirm relations, internal dynamics, ownership and governance, social capital, state	Socialist economies, emergent economies, Arab oil-based economies, advanced city economies, advanced emerging economies, European peripheral economies, liberal market economies, coordinated market economies, highly coordinated economies
Human Development Index	United Nations Development Program [79]	2018–2019	126	Health, education, income	Very high, high, medium, low
GINI Index	World Bank [80]	2019	117	Distribution of income across a population	GINI coefficient: 1 = perfect inequality, 0 = perfect equality (lowest = 0.25, highest = 0.63)
Effective coverage of health services index	Global Burden of Diseases 2019 Universal Health Coverage Collaborators (2020) [81]	2019	127	Promotion, prevention, treatment	Scale (lowest = 32 highest = 96)
Colonization history	N/A	N/A	129	History of being colonized by a Western power (excluding settler colonies such as Australia)	Yes, no

Table 4 (continued)

Name of the index	Organization/individual that compiled the index	Year of data collection	Number of countries represented	Indicators	Assigned categories
United Nations Geoscheme	United Nations Statistics Division [82]	N/A	129	Geographic subregions	Northern America, Central America, Caribbean, South America, Northern Europe, Western Europe, Eastern Europe, Southern Europe, Western Asia, Central Asia, Eastern Asia, Southern Asia, Southeastern Asia, Melanesia, Polynesia, Australia and New Zealand, Northern Africa, Western Africa, Central Africa, Southern Africa, Eastern Africa

Part II: An explanation on network similarity formulas

A closer elaboration of the formulas has been taken from a chapter in *Financial Data Analytics: Theory and Applications* [131]:

“Frobenius Distance [103] computes the similarity between two graphs by “locally” comparing the individual connections between pairs of nodes. This is considered a known node correspondence (KNC) method (an algorithm that needs information about which nodes should be compared to each other). If a_{ij} and b_{ij} represent the connections between two nodes i and j that belong to two different graphs $G1$ and $G2$ such that a_{ij} is in $G1$ and b_{ij} is in $G2$, Frobenius distance $d(G1, G2)$ is the following:

$$d(G1, G2) = \sqrt{\sum_{ij} ||a_{ij} - b_{ij}||}.$$

Quantum-JSD distance [103] compares the spectral entropies of the density matrices. This is done by calculating the ‘Quantum’ Jensen–Shannon divergence between two graphs. The authors create a connection-based density matrix to calculate the von Neumann entropy of a network. The algorithm proposed uses the whole network, instead of a subset of network features. Most importantly, the algorithm allows the authors to quantify the distance between ‘complex’ networks. Classical algorithms attempt to quantify the amount of information about a probability distribution (entropy), and quantum JSD expands this definition by introducing divergences (also known as quantum relative entropy). The distance is calculated by using a generalized Jensen–Shannon divergence between two graphs:

$$J_q(\rho||\sigma) = S_q\left(\frac{\rho + \sigma}{2}\right) - \frac{1}{2}[S_q(\rho) + S_q(\sigma)]$$

In the equation above, ρ and σ represent the density matrices and q represents the order parameter. The density matrix looks like the following:

$$\rho = \frac{e^{-\beta L}}{Z},$$

where

$$Z = \sum_{i=1}^N e^{-\beta \lambda_i(L)}$$

and $\lambda_i(L)$ represents an imaginary diffusion process i , over the network with time parameter $\beta > 0$.

Part III: Content similarity comparison graphs

A list of the visuals that demonstrate insignificant comparison results have been listed below (Figs. 16, 17, 18, 19, 20, 21, 22, 23, 24).

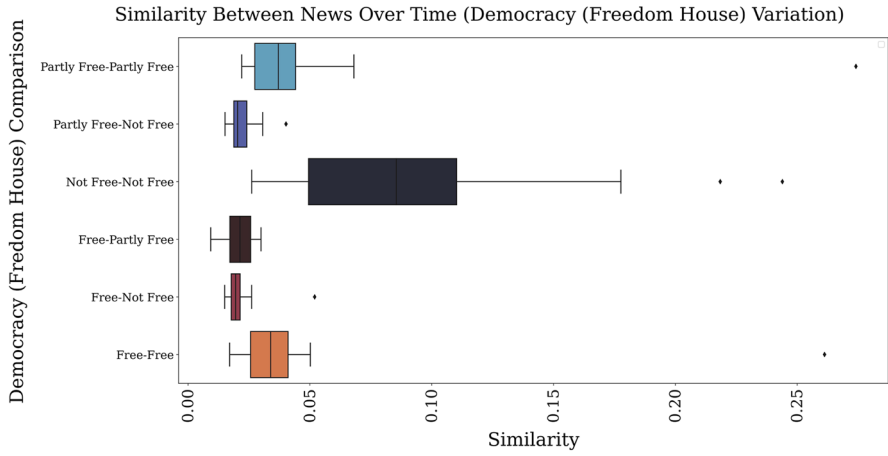


Fig. 16 Similarity between news over time, Freedom House Index comparison

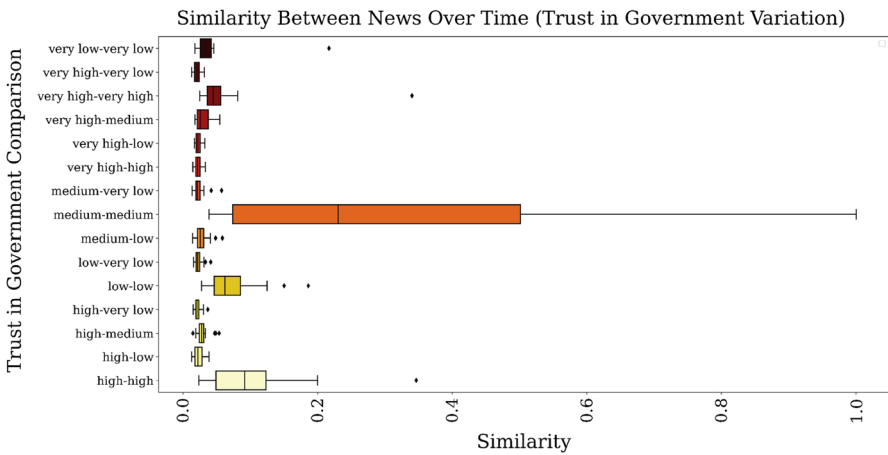


Fig. 17 Similarity between news over time, trust in government comparison

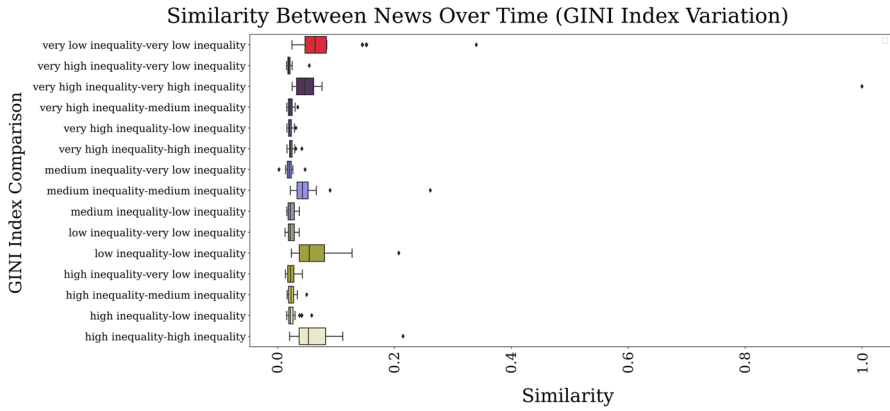


Fig. 18 Similarity between news over time, GINI Index comparison

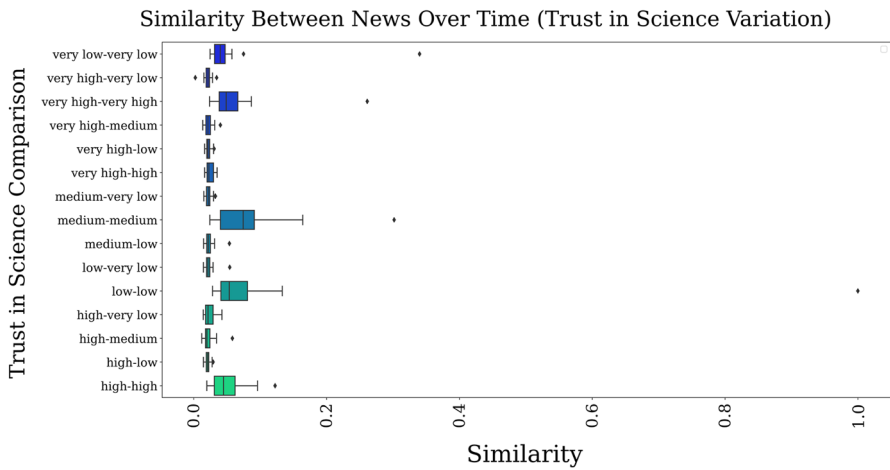


Fig. 19 Similarity between news over time, trust in science comparison

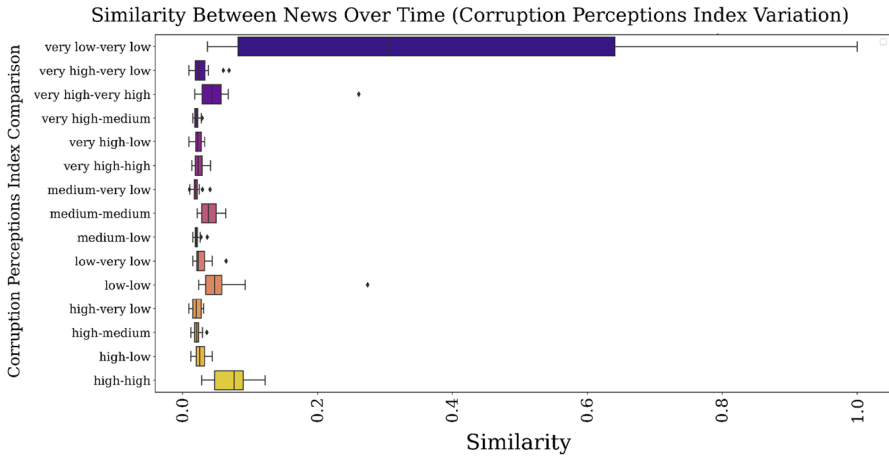


Fig. 20 Similarity between news over time, Corruption Perceptions Index comparison

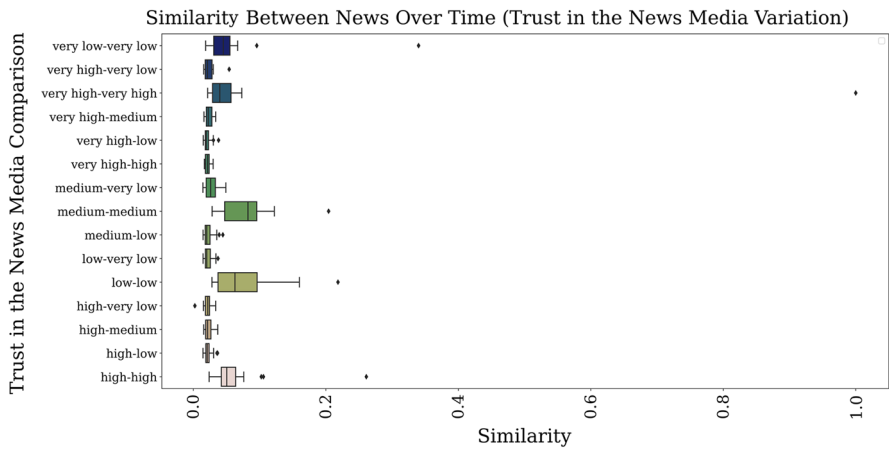


Fig. 21 Similarity between news over time, trust in news media comparison

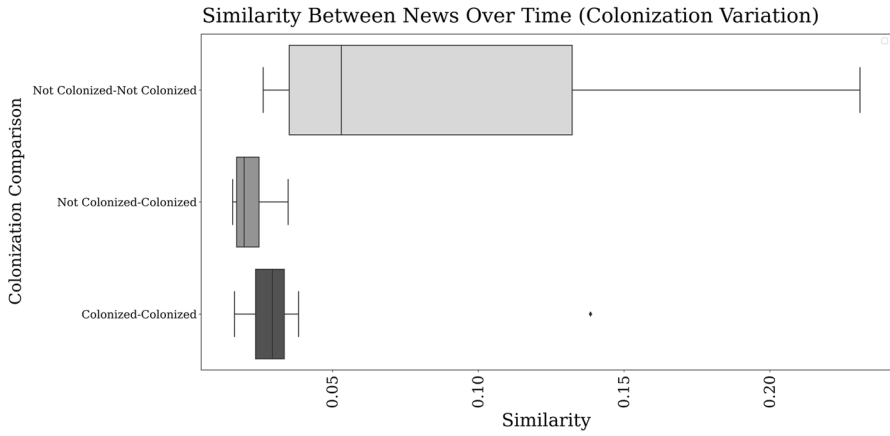


Fig. 22 Similarity between news over time, colonization comparison

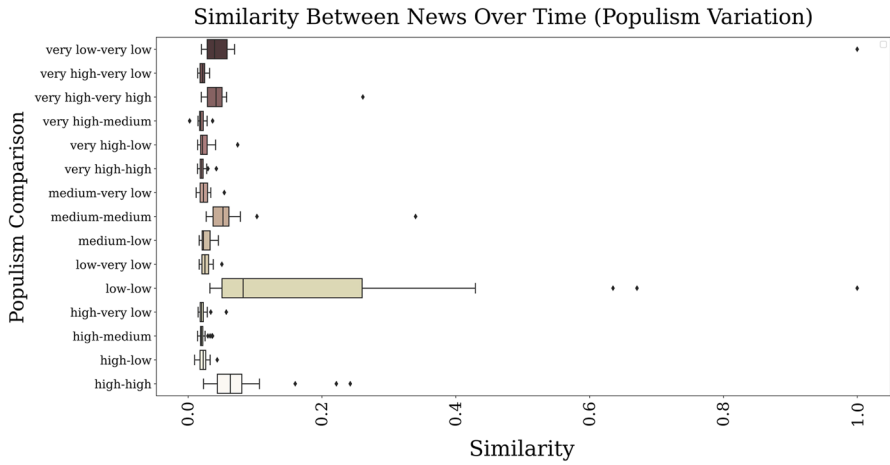


Fig. 23 Similarity between news over time, populism comparison

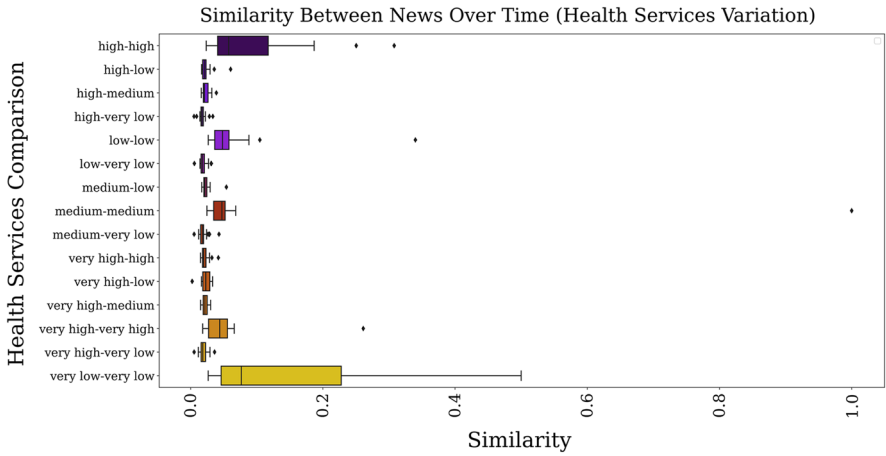


Fig. 24 Similarity between news over time, Health Services Index comparison

Part IV: Unusualness differences over time

A list of the visuals that demonstrate the comparison results that are insignificantly different from each other have been listed below (Figs. 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36)

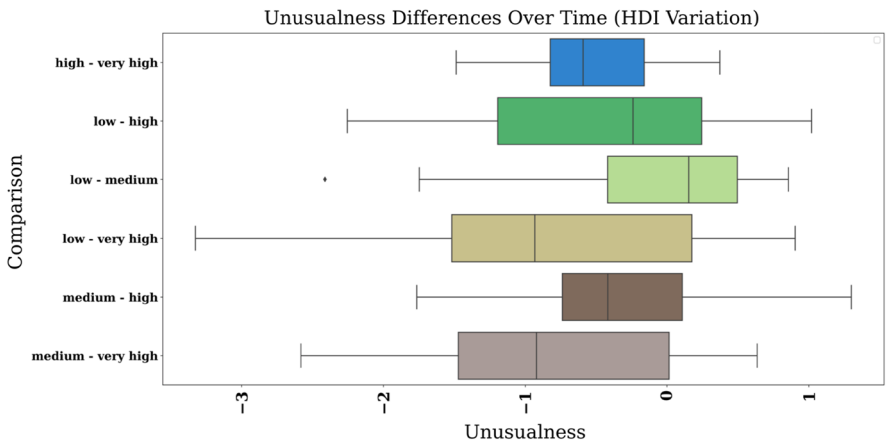


Fig. 25 Unusualness differences between varieties of countries over time, HDI comparison

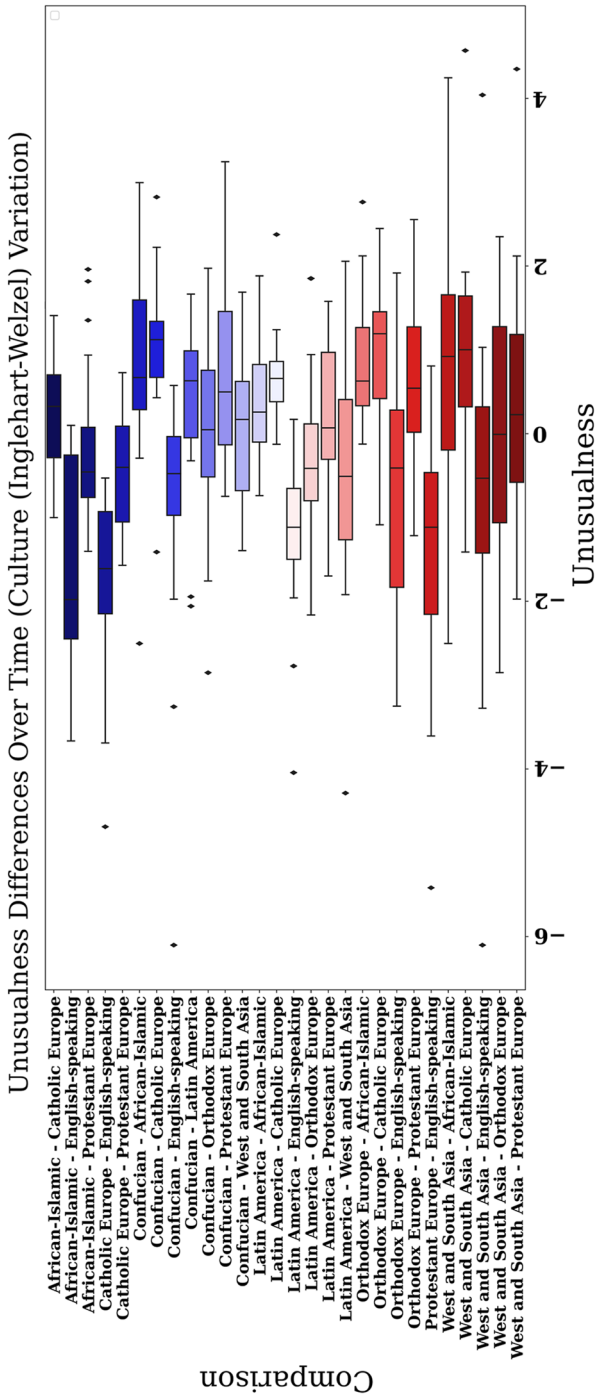


Fig. 26 Comparison of unusualness scores between clusters of cultures

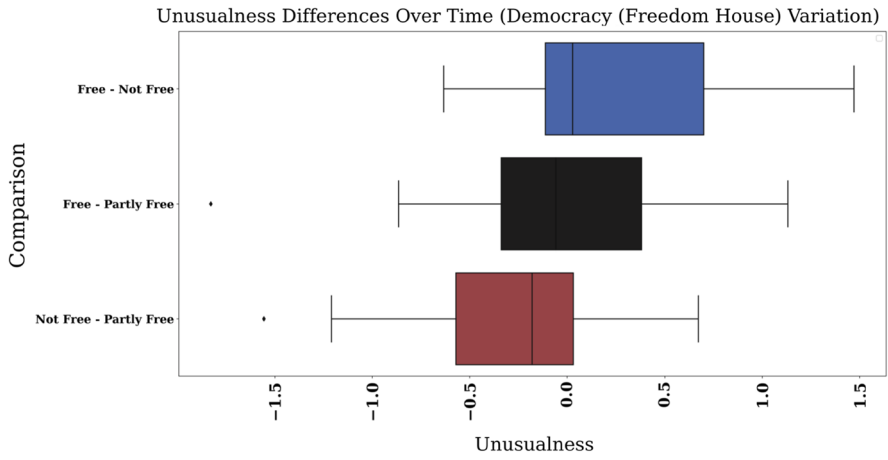


Fig. 27 Unusualness differences between varieties of countries over time, Freedom House Index comparison

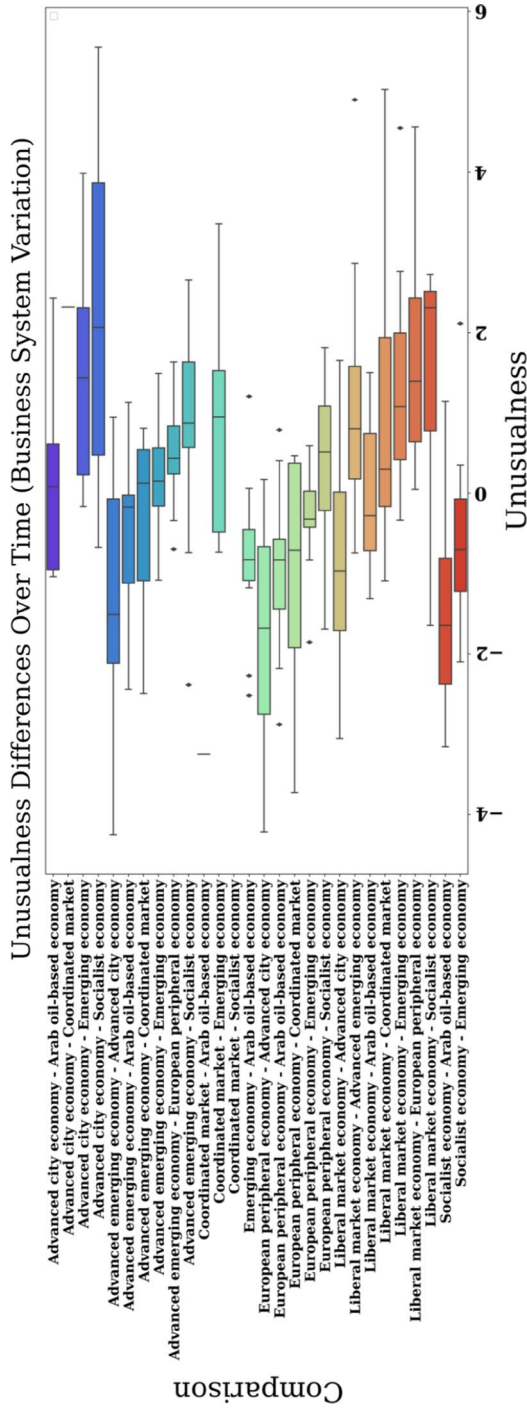


Fig. 28 Unusualness differences between varieties of countries over time, business systems comparison

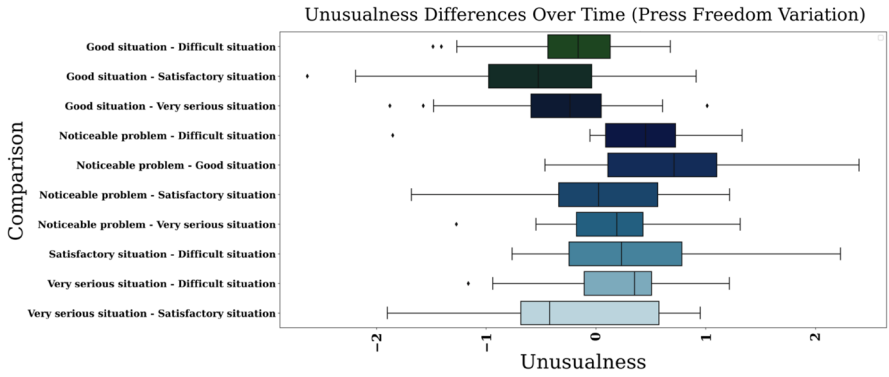


Fig. 29 Unusualness differences between varieties of countries over time, press freedom comparison

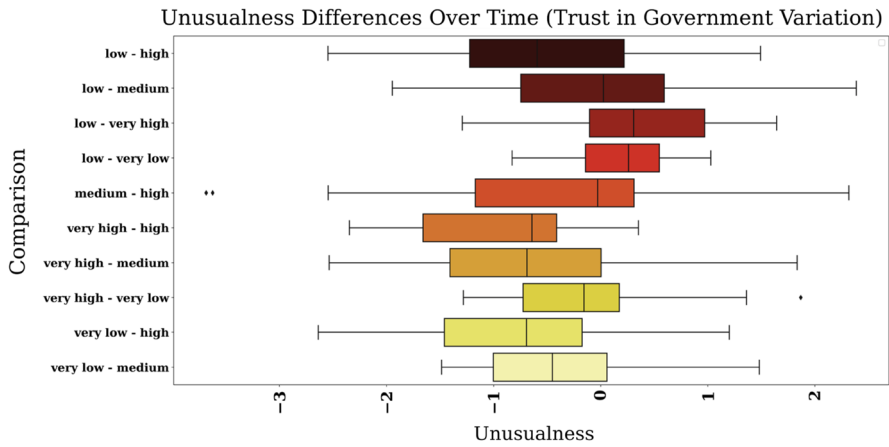


Fig. 30 Unusualness differences between varieties of countries over time, trust in government comparison

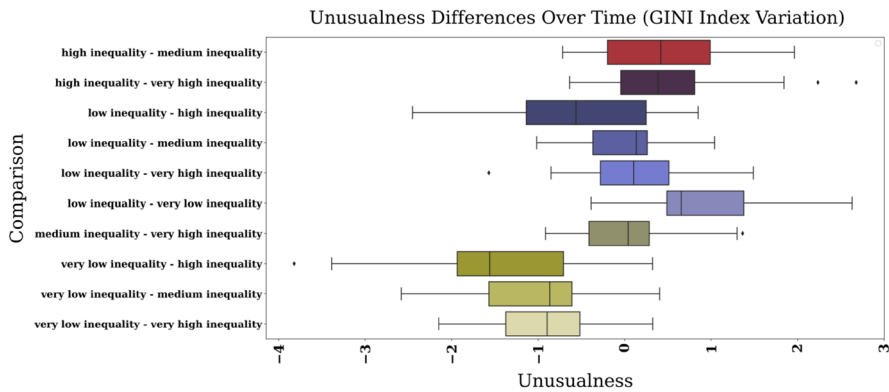


Fig. 31 Unusualness differences between varieties of countries over time, GINI Index comparison

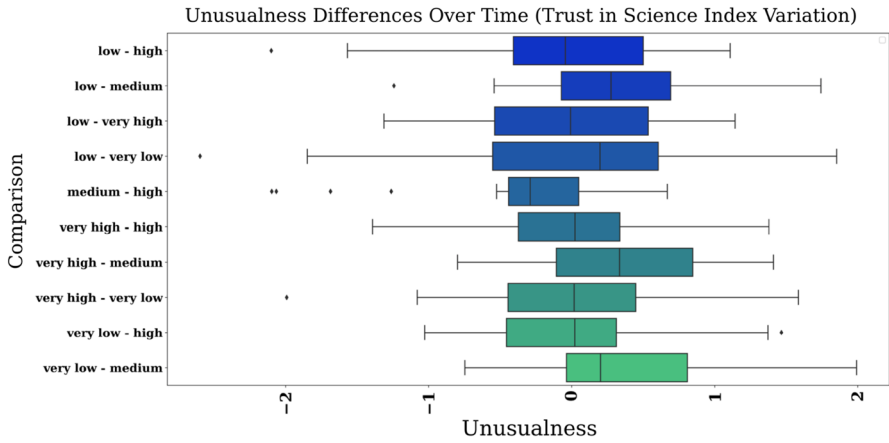


Fig. 32 Unusualness differences between varieties of countries over time, trust in science comparison

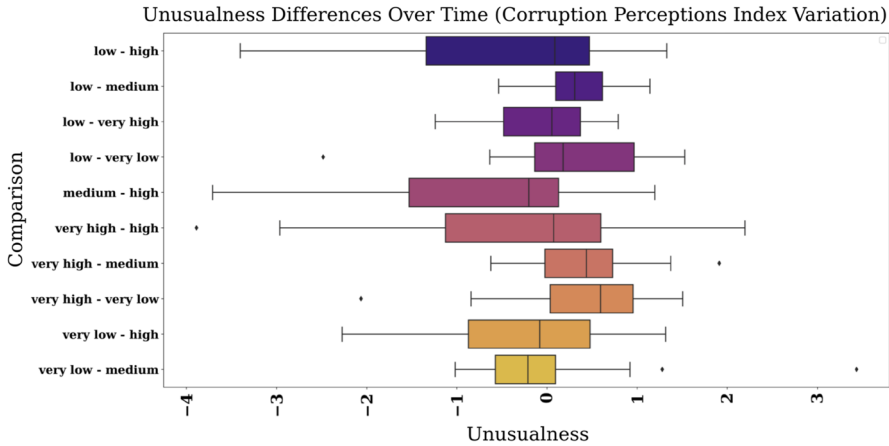


Fig. 33 Unusualness differences between varieties of countries over time, Corruption Perceptions Index comparison

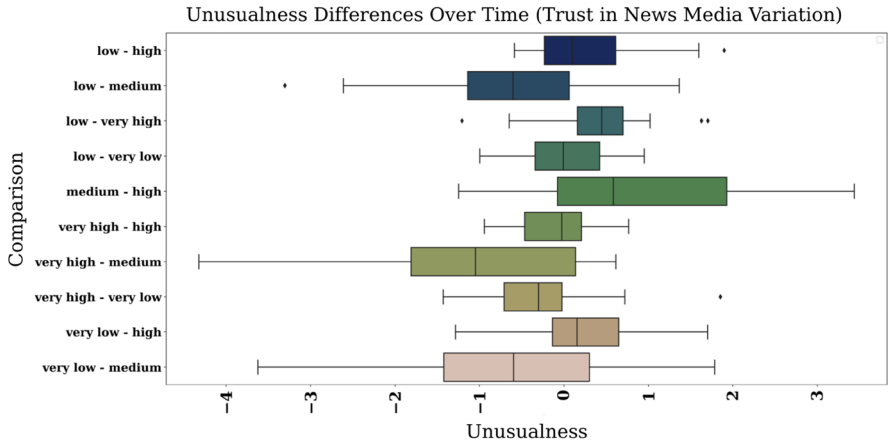


Fig. 34 Unusualness differences between varieties of countries over time, trust in news media comparison

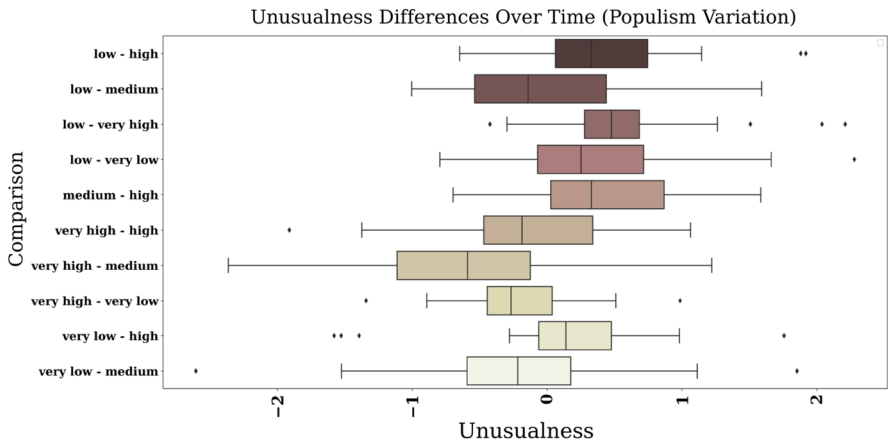


Fig. 35 Unusualness differences between varieties of countries over time, populism comparison

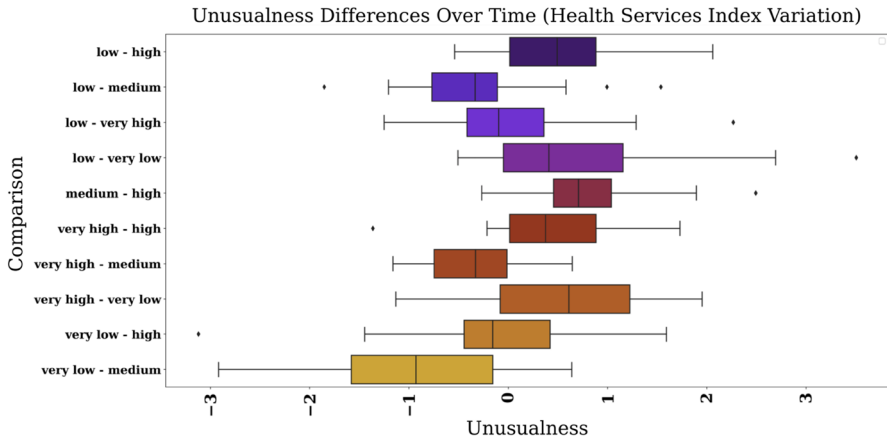


Fig. 36 Unusualness differences between varieties of countries over time, Health Services Index comparison

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s42001-022-00193-5>.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Data availability statement The authors confirm that all data generated or analyzed during this study will be made available as a supplement with the publication of this article.

References

1. Poynter Institute. IFCN Covid-19 Misinformation [Internet]. Poynter. [cited 2022 Feb 25]. <https://www.poynter.org/ifcn-covid-19-misinformation/>. Accessed 13 Nov 2022.
2. Hall, P. A., & Soskice, D. (2001). *Varieties of capitalism: The institutional foundations of comparative advantage*. Oxford University Press.
3. Benz, A., & Broschek, J. (2013). *Federal dynamics: Continuity, change, and the varieties of federalism*. OUP Oxford.
4. Bochsler, D., & Kriesi, H., et al. (2013). Varieties of democracy. In H. Kriesi (Ed.), *Democracy in the age of globalization and mediatization*. Palgrave Macmillan.
5. Devinney, T. M., & Hartwell, C. A. (2020). Varieties of populism. *Global Strategy Journal*, 10(1), 32–66.
6. Dorman, S. R. (2015). The varieties of nationalism in Africa. *Current History*, 114(772), 189–193.
7. Rothstein, H., Demeritt, D., Paul, R., et al. (2019). Varieties of risk regulation in Europe: Coordination, complementarity and occupational safety in capitalist welfare states. *Socio-Economic Review*, 17(4), 993–1020.
8. Saraceno, C. (2016). Varieties of familialism: Comparing four southern European and East Asian welfare regimes. *Journal of European Social Policy*, 26(4), 314–326.
9. Zubiaga, A., Liakata, M., Procter, R., Hoi, G. W. S., & Tolmie, P. (2016). Analysing how people orient to and spread rumours in social media by looking at conversational threads. *PLoS ONE*, 11(3), e0150989.

10. Kumar, S., West, R., Leskovec, J. (2016). Disinformation on the web: Impact, characteristics, and detection of Wikipedia hoaxes. In: Proceedings of the 25th international conference on World Wide Web [Internet]. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee. [cited 2022 Apr 23]. p. 591–602. (WWW '16). <https://doi.org/10.1145/2872427.2883085>.
11. Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives.*, 31(2), 211–236.
12. Mejias, U. A., & Vokuev, N. E. (2017). Disinformation and the media: The case of Russia and Ukraine. *Media, Culture and Society.*, 39(7), 1027–1042.
13. Merriam-Webster. The dictionary [Internet]. [cited 2022 Apr 23]. <https://www.merriam-webster.com/>. Accessed 13 Nov 2022.
14. Shahi, G. K., Dirkson, A., & Majchrzak, T. A. (2021). An exploratory study of COVID-19 misinformation on Twitter. *Online Social Networks and Media.*, 1(22), 100104.
15. Wardle, C., Derakhshan, H. (2017). Information disorder: Toward an interdisciplinary framework for research and policy making. Council of Europe Report. [cited 2022 Apr 23]. <https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-research/168076277c>. Accessed 13 Nov 2022.
16. Guess, A. M., Lyons, B. A. (2020). Misinformation, Disinformation, and Online Propaganda. In *Social Media and Democracy: The State of the Field, Prospects for Reform*, edited by Joshua A. Tucker and Nathaniel Persily, 10–33. SSRC Anxieties of Democracy. Cambridge: Cambridge University Press. <https://www.cambridge.org/core/books/social-media-anddemocracy/misinformation-disinformation-and-online-propaganda/D14406A631AA181839ED896916598500>.
17. World Health Organization. (2020). Coronavirus disease (COVID-19): situation report, 162 [Internet]. World Health Organization. [cited 2022 Apr 23]. <https://apps.who.int/iris/handle/10665/332970>. Accessed 13 Nov 2022.
18. Lewis, T. (2020). Eight persistent COVID-19 myths and why people believe them [Internet]. Scientific American. [cited 2022 Apr 23]. <https://www.scientificamerican.com/article/eight-persistent-covid-19-myths-and-why-people-believe-them/>. Accessed 13 Nov 2022.
19. Lynas, M. (2020). COVID: Top 10 current conspiracy theories [Internet]. Alliance for Science. [cited 2022 Apr 23]. <https://allianceforscience.cornell.edu/blog/2020/04/covid-top-10-current-conspiracy-theories/>. Accessed 13 Nov 2022.
20. Spring, M., Wendling, M. (2020). How Covid-19 myths are merging with the QAnon conspiracy theory. BBC News [Internet]. [cited 2022 Apr 23]. <https://www.bbc.com/news/blogs-trending-53997203>. Accessed 13 Nov 2022.
21. Islam, M. S., Sarkar, T., Khan, S. H., Kamal, A.-H.M., Hasan, S. M., Kabir, A., et al. (2020). COVID-19-related infodemic and its impact on public health: A global social media analysis. *The American Journal of Tropical Medicine and Hygiene.*, 103(4), 1621.
22. Murphy, J. (2021). Biden has new vaccination goals. See if the country is on pace to hit them. [Internet]. NBC News. [cited 2022 Apr 23]. <https://www.nbcnews.com/politics/white-house/graphic-track-biden-fourth-july-vaccination-goals-n1268803>. Accessed 13 Nov 2022.
23. Centers for Disease Control and Prevention. (2021). Delta variant: What we know about the science. 2021. <https://www.cdc.gov/coronavirus/2019-ncov/variants/delta-variant.html>. Accessed 13 Nov 2022.
24. Obiała, J., Obiała, K., Mańczak, M., Owoc, J., & Olszewski, R. (2020). COVID-19 misinformation: Accuracy of articles about coronavirus prevention mostly shared on social media. *Health Policy and Technology.*, 10, 182–186.
25. Singh, L., Bansal, S., Bode, L., Budak, C., Chi, G., Kawintiranon, K., et al. (2020). A first look at COVID-19 information and misinformation sharing on Twitter. arXiv:2003.13907v1.
26. Chen, E., Chang, H., Rao, A., Lerman, K., Cowan, G., & Ferrara, E. (2021). *COVID-19 misinformation and the 2020 US presidential election*. The Harvard Kennedy School Misinformation Review.
27. Memon, S. A., Carley, K. M. (2020). Characterizing COVID-19 misinformation communities using a novel Twitter dataset. arXiv:200800791 [cs] [Internet]. [cited 2022 Apr 23]. <http://arxiv.org/abs/2008.00791>. Accessed 13 Nov 2022.
28. Bridgman, A., Merkley, E., Loewen, P. J., Owen, T., Ruths, D., Teichmann, L., et al. (2020). The causes and consequences of COVID-19 misperceptions: Understanding the role of news and social media. *Harvard Kennedy School Misinformation Review*, 1(3).

29. Brennen, J. S., Simon, F. M., Howard, P. N., Nielsen, R. K. (2020). Types, sources, and claims of COVID-19 misinformation. Ph.D. thesis. University of Oxford.
30. Enders, A. M., Uscinski, J. E., Klofstad, C., Stoler, J. (2020). The different forms of COVID-19 misinformation and their consequences. *The Harvard Kennedy School Misinformation Review*.
31. Stecula, D. A., & Pickup, M. (2021). How populism and conservative media fuel conspiracy beliefs about COVID-19 and what it means for COVID-19 behaviors. *Research and Politics*, 8(1), 2053168021993979.
32. Uscinski, J. E., Enders, A. M., Klofstad, C., Seelig, M., Funchion, J., Everett, C., et al. (2020). Why do people believe COVID-19 conspiracy theories? *Harvard Kennedy School Misinformation Review* [Internet]. [cited 2022 Apr 23];1(3).
33. Lobato, E. J. C., Powell, M., Padilla, L. M. K., Holbrook, C. (2020). Factors predicting willingness to share COVID-19 misinformation. *Front. Psychol.* [Internet]. [cited 2022 Apr 23];11. <https://doi.org/10.3389/fpsyg.2020.566108>. Accessed 13 Nov 2022.
34. Motta, M., Stecula, D., & Farhart, C. (2020). How right-leaning media coverage of COVID-19 facilitated the spread of misinformation in the early stages of the pandemic in the US. *Canadian Journal of Political Science*, 53(2), 335–342.
35. Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G., & Rand, D. G. (2020). Fighting COVID-19 misinformation on social media: experimental evidence for a scalable accuracy-nudge intervention. *Psychological Science*, 31(7), 770–780.
36. Sallam, M., Dababseh, D., Yaseen, A., Al-Haidar, A., Taim, D., Eid, H., et al. (2020). COVID-19 misinformation: Mere harmless delusions or much more? A knowledge and attitude cross-sectional study among the general public residing in Jordan. *PLoS ONE*, 15(12), e0243264.
37. Agle, J., & Xiao, Y. (2021). Misinformation about COVID-19: Evidence for differential latent profiles and a strong association with trust in science. *BMC Public Health*, 21(1), 1–12.
38. Laato, S., Islam, A. K. M. N., Islam, M. N., & Whelan, E. (2020). Why do people share misinformation during the COVID-19 pandemic? *European Journal of Information Systems*, 29(3), 288–305.
39. Roozenbeek, J., Schneider, C. R., Dryhurst, S., Kerr, J., Freeman, A. L. J., Recchia, G., et al. (2020). Susceptibility to misinformation about COVID-19 around the world. *Royal Society Open Science*, 7(10), 201199.
40. Hornik, R., Kikut, A., Jesch, E., Woko, C., Siegel, L., & Kim, K. (2021). Association of COVID-19 misinformation with face mask wearing and social distancing in a nationally representative US sample. *Health Communication*, 36(1), 6–14.
41. Imhoff, R., & Lamberty, P. (2020). A bioweapon or a hoax? The link between distinct conspiracy beliefs about the Coronavirus disease (COVID-19) outbreak and pandemic behavior. *Social Psychological and Personality Science*, 11(8), 1110–1118.
42. Loomba, S., de Figueiredo, A., Piatek, S. J., de Graaf, K., & Larson, H. J. (2021). Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. *Nature Human Behaviour*, 5(3), 337–348.
43. Romer, D., & Jamieson, K. H. (2020). Conspiracy theories as barriers to controlling the spread of COVID-19 in the US. *Social Science and Medicine*, 263, 113356.
44. Teovanović, P., Lukić, P., Zupan, Z., Lazić, A., Ninković, M., Žeželj, I. (2021). Irrational beliefs differentially predict adherence to guidelines and pseudoscientific practices during the COVID-19 pandemic. *Applied Cognitive Psychology*, 35(2), 486–496. <https://doi.org/10.1002/acp.3770>.
45. Bertin, P., Nera, K., Delouève, S. (2020). Conspiracy beliefs, rejection of vaccination, and support for hydroxychloroquine: A conceptual replication-extension in the COVID-19 pandemic context. *Frontiers in Psychology*, 11, 565128. <https://doi.org/10.3389/fpsyg.2020.565128>.
46. Chen, Y., Biswas, M.I. (2022). Impact of national culture on the severity of the COVID-19 pandemic. *Current Psychology*. <https://doi.org/10.1007/s12144-022-02906-5>.
47. Shirish, A., Srivastava, S. C., & Chandra, S. (2021). Impact of mobile connectivity and freedom on fake news propensity during the COVID-19 pandemic: A cross-country empirical examination. *European Journal of Information Systems*, 30(3), 322–341.
48. Hammes, L. S., Rossi, A. P., Pedrotti, L. G., Pitrez, P. M., Mutlaq, M. P., & Rosa, R. G. (2021). Is the press properly presenting the epidemiological data on COVID-19? An analysis of newspapers from 25 countries. *Journal of Public Health Policy*, 42(3), 359–372.
49. Al-Zaman, M. S. (2021). Prevalence and source analysis of COVID-19 misinformation of 138 countries [Internet]. medRxiv. [cited 2022 Feb 25]. p. 2021.05.08.21256879. <https://doi.org/10.1101/2021.05.08.21256879v1>.

50. Cha, M., Cha, C., Singh, K., et al. (2016). Prevalence of misinformation and factchecks on the COVID-19 pandemic in 35 countries: observational infodemiology study. *JMIR Human Factors*, 8(1), e23279.
51. De Coninck, D., Frissen, T., Matthijs, K., d'Haenens, L., Lits, G., Champagne-Poirier, O., et al. (2021). Beliefs in conspiracy theories and misinformation about COVID-19: Comparative perspectives on the role of anxiety, depression and exposure to and trust in information sources. *Frontiers in Psychology [Internet]*. [cited 2022 Feb 26];12. <https://doi.org/10.3389/fpsyg.2021.646394>.
52. Sharma, K., Seo, S., Meng, C., Rambhatla, S., Liu, Y. (2020). COVID-19 on social media: Analyzing misinformation in Twitter conversations. arXiv:200312309 [cs] [Internet]. [cited 2022 Apr 23]. <http://arxiv.org/abs/2003.12309>.
53. Xaudiera, S., Cardenal, A. S. (2020). Ibuprofen narratives in five European countries during the COVID-19 pandemic. *Harvard Kennedy School Misinformation Review [Internet]*. [cited 2022 Apr 23];1(3).
54. Statista. Most popular social networks worldwide as of October 2021, ranked by number of active users. [Internet] [cited 2022 Apr 23]. <https://lb-aps-frontend.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>.
55. Apuke, O. D., & Omar, B. (2020). Fake news and COVID-19: Modelling the predictors of fake news sharing among social media users. *Telematics and Informatics*, 56, 101475.
56. Shehata, A., & Eldakar, M. (2021). An exploration of Egyptian Facebook users' perceptions and behavior of COVID-19 misinformation. *Science and Technology Libraries*, 40(4), 390–415.
57. Champion, V. L., Skinner, C. S. (2008). The health belief model. In *Health behavior and health education: Theory, research, and practice*, Glanz, K., Rimer, B.K., & Viswanath, K. Eds., (4th ed., pp. 45–65). San Francisco: Jossey-Bass.
58. Pranesh, R. R., Farokhnejad, M., Shekhar, A., Vargas-Solar, G. (2021). Looking for COVID-19 misinformation in multilingual social media texts arXiv:2105.03313 [Internet]. [cited 2022 Apr 23]; <https://arxiv.org/abs/2105.03313>.
59. Caliskan, C. (2021). How does “A Bit of Everything American” state feel about COVID-19? A quantitative Twitter analysis of the pandemic in Ohio. *Journal of Computational Social Science [Internet]*. [cited 2022 Apr 23]; <https://doi.org/10.1007/s42001-021-00111-1>.
60. Himelboim, I., McCreery, S., & Smith, M. (2013). Birds of a feather tweet together: Integrating network and content analyses to examine cross-ideology exposure on Twitter. *Journal of Computer-Mediated Communication*, 18(2), 154–174.
61. Kušen, E., & Strembeck, M. (2018). Politics, sentiments, and misinformation: An analysis of the Twitter discussion on the 2016 Austrian presidential elections. *Online Social Networks and Media*, 5, 37–50.
62. Lai, M., Bosco, C., Patti, V., Virone, D. (2015) Debate on political reforms in Twitter: a hashtag-driven analysis of political polarization. In *2015 IEEE international conference on data science and advanced analytics (Dsaai)* (pp. 1–9). IEEE.
63. Lansdall-Welfare, T., Dzogang, F., Cristianini, N. (2016). Change-point analysis of the public mood in UK Twitter during the Brexit Referendum. In *2016 IEEE 16th international conference on data mining workshops (ICDMW)* (pp. 434–439). IEEE.
64. Sharma, S. (2013). Black Twitter? Racial hashtags, networks and contagion. *New Formations*, 78(78), 46–64.
65. World Health Organization. WHO Coronavirus (COVID-19) Dashboard [Internet]. [cited 2022 Apr 23]. <https://covid19.who.int>. Accessed 13 Nov 2022.
66. CNN BMM Hayes, M., Alfonso III, F., Rocha, V. (2020). US coronavirus news [Internet]. CNN. [cited 2022 Apr 23]. <https://www.cnn.com/us/live-news/us-coronavirus-update-05-27-20/index.html>. Accessed 13 Nov 2022.
67. Furlong, H. (2020). WHO ends hydroxychloroquine study. Politico. [Internet]. [cited 2022 Apr 23]. <https://www.politico.com/news/2020/06/17/who-ends-hydroxychloroquine-study-326238>. Accessed 13 Nov 2022.
68. WHO officials are reviewing new evidence of airborne transmission, importance of ventilation in fighting coronavirus [Internet]. CNBC. 2020 [cited 2022 Apr 23]. <https://www.cnbc.com/2020/07/07/who-officials-are-reviewing-new-evidence-of-airborne-transmission-importance-of-ventilation-in-fighting-coronavirus.html>. Accessed 13 Nov 2022.
69. Freedom House. Freedom in the World: 2019 Scores [Internet]. [cited 2022 Apr 23]. <https://freedomhouse.org/report/freedom-world/2019/scores>. Accessed 13 Nov 2022.

70. Freedom House. Freedom on the Net [Internet]. [cited 2022 Apr 23]. <https://freedomhouse.org/report/freedom-net>. Accessed 13 Nov 2022.
71. Edelman. 2019 Edelman Trust Barometer [Internet]. [cited 2022 Apr 23]. <https://www.edelman.com/trust/2019-trust-barometer>. Accessed 13 Nov 2022.
72. Transparency International. 2020 Corruption Perceptions Index [Internet]. [cited 2022 Apr 23]. <https://www.transparency.org/en/cpi/2020>. Accessed 13 Nov 2022.
73. Reporters Without Borders. (2020). 2020 World Press Freedom Index: “Entering a decisive decade for journalism, exacerbated by coronavirus” [Internet]. [cited 2022 Apr 23]. <https://rsf.org/en/2020-world-press-freedom-index-entering-decisive-decade-journalism-exacerbated-coronavirus>. Accessed 13 Nov 2022.
74. Newman, N., Fletcher, R., Kalogeropoulos, A., Nielsen, R. K. (2019). Reuters Institute Digital News Report 2019 [Internet]. Rochester, NY: Social Science Research Network. [cited 2022 Apr 23]. Report No.: ID 3414941. <https://papers.ssrn.com/abstract=3414941>. Accessed 13 Nov 2022.
75. Wellcome Trust. Trust in science and health professionals | Wellcome Global Monitor 2018 [cited 2022 Apr 23] [Internet]. Wellcome. <https://wellcome.org/reports/wellcome-global-monitor/2018/chapter-3-trust-science-and-health-professionals>. Accessed 13 Nov 2022.
76. Curini, L. (2019). Can the market stop populism? [Internet]. IREF Europe EN. [cited 2022 Apr 23]. <https://en.irefeurope.org/publications/online-articles/article/can-the-market-stop-populism/>. Accessed 13 Nov 2022.
77. Haerpfner, C., Inglehart, R., Moreno, A., Welzel, C., Kizilova, K., Diez-Medrano, J., et al. (2021). World values survey time-series (1981–2020) Cross-National Data-Set [Internet]. World Values Survey Association. [cited 2022 Apr 23]. <https://www.worldvaluessurvey.org/WVSEVStrend.jsp>. Accessed 13 Nov 2022.
78. Witt, M., de Castro, L. R. K., Amaeshi, K., Mahroum, S., Bohle, D., Saez, L. (2018). Mapping the business systems of 61 major economies: a taxonomy and implications for varieties of capitalism and business systems research. *Socio-Economic Review*, 16(1), 5–38.
79. United Nations Development Programme. Latest Human Development Index Ranking. [cited 2022 Apr 23]. <https://hdr.undp.org/en/content/latest-human-development-index-ranking>. Accessed 13 Nov 2022.
80. World Bank. Gini index (World Bank estimate) | Data. [cited 2022 Apr 23]. <https://data.worldbank.org/indicator/SI.POV.GINI>. Accessed 13 Nov 2022.
81. Lozano, R., Fullman, N., Mumford, J. E., Knight, M., Barthelemy, C. M., Abbafati, C., et al. (2020). Measuring universal health coverage based on an index of effective coverage of health services in 204 countries and territories, 1990–2019: A systematic analysis for the global burden of disease study 2019. *The Lancet*, 396(10258), 1250–1284.
82. United Nations Statistics Division. Methodology: Standard country or area codes for statistical use. [Internet] [cited 2022 Apr 23]. <https://unstats.un.org/unsd/methodology/m49/>. Accessed 13 Nov 2022.
83. Ritchie, H., Mathieu, E., Rodés-Guirao, L., Appel, C., Giattino, C., Ortiz-Ospina, E., et al. (2020). Coronavirus pandemic (COVID-19). Our world in data [Internet]. [cited 2022 Apr 23]. <https://ourworldindata.org/coronavirus>. Accessed 13 Nov 2022.
84. Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022.
85. Shahnaz, F., Berry, M. W., Pauca, V. P., & Plemmons, R. J. (2006). Document clustering using nonnegative matrix factorization. *Information Processing and Management*, 42(2), 373–386.
86. Sauvy, A. (1952). Trois Mondes, Une Planète. *L'Observateur*, 118, 14.
87. Huntington, S.P. (2000) The clash of civilizations? In *Culture and politics* (pp. 99–118). Springer.
88. Wallerstein, I. (1979). *The capitalist world-economy*. Cambridge University Press.
89. Vassilvitskii, S., & David, A. (2006). K-Means++: The Advantages of Careful Seeding. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms* (pp. 1027–1035).
90. van Buuren, S., & Groothuis-Oudshoorn, K. (2011). Mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 12(45), 1–67.
91. Ahn, H., Sun, K., & Kim, K. (2021). Comparison of missing data imputation methods in time series forecasting. *Computers, Materials and Continua*, 70(1), 767–779. <https://doi.org/10.3260/cmc.2022.019369>.

92. Daberdaku, S., Tavazzi, E., & Di Camillo, B. (2020). A Combined interpolation and weighted K-nearest neighbours approach for the imputation of longitudinal ICU laboratory data. *Journal of Healthcare Informatics Research*, 4(2), 174–188. <https://doi.org/10.1007/s41666-020-00069-1>.
93. Sun, B., Ma, L., Cheng, W., Wen, W., Goswami, P., Bai, G. 2017. An improved K-nearest neighbours method for traffic time series imputation. In *2017 Chinese Automation Congress (CAC)* (pp. 7346–7351). <https://doi.org/10.1109/CAC.2017.8244105>.
94. Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., Botstein, D., & Altman, R. B. (2001). Missing value estimation methods for DNA microarrays. *Bioinformatics*, 17(6), 520–525.
95. Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9, 2579–2605.
96. Belkin, M., & Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6), 1373–1396.
97. Kusner, M. J., Sun, Y., Kolkin, N. I., Weinberger, K. Q. (2015). From word embeddings to document distances. In *Proceedings of the 32nd international conference on international conference on machine learning—volume 37 (ICML'15)* (pp. 957–966). JMLR.org.
98. Cer, D., Yang, Y., Kong, S., Hua, N., Limtiaco, N., John, R. S., et al. (2018). Universal sentence encoder for English. In *Proceedings of the 2018 conference on empirical methods in natural language processing: system demonstrations* (pp. 169–174).
99. Devlin, J., Chang, M.-W., Lee, K., Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
100. Mihalcea, R., Corley, C., Strapparava, C. (2006). Corpus-based and knowledge-based measures of text semantic similarity. In *Proceedings of the 21st national conference on artificial intelligence—volume 1 (AAAI'06)* (pp. 775–780). AAAI Press.
101. Bun, K. K., & Ishizuka, M. (2006). Emerging topic tracking system in WWW. *Knowledge-Based Systems*, 19(3), 164–171.
102. Sankowski, P., Wegrzycki, K. (2019). Improved distance queries and cycle counting by frobenius normal form. *Theory of Computing Systems*, 63, 1049–1067. <https://doi.org/10.1007/s00224-018-9894-x>.
103. Majtey, A. P., Borrás, A., Casas, M., Lamberti, P. W., Plastino, A. (2008). Jensen Shannon divergence as a measure of the degree of entanglement. arXiv:08043662 [quant-ph] [Internet]. [cited 2022 Apr 23]. <http://arxiv.org/abs/0804.3662>.
104. Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Australian Ecology*, 26(1), 32–46.
105. Clarke, K. R. (1993). Non-parametric multivariate analyses of changes in community structure. *Australian Journal of Ecology*, 18(1), 117–143.
106. Bandeira, L., Aleksejeva, N., Knight, T., & Le Roux, J. (2021). *Weaponized: How rumors about Covid-19's origins led to a narrative arms race*. Atlantic Council.
107. Jamison, A. M., Broniatowski, D. A., Dredze, M., Wood-Doughty, Z., Khan, D., & Quinn, S. C. (2020). Vaccine-related advertising in the facebook ad archive. *Vaccine*, 38(3), 512–520. <https://doi.org/10.1016/j.vaccine.2019.10.066>.
108. Cooper, C. L. (1982). *Culture's consequences: International differences in work related values*, Geert Hofstede, Sage Publications, London and Beverly Hills, 1980. No. of Pages: 475. Price £18.75. *Journal of Organizational Behavior*, 3(2), 202–204. <https://doi.org/10.1002/job.4030030208>.
109. Schwartz, S. (1994). Beyond individualism/collectivism: New cultural dimensions of values. In: *Cross-cultural research and methodology* (pp. 85–119).
110. Mättö, M., & Niskanen, M. (2019). Religion, national culture and cross-country differences in the use of trade credit: Evidence from European SMEs. *International Journal of Managerial Finance*, 15(3), 350–370.
111. Jung, H. J., Oh, K. W., & Kim, H. M. (2021). Country differences in determinants of behavioral intention towards sustainable apparel products. *Sustainability*, 13(2), 558.
112. Hong, W., Liu, R.-D., Ding, Y., Hwang, J., Wang, J., & Yang, Y. (2021). Cross-country differences in stay-at-home behaviors during peaks in the covid-19 pandemic in China and the United States: The roles of health beliefs and behavioral intention. *International Journal of Environmental Research and Public Health*, 18(4), 2104.
113. Nguyen, A., & Tran, M. (2019). Science journalism for development in the global south: A systematic literature review of issues and challenges. *Public Understanding of Science*, 28(8), 973–990.

114. Hase, V., Mahl, D., Schäfer, M. S., & Keller, T. R. (2021). Climate change in news media across the globe: An automated analysis of issue attention and themes in climate change coverage in 10 countries (2006–2018). *Global Environmental Change*, 70, 102353.
115. McNeil, R., Karstens, E. (2018). Comparative report on cross-country media practices, migration, and mobility [Internet]. The Reminder Project. [cited 2022 Apr 23]. <https://www.reminder-project.eu/publications/reports/comparative-report-on-cross-country-media-practices-migration-and-mobility/>. Accessed 13 Nov 2022.
116. World Health Organization Regional Office for Africa. (2021). On the frontlines in the fight against dangerous misinformation [Internet]. [cited 2022 Apr 23]. <https://www.afro.who.int/news/frontlines-fight-against-dangerous-misinformation>. Accessed 13 Nov 2022.
117. Okereke, M., Ukor, N. A., Ngaruiya, L. M., Mwansa, C., Alhaj, S. M., Ogunkola, I. O., et al. (2021). COVID-19 misinformation and infodemic in Rural Africa. *American Journal of Tropical Medicine and Hygiene*, 104(2), 453–456.
118. Larson, H., Tajudeen, R. (2022). Vaccinating Africa against COVID-19: riding a roller coaster of poor information [Internet]. The conversation. [cited 2022 Apr 23]. <http://theconversation.com/vaccinating-africa-against-covid-19-riding-a-roller-coaster-of-poor-information-159716>. Accessed 13 Nov 2022.
119. Menezes, N. P., Simuzingili, M., Debebe, Z. Y., Pivodic, F., Massiah, E. (2021). What is driving COVID-19 vaccine hesitancy in Sub-Saharan Africa?. World Bank Blogs. [Internet]. [cited 2022 Apr 23]. <https://blogs.worldbank.org/africacan/what-driving-covid-19-vaccine-hesitancy-sub-saharan-africa>. Accessed 13 Nov 2022.
120. Farahat, M. (2021). Coronavirus trials in Egypt: Blurring the lines between fake news and freedom of expression [Internet]. SMEX. [cited 2022 Apr 23]. <https://smex.org/coronavirus-trials-in-egypt-blurring-the-lines-between-fake-news-and-freedom-of-expression/>. Accessed 13 Nov 2022.
121. Lipps, J., & Schraff, D. (2021). Regional inequality and institutional trust in Europe. *European Journal of Political Research*, 60(4), 892–913.
122. Habibov, N., Afandi, E., & Cheung, A. (2017). Sand or grease? Corruption-institutional trust nexus in post-Soviet countries. *Journal of Eurasian Studies*, 8(2), 172–184.
123. Rodríguez, A. B. (2021). Social networks: A source of lexical innovation and creativity in contemporary peninsular Spanish. *Languages*, 6(3), 138. <https://doi.org/10.3390/languages6030138>.
124. Herring, S. C. (2012). Grammar and electronic communication. The encyclopedia of applied linguistics. (pp. 1–9).
125. Asr, F. T. (n.d). The language gives it away: How an algorithm can help us detect fake news. The conversation. <http://theconversation.com/the-language-gives-it-away-how-an-algorithm-can-help-us-detect-fake-news-120199>. Accessed 2 Dec 2022.
126. Rashkin, H., Choi, E., Jang, J., Volkova, S., Choi, Y. (2017). Truth of varying shades: Analyzing language in fake news and political fact-checking. (pp. 2931–2937).
127. Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 8(1), 1146–1151.
128. Center for Countering Digital Hate. The Disinformation Dozen [Internet]. [cited 2022 Apr 23]. <https://www.counterhate.com/disinformationdozen>. Accessed 13 Nov 2022.
129. Robertson, R. (2014). *European glocalization in global context*. Palgrave Macmillan.
130. Roudometof, V. (2016). *Glocalization: A critical introduction*. Routledge & CRC Press.
131. Caliskan, C. (2022). Network modeling: Historical perspectives, agent-based modeling, correlation networks, and network similarities. In S. Derindere-Koseoglu (Ed.), *Financial data analytics: Theory and application*. Springer.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.