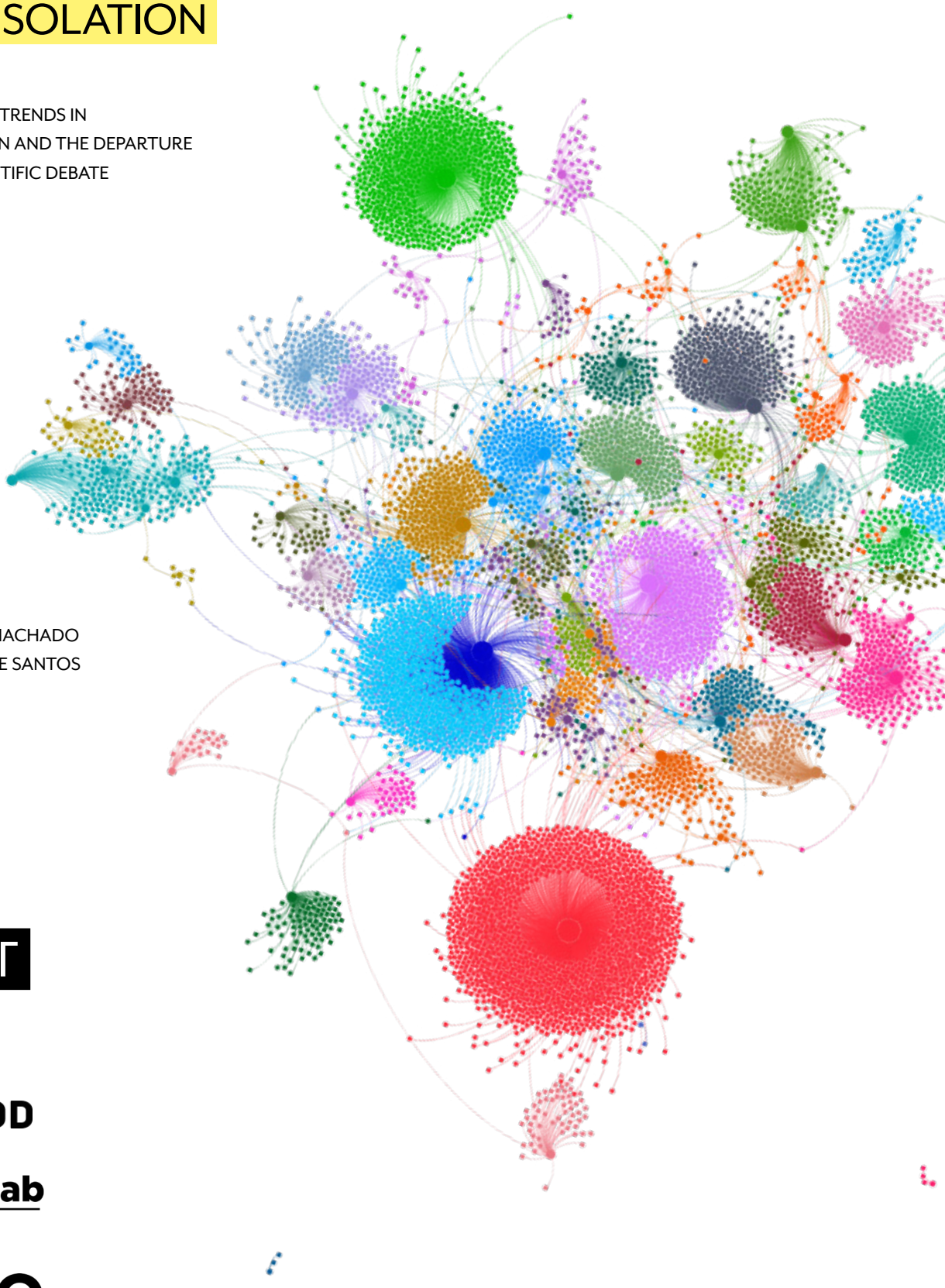
SCIENTIFIC [SELF] ISOLATION

INTERNATIONAL TRENDS IN MISINFORMATION AND THE DEPARTURE FROM THE SCIENTIFIC DEBATE



CAIO C. VIEIRA MACHADO
JOÃO GUILHERME SANTOS
NINA SANTOS
LUIZA BANDEIRA

NOVEMBER 2020

LAUT

INCT DD
INSTITUTO NACIONAL DE CIÊNCIA & TECNOLOGIA EM DEMOCRACIA DIGITAL

DFRLab

vero

ABOUT THE PROJECT

The **Infected Democracy project** analyses freedom of democratic expression and public debate dynamics in digital environments. The objective is to produce studies on how society incorporates online social networks, their platforms and instant messaging applications, and other internet services in the context of democratic participation and formulation of public policy.

Scientific [Self] Isolation is the second report in the series. It was produced using the International Fact-Checking Network's database, the largest database on Covid19 fact-checks available. It combines the fact-checks produced by the CoronaVirusFacts/DatosCoronaVirus alliance, produced by agencies in over 70 countries.

Email address for questions and requests regarding the report:
desinfo@laut.org.br

ISBN
978-65-980829-9-4

THE TEAM

CAIO C. VIEIRA MACHADO

[*@caiocvm*]

A lawyer and social scientist, he is a PhD candidate at the University of Oxford's Centre for Socio-Legal Studies and has previously carried out research at the Oxford Internet Institute. He is a researcher at LAUT and at the VERO Project. Before that, Caio was a Google Public Policy Fellow at the Institute for Technology & Society, Rio de Janeiro and a research associate at CEPEDISA. He holds an MSc degree from the University of Oxford, a Master of Law from the École de Droit de la Sorbonne and a degree in Law from the University of São Paulo. He has published many papers on disinformation, which have included studies of Whatsapp, Twitter and YouTube political and scientific disinformation, and specifically Covid19 disinformation.

JOÃO GUILHERME BASTOS DOS SANTOS

[*@Jgb_santos*]

Coordinator of Data Science for the Digital Communication Laboratory and postdoctoral Internet and Politics researcher at the Brazilian National Institute of Science and Technology for Digital Democracy (funded by Brazilian Coordination for the Improvement of Higher Level Personnel, CAPES). He is also a member of the Brazilian Association of Political Communication Scholars' Scientific Committee (Compolítica). Doctor in Communication at Rio de Janeiro State University School of Social Communication, Brazil, which included a doctoral visit (Misc. PG Research) supervised by Stephen Coleman, at the School of Media and Communication, University of Leeds, United Kingdom (UK). He has published many papers on internet and politics, which have included studies of Whatsapp, Twitter and YouTube political and scientific disinformation, and specifically Covid19 disinformation.

NINA SANTOS

[*@ninocasan*]

Post-doctoral Fellow at the Brazilian National Institute of Science and Technology for Digital Democracy and an Associate Researcher at the Analysis and Interdisciplinary Research Centre for Media at the Université Panthéon-Assas. Her research projects have been mainly focused on methodological frameworks for information dissemination via social networking sites and instant messaging apps, new information mediators and the political impacts of the transformation of media systems. She has published multiple studies on disinformation, which have included studies of Whatsapp, Twitter and YouTube political and scientific disinformation, and specifically Covid19 disinformation.

LUIZA BANDEIRA

[*@luizabandeira*]

Research Associate with the DFRLab (Digital Forensics Research Lab), a centre which is a part of the Washington DC-based think-tank Atlantic Council, where she specialises in investigating disinformation in Latin America. She has worked as a multimedia journalist for ten years in some of the most influential media outlets in Brazil, such as Folha de S. Paulo and the Brazilian service of the BBC. She graduated in journalism at the Universidade Federal do Rio de Janeiro and obtained an MSc in Social Policy and Development from the London School of Economics and Political Science (LSE).

ORGANISATIONS

The **Center for the Analysis of Liberty and Authoritarianism** (**LAUT**) observes and monitors the manifestations of authoritarianism by generating and sharing knowledge on the quality of rule of law and democracy. Its events, research reports and publications in national and international media aim to ground civil society mobilisation and the protection of civil liberties.

The **National Institute of Science and Technology in Digital Democracy** (**INCT.DD**) is composed of a network of more than 50 high-level Brazilian and foreign research centres in the area of digital democracy. It is based at the Federal University of Bahia (UFBA) and its Data Science Laboratory for Digital Communications (C2D2) develops methodologies for analysing online platforms, social networks, instant messaging and government applications.

The **Atlantic Council's Digital Forensic Research Lab** (**DFRLab**) is a leading centre at the intersection between government, media and technology. It aims to identify, expose and explain disinformation when and where it occurs, promote objective fact as a foundation in free and open societies and protect democratic institutions and norms from those who would undermine them.

The **VERO Project** is an initiative from digital communicators committed to protecting democracy, fostering online speech and building solutions to combat misinformation. The project was built by internet natives and is focused on establishing a healthy internet environment for individual and collective development. With a vocation for the digital public sphere, VERO promotes online communication through research and education initiatives.

INDEX

EXECUTIVE SUMMARY

ABSTRACT

The digital environment became a central arena for political disputes around the world and, as such, has been subject to various disinformation campaigns. False or misleading information has populated online discussions and has circulated across different countries. Disinformation is both a global and regional phenomenon; while some false claims travel through the internet, other claims are directly related to regional beliefs and political narratives. This report analyzes trends in disinformation across different countries, evaluating different patterns to understand how countries position themselves with regards to mis and disinformation related to the pandemic. To this report, we have used the International Fact-Checking Network's CoronaVirusFacts Alliance database and, subsidiarily, the Portuguese and Spanish subset of the Latam Chequea database. These databases are composed of information verified by over 70 fact-checking outlets affiliated to IFCN around the world. The datasets comprise thousands of debunking articles, translated to English, which we used to evaluate the content of misleading claims in the 134 countries and regions represented in the database. We ran two different analyses. First, a Factorial-analysis grouping the countries in the database based on different themes and word-usage. Second, a hypergeometric analysis that compares the incidence of chosen word-groups between different countries. We have analyzed the incidence of terms related to medical and non-medical treatments and the incidence of words that refer to public officials at varying administration levels. We found that, although COVID-related disinformation followed local trends, most country narratives were related to other countries, except in the cases of India and Brazil, which seemed more isolated than other nations. We confirmed this isolation using hypergeometric analysis and qualitative analysis, which showed that Brazil is even more isolated than India, as drugs such as Chloroquine and Ivermectin appeared more prominently, despite the lack of scientific evidence of their effectiveness against Covid19. We also found that, in contrast with other countries, which saw waves of disinformation, in Brazil, these drugs continue to be present in Covid misinformation throughout time, indicating that these claims are not being dismissed in the Brazilian public debate. Further analysis has indicated that Brazil is also isolated when it comes to references to public authorities in varying administration levels, showing that disinformation is largely associated with internal political disputes within the country. These findings are relevant because they offer comparative insights to mis and disinformation patterns and vocabularies circulating around the globe.

KEY FINDINGS

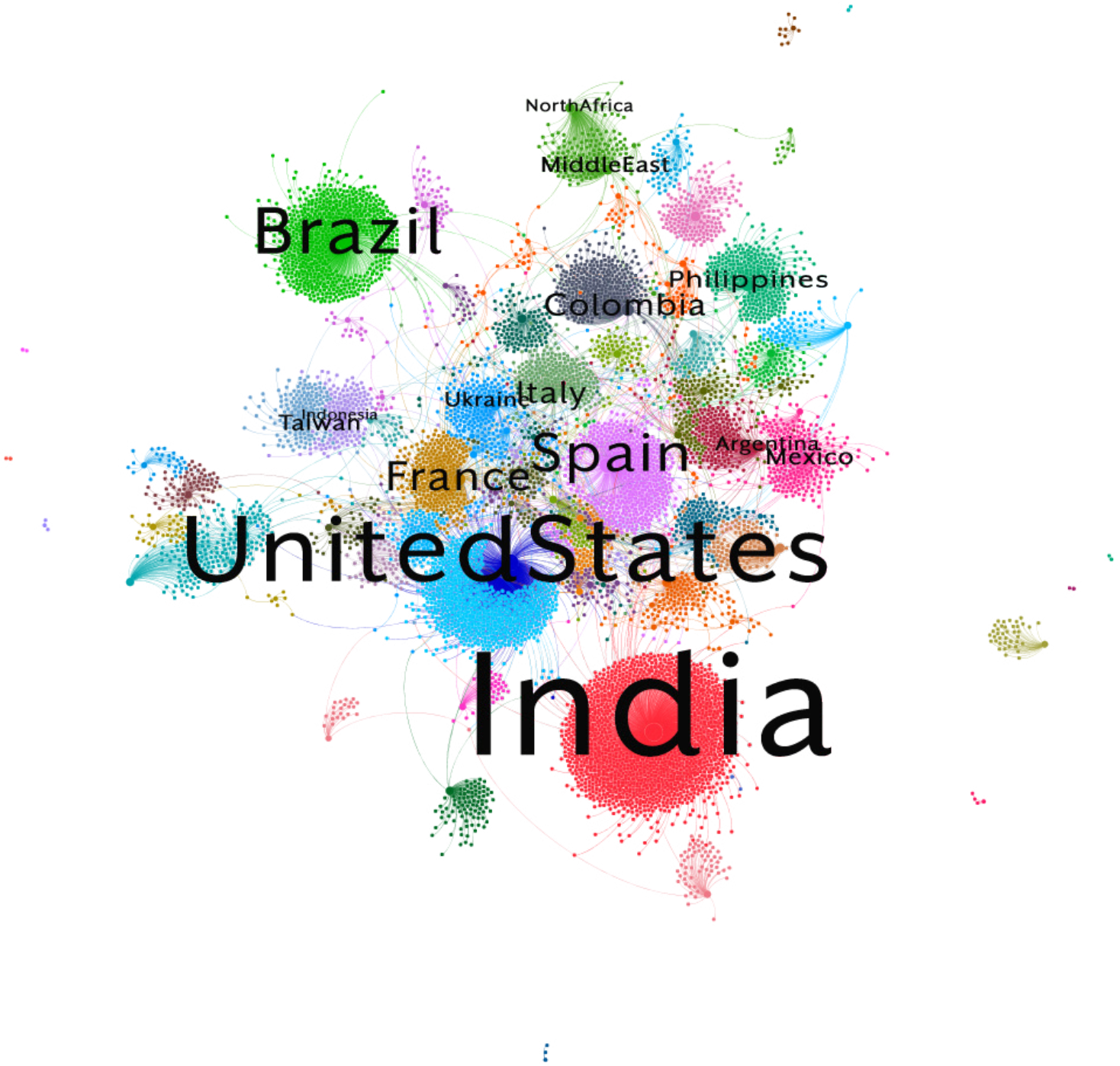
1 Using an international database, this research was able to make comparative analyses on disinformation patterns across different countries, instead of relying solely on the internal perspectives of each nation.

2 The research has identified that the falsehoods on Covid19 drug treatments follow regional patterns, and that there is some evidence that similar themes are repeated across different countries. Overall, countries seem to have common ground when it comes to disinformation topics, and the flow of disinformation across borders is likely to be due to cultural, geographical, and political factors.

3 Brazilian and Indian misinformation and disinformation trends related to Covid19 seem to be isolated from those from other countries on the database. This isolation is emphasised in the Brazilian case. This conclusion has been reached using two different methods.

4 The case of Chloroquine and Ivermectin are particularly concerning. News of these two drugs appeared as a result of scientific speculation but persist in Brazilian public debate, even after scientific research has not supported them. This trend has not been followed in other countries.

5 Regarding the appearance of public authorities in disinformation discourses, among the five countries appearing most often in the database, Brazil is the only one that has the term "Governor" mentioned more frequently than the title of the national leader, such as "President" or "Prime Minister". This suggests that disinformation is being used in the context of local power disputes at the different levels of government.



This image was made by mapping every country as a node which was connected to every claim present in the database. It is a network visualisation of the findings of this report, where the colours represent clusters of claims associated with different countries. This network structure displays shared fake news between countries, providing insights into the flow of false claims across countries. The image was produced using Gephi 0.9.1 and Adobe Photoshop.

INTRODUCTION

The Covid19 pandemic has encouraged a global effort towards scientific development. This international crisis has prompted scientists from all fields to propose solutions to this complex problem, ranging from drug and vaccine development to rethinking supply chains.

The pandemic has happened at a challenging time, as societies have never been more connected in any other time in history. Because of that, the virus was able to spread fast, travelling from one country to another via airplanes and other forms of transportation. But it is not only physically that the world is more connected; with the internet, (dis)information about the pandemic also spread faster than in any other pandemic in the past.

The intensity of contemporary information flow appears to be both a remedy and a catalyst for the serious public health calamity we are living in. On the one hand, fast communication allows for scientific efforts to be more easily coordinated and for information to be exchanged between different groups more quickly. Fighting the pandemic becomes a joint effort across research organisations, journalistic outlets, civil society, and governments around the world. On the other hand, coordinating social behaviour is crucial for responding to the pandemic and implementing public health policies. Enabling verified scientific information to reach the general public and generate trust is increasingly difficult. As the virus spreads, society seeks answers to the pressing questions that appear with the rise of a new disease: Which drugs are effective? Which types of behaviour make people vulnerable to the virus? In order to answer the questions appropriately and rigorously, science needs time and protocols for action.

While science searches for answers, disinformation spreads rapidly, motivated by a diversity of political and economic interests. This means that, by the time scientific information is out in the public sphere, it has to compete with other types of information that are already circulating, coming from different sources and on different platforms, some with high levels of adherence. Thus, in many respects, fighting the virus entails making high quality information reach the largest audience possible, in order to coordinate social behaviour. Unfortunately, this also means fending off falsehoods that compete in the information space, and the behaviours they encourage.

As it has been shown, Covid19 disinformation is produced due to a variety of motivations, which are mainly economic and political (Recuero, Soares, 2020; Ajzenman, Cavalcanti, Da Mata, 2020; Recuero, Soares, Zago, 2020). Our previous work has shown that a number of communities that are not scientific in nature have not only engaged in the policy debate, but have also attempted to assume a role in producing and analysing scientific evidence. This is particularly worrying, since it is a signal that scientific output is being politicised and brought into spheres with different values and methodologies (Machado, Dourado, Santos, Santos, 2020).

Since the beginning of the pandemic, two false narratives have been particularly prominent in the public health debate. The first targeted social distancing, with disinformation being particularly damaging to efforts aimed at coordinating social behaviour for public health reasons, such as the use of masks, self-isolation, and other non-pharmaceutical measures. The second related to substances that allegedly could prevent or treat Covid19. Claims have emerged that different drugs could cure or prevent SARS-CoV2, but most have not been backed by science. The most prominent examples have been Chloroquine, Hydroxychloroquine, and Azithromycin.

This report uses a database of false claims and fact-checks provided by the International Fact-Checking Network (IFCN) to analyse geographical trends in the spread of medical disinformation around the globe during the first months of the Covid19 pandemic. After mapping general patterns regarding disinformation, this work focuses on narratives involving medical and pseudo-medical treatments, most notably Chloroquine, Hydroxychloroquine, and Azithromycin.

METHODOLOGY

In this work, we have used the Coronavirus Fact-Check Alliance database, produced by a global coalition of fact-checkers associated with, and accredited by, the International Fact-Checking Network. This database gathers false claims and debunking articles produced by fact-checkers with common elements for comparison and in the same language. This common denominator enables the evaluation of trends across different countries.

The database is the result of a joint effort of over 70 fact-checking agencies around the world, which have checked claims relating to the Covid19 pandemic in their countries and regions. The agencies have provided an English description of the claims, along with their analysis, to the database, thus providing a common denominator for comparing Covid19 trends around the globe. The database contains over 8,600 entries and includes reports from January to late August 2020. Although this has demanded the translation of several languages into English, potentially posing challenges for making effective comparisons across countries, it was possible for the key terms relating to Covid19 disinformation to be translated the same way across all languages, making it safe to draw comparisons on the incidence of specific words such as "Chloroquine" or "Ivermectin". Only terms that were highly unambiguous were analysed, and errors and possible synonyms were also identified. To complement the analysis when discussing the Brazilian and the Latin American context, reference has also been made to the Latam Chequea database, a subset of the global database, which contains information in Portuguese and Spanish.

Two methods were used for evaluating patterns across the different countries and regions in the dataset. These methods are described in detail in the Appendix but are summarised as follows. First, similarities and differences between countries were identified and represented visually in a two-dimensional plane. Correspondence Factor Analysis (CFA) was used to analyse the claims' texts, in each country, in terms of similarities and differences in the use of words. This method maps countries on a two-dimensional graph in which the relative proximity among countries represents the similarity of word usage between them. The graph summarises how countries in the sample relate to one another, facilitating the development of hypotheses, and suggesting questions to be answered, through qualitative analysis. It is important to note that the axes in these graphs are not cartesian, but factorial, which means they serve solely for comparison, since the distances in the graph are not homogenous. The proximity in the graph represents similarity between the patterns identified, and not a result actual of values x and y (understood on a scale of entire number s). This difference highlights the importance of qualitative analyses done in this work.

The vocabulary collected from all countries in the sample was analysed using algorithms that categorised words in terms of their patterns of association (for example, one set involved political terms, another words relating to economical debate, etc.). When needed, the algorithm then broke each group into two again, to create a sort of genealogical tree of vocabulary. For example, conspiracy theories could be categorised in terms of those relating to institutional political debate or in terms of theories associating Covid19 with a planned economic warfare and others related to economic measures to fight Covid19). This technique was a useful aid in identifying the associations between fake news vocabularies and recognizing specific patterns within the sample. The divisive hierarchical clustering algorithm that was used to create this 'family tree' of words was automated.

In addition, since the number of entries per country was very disparate, a technique was used to calculate the probability of a given term being associated with a specific country. The negative values that appear on images 3 and 4 simply indicate the likelihood of a specific term not being associated with a given country. This establishes a common element for comparison across the countries. This so-called hypergeometric analysis enables (i) an evaluation of the terms that are most often used in each country, to produce an overall comparison across countries, and (ii) a means of grouping countries according to similarities in their word patterns.

Based on this method, the entire dataset was analysed and a set of selected terms was compiled that could be used for comparison across all countries. Graphs were compiled which could be used understand the relative occurrence of these key terms. By selecting from lists of medical terms, terms associated with pseudo-medical treatments, and terms associated with public administration occupations, it was possible to compare countries' behaviour in relation to these three specific fields.

FINDINGS

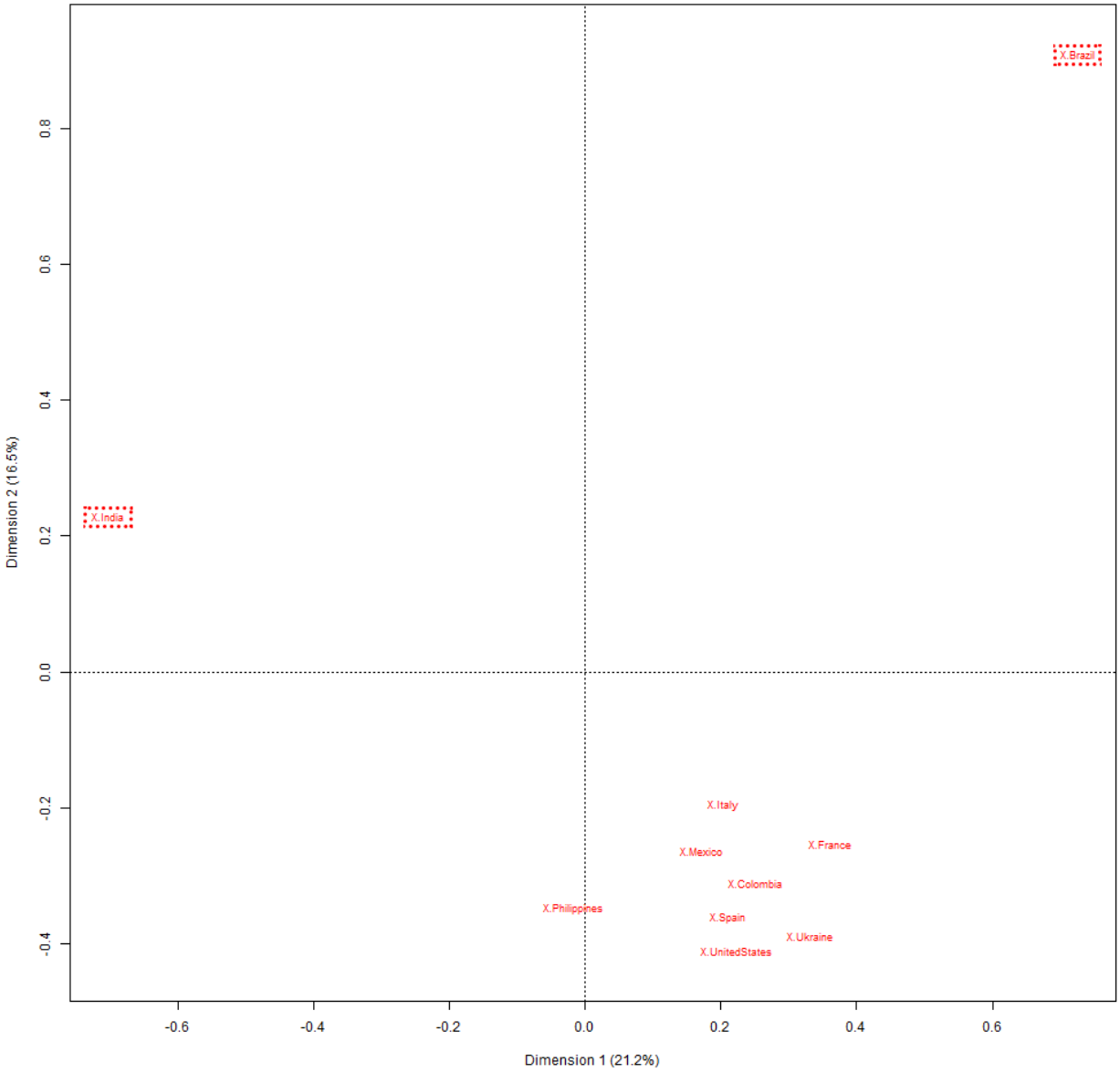
GEOGRAPHIC TRENDS RELATING TO DRUG MISINFORMATION

The analysis compared the co-occurrence of key terms in disinformation claims across countries in the database, aiming to identify geographic trends regarding the circulation of falsehoods. Countries were plotted using the methods described above to represent similarities and differences in the use of words in disinformation campaigns in each country. To facilitate visualisation, this analysis in presented in two graphs: one that includes only the ten countries with the most entries in the database (listed below) and one that includes all countries in the database.

The number of fact-checks carried out in each country did not have a direct effect on the methods used, although there might have been sampling issues where there were too few entries for a given country. For this reason, attention was focused on the comparison between countries associated with a very high number of entries in the database.

Country / Region	Number of Fact-checks
India	1.715
United States	1.018
Brazil	703
Spain	632
Colombia	349
France	342
Philippines	308
Italy	267
Mexico	233
Ukraine	202
...	...
Others	Total 8621

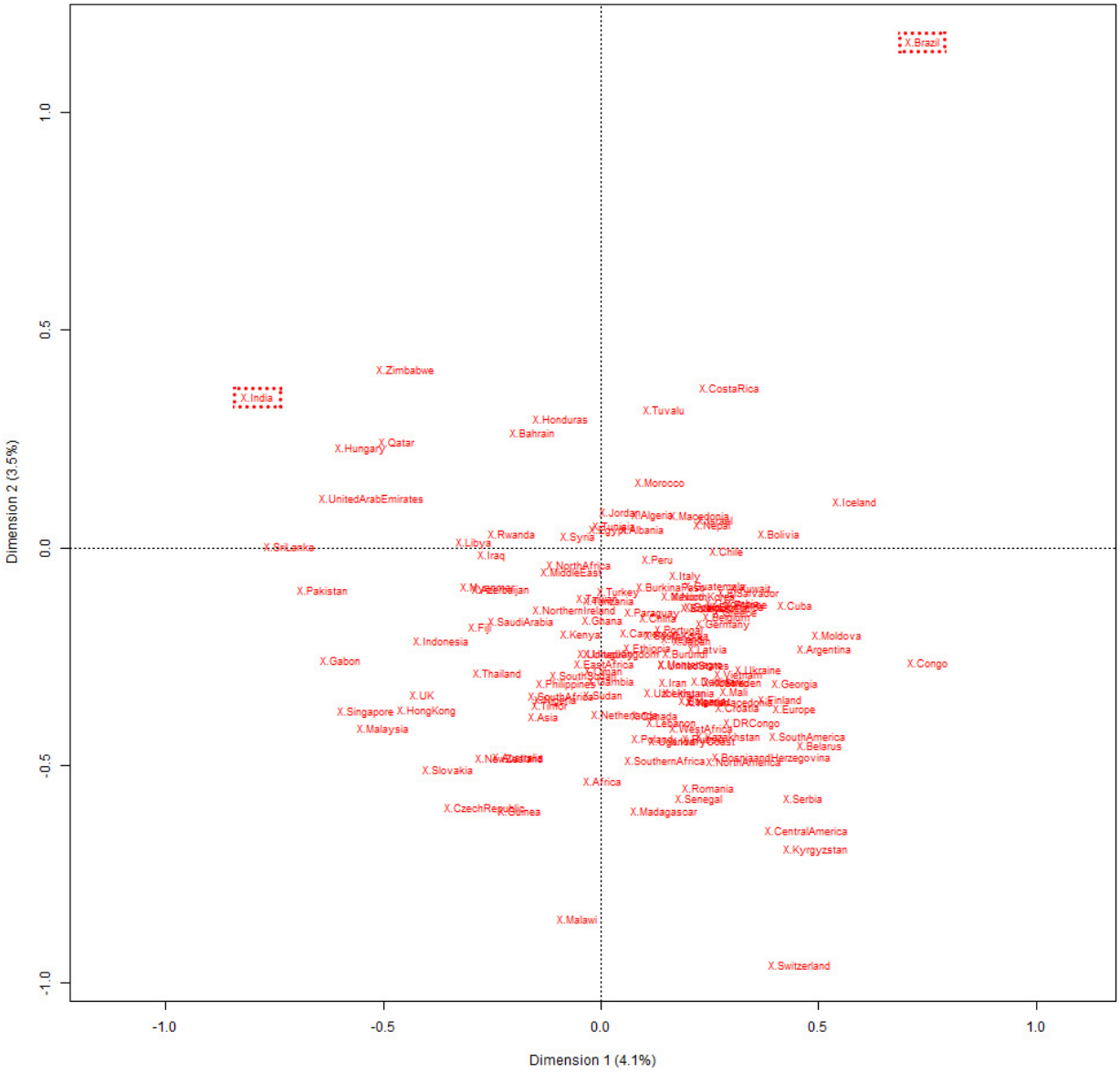
[Table 1] Ten countries with the highest number of entries in the IFCN CoronaVirus Database.



[Image 1] Relative proximity of words and phrasal structures in the ten countries with the highest number of checks registered.

The proximity of countries on the graph suggests that a similar vocabulary is being used. Conversely, the further away the countries appear, the more unrelated their word usage is. Image 1 shows the result of that analysis considering the ten countries with the highest number of checks registered in the database, while Image 2, displayed below, is the same analysis but including all countries.

The isolated position of India (upper left quadrant) and Brazil (upper right quadrant) compared with the other eight of the ten countries with the greatest number of entries stands out in Image 1. Brazil's outlier status becomes more evident when all of the countries in the graph are plotted, as in Image 2. As can be seen in the second image, while India's isolation is reduced, Brazil's isolation from the rest of the world is accentuated. This indicates that Brazil and India are using words that are very different from the rest of the world. In other words, what can be inferred is that narratives on the Covid19 pandemic circulating in these countries are very different from those circulating in the countries grouped together.



[Image 2] Relative proximity, based on word co-occurrence in all of the countries with registered fact-checks.

A few things can be derived from Image 2. For most countries, there is a high degree of similarity between the topics and narratives that appear in their public debate regarding the Covid19 pandemic. This suggests that narratives eventually circulate among countries that are closer, geographically or otherwise, and that there is some element of shared topics within this discussion.

Plotting all data points on the graph made it possible to identify countries that have more in common with India and that bridge the gap between India and the 'central group' of countries. India remains clearly isolated, but it is possible to find countries that share topics with India while also sharing topics with countries closer to the concentration of datapoints.

This makes the Brazilian case particularly interesting, since the plotting of all data points does not have the same effect with Brazil. The country remains isolated from the group and, in fact, its distance from the group appears to be greater than it was in Image 1.

It is intuitive that cultural and geographic elements should group countries around similar themes. This intuition can be confirmed in some cases. For example, Latin American countries such as Venezuela, Ecuador, Peru, and Bolivia, which share boundaries and a common language, are close to one another in the graph. Similarly, some European countries, such as the UK, Spain, France and Poland, are close together, as are North African countries. Image 1 suggests, however, that the two countries, India and Brazil, are isolated. This isolation was confirmed using subsidiary methods, which are discussed in detail below.

In terms of the detail of the themes that make Brazilian disinformation unique, one characteristic is an intense use of specific medical disinformation vocabulary that is not used in other countries, (which is discussed below, with reference to Images 3 and 4). Of course, this does not mean that other countries did not produce their own pseudo-medical disinformation, but rather that the themes that have gained force in Brazil are somehow different from those used in the rest of the world. For this reason, subsequent analysis will focus on the Brazilian case.

BRAZILIAN ISOLATION

In the case of Brazil, it is evident that the country has given exceptional emphasis to three drugs that did not have any proven effects in the treatment of Covid19: Chloroquine, Azithromycin, and Ivermectin. The former two drugs have been debated in many countries and have even been part of experiments endorsed by the World Health Organisation (WHO). As none of these drugs proved to be effective, the discussion waned in most countries. In Brazil, however, these topics persisted, to the extent that it has become a different environment for discussion than the rest of the world.

Ivermectin is a Brazilian peculiarity. The drug is a vermifuge used to kill certain kinds of intestinal worms and lice. Many Brazilians have been consuming this medication as a supposedly preventive measure[1]. Of the 36 times the term "Ivermectin" appears in the database, 23 were from Brazilian fact-checkers. The country with the second-highest incidence of the term was Colombia, a neighbouring country, with seven entries, with Mexico having the third-highest number of entries (three).

There are two interesting aspects of these data. First, this topic was very much restricted to only a few countries, with the vast majority of the discussion of Ivermectin being Brazilian. Second, this is evidence that fact-checkers in Brazil are repeatedly challenging this falsehood. This might point towards an issue in fact-checking itself, but the data suggest that something else might be involved.

In Colombia, fact-checking agencies evaluated the claim in April, and they analysed it again in July-August. In Brazil, we see that three fact-checking outlets were debunking narratives related to Ivermectin throughout all of the months of the pandemic, from April until the end of the database, in late August. This suggests that the narrative was persistent in Brazil, showing up in many forms, and that fact-checking allied with science was struggling to tackle the presence of this topic in the public debate.

1 For example, the Brazilian State of Amapá was instructing the population to take Ivermectin and Chloroquine preventively, following instructions from the president. Available a apublica.org, last accessed 05 November 2020.

There is a similar effect in relation to Chloroquine and Hydroxychloroquine, the falsehoods appear 176 times in the database and in 24 different countries. In 75 cases, the falsehoods occurred in Brazil, with India having the second-highest number of occurrences (29); France had 26 cases and the United States had 24. The discussion around Chloroquine was much larger than that of Hydroxychloroquine and Ivermectin, with many more countries presenting falsehoods in this field, but Brazil leading the number of false claims by a large margin.

Chloroquine appears to be a national success in Brazil, with a high incidence in France as well. It is worth noting that the bogus research that supported Chloroquine in the treatment of Covid19 was published in France, but work was considered of low quality and was rejected by the scientific community[2]. France abandoned studies with Chloroquine, after which the false claims on the topic waned, as can be seen from the declining incidence of checks in the database. In Brazil, however, the topic persisted.

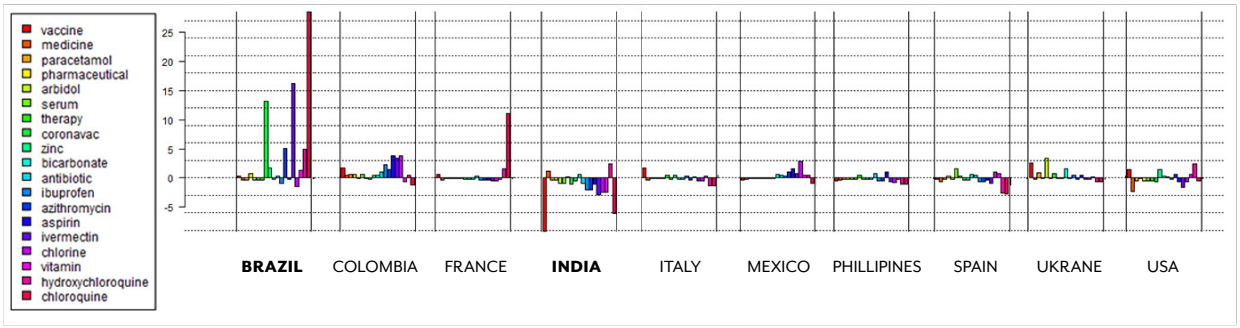
Similar to Ivermectin, after a given date, Brazil was almost alone in terms of fact-checks in relation to Chloroquine. From the May 19, 2020 onwards, debunkings ceased to be made in a variety of countries, with Brazil being virtually alone in reporting on this narrative. Of the 60 fact-checks carried out from mid-May until the end of the database, 43 originated from Brazil. This may be associated with the fact that high-ranking politicians in Brazil, including President Jair Bolsonaro, announced that they endorsed Chloroquine as a measure to fight Covid19, regardless of the lack of scientific evidence[3].

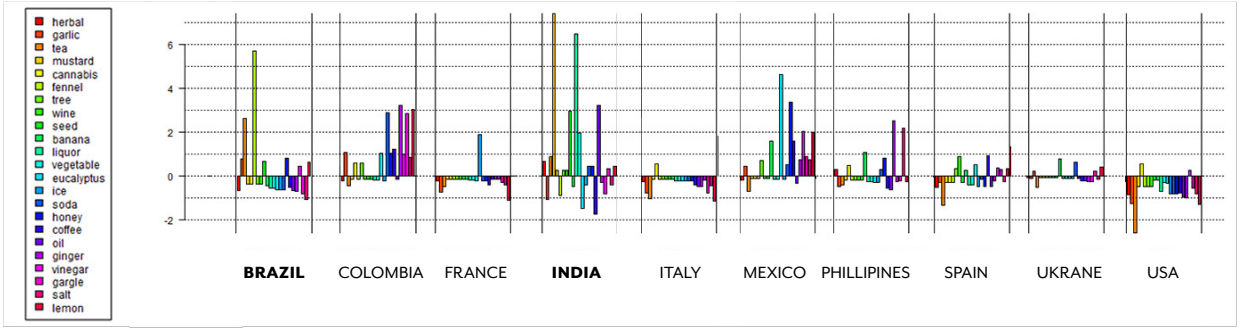
In order to validate the patterns described above, a hypergeometric analysis of the incidence of pharmaceutical terms, such as vaccine or Chloroquine, was carried out, as well as of the incidence of the non-pharmaceutical treatments, such as honey or garlic, that appeared most often in the data. The approach is described in detail in the "Methodology" section, and the results are displayed below, in Images 3 and 4, respectively.

In many countries, supposedly natural cures, such as herbal medicines, teas, and other non-pharmaceutical treatments, are common. India has been very prominent as an origin of such falsehoods but is by no means unique. The use of such terms and their distribution among the ten countries with the most entries in the database can be seen below.

2 There were several responses to the papers supporting Chloroquine and eventually the World Health Organisation decided to suspend Chloroquine testing altogether in June 2020. Available at who.int,last accessed 05 November 2020.

3 President Bolsonaro issued conflicting statements about the drug in Brazil. He would advertise it as the cure, despite he lack of scientific evidence and he even filmed himself taking a pill he claimed to be Chloroquine. He stated, at one point, that Chloroquine was 100% effective in curing Covid19 in Brazil. Nonetheless, when questioned if he was actively prescribing the drug, Bolsonaro claimed that only doctors could do that. His statements raised the sales of anti-malarial drugs in the country. For more information, see em.com.br and bbc.co.uk, last accessed 05 November 2020.

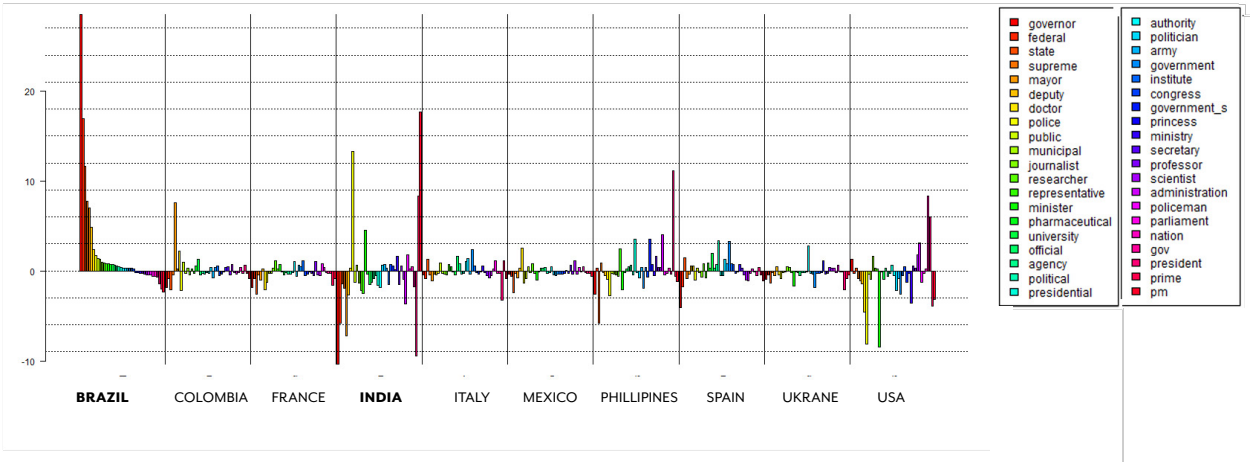




[Images 3 and 4] Comparative incidence per country of the main pharmaceutical treatments (above) and the main non-medical treatments (below) in the data set. The bars represent the probability of a word being associated with a given country; a positive probability means that there is a likelihood of the word being associated with the country, and a negative probability indicates the likelihood of the word and the country not being associated. The relevant words were selected from the sample.

INTERNAL POLITICAL DISPUTES

The analysis of the dataset reveals another curious aspect which differentiates Brazil from other countries. When comparing the vocabulary used in relation to public authorities, among the five countries most frequently mentioned in the database (India, the United States, Brazil, Spain, and Colombia), Brazil is the only country in which the term “Governor” is mentioned more frequently than the title of the national leader, (such as “President” or “Prime Minister”). In Brazil, the term “Governor” appears twice as many times as “President”. Additionaly, a prominent target is the Governor of the State of São Paulo, João Doria, who has engaged in political disputes with the national President, Jair Bolsonaro.



[Image 5] Relative incidence of terms relating to political and administrative positions in the data obtained from the hypergeometric analysis. The bars represent the probability of a term being, or not being, positively or negatively associated with a given country.

Of the 102 claims identified in the whole database that mention the word "Governor", which occurred in 19 countries, Brazil accounts for 52 of these falsehoods, followed by the 21 false claims pertaining to the United States and six from Argentina. Regarding the US, the states referenced are Michigan, Florida and California.

There were also three Brazilian states whose authorities have engaged in public health policy disputes with the federal government, namely Bahia, Rio de Janeiro and São Paulo. All of Brazilian entries that included the word "Governor" were targeting local administrators negatively and were related to a dispute between the administration and the federal government on how to protect public health in the context of the pandemic.

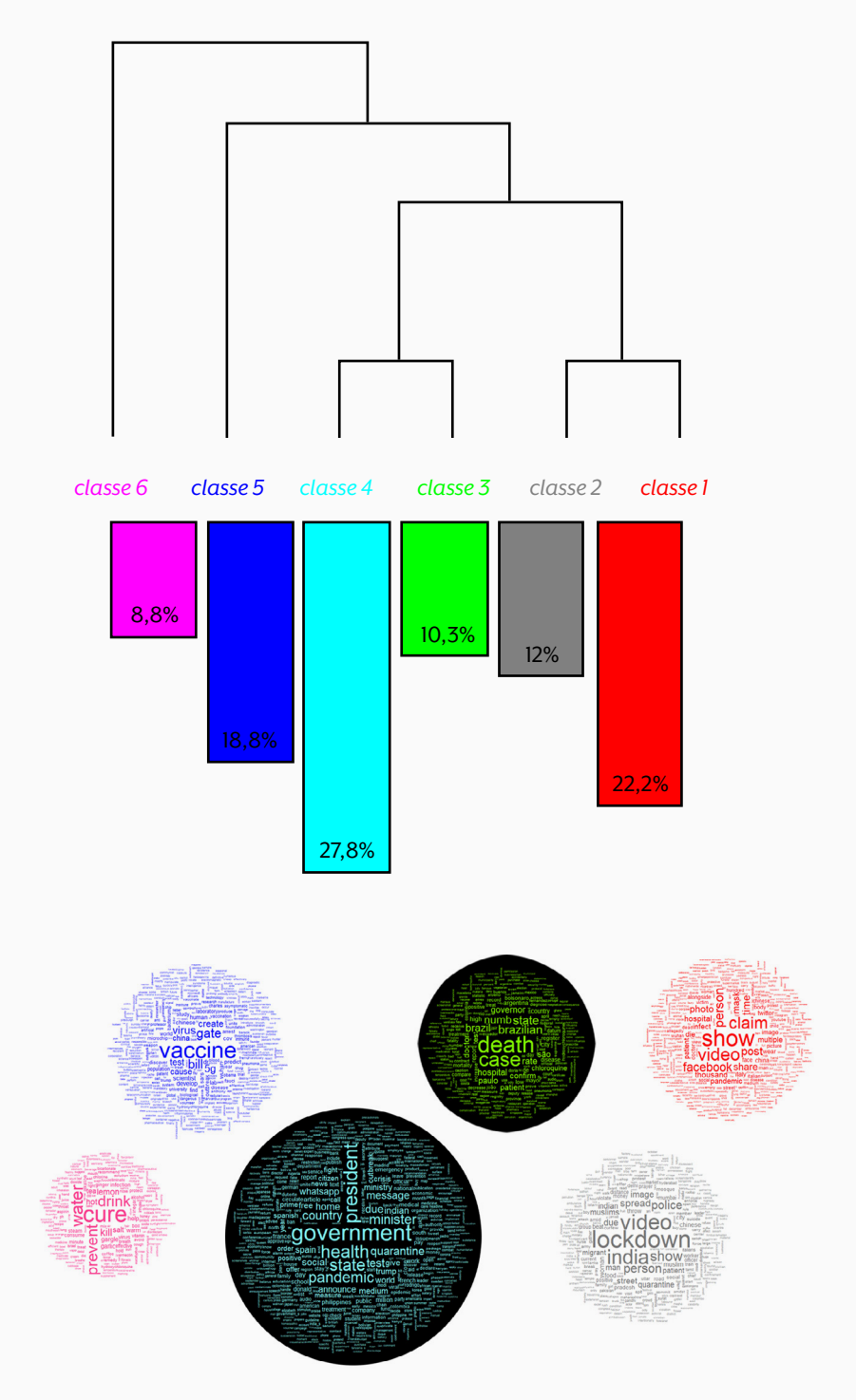
In the US, tehe main targets of disinformation campaigns have been Nancy Pelosi (Speaker of the House of Representatives), Joe Biden (presidential candidate), Barack Obama (former President) and Anthony Fauci (Head of the Center for Disease Control), all of whom are adversaries of, or have had disputes with, Donald Trump.

The prominence of these political actors in Brazil and the US suggests that many false claims in these countries have been directly related to ongoing internal political disputes[4]. Even though there are several economic incentives for producing disinformation, especially in the drug treatment market[5], the clear reference to public authorities and the systematic defamation of specific targets, aligned with political narratives, strongly suggests that these disinformation campaigns have been launched with political intentions.

4 See, for example, Trump's dispute with American Governors over public health or even Bolsonaro's dispute with governors over how to account for mortality in the country. For more information, see bbc.co.uk and noticias.uol.com.br, last accessed 05 November 2020.

5 Machado, C.; Dourado, D.; Santos, J. G.; Santos, N. (2020). Ciência Contaminada: Analisando o Contágio de Desinformação Sobre o Coronavírus via YouTube. LAUT, INCT-DD, CEPEDISA. Available at bit.ly/CienciaCont, last accessed 05 November 2020.

The database cannot reveal the overall extent of disinformation in the public debate of these countries, but the fact-checked claims nevertheless indicate an ongoing political narrative in both Brazil and the US. One can assume that fact-checkers tend to verify claims that have greater reach and that are most relevant to the ongoing political debate, which instigated the carrying out of a Factorial Correspondence Analysis of the main themes that have appeared in the political narrative. The results of this analysis are presented below.



[image 6] Dendrogram ('family tree') of the classes created after our Factorial Analysis. To characterise the content of each class, a word cloud, in corresponding colours, displays the predominant term in each class.

The dendrogram shows an expected separation based on topics, such as the associated parallel geo-political, public health, medical, and political discussions. However, it is interesting to see that the Brazilian and Indian debates each establish a class of their own. (Classes 3 and 2 respectively). This further reinforces the impression that the national discussions on Covid19 have strayed from the global discussion, and have focused on internal issues, as identified in Images 1 and 2. This is surprising, because the pandemic is a global phenomenon and most of the public health discussion is informed by scientific information being announced through international media, scientific journals and international organisations such as the WHO.

Based on the presence or absence of the co-occurrence of words and the similarity of vocabulary in the database, a set of lexical characteristics was defined. By applying the Reinert algorithm to the data, various "families" of words were derived, considering how closely those words' usage structure appear together[6]. The results of this analysis are illustrated using a dendrogram (a 'family tree'). To facilitate comprehension of this analysis, 'vocabulary families' have been placed in word clouds corresponding to specific branches of the dendrogram. The creation of these groups within this 'family tree' denotes a degree of specification within the data. The earlier the 'branching', the higher the difference between the groups in their vocabulary and lexical structures.

The first, and earliest, separation (Class 6) refers to the claims of home-made treatments and non-medical cures for Covid19. Terms relating to these claims have been very quickly isolated from other discussions. The second separation (Class 5) illustrates another category of our corpus which did not separate into further specific classes. Class 5 contains terms related to conspiracy theories on 5G, vaccines and China, with specific reference to individuals such as Anthony Fauci, Barack Obama, Bill Gates, and the Pope.

Then, the graph separates into two branches with two pairs of subcategories each. Class 4 is characterised by the debate on institutional politics, including terms such as "government", "President", "Minister", "Governor", "state", "emergency" and others. Class 3 denotes a group of words that is very close to its bifurcation (class 4), and these words are dominant in the Brazilian discussion; these words include "death", "States", "Bolsonaro", "Governor", "São Paulo", "Chloroquine", some of which are highly specific to the Brazilian debate. It is important to note, here, that this analysis includes the whole dataset of 137 countries and regions, and so it is an interesting finding that Brazilian particularities have such prominence.

It is worth remembering that India, the United States are the only three countries of a database of 134 different countries and regions that have references to local administrators. This suggests that disinformation in the countries is different in the sense that it is linked to internal disputes. Also, India and the United States, summed together, have almost four times as many inputs to the database as Brazil. This makes the Brazilian case even more salient; it is by far the country with the greatest amount of references to local administrator, indicating that the relative intensity of the dispute in Brazil is accentuated.

6 Refer to the appendix for a detailed explanation on how this methodology works.

The final two groups of words, in Class 1, include terms relating to public health measures, such as “lockdown”, “quarantine”, “police”, and “spread”. Class 2 includes “video”, and some vocabulary relating to the Indian discussion, such as “India”, “muslim” and “migrant”. Closer examination of this group of words reveals terms such as “Facebook”, “Twitter”, “share” and “post”, as well as “masks”, “hospital” and “China”.

In sum, the classes correspond to six groupings elements we have identified in the whole corpus: Home-made treatments (Class 6), Conspiracy theories (Class 5), Institutional politics (Class 4), Brazilian politics (Class 3), Public Health measures (Class 2), and Indian politics (Class 1).

In conclusion, we have identified a very particular trend of disinformation that relates to internal power disputes of different public officials. Considering the size of the corpus, it is interesting to see how intense these narratives were present in the disputes in Brazil, to the extent that they appear much more than in other countries. It is also worth mentioning that our methods of grouping word-usage in classes has identified a separate class almost entirely for the Brazilian discussion, out of the corpus composed of 134 countries and regions, showing how the country is straying from the global discussions on Covid19 (dis)information.

DISCUSSION

Similarities among countries show that disinformation is not confined to specific nations and territories. This means that falsehoods, as well as news of scientific discoveries, quickly travel across different countries (Nielsen et al., 2020). Disinformation cannot be treated as a country-specific issue. The isolation of a country, on the other hand, indicates a general distancing from science itself. Since the scientific debate occurs at a global level, especially in the context of the current pandemic, the interpretation must be that rejecting this debate is a form of rejecting the sources of scientific knowledge responsible for producing evidence and consensus in relation to Covid19.

The findings of the research suggest that there are regional patterns regarding drugs treatments for Covid19. It appears that countries with similar languages, cultures and geographic locations experience similar narratives. In this regard, the yardstick for measuring public policy performance remains the same for most countries and there is no reason yet to believe in region-specific solutions in relation to drugs and vaccines.

In other words, the scientific debate on drugs and the pharmacological treatment of Covid19 is highly internationalised, and it is generated through exchanges between expert organisations across the world. However, bridging the gap between the production of science, society, and policy making has been particularly challenging during the pandemic.

Awareness of the science behind Covid19 is driven by formal and informal institutions, ranging from public health agencies and international organisations, to media outlets and even 'influencers'. Digital communication has become a very diverse space, in which all of these actors have a dialogue with different audiences, in many instances competing for this space. This is where bogus content can obfuscate access to, understanding of or belief in factual information and knowledge derived from scientific research. The high incidence of falsehoods indicates how scientific progress can be hampered simply by impairing communication in society.

As a result, the problem of disinformation is less related to scientific quality and rigour, and rather more related to the narrative disputes that occur in the public sphere through digital communication. Science has its methods for producing knowledge and consensus-building, but the current pandemic is occurring within a scenario of intensified political disputes, high polarisation and as yet embryonic norms to regulate the flow of information in the digital space.

It is quite evident that, in many regards, political disputes have overshadowed and interfered with scientific consensus-building and scientific communication. The digital public sphere has become, at the same time, the medium used to irrigate society with information, but also the medium for disputes relating to political narratives. The association of drug treatments with specific parties, authorities and ideologies is quite symptomatic of this reality. As has been shown, Brazil, India and the United States are countries immersed in strong disinformation campaigns regarding Covid19 drugs, but these campaigns are also strongly permeated by clashes between local powers.

The politicisation of expert systems impairs scientific progress, be it by interfering in the scientific debate itself, or by affecting the way in which science is communicated to the public and informs public policy. From the analysis, it can be inferred that, in some cases, scientific positions have been associated with political preferences. This has been the case in the United States and Brazil, and in these countries there have been instances where disinformation has emanated from public authorities.

On this subject, the Brazilian isolation identified by both methodologies applied in the research is of great concern. By examining the different terms being used, the persistence of certain topics over time and the political disputes that persistently permeate the discussion in Brazil, there is strong evidence that the country is distancing itself from the ongoing scientific debate. Certain false claims seem to persist in Brazil, despite there being a scientific consensus discarding a certain drug as a possible treatment.

The cases of Chloroquine and Ivermectin are particularly concerning. These are not regional treatments that have a strong cultural component to their belief as cures for Covid19, as is the case with certain herbs, teas, honey and other home-made cures and recipes. These two drugs have risen to prominence as a result of scientific speculation, but persist in Brazilian public debate even after scientific research has not supported them[7]. Fact-checkers having to debunk such claims repeatedly throughout the pandemic is a sign that the element of trust in scientific output is missing. Moreover, it also entails a huge amount of time and effort in attempting to fend off damaging narratives from the public sphere, with disinformation nevertheless being successful in the sense that it is occupying space in the public debate. This space could be used for circulating new scientific findings, generating awareness or discussing public policies. A finding of the research is that political communication is capturing the public health and medical debate at a wide scale, and that is leading to the rejection of scientific consensus (which is the basis for fact-checking).

7 See footnotes 2 and 3 above.

CONCLUSION

In different countries, the use of medications has been largely politicised and associated with political parties and ideologies. The politicisation of science is a grave threat to science, which is the effective and credible means of developing solutions for the pandemic. Scientific understanding is built and upheld through its own set of methods and consensus-building spaces.

In this report, the IFCN CoronaVirusFacts Alliance's database has been analysed using a variety of language analysis methodologies. These methods have enabled countries to be grouped based on the co-occurrence of terms and the relationships between certain topics in the data. Through the use of algorithms, it has been possible to identify an overall cohesion between countries, showing that disinformation is similar throughout the different countries analysed. There are, however, important exceptions to this. Data obtained from India is very different to the rest of the dataset and Brazil, especially, is completely isolated in terms of the themes and words that appear in the data.

A hypergeometric analysis of the incidence of terms from each country has revealed that Brazil has a well above average incidence of terms related to three drugs: Chloroquine, Azithromycin, and Ivermectin. These drugs have received some attention in international debate but have not been scientifically supported as effective treatments for Covid19. These terms seem to linger in Brazil and continue to demand much attention from national fact-checkers, which is concerning. While other fact-checkers around the globe have tackled the issue, Brazil seems to dedicate enormous resources to debunking false narratives, even after the rest of the world has moved on to verifying other false narratives.

One possible reason for this is that these terms, especially Chloroquine, have become highly politicised in Brazil and there has been much interference from politicians in the scientific discussion. As evidence of this, Brazil has also been associated with a high incidence of terms relating to institutional politics, which are related to the campaign management of internal power disputes between administrations.

Although these findings need to be set in the wider context of the consumption of accurate scientific information also gaining in importance during the pandemic (Pulido et al., 2020), they are, nonetheless, extremely concerning. They suggest that scientific output is becoming associated with political, economic, and religious values and that the scientific method is being tampered with by external forces. While it appears wise for evidence to be discussed during decision-making, the digital public sphere has become subjected to a great deal of competing information on what scientific consensus means. Expert networks are unable to irrigate society with information, since disinformation narratives – sometimes fostered by public authorities – are occupying these spaces and competing for the same audiences.

ACKNOWLEDGEMENTS

The authors would like to thank the International Fact-Checking Network for providing the CoronaVirusFacts Alliance database for this research, the richest international dataset of fake news concerning Covid19. Moreover, we would like to thank the Shuttleworth Foundation for providing us with the flash grant to run this project.

We would also like to thank Chico Camargo for his valuable comments and contributions. Finally, we are also very grateful for the community of fact-checkers and influencers who have continuously engaged with the issues of disinformation and provided insights that have led to this report.

REFERENCES

Ajzenman, N.; Cavalcanti, T.; Da Mata, D. (2020). More Than Words: Leaders' Speech and Risky Behavior during a Pandemic. *SSRN Electronic Journal*.

Machado, C.; Dourado, D.; Santos, N.; Santos, J. G. (2020). Ciência Contaminada: Analisando o Contágio de Desinformação Sobre o Coronavírus via YouTube. *LAUT, INCT-DD, CEPEDISA*. Available at bit.ly/CienciaCont.

Nielsen, R. K.; Fletcher, N. N.; Brennen, J. S.; Howard, P. N. (2020). Navigating the 'Infodemic': *How People in Six Countries Access and Rate News and Information about Coronavirus*. Reuters Institute, University of Oxford.

Pulido, C. M. et al. COVID-19 infodemic: More retweets for science-based information on coronavirus than for false information. International Sociology, v. 35, n. 4, p. 377–392, 2020.

Recuero, R.; Soares, F. B. (2020). O Discurso Desinformativo sobre a Cura do COVID-19 no Twitter: Estudo de caso. *E-Compós*.

Recuero, R.; Soares, F.; Zago, G. (2020) Polarização, Hiperpartidarismo e Câmaras de Eco : Como circula a Desinformação sobre Covid-19 no Twitter. *Revista Contracampo*.

Reinert, M. Alceste, une méthodologie d'analyse des données textuelles et une application: Aurelia de Gerard de Nerval. Bulletin de Methodologie Sociologique, v. 26, p. 24-54, 1990. doi.org.

APPENDIX

To understand how fake news spreads in different countries, combining global trends and unique regional features, we combined different textual clustering methods able to group or distinguish countries based on lexical similarities and discrepancies, (all texts from the IFCN Database used were in English). These methods were used to identify trends to be checked and analysed qualitatively, and their results are not considered meaningful before qualitative validation.

One of them is the Reinert method (Reinert, 1990), used for the identification of general vocabularies (classes), their subdivisions and genealogies based on lexical resemblance (similarities being measured by Chi-squared distance). In this approach: (i) large texts are divided into smaller sentences or segments (elementary context units), (ii) the presence or absence of words and their co-occurrences in each segment are turned into a matrix, without taking frequency into account, (iii) the similarities and disparities in their lexical features are used to group similar segments and identify the ones that diverge (inter-cluster Chi-squared distance), and (iv) vocabularies are progressively split into two sub vocabularies, building a structure of hierarchical bifurcations (dendrogram) similar to a genealogical tree (see Image 6). This process is automated through a divisive hierarchical clustering algorithm in R language, using the R Interface for Multidimensional Analysis of Texts and Questionnaires (IRaMuTeQ) written by Pierre Ratineau.

The interface is also used to deploy Factorial Correspondence Analysis, which also uses lexical features to define distances and proximity among the segments evaluated, but, instead of identifying relations among vocabularies, it compares the texts' sources, generating a graphical representation in which all sources analysed can be located in relation (closer or far) to all others. In our analysis, we considered the country in which the fake news was registered as the source of the text segment, and all countries are plotted in groups or isolated, according to the lexical similarities of the fake news found on them (see Images 1 and 2).

We also present measurements related to the sub representation or super representation of specific words in each country's dataset of fake news texts, using a hypergeometric test. The test is used to build a chart in which the differences between the probability of finding a specific word in the entire sample and its presence in each country results in positive (super represented) or negative (sub represented) values. It highlights recurrently used or ignored words in each country and is useful for understanding their specificities, avoiding words that are common and frequent in all countries (see Images 3, 4 and 5).

The image on our cover page was made mapping every country as a node which was connected to every claim present in the database. The colours are associated with the interconnectedness of the claims and countries, used to identify the different clusters. The result is a network structure in which shared fake news connects countries, providing insights into the flow of false claims across countries. The image was produced using Gephi 0.9.1 and Adobe Photoshop.

It is important to consider that the dataset was made by fact-checkers. This means that the selection processes for misinformation-chekcs across the multiple fact-checking agencies might not be the same and may introduce divergences. We consider this to be the richest international dataset of fake news concerning Covid19.