At the onset of an infodemic: Geographic and disciplinary boundaries in researching problematic COVID-19 information
by Nicola Righetti, Luca Rossi, and Giada Marino

Abstract
This paper analyzes the literature on problematic COVID-19 information published at the onset of the pandemic in 2020. It explores how scientific research has addressed this issue from a disciplinary, methodological and substantive perspective in different world regions. Three hundred seventy-eight articles were analyzed using content analysis and computational methods, including social network analysis and text mining. The study revealed a multidisciplinary field characterized by substantial contributions from medicine and social sciences and with a certain degree of interdisciplinarity and international collaborations. Research devoted particular attention to infodemic and conspiracy theories and their impact on compliance with health-protective behaviors, and showed a general preference for quantitative methods such as surveys. Most contributions focused on European and Americas regions and were from authors working in the same areas. Attention to various topics was also geographically differentiated. For example, conspiracy theories and informational factors that may influence COVID-compliant behaviors characterize the whole corpus, but Western research did the lion’s share. Similarly, the attention paid to different social media platforms differed geographically. Worldwide attention was dedicated to global social media platforms such as Twitter, WhatsApp, Facebook and Instagram. In contrast, geographically specific platforms (e.g., VKontakte or WeChat) have been studied less and mainly in regions where they are used the most.

Focusing on a crucial period of the COVID-19 infodemic such as the first year of the pandemic, this study contributes to the literature on problematic information by highlighting how misinformation and disinformation research was carried out within and at the crossroads of disciplines and geographic regions. It suggests under-explored areas by contrast with the most relevant trends and establishes a foundation for benchmarking with post-2020 research on this topic.

Contents
Introduction
Problematic information amidst a pandemic
Method
Results
Discussion and conclusions
Introduction

The COVID-19 pandemic marked a turning point in academic attention to problematic information, which has become a central topic in communication science and has attracted more and more attention from other disciplines, such as health sciences. The pandemic has exposed people to unprecedented levels of uncertainty and prompted them to seek information. Information needs have been met not only by news outlets (Bermejo, 2020; Zarocostas, 2020) and institutions (Lovari, 2020). Disinformation motivated by ideological and economic reasons has also flourished (Brennen, et al., 2020; Ferrara, et al., 2020; Giglietto, et al., 2021), along with state-backed propaganda and conspiracy theories (DiResta, et al., 2020; Molter and DiResta, 2020; Serrato and Schafer, 2020; Shahsavari, et al., 2020).

When individual behaviors are the key to containing the spread of a problem, reaching people with the correct information at the right time is one of the most critical assets. Pandemic communication is an area where policy, communication, and epidemiology are closely intertwined. Therefore, the interaction of relevant expertise is fundamental, as a multidisciplinary and even interdisciplinary approach capable of tackling the complexity of the phenomenon. At the same time, the global nature of the phenomenon raises the question of how it has been addressed in different parts of the world. This is particularly important given the possible structural differences in the resources available to researchers in different countries. Nothing like a global issue like a pandemic, and the disinformation that floods the World Wide Web during a pandemic, can show the interconnectedness of a global society.

This paper explores how research responded to information disorder at the onset of the pandemic in 2020. It examines levels of attention from disciplinary and geographic perspectives and the breadth of interdisciplinary and transnational dialogue. It also uses a geographic and disciplinary perspective to explore the methodological approaches to address the issue and the main regional focus of interest. The particular attention paid to the disciplinary and geographical dimension makes this review different from the others on the same topic (e.g., Ali, 2022; Gabarron, et al., 2021; Janmohamed et al., 2021; Walter, et al., 2021).

Problematic information amidst a pandemic

Problematic information was a problem well before the COVID-19 outbreak. It has been observed that the crisis of legitimacy of expert knowledge paved the way to the spread of unscientific beliefs (Kata, 2012), a process favored by digital media. Digital media also lowered the barriers to entry into the information market and created an overloaded environment (Lovari and Righetti, 2021) competing for people’s attention (Goldhaber, 1997). The use of controversial and polarizing content, and even fake news (Righetti, 2021), has become common in order to attract readers and generate profits (Stanyer, 2008; Sunstein, 2017). On this basis, it shouldn’t be surprising over a correlation between COVID-19 and the information pandemic (Bermejo, 2020), which flooded the Web, and other media as well, along with anti-scientific, false, sensationalistic, and otherwise problematic content, besides legitimate news. The complexity of the issue is reflected in a kaleidoscopic vocabulary used to describe such problematic content (Jack, 2017): fake news, misinformation, disinformation, information disorder, conspiracy theories, and infodemic. In this paper, this terminology has been used for identifying a comprehensive corpus of studies dealing with the issue of problematic information during the COVID-19 pandemic in 2020. A brief description of the meaning of these terms may thus prove helpful.

Fake news is one of the most popular ways of referencing mis/disinformation (Righetti, 2021). It has been defined as “fabricated information that mimics news media content in form but not in organizational process or intent” [1]. Although it has become synonymous with false information, it is just one kind of problematic information among many. In particular, the experts differentiated problematic information based on the different intentions behind their dissemination and creation. Misinformation indicates the
unintentional spread of false news, whereas disinformation refers to the deliberate fabrication and spread of falsehoods (Jack, 2017). This simple conceptual distinction conceals more vague empirical boundaries (Giglietto, et al., 2019). More general concepts have also been suggested, such as pseudo-information (Kim and de Zúñiga, 2021).

Infodemic has been one of the buzzwords of the pandemic. The term was used initially in a Washington Post commentary about SARS by David J. Rothkopf (2003): “A few facts, mixed with fear, speculation and rumor, amplified and relayed swiftly worldwide by modern information technologies”. Likewise, the World Health Organization used it to refer to an overly prolific, confusing, and potentially harmful flow of information about COVID-19:

“To much information including false or misleading information in digital and physical environments during a disease outbreak. It causes confusion and risk-taking behaviors that can harm health. It also leads to mistrust in health authorities and undermines the public health response. An infodemic can intensify or lengthen outbreaks when people are unsure about what they need to do to protect their health and the health of people around them. With growing digitization — an expansion of social media and Internet use — information can spread more rapidly. This can help to more quickly fill information voids but can also amplify harmful messages.” [2]

As the flow of content online depends on algorithms (van Dijck and Poell, 2013), digital media companies play a fundamental role in fostering and countering an infodemic. Social media platforms have put in place several initiatives to counter misinformation and disinformation related to COVID-19: they have monitored keywords and tagged related posts, removed false news, and supplied users with verified sources of information (Lovari, 2020). Nonetheless, the uncertainty of the pandemic has been fertile ground for problematic ideas. Conspiracy theories are a case in point (Douglas, et al., 2019; Grey Ellis, 2020; Sunstein and Vermeule, 2009). Their spread has been linked to some undesirable consequences: low compliance with social distancing measures (Swami and Barron, 2021), racist attitudes towards social groups and minorities (Grey Ellis, 2020; Shimizu, 2020), diminished trust in science and in institutions (Douglas, et al., 2019), reduced political participation (Butler, et al., 1995; Uscinski and Parent, 2014), and disruptive and violent behavior in general (Ahmed, et al., 2020).

The present study was conceived to investigate how scientific research tackled the complexity of information disorder at the onset of the pandemic, with particular attention to geographical and disciplinary dimensions. It was guided by three main research questions. The first question dealt with interconnections, both geographical and disciplinary, and stemmed from the importance of an interdisciplinary and transnational approach to such a complex and global problem. It asked: what was the structure of this scientific domain in terms of the most represented disciplines, their interdisciplinary interconnections, extent of international research collaborations, and geographic focus of interest? The second research question examined the methods used to address the issue. It asked what types of methods had been adopted and whether there were differences in methodological approaches between geographies and disciplines. Particular attention was also paid to social media platforms used to analyze the problem. The third question addressed the main themes covered in the studies and potential differences between geographical and disciplinary areas.

Method

To address these questions, scientific literature was collected and screened following PRISMA guidelines
At the onset of an infodemic: Geographic and disciplinary boundaries in researching problematic COVID-19 information (Page, et al., 2021), a series of methodological steps for conducting and reporting literature research (Figure 1).

Figure 1: PRISMA flow chart of the different phases of the systematic review.

Note: Larger version of Figure 1 available here.
The identification phase relied on four leading academic databases (Scopus, Web of Science, EBSCOhost Communication & Mass Media Complete, EBSCOhost Communication Abstracts), to search papers matching a set of keywords related to COVID-19 and problematic information [3]. The search was conducted in the title and abstract of articles published in 2020 and written in English. Commentaries, letters, etc., were excluded, to concentrate on empirical research. This resulted in a total of 573 unique articles. The completeness of this dataset was assessed based on the number of records returned by the World Health Organization’s literature search service dedicated to COVID-19 [4], which yielded 484 results for the same search string.

During the second phase (screening), the papers were closely inspected by the authors. In the third phase (eligibility), non-empirical articles (e.g., introductions to special issues, articles in magazines, and letters) and those dealing only incidentally with problematic information were removed, resulting in 378 records. The eligibility analysis was performed by the three authors, whose intercoder agreement was measured with Krippendorff’s alpha on a random sample of 50 records (over 10 percent of the dataset) and resulted equal to 0.885. According to Krippendorff, this is a high level of reliability, as he suggested to rely on variables with reliability equal to or above .800, except when tentative conclusions are tolerable and reliability equal to or above .667 could be acceptable [5]. The 378 papers were further analyzed by the three coders based on a predefined number of categories and included in a first quantitative synthesis. Intercoder reliability was assessed on a random sample of 50 items (see Table 1).

<table>
<thead>
<tr>
<th>Category</th>
<th>Krippendorff’s alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methodological approach (e.g., quantitative, qualitative, and discursive)</td>
<td>0.924</td>
</tr>
<tr>
<td>Methods (e.g., survey, qualitative content analysis, social network analysis)</td>
<td>0.820</td>
</tr>
<tr>
<td>Unit of observation (e.g., health professionals, general population, social media accounts, etc.)</td>
<td>0.777</td>
</tr>
<tr>
<td>Countries representing the focus of the analysis</td>
<td>0.863</td>
</tr>
<tr>
<td>Social media platforms taken into account</td>
<td>1.00</td>
</tr>
<tr>
<td>Online services and social media used to collect data (e.g., Amazon Mechanical Turk, survey shared on Facebook, etc.)</td>
<td>0.794</td>
</tr>
<tr>
<td>Sampling strategy (e.g., probabilistic sampling, convenience sampling, quota sampling, etc.)</td>
<td>0.812</td>
</tr>
<tr>
<td>Sample size</td>
<td>0.798</td>
</tr>
</tbody>
</table>
At the onset of an infodemic: Geographic and disciplinary boundaries in researching problematic COVID-19 information

Computational analysis included network and text mining analysis. Network analysis was used to answer the first research question. It made use of metadata available only from papers retrieved from *Scopus* and *Web of Science* and was therefore based on the 353 papers including the necessary information (93.4 percent of the entire dataset). This subset of data was augmented with the disciplinary areas of the journals in which the articles were published, retrieved from *SCImago Journal & Country Rank* (https://www.scimagojr.com). The *bibliometrix* R package (Aria and Cuccurullo, 2017) was used to analyze interdisciplinary connections through bibliographic coupling (Kessler, 1963). Papers were bibliographically coupled when two works referenced a common third work in their bibliographies; it was customarily used as a sign that the two works deal with a common subject matter. Moreover, identifying a common set of references could shed light on the disciplinary background of the authors of the publications. This was an important piece of information since it was not possible to assume a full alignment between the scientific area of the journal that hosted the publication and that of the authors.

The network of scientific collaborations, as well as their relations with geographical regions, were also studied. An index of international collaboration (IC) within the academic discipline was defined for exploring these dynamics. It was computed as the sum of the number of distinct countries that participated in the papers within a given discipline, divided by the number of papers in that discipline. The IC score would be one if all the papers were written only by authors affiliated with the same country.

Data extracted from *bibliometrix* were also used to respond to the second and third research questions with specific reference to the geographical dimension of methods and themes. The most useful information for answering this question came from the affiliation of the first author of the study. This information was complemented by the geographic focus of the research that was annotated manually. Together, these data provided a more comprehensive picture of the geographical dimension of research. This analysis was carried out on the subset of the original dataset for which this information was available (N = 174, 46 percent of the original corpus). Data on the geographic dimension were analyzed by considering individual nations and aggregating them within the larger regions of the World Health Organization (WHO). There are six WHO regions, consisting of 194 member states. African Region (AFR), Region of the Americas (AMR), South-East Asian Region (SEAR), European Region (EUR), Eastern Mediterranean Region (EMR), and Western Pacific Region (WPR). We also distinguished between the United States (USA) and the other countries that are part of the Americas region. The European Region (EUR) was also divided into Eastern Europe countries (EASTEU) and Western Europe countries (WESTEU).

Topics and their prevalence across the research corpus were analyzed through text mining techniques, a useful approach for extracting information from a large number of documents. The analysis focused on the abstracts, as they summarized the main content of a paper. Two different methods were employed to get more robust results: descending hierarchical classification (Reinert, 1990, 1983; Ratinaud, 2008), and topic modeling (Roberts, *et al.*, 2019). The first method identifies lexical clusters representing distinct semantic dimensions. The second identifies topics and estimates their prevalence. The first method automatically extracts the optimal number of clusters, the second makes the researcher specify the desired number of topics. Several models were fitted, varying systematically the number of topics, and similar themes at different levels of granularity tended to emerge. The model with 15 topics was finally chosen to capture key trends and avoid excessive redundancy. The topics were analyzed for their overall corpus prevalence and differences across geographic regions. A lexical correspondence analysis (Benzécri, 1992) was also carried out, to analyze thematic differences between the regions.
Results

Scientific structure

Figure 2 shows the number of articles published in journals associated with a specific field. While 17 academic areas are represented (including journals listed explicitly as “multidisciplinary”), medicine and social sciences took the lion’s share, accounting for 214 papers out of 353 (60.6 percent). A look at the top five most represented fields confirmed that the problem was clearly perceived as relevant in a diverse group of scientific disciplines that also included psychology, humanities, and computer science.

![Figure 2: Number of articles published in journals associated with a specific field.](image)

Note: Larger version of Figure 2 available [here](#).

Within this perspective, it might be surprising that journals that defined themselves as multidisciplinary (e.g., *PLOS One*) collected only 10 articles. However, despite research being mainly published within the boundaries of disciplinary journals, interdisciplinary influences could also be observed through the sources and references shared by the articles. The construction of a network of bibliographic coupling, where an edge connects two papers if they share at least one reference, resulted in a graph with 353 nodes (papers), connected by 18,995 edges, and differentiated into 34 components. These components differed widely in
size, with the largest one ("giant component") counting 282 articles (80 percent of the analyzed corpus) and the second-largest for 38 (11 percent). While this does not mean that 282 articles share the same references, it was possible to trace a trail of shared references across these articles. By looking more in detail at the fields represented by the two largest components, the interconnection emerged between social sciences and medicine, the two more represented disciplines in the corpus, and between other scientific fields. Overall, the analysis suggested a certain degree of dialogue between the scientific areas that addressed information problems in the context of the pandemic.

The interconnection between disciplines follows that between countries. For example, 932 authors wrote papers in the dataset. Although these authors were prevalently affiliated with institutions located in the United States (199 authors), U.K. (68), China (58), and India (56) (Figure 3), there was also a certain degree of international collaboration. We defined international collaboration when the authors contributing to a paper were affiliated with institutions in different countries.

![Figure 3: Distribution of papers across countries.](image)

Note: Larger version of Figure 3 available [here](image).

Most of the disciplines illustrated some degree of international collaboration/cooperation. Therefore, we measured a global collaboration index by field as the number of countries of affiliation of authors over the number of articles in the discipline. Within the three dominant disciplines in our dataset (medicine, social sciences, and psychology), medicine featured a higher level of international collaboration. In contrast, social sciences scored one of the lowest values in the whole dataset. Figure 4 shows the IC index for all
At the onset of an infodemic: Geographic and disciplinary boundaries in researching problematic COVID-19 information

disciplinary areas (the size of the bubble is proportional to the number of papers in the discipline).

<table>
<thead>
<tr>
<th>Social Sciences</th>
<th>Psychology</th>
<th>Pharmacology, Toxicology and Pharmaceutics</th>
<th>Nursing</th>
<th>Multidisciplinary</th>
<th>Medicine</th>
<th>Mathematics</th>
<th>Health Professions</th>
<th>Environmental Science</th>
<th>Engineering</th>
<th>Economics, Econometrics and Finance</th>
<th>Decision Sciences</th>
<th>Computer Science</th>
<th>Business, Management and Accounting</th>
<th>Biochemistry, Genetics and Molecular Biology</th>
<th>Arts and Humanities</th>
<th>Agricultural and Biological Sciences</th>
</tr>
</thead>
</table>

![Graph showing International Collaboration Index per discipline.](image)

**Figure 4:** International Collaboration Index per discipline.

Note: Larger version of Figure 4 available [here](#).

An alternative way to explore this dynamic is illustrated in **Figure 5** — filtered only for countries with at least 10 papers within a discipline. Researchers from the largest majority of countries published their articles in medical journals. Venues not belonging to the broader life sciences area (e.g., social sciences and psychology) were chosen only by researchers from the U.S., U.K., and Spain. This suggests that, while there is a certain level of interdisciplinarity in the approach to the problem, this interdisciplinary approach was not equally shared across the world. There may be disciplinary compartmentalization entrenched in geographical location, which may affect the way the problem was considered from certain parts of the world. This hypothesis was explored by building a country-collaboration network where two countries were connected if two authors affiliated with institutions based in those countries were co-authors of a paper. The network contained 61 nodes (countries) connected by 515 edges. **Figure 6** shows the networks with the nodes colored according to their geographic region (following the WHO organizational division into “regional offices” [6]).
While some specific collaborations between countries were visible (e.g., between the United States and United Kingdom) there was no visible preference for collaborating within geographic regions. This may suggest a collaboration network between research centers geographically located in very different parts of the world. This was also tested by using the value of nominal assortativity (Newman, 2002) which measures the tendency of nodes to connect with other nodes given a specific attribute. The nominal version of assortativity is commonly used to measure the level of homophily existing within network structures. The result (-0.01) suggests a network that is neither assortative nor disassortative for the regional variable and where the regional dimension plays no role. As an additional way to explore the possible collaboration within geographic regions we compared the results of a modularity-based community detection method (Newman, 2006) with the geographic clusters of countries as described above. Community detection methods based on modularity optimization will identify clusters of countries that have an internal level of connectivity (collaboration) higher than might expect from a null model. If the clusters identified through the modularity-based method are very similar to clusters based on geographical proximity this would suggest that cooperation happens within those groups in a higher-than-expected way. We compared the two using normalized mutual information (NMI) (Kvålseth, 2017). The NMI value of 0.1 shows a low similarity.
confirming that cooperation between countries did not take place within the geographical regions. Therefore, the analysis showed no visible tendency to co-operate in similar parts of the world, nor to co-operate with regions in very different geographical areas.
On the other hand, there was a strong and direct correlation between the geographical region of the first author and the region on which the research was focused (Figure 7). Some overlap was also observed between Western Europe (WESTEU, 45 papers out of the 174 stating an explicit geographical focus, corresponding to 26 percent of the mentioned regions, considering that papers can have a comparative and inter-regional focus) and the United States (USA, 41 papers, 24 percent). The Western Pacific Region (WPR, 37 papers, 18 percent) also appeared of more global interest. It includes China, which was identified as the source of the pandemic and thus was of particular interest. Less overlap was observed between the WPR region and the South-East Asian Region (SEAR, 23 papers, 11 percent), the African Region (AFR, 15 papers, 7 percent), and the Region of the Americas (AMR), or between the Eastern Mediterranean Region (EMR, 16 papers, 8 percent) and the Region of the Americas (AMR).
At the onset of an infodemic: Geographic and disciplinary boundaries in researching problematic COVID-19 information

Figure 7: Relationship between the regions at the center of COVID-19 research (rows) and the regions of the first author’s institution (columns). The figure shows how many papers covering a certain region were published by first authors located in the same or other regions. Darker colors represent a higher proportion of papers focusing on a certain region published by first author located in a certain region. The numbers in each cell indicate the absolute number of articles. The analysis was made on cases with both an explicit regional orientation and without missing data regarding the region of the first author.

Note: Larger version of Figure 7 available [here](#).

There were also connections between geographical regions, representing the subject of research, relative to
At the onset of an infodemic: Geographic and disciplinary boundaries in researching problematic COVID-19 information

the attention of different disciplines. When looking at this information in the light of the analysis presented earlier, it should not be surprising that medical research considered a wide range of countries, while social sciences and psychology concentrated their efforts on a lower number of countries. Using the Gini coefficient — a commonly used measure of inequality, ranging from 0 to 1 (the higher the value, the more unequal the distribution) — as a measure of the geographical concentration of research interest, medicine turns out to have, in fact, a Gini equal to 0.376, social sciences equal to 0.430, and psychology to 0.479. The results depended partly on different levels of scientific productivity since a discipline cannot produce just a few papers to analyze all of the world’s countries. Still, even within this limit, the unequal attention to some geographical regions, such as the USA, is evident (Figure 8).

Figure 8: Disciplinary focus on problematic COVID-19 information across regions. The figure shows how many papers from a specific discipline (rows) focus on a particular region (columns). Darker colors represent more papers from a discipline focusing on a specific region. The number in each cell indicates the absolute number of papers.

Note: Larger version of Figure 8 available here.

Methodological approaches

The analysis of methods showed that most of the studies were based on a quantitative approach (36 percent), a considerable part relied on a discursive approach (30.2 percent), and just a minor part was based
At the onset of an infodemic: Geographic and disciplinary boundaries in researching problematic COVID-19 information on a qualitative approach (16.6 percent). A residual number of papers present datasets shared with the scientific community (0.7 percent).

By focusing on the set of papers based on a qualitative or quantitative approach, it can be noticed that the survey was the most popular method (120 papers, 49 percent of the empirical-based papers). Surveys focused more on the general population (75.4 percent) than specific social categories. For example, only a residual percentage analyzed students (9 percent) or health professionals (5.7 percent). They employed convenience samples (77.5 percent) — for instance, collected via online services like Amazon Mechanical Turk (8.3 percent) or by sharing the link to the survey on social media — followed by quota samples (19.2 percent), and probability samples (3.3 percent). Survey-based research that explicitly referenced power analysis — a commonly employed method to calculate sample size (Cohen, 1977) — was only 21.7 percent. Considering the accessible survey-based papers that specified the sample size (113 out of 120, 94 percent), 209,008 interviews were carried out. Averaging the sample size used in the studies (certain studies ran more than one survey), they ranged from a minimum of 51 to a maximum of 40,157 interviewees, with a median value of 673.5 and a mean value of 1,503 ($SD = 3,953$). The second most frequent method was content analysis (20.8 percent), followed by natural language processing (10.2 percent), interviews (7.3 percent), time series analysis (4.9 percent), and social network analysis (3.3 percent). Agent-based modeling (1.2 percent), case studies (1.2 percent), ethnographies (1.2 percent), and round tables (0.8 percent) were used in just a few examples.

There was a disparity in how the methodologies were distributed across papers that focused on different regions. First of all, most of the surveys were conducted in the U.S. and Western Europe. Interestingly, most of the few probability-based surveys were conducted in the African region (three of four), particularly in Ethiopia and Nigeria. Furthermore, only one was led by one European researcher with Nigerian co-authors, while the other two were authored by researchers working in Africa.
In their research design, only 124 papers (32.8 percent) included at least one social media platform. While this might appear surprising, since social media platforms have proven to be deeply intertwined with the spread of problematic information during critical social and political periods (Benkler, et al., 2018), the problem was multifaceted and social media represents just one of its noteworthy aspects. The list of social media platforms was dominated by Twitter and Facebook. Mainstream social media platforms such as Instagram, YouTube, and instant messaging platforms received less attention (Figure 10).
At the onset of an infodemic: Geographic and disciplinary boundaries in researching problematic COVID-19 information

Figure 10: Social media platforms included in the research design of articles on problematic information related to COVID-19 published in 2020 (more than one platform can be used in the same study).

Note: Larger version of Figure 10 available [here](#).

From a regional perspective, the most widely used platforms, such as Facebook, Twitter, and WhatsApp, attracted attention from various geographic regions. However, others were of more regional interest, such as Reddit (U.S. and Western Europe), WeChat, or Weibo (Western Pacific regions, mainly China).
At the onset of an infodemic: Geographic and disciplinary boundaries in researching problematic COVID-19 information

Figure 11: Social media platforms included in the research design across regions. The figure shows how many papers including a certain social media platform (rows) focus on a certain region (columns). Darker colors represent a higher proportion of papers studying a certain platform that focuses on a certain region. The number in each cell indicates the absolute number of papers.

Note: Larger version of Figure 11 available here.

Topics

The analysis of the most frequent terms for problematic information was performed on the title of the papers (176 papers, 47 percent of the corpus, mentioned at least one term in the title). It revealed a few noticeable trends in literature. By considering the whole corpus (353 papers), a preference emerges for widely used words like “misinformation” (27 percent of the analyzed papers mentioning a keyword in the title, and 15 percent of the papers in the corpus) and “fake news” (20 and 11 percent respectively). The second most common keyword was “infodemic” (23 percent of the papers using a keyword in the title and 12 percent of the corpus), confirming the considerable influence of the World Health Organization in setting the research agenda on the topic. The popularity of the term in scientific research skyrocketed in 2020: searching for every mention of the term (including in papers’ references) in the scientific database Scopus, there were no occurrences before 2019, just two entries in 2019, and 792 in 2020.
Conspiracy was the other buzzword of the first year of research, with 18 percent of the keyword-in-title papers (10 percent of the whole corpus). Its relevance was further confirmed by the results of the lexical analysis (Reinert, 1990, 1983), which distinguished three main semantic clusters (Figure 12). The one in the middle of the lower part of the Cartesian plan turns the spotlight on the centrality of conspiracy theories, along with their consequences for compliance with health-protective measures such as social distancing and wearing masks. Among the representative words in this cluster were: “conspiracy theory”, “belief”, “perceive”, “5G”, “behavior”, “distance”, “motivation”, “negatively”, “positively”, “predict”, “compliance”, and “wear”. The other two clusters dealt with the more general context of the health emergency (e.g., “pandemic”, “public”, “health”, “disease”, “spread”, “challenge”, “society”, “crisis”, “global”, “medium”, “information”, and “infodemic”), and research methods (e.g., “survey”, “analysis”, “search”, “participant”, “questionnaire”, “method”, and “collect”).
The attention to conspiracy theories and their impact on adherence to health-protective behaviors was further confirmed by topic modeling analysis (Figure 13), which singled out this theme as the most prevalent in the corpus (Topic 3). Comparatively, other high-frequency topics (e.g., Topic 12 and 11) were
marked by far more blurred boundaries, and their prevalence therefore could be considered as a side effect of their generality. On the contrary, the topic of conspiracy was both specific and widespread. Representative papers are “Looking out for myself: Exploring the relationship between conspiracy mentality, perceived personal risk, and COVID-19 prevention measures” (Marinthe, et al., 2020), “Cultural orientation, power, belief in conspiracy theories, and intentions to reduce the spread of COVID-19” (Biddlestone, Green, and Douglas, 2020), and “Association of COVID-19 misinformation with face mask wearing and social distancing in a nationally representative US sample” (Hornik, et al., 2021), just to cite a few examples.

Concern over conspiracy theories was primarily in the regions of the Americas and Europe and partly in the eastern Mediterranean region (EMR). Considering research stating an explicit geographical focus (N = 174), 50 papers (29 percent) mentioned the term “conspiracy”, of which 56 percent were focused on Western Europe (35 percent of the papers about this region) and the U.S. (37 percent). “Conspiracy” was mentioned in 47 percent of the papers originating in Eastern Europe and 38 percent of those from the Eastern Mediterranean region, demonstrating relative interest in the topic. Still, the weight of these regions was lower, since only 19 percent of the corpus concentrated on these areas, while 44 percent of the corpus of papers stating an explicit geographical focus was about USA and Western European regions. The same was true for the African region (27 percent of articles in that region cited conspiracy theories on COVID-19, but only eight percent of the corpus), where some of the most infamous conspiracy theories were also discussed. This is the case, for instance, with the 5G conspiracy theory in Nigeria (Ndinojuo, 2020).

Consequences of problematic information have been considered about behaviors capable of limiting the contagion, like social distancing and the use of face masks, and attitudes such as vaccine skepticism which obstruct measures capable of stemming the pandemic (Topic 14). Terms characterizing this topic, such as “vaccine” or “hesitancy,” were found mainly in the Eastern Mediterranean Region (19 percent of total papers focusing on this region, which are 16) and Western European corpus (14 percent), but also in the Americas (25 percent) and African one (7 percent). Other regions showed less attention to the topic. Also in this case, most of the contributions came from the scientifically most prolific regions of USA and Western Europe (53 percent of the total contributions mentioning these terms).
Account has been taken of behaviors affecting the availability of goods on the market (Topic 10), investigating phenomena such as panic stockpiling (Naeem, 2020) and the fluctuating availability and prices of medicines. For instance, specific treatments like hydroxychloroquine raised controversies and misinformation flows, which led to shortages and price rises (Godman, et al., 2020). Understandably, this topic was seized upon especially by economists. Research focusing on countries of the South Eastern Asian region, such as Bangladesh (Haque, et al., 2020), Pakistan, Malaysia, and Vietnam (Godman, et al., 2020), showed some interest in the topic (3 papers out of 23, or 13 percent). However, overall, its global weight in the international scientific production was low.

Another strand of research was about drivers behind sharing false information (Topic 9). Representative papers are: “What drives unverified information sharing and cyberchondria during the COVID-19 pandemic?” (Laato, et al., 2020), or “The spread of fake science: Lexical concreteness, proximity, misinformation sharing, and the moderating role of subjective knowledge” (Kirkpatrick, 2021), to mention but a few. A relevant and connected strand of research focused on social media (e.g., Topic 4, 6, 7, and 9), which was considered a major means of enabling an infodemic (Cinelli, et al., 2020). Studies have described the main myths about the pandemic and analyzed the process of misinformation spreading (Gruzd and Mai, 2020), differentiated false claims into separate categories (Islam, et al., 2020), and tackled specific types of health misinformation for economic purposes, like that related to immune-boosting products (Rachul, et al., 2020) and other questionable “cures” (Mackey, et al., 2020). They have tried to assess the prevalence of misinformation (Li, et al. 2020; Prabhakar and Prasad, 2020), and to compare the popularity of scientific and fact-checked content (Pulido, et al., 2020). This research stream, as the focus on social media platforms, was unevenly distributed among countries (see Figure 11).

Psychological and psychiatric consequences of the pandemic and misinformation, often in relation to social media, were also investigated (Topic 8). Some research pointed out the positive role of social media in pandemic communication: celebrities can easily reach many followers with quality information (Mututwa and Matsilele, 2020). While higher levels of social media use for both social support and information seeking were associated with higher levels of anxiety and depression (Drouin, et al., 2020; Gao, et al., 2020; Ni, et al., 2020), social media also acted as an institutional ally to reach people with correct information, becoming protective agents supporting psychological well-being (Liu and Tong, 2020).

Considering interest by geographical areas, most of the attention on the topic focused on the Eastern Europe (32 percent), Eastern Mediterranean Region (31 percent of the paper about this region), Western Pacific Region and Americas region (25 percent each), and South-East Asian Region (18 percent of the papers on this area). No papers in the corpus mentioned these psychological terms in the context of African countries (although some of them dealt with fear associated with the pandemic).

Overall, topics were distributed differently across world regions (Figure 14). The USA and Western European regions (WESTEU) appeared to be closer together and different from the Western Pacific (WPR), Eastern Mediterranean (EMR), and South-East Asian (SEAR) regions. These groups appeared to be more similar concerning topical focus. The African region (AFR) seemed distant from all the other ones. However, it also had only a few contributions; more data will be needed to achieve more robust results.
Discussion and conclusions

This study investigated 378 papers on the COVID-19 information disorder published in 2020, providing an overview of the international scientific community’s response to infodemic issues at the onset of the COVID-19 crisis. This comprehensive corpus of research was analyzed from a tridimensional perspective, taking into account the disciplinary structure of the field and collaboration among scholars from different parts of the world, methods used in research, and main trends in substantive content. In addition, special attention was paid to geographical differences in research between world regions.

The complex and transnational nature of the phenomenon required interdisciplinary cross-national approaches, and the analysis examined the degree to which the early COVID-19 research integrated these aspects. This study demonstrated that many disciplines worked on the problem, particularly medicine and social sciences, confirming a presence of a multidisciplinary effort. A network analysis approach further investigated the degree of interdisciplinary and international connections. That analysis illustrated that papers are linked together beyond disciplinary boundaries by a common set of references, suggesting that disciplinary confines did not strictly bind the scientific horizons of scholars. These findings suggest promising developments in the field. However, assessing interdisciplinarity is complex. For example, the common references connecting the papers could be traced back to disciplinary contributions exerting influence beyond their disciplines, like interdisciplinary sources of general interests or general content such as WHO reports. Therefore, more nuanced analysis will be required to investigate this aspect further.
Different degrees of international collaboration between scholars was also shown, from disciplines that relied strongly on international networks, such as medicine, to more minor internationally structured fields, such as social sciences. This difference could, in some ways, depend on the fact that cultural boundaries inherently circumscribe the object of the social sciences. However, it also emerged that, besides stronger links such as those between the U.S. and U.K., regional factors did not impact the actual structure of international networks. Except for a few cases, international networks were not structured based on a strong and discernible geographical dimension. This analysis, however, suggested that there was a skewed distribution of geographic interest across countries, with a stronger focus on the U.S., although several papers also dealt with China, where the virus first originated. Important disciplines in the dataset, like social sciences and especially medicine, took a more extensive range of countries into account in their analyses. Nevertheless, there was a very strong correlation between research regions and the countries representing authors’ institutions. This meant that regional boundaries often delimited research, despite more or less international networks.

From a methodological standpoint, there was a general preference for quantitative methodologies. This is unsurprising since they traditionally represent one of the most widespread approaches in the social sciences and health-related fields (e.g., epidemiology). Moreover, the restrictions on the movement during the pandemic made methods, like self-administered surveys, more suitable than qualitative approaches, such as interviews or ethnographic observation. Qualitative approaches tend to be more time-consuming and difficult to adopt when time is scarce, like during an emergency. Most survey-based research focused on the United States and European regions. Overall, they also represented the regions most studied by the contributions. Curiously, most probabilistic surveys focused on the African region, the last in terms of the total number of contributions (15 papers, 7 percent).

Discursive papers provided informed analyses, sometimes proposing strategies, recommendations, and possible solutions to help tackle field-specific problems created by the pandemic. Considering the need for expert views in the extraordinary circumstances of the pandemic, it was understandable that this was a significant category in the corpus. Specific methodological observations could also be made. Many surveys exploited social media or online services. One was Amazon Mechanical Turk for collecting data, an approach that met the need to gather evidence as soon as possible in the middle of the pandemic. Still, the self-selected samples collected in this way risked being unrepresentative and potentially biased. Just 21.7 percent of the survey-based papers referred to power analysis, although the samples used in the studies were usually not too small (1,503 cases on average, with a median of 673.5). The development of the field would undoubtedly profit from a systematic meta-analysis, capable of synthesizing this wealth of information and the variety of results. Up-to-date meta-analysis would also be helpful to accelerate research in times of crisis. To facilitate this kind of synthesis in times of emergency, a structured approach to international collaboration may be beneficial. An example is a database hosted on the Web site of WHO. Since March 2022, this repository has collected more than 500,000 records of the most recent international multilingual research on COVID-19 ([https://covid19.who.int](https://covid19.who.int)).

Conspiracy theories and their impact on compliance with health-protective behaviors emerged as the most significant trend in the first year of studying problematic information around COVID-19. Regional analysis noted that this topic was discussed worldwide but was especially influential in Western Europe and the U.S. Research also considered the problem from the perspective of its consequences on other spheres of social and individual life, such as psychological well-being or market dynamics. Besides consequences, motives for sharing false or misleading information, analyses of the processes of misinformation spreading online, estimation of the overall prevalence of the problem, typologies of deceptive content, and a wide range of possible solutions were also proposed. Overall, the textual analysis showed several differences in the focus on different topics in different world areas. In particular, Western Europe and U.S. were more similar in their substantive interest and separate from areas of the Western Pacific Region (WPR), Eastern Mediterranean Region (EMR), and South-East Asian Region (SEAR). The African region (AFR) appeared to be distant from all others.

Besides conspiracy theories, a noticeable trend was the so-called infodemic. The term ranked second among
the most popular keywords in the titles of papers on the topic. Despite being coined several years ago, it has skyrocketed in academic research only in 2020, after the WHO defined COVID-19 as a pandemic and an infodemic. The term’s popularity points to the effectiveness of the World Health Organization in setting a research agenda. Terms employed to define a situation often carry assumptions and frames, such as an epidemiological paradigm in the case of infodemic. Scholars might find this approach more or less stimulating and valuable, and a thorough analysis of terms’ uses and related theoretical assumptions could be helpful to reflect critically on the direction given to the field (Simon and Camargo, 2021).

The analysis presented here allows researchers to critically consider several aspects of the academic construction of discourse on problematic information about COVID-19 (Neumayer and Rossi, 2016), taking into account the often-overlooked geographical dimension. Besides, it suggests some underexplored areas by contrast with the most significant trends, which have also been shown. To mention a few, qualitative studies could receive more attention since they are essential to exploring and defining new theoretical and empirical dimensions of the phenomenon. Meta-analysis would be key to evaluating the robustness of the results of different surveys. Comparative analyses could clarify regional factors that influence information issues. Finally, a transnational and more balanced focus on the world’s regions could shed some light on the global trajectories that characterize the spread of misinformation and disinformation.

This study also has limitations. Determining interdisciplinarity and the nature of international cooperation using computational methods is not an easy task. A more systemic approach, combining various dimensions of research (methodological, substantive, disciplinary, etc.) in a global synthesis, could improve our understanding without losing a panoramic view of the field. Geographical analysis could also be enhanced by considering the relationship between the biographies of researchers and their collaborative networks, and the geographical focus of their research. Finally, although this study focuses on a crucial period of the COVID-19 infodemic, the first year of the pandemic in 2020, a comparative analysis with research conducted on the topic after 2020 would be necessary to follow the evolution of the research field and provide further clues on its development.

**About the authors**

*Nicola Righetti* is a researcher in computational communication science in the Department of Communication at the University of Vienna and research associate at the Computational Communication Lab of the University of Vienna.  
E-mail: nicola [dot] righetti [at] univie [dot] ac [dot] at

*Luca Rossi* is an associate professor in the Department of Digital Design at IT University of Copenhagen.  
E-mail: lucr [at] itu [dot] dk

*Giada Marino* is a postdoctoral research fellow in the Department of Economics and Business at the University of Sassari, where she is part of the research team of I-POLHYS, studying citizens and polarization.  
E-mail: gmarino [at] uniss [dot] it

**Acknowledgements**

We thank *First Monday*’s anonymous reviewers for their careful reading of our manuscript and their many constructive and insightful comments and suggestions.

**Notes**
At the onset of an infodemic: Geographic and disciplinary boundaries in researching problematic COVID-19 information


2. https://www.who.int/health-topics/infodemic#tab=tab_1.

3. COVID OR COVID-19 OR coronavirus AND misinformation OR disinformation OR malinformation OR “problematic information” OR “fake news” OR “information disorder” OR infodemic OR conspiracy OR rumor OR hoax.


6. https://www.who.int/about/who-we-are/regional-offices.

References


At the onset of an infodemic: Geographic and disciplinary boundaries in researching problematic COVID-19 information


At the onset of an infodemic: Geographic and disciplinary boundaries in researching problematic COVID-19 information


doi: https://doi.org/10.2196/22142, accessed 11 June 2022.


At the onset of an infodemic: Geographic and disciplinary boundaries in researching problematic COVID-19 information


doi: https://doi.org/10.2196/19009, accessed 11 June 2022.


At the onset of an infodemic: Geographic and disciplinary boundaries in researching problematic COVID-19 information


At the onset of an infodemic: Geographic and disciplinary boundaries in researching problematic COVID-19 information


Editorial history

Received 18 March 2022; revised 29 April 2022; revised 3 June 2022; accepted 3 June 2022.

This paper is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

At the onset of an infodemic: Geographic and disciplinary boundaries in researching problematic COVID-19 information
by Nicola Righetti, Luca Rossi, and Giada Marino.
First Monday, volume 27, number 7 (July 2022).
doi: https://dx.doi.org/10.5210/fm.v27i7.12557