

On the global health science response to COVID-19

Moritz Müller · Kevin Wirtz · Pierre Pelletier ·
Stefano Bianchini

Received: date / Accepted: date

Abstract How has the global health science system reacted to the COVID-19 pandemic? Here we investigate how — over the first year of the pandemic — national output and international collaboration on coronavirus related research (CRR) correlates with prior activity in the health sciences, pandemic related factors, and the broader socio-economic context. We find that prior CRR experience is influential in (inter-)national CRR in particular in the first two months of the pandemic. Subsequently, more general health science capacity becomes the dominating factor of CRR output. National COVID-19 incidences, national confinement measures, and broader socio-economic conditions turn out to be only weakly correlated with (inter-)national CRR. Consequently, the rapid expansion of global CRR followed mostly the structure laid out by the global health science system. However, the international CRR network experienced a significant decrease in hierarchy accompanied by increasing collaboration within pre-established regional science communities.

Keywords COVID-19 · Scientific Networks · International Collaboration · Health Sciences

1 Introduction

The paper at hand treats the outbreak of the novel coronavirus Sars-CoV-2 in January 2020 as an exogenous shock to the international health science system. The subsequent COVID-19 pandemic has been global in the sense that, within relatively short time, most countries worldwide have been directly concerned. Consequently, coronavirus related research (CRR) became a top priority for health science on national and international levels worldwide.

The burst of CRR across countries and the associated emergence of a large international CRR network have been documented since the early phases of the pandemic (Aviv-Reuven and Rosenfeld,

M. Müller
BETA, Université de Strasbourg, France.
E-mail: mueller@unistra.fr

K. Wirtz
BETA, Université de Strasbourg, France.

P. Pelletier
BETA, Université de Strasbourg, France.

S. Bianchini
BETA, Université de Strasbourg, France.

2020; Cai et al., 2020; Chahrour et al., 2020; Fry et al., 2020; Haghani and Bliemer, 2020; Radanliev et al., 2020; Zhang et al., 2020). In the rankings of national scientific output as well as network centrality, the usual suspects tend to score high; the US takes the lead, followed by China and the most developed countries. Most scholars would probably agree that the global distribution of (international) health science capacity given at the time of crises is the obvious explanation. Yet, quantitative evidence is lacking because most studies focus exclusively on CRR. Therefore, extant research captures system dynamics mostly by considering pre-pandemic CRR as initial conditions from which the CRR network strides away during the pandemic as it expands. The existing broader health science system in which the international CRR network expands has not been explicitly accounted for as a potential attractor. So far, the mirroring of global CRR and health science capacity is merely hypothesized. Further note that the (hypothesized) mirroring is unlikely to be perfect. The global science system is shaped by an interwoven web of processes playing inside and outside the science system, and the COVID-19 pandemic shifted some of the relevant parameters.

This paper investigates the dynamics of global CRR during the COVID-19 pandemic within the global health science system, taking into account initial socio-economic conditions as well as pandemic related factors. We conceive the global science system as a network of interconnected national science systems. The analysis proceeds in three steps. In a first step, we model national CRR conditional on initial conditions and pandemic development at the national level. The second analysis models to bi-national collaboration on CRR. A third analysis looks at the convergence of global CRR towards global health science in terms of national scientific output, network centrality of countries, and the overall network structure. Thus, whereas analyses one and two reveal driving factors of national and bi-national CRR respectively, analysis three deals with global CRR dynamics at the macro level.

This analysis aims at a better understanding of the dynamics of the global science system, with a focus on internal science dynamics triggered by the external COVID-19 shock. How the global science system responds to external shocks is certainly relevant given the global social and environmental challenges ahead (Schot and Steinmueller, 2018). Knowledge about the patterns of response to global crises may guide organization of science in the future. For one, scientific capacity needs to be build on national and international levels that facilitates changes in the direction of scientific development. Another issue is how the existing scientific capital can be efficiently orchestrated during crises. Our paper does not provide definite answers to these questions. However, we hope that our empirical description contributes to the rationalization of global science dynamics during the pandemic.

Our empirical analyses show that the pre-pandemic, broader health science system is the dominating factor in the dynamics of global CRR during the pandemic. Other factors that are currently debated, and we explicitly account for in our analysis, have more limited influence. Consistently, at the macro level, we find that global CRR rapidly converged towards the global structure of the broader health sciences. But convergence is not perfect. In the first two months, network hierarchy overshoots such that it significantly exceeds the ‘typical’ levels of global health science. Subsequently, network hierarchy decreases significantly below the hierarchy of global health whereas collaboration activity within (world) regions increases significantly. Thus, the (science) world responds to the global crises with increased regionalization.

For science policy our findings provide an empirical argument for a strategy that emphasizes generic scientific capability and a broad research spectrum. Furthermore, active participation in global knowledge flows during ‘normal’ times will be key for handling the global crises in the future. How exactly global science efforts should be orchestrated remains an open issue though. Future research could investigate to what extent our observations on the health system dynamics — global concentration followed by global diffusion and regional interaction — constitute an efficient response to global crises. Another question raised is whether the observed regionalization is due to (transient) pandemic circumstances or accentuates a long-term trend in the sciences and other socio-economic spheres.

2 Background

2.1 Overview

This section starts with a birds-eye-view on the national context of science systems and (dynamics of) the international science system. The second part turns to the global health science system where our focus is on how the COVID-19 pandemic affects determinants of international research collaboration (IRC).

2.2 The ‘national’ in global science

Science does not exist in isolation. The global science system evolves within the overall social and economic system with strong mutual feedbacks. On the one hand, science contributes to the global pool of knowledge that drives social and economic development (see e.g. Kuznets, 1960). On the other hand, broader society strongly conditions scientific activity.

The national context is particularly influential (Ben-David, 1971). In Europe from around 1800 on recognition of the utility of science by national governments and industry led to increased support and autonomy of the scientific community (Beaver and Rosen, 1978). The establishment of formal and informal institutions has been mostly a national effort, with France leading and serving as a role model for Germany and Great Britain (Beaver and Rosen, 1978). Also the expansion of science in the late 19th century and first half of the 20th century has been born out of an explicitly national rationale. Building on arguments of militaristic and economic competition among nations, and national reputation, governments of advanced economies expanded and shaped purposefully their national science systems. This gave rise notably to the foundation of today’s national research institutions in advanced economies such as the CNRS in France, the Kaiser-Wilhelm-Gesellschaft (succeeded by Max-Planck-Gesellschaft after the Second World War) in Germany, or NSF in the US (Ben-David, 1971).¹ In the second half of the 20th century, the contribution of science to (national) economic growth has become a major justification for national investments in higher education and scientific research (Pavitt, 1991). It still is today.

To this end, national governments set strategic targets for Gross Domestic Expenditure on R&D (GERD) of around three percent. Government funding of science specifically may be proxied best by research expenditures of higher education institutions, i.e. higher education R&D expenditure (HERD) Stephan (2010).² In the past 20 years, HERD in OECD countries has been between 0.2 to 0.6 percent of GDP, fairly stable within countries and across, with an average of 0.4 percent. For non-OECD countries, HERD is systematically lower, around 0.2 percent of GDP (OECD, 2019).³

Public science expenditures also respond to national and global events. In the USA, for example, the Sputnik shock in the late 1950s led to a considerable increase in government expenditures, and

¹ President’s Roosevelt’s letter to the director of the ‘Office of Scientific Research and Development’ in 1945, initiating the foundation of the National Science Board in the US, is a good example on that point: “*DEAR DR. BUSH: The Office of Scientific Research and Development, of which you are the Director, represents a unique experiment of team-work and cooperation in coordinating scientific research and in applying existing scientific knowledge to the solution of the technical problems paramount in war. [...] its tangible results can be found in the communiques coming in from the battlefronts all over the world. [...] There is, however, no reason why the lessons to be found in this experiment cannot be profitably employed in times of peace. The information, the techniques, and the research experience developed by the Office of Scientific Research and Development and by the thousands of scientists in the universities and in private industry, should be used in the days of peace ahead for the improvement of the national health, the creation of new enterprises bringing new jobs, and the betterment of the national standard of living.*” https://www.nsf.gov/about/history/nsf50/vbush1945_roosevelt_letter.jsp

² Note that not all government funded science is performed in the higher education system, not all research in higher education is science, and government is the main but not the only funding source of university research HERD.

³ Supra-national entities such as the EU also define science targets, but funding remains national to a large extent.

during the Vietnam War (relative) expenditures decreased again (Stephan, 2010). In 1998, the National Institute of Health (NIH) doubled its funding to strategically support the high-growth of the biotechnology industry (Zucker et al., 1998). The American Recovery and Reinvestment Act of 2009 increased considerably public science expenditures as a countercyclical measure (Stephan, 2010). Most recently, the US congress discussed the proposition to increase the NSF budget by 100 billion US\$ — the largest increase in the agencies history — with the explicit goal to maintain global innovation leadership as a response to China’s national efforts (and great success) in science (Mervis, 2021; Rimmel, 2021). Somewhat ironically, an integral part of China’s science strategy in the past 30 years has been scientific collaboration — bolstered through large-scale student and scientist exchange programs — notably with the US but also Europe (Wang and Wang, 2017). In parallel, European wide research programs, notably Horizon 2020, now Horizon Europe (95 billion Euro), are relevant for national science systems and support their integration on a regional level. Similar tendencies can be observed in practically all world regions.

The phenomenon of increasing international research collaboration (IRC) is one aspect of globalization in science. It accelerated in the early 1980s (Adams, 2013), and may have reached a saturation level in some more advanced economies by now (Ponds, 2009). International collaboration is observed in particular among productive researchers from top-tier universities located in advanced national scientific systems (Pan et al., 2012; Jones et al., 2008). The gain is (more) excellent research (Adams, 2013; Pan et al., 2012). The tendency of ‘excellence-attracting-excellence’, however, entails the risk of increasing hierarchical stratification not only within but also between national science systems (Beaver and Rosen, 1979; Jones et al., 2008; Horlings and van den Besseelaar, 2013). In order to catch-up scientifically, or at least not to fall behind, being well connected to the global knowledge flows has become a science policy imperative in most countries.

Considering national and international scientific activity jointly as a global science system is therefore warranted.

2.3 Structure and processes in global science

The global science system is characterized by two salient features: First, worldwide scientific activity is highly concentrated at some places. In other words, the global science system is highly hierarchical. Second, countries exhibit a certain ‘preference’ to collaborate with certain other countries. Such national tendencies in IRC create clusters of scientific activity. We discuss each in the following.

The uneven distribution of scientific capacity in the world has been discussed early on in the literature. Frame et al. (1977) counted country affiliations of scientific articles published in an early 1973 ISI collection, and found that the concentration of science production exceeds economic concentration in the world. Subsequent studies repeated and varied the exercise. The extent of scientific concentration is immediately grasped by looking at the global share of the most prolific countries; shown in Table 1 for four time periods (based on Frame et al., 1977; May, 1997; King, 2004; Allik et al., 2020). In the period from 1970 to 2000, the ten most productive countries contribute around 80 percent to worldwide scientific output (‘Total share’ in Table 1). In that period, we see the fall of the USSR in the 1980s and 1990s in the data.⁴ Somewhat less dramatic, but still noticeable, is the (scientific) rise and fall of Japan. Besides movements of individual countries, worldwide concentration remained rather stable throughout the fourth quarter of the last century.

In the most recent period, i.e. 2008 to 2018 in the last column of Table 1, the global share of top ten countries dropped to around 60 percent. The lower share may be partly explained by the larger sample of journals which includes more non-English and less established (or mainstream) journals. However, the drop reflects a significant real-world development. In the last twenty years, in particular Eastern European and Asian emerging economies, notably China, experienced high

⁴ May (1997) excluded USSR (and Russia) due to issues in assigning research output.

Table 1 World’s largest science countries.

	1973 (Frame et al., 1977)	1981-1994 (May, 1997)	1997-2001 (King, 2004)	2008-2018 (Allik, 2020)
1	US (38.2)	US (34.6)	US (34.6)	US (19.9)
2	UK (9.2)	UK (8.0)	UK (8.5)	China (11.7)
3	USSR (9.0)	Japan (7.3)	Japan (8.0)	UK (6.0)
4	West Germany (6.0)	Germany (7.0)	Germany (7.4)	Germany (5.3)
5	France (5.5)	France (5.2)	France (5.6)	Japan (4.1)
6	Japan (5.2)	Canada (4.5)	Canada (4.6)	France (3.7)
7	Canada (4.4)	Italy (2.7)	Italy (3.3)	Canada (3.3)
8	India (2.5)	India (2.4)	Russia (3.3)	Italy (3.2)
9	Australia (1.9)	Australia (2.1)	India (2.8)	India (2.8)
10	Italy (1.7)	Netherlands (2.0)	Netherlands (2.3)	Australia (2.8)
Total share	84%	76%	80%	63%

Frame et al. (1977) covers 2,300 journals indexed by SCI, May (1997) covers 4,000 journals indexed by ISI, King (2004) covers 8,000 journals indexed by ISI, Alik (2020) covers 12,000 journals indexed by ESI.

growth rates that have been much higher than growth rates in the (economic and scientifically) more advanced countries (Horlings and van den Besselaar, 2013). This results in lower concentration. On the other hand, less developed countries had very little or no growth during that period, which increases the global divide (Horlings and van den Besselaar, 2013). In sum, advanced and developing economies became more equal in the past two decades, but not the world as a whole.

Patterns in IRC are coupled to the size and dynamics of national science systems, and are tied to the same geo-political and economic processes. For example the fall of the USSR has been associated with a significant west orientation in international scientific collaboration of the former east-block countries between 1985 and 1995 (Braun and Glänzel, 1996). And China’s growth of scientific output has been coupled with an increase of internationally co-authored papers (Niu and Qiu, 2014). More accurately, China’s IRC intensity, i.e. internationally co-authored papers over total papers, remained stable at around one fourth (Niu and Qiu, 2014). The same pattern of proportional growth of IRC and total scientific output has been observed also for other (larger) emerging economies (China, India, South Korea, Brazil) over the growth period 1980 to 2010 by Adams (2013). During the same period, advanced economies have grown much more through internationally co-authored papers rather than through domestic papers, implying an increasing IRC intensity over time. Another observation is that IRC intensity decreases with the overall size of the (advanced) national science system. In other words, the absolute number of internationally co-authored papers tends to increase with the total number of papers with an elasticity below one (Frame and Carpenter, 1979; Luukkonen et al., 1992; Pan et al., 2012). Thus, larger (science) nations tend to have more international collaborations in total, but less international collaborations per paper than smaller (science) nations.

National science output together with variations in national IRC intensity translate into a natural order in the hierarchy of the worldwide IRC network. Gui et al. (2018) investigate hierarchy in the IRC network by looking at the countries’ network centrality, i.e. degree, closeness, betweenness, and eigenvector centrality.⁵ The different centrality measures tend to be all highly correlated, which is typical for real-world networks.

We note that the ranking by network centrality corresponds largely to the ranking by countries’ science production as e.g. seen in Table 1. This is due to the coupling of national size and the amount of international collaborations. In the few cases where rankings disagree, national IRC intensity is the obvious explanation. Japan is less central in the IRC network than expected given

⁵ The study is based on papers in the Web of Science Core Citation Database in the years 2000 and 2015.

overall science production. The reason is Japan’s relatively low IRC intensity, which in turn may be explained by geography, culture, and language (Yonezawa, 2011). Also Russia tends to be less central in the IRC network than in terms of national science production. On the other hand, small and advanced western countries, in particular the Netherlands and Switzerland, have high IRC intensity and are consequently among the ten most central countries in the network. The question of partner choice, i.e. national tendencies to collaborate with certain other countries, is most likely a secondary effect that is dominated by overall collaboration activity in the creation of the hierarchy.⁶

The hierarchical structure of the IRC network suggests that positive feedback drives the local accumulation of science capacity. Such rich-get-richer, or accumulating advantage effects, are central in economic and sociological theories. The core-periphery theory introduced by Friedmann (1967) describes positive feedbacks in the structuring of (international) economic and political relations. This theory puts forward, among others, that the system’s core is marked by a high density of creative potential leading to high creative interaction compared to the periphery, which attracts further inflow of creative potential from the periphery. Observed migration flows of scientists are in line with that theory (Stephan and Levin, 2001), and the literature on international migration of scientists provides a close connection to international research collaboration that is arguably causal (Jonkers and Tijssen, 2008; Kato and Ando, 2017). Accumulating-advantage in science through social effects of scientific collaboration have been emphasized by Beaver and Rosen (1979). They noted that scientific collaboration is not only about resolving inter-dependence of physical or cognitive resources, but also instrumental for gaining recognition, reputation, and status. As a result, differential access to resources drives international scientific status, and vice versa. The same logic goes through in particular for IRC (Melkers and Kiopa, 2010), susceptible of creating a rich-get-richer phenomenon (Hâncean et al., 2021; Katz, 2000; Wagner and Leydesdorff, 2005; Wagner et al., 2019).

Once country size effects are accounted for, national preferences in IRC partner choice are clearly revealed. Collaboration preferences are shaped along multiple dimensions that are often not purposefully created at all (e.g. geographical distance), somehow inherited (e.g. culture and language, common cognitive basis, and historical connections from colonial times or past migration flows), or subject to (inter-)national policy (e.g. efforts to structure the European Research Area, exchange agreements in higher education, migration policies to attract scientists, joint research mega-projects such as CERN) (Frame and Carpenter, 1979; Luukkonen et al., 1992; Zitt et al., 2000; Frenken et al., 2009, and the literature cited therein). Most of these factors can be conveniently couched in terms of some kind of (dyadic) distance (Frenken et al., 2009). Hence, countries that are close in (some dimensions of) space tend to form (mostly regional) science clusters (Wagner and Leydesdorff, 2005).

The role of the evolving technological regime, notably the ICT revolution, for IRC is still somewhat ambiguous in the empirical literature (Gui et al., 2018, and the literature cited therein). ICT developments facilitate fast and massive information exchange. This has probably contributed to the rise of international scientific collaboration since 1980. However, ICT impacts are not likely to be uniform because developed countries have still preferential access to ICT, and ICT may help more to overcome geographical than other kinds of distances.

Let us note the key ideas on the global science system laid out so far. First, national scientific activity is highly concentrated in advanced economies and (some) emerging economies. And the international research collaboration network exhibits essentially the same hierarchy. Countries have

⁶ One argument is the strong alignment between network centrality and IRC activity (number of collaborations). Another argument is that the different network measures applied by Gui et al. (2018) produce essentially the same ordering. To see the argument, consider India as the exception where partner choice makes a difference. India has become a regional science leader in 2015. As a hub in the (semi-)periphery, India has high closeness centrality (well connected to all other countries worldwide), but low betweenness and eigenvector centrality (many knowledge flows in the network sidestep India). The fact that for most countries these measures closely align, indicates that differences in partner choice (which are arguably there, see below) do not heavily influence the hierarchy dimension of the network position.

a certain propensity to collaborate depending on intra- and extra-science factors, creating regional ‘science clusters’. Dynamics in terms of national scientific activity and network density and structure tend to be on a time scale of decades, which can be explained by the stability of underlying factors but also the process of developing (international) science capacity characterized by positive feedback effects.

2.4 COVID-19 and global health sciences

Global *health* science has essentially the same structure, and is subject to the same processes, as global sciences (Cantner and Rake, 2014; Wagner et al., 2017; Gazni et al., 2011). Some specificities are to be noted though. First, health sciences include a variety of scientific fields, and IRC intensity varies substantially across these fields: Most international is Molecular Biology and Genetics with 25 percent of internationally co-authored publications, while in Clinical Medicine IRC is least common with an IRC share of around 15 percent (Gazni et al., 2011). Second, countries differ in their scientific profiles. In terms of investments, while public GERD expenditures are in general highly correlated with GDP, richer countries tend to spend a higher share of their GERD on health sciences.⁷ Looking at scientific publications — some exceptions allowed — eastern countries (former USSR and Asia) tend to focus more on engineering and technologies, while western countries (USA, west Europe) tend to be particularly strong in health sciences (Glänzel, 2001). Third, health sciences are particularly embedded in the national social and economic system. Medical practice is carried out within the broader health infrastructure (e.g. hospitals). And health science is highly connected to the research intensive pharmaceutical industry (Zucker et al., 1998). All this suggests that global health sciences may exhibit a pronounced (hierarchical) structure and strong inertia in (inter-)national development.

Prior epidemics caused by a coronavirus (MERS and SARS) led to the formation of CRR specific national capacity and IRC network structures (Haghani and Bliemer, 2020; Mendes and Carvalho, 2020; Zhang et al., 2020). These epidemics took place mostly in the developing world and have been regionally confined. Countries that have been directly concerned had a strong incentive to increase relevant knowledge production and to shape research priorities and the research agenda. One way forward has been research collaboration among concerned country as well as developed countries to which strong ties existed before (mostly due to historical, originally non-scientific relationships) or who had strong competences in a given field (Zhang et al., 2020; Haghani and Bliemer, 2020). This resulted into regional CRR networks in the Middle East and Asia connect to advanced economies. In particular China (after SARS in 2002), Saudi Arabia (after MERS epidemic in 2012), and some developed countries (US, UK, Germany and Netherlands involved in both) built CRR specific competences (Mendes and Carvalho, 2020; Zhang et al., 2020). These regional efforts have been continued after respective epidemics and resulted in regional specialization patterns. Notably Saudi Arabia, particularly affected by the MERS crisis in 2012, had still a strong focus on CRR in 2019; contributing 6% of CRR compared to 0.6% of non-CRR research output in our sample (Section 3.2).

CRR before the COVID-19 pandemic naturally yields CRR relevant scientific capacity. This footprint may influence the dynamics of international CRR during the COVID-19 pandemic. However, health research on any pandemic necessarily deals with a vast array of research fields and topics (Zhang et al., 2020). Countries therefore may still differ in their relative research focus (e.g. public health in the US and biochemistry in China) (Zhang et al., 2020). Hence, many countries that did not research specifically the coronavirus before the pandemic nevertheless will have CRR capacity (Lee et al., 2020). This suggests a strong alignment of global CRR during the global

⁷ Own calculation based on the ‘Research and Development’ dataset of UNESCO Institute of Statistics (UIS), release date March 2021.

COVID-19 pandemic with the existing scientific, technological and human capital in the broader international health sciences.

The global spreading of the coronavirus implies that observed cases and, hence, national research needs and opportunities varied over time. The spreading of the pandemic is well documented by the Johns Hopkins Coronavirus Resource Center (CRC) (Dong et al., 2020). The origin of the pandemic has been in China where the number of cases strongly increased until mid February. The first cases in neighboring Asian countries have been confirmed mid January, closely followed by first confirmations in the US, Europe and Australia end of January. In Europe, Italy has been particularly concerned by the end of February 2020. At around the same time, the virus has been confirmed in the Middle East and North Africa. By end of March 2020 the virus has been confirmed all around the world with varying intensity from then on. The global infection process provides a compelling narrative for explaining certain aspects of the observed dynamics, in particular the early centrality of USA, China, and later Italy in CRR research (as put forward in Fry et al., 2020).

Governments started a first wave of research programs on COVID-19 from March 2020. A survey of the OECD provides some information in that direction (EC-OECD, 2021). However, the additional investment into health sciences due to these programs is difficult to tell. Budget data is missing in 60 percent of the cases, and it is not clear to what extent pre-pandemic budget plans have been rather redirected or even re-labeled. Thus, this data gives some idea on national research efforts but is too noisy to be included in a systematic analysis. There are essentially three arguments though why (early) Covid-19 specific national funding may have no strong effect on the overall pattern of international CRR. First, CRR specific funding is likely to be strongly aligned with national funding of health sciences in general. Second, national investments can not, in the short-run, scale up the national health science because the formation of scientific, technological and human capital takes years or rather decades. Third, there is anecdotal evidence that national science agencies have been rather flexible regarding redirection and prolongation of funds during the pandemic. Thus, the redirection of already granted funds has been probably the most important source of funding for (early) CRR.

Existing bibliometric studies on the impact of the COVID-19 pandemic on global health science tend to focus only on coronavirus related research. Several empirical studies show that — within the first three months of the pandemic — global scientific output increased significantly through a highly uneven contribution of individual countries; often framing it as a scientific race (Aviv-Reuven and Rosenfeld, 2020; Chahrour et al., 2020; Haghani and Bliemer, 2020; Radanliev et al., 2020; Zhang et al., 2020). A closer look at the content of CRR papers shows remarkable differences across countries in terms of disciplinary focus. Zhang et al. (2020) shows for the ten most prolific countries that they contribute in the scientific field in which they ‘specialized’. Guleid et al. (2021) show for African countries that only one percent of CRR focuses on therapeutics or vaccines, while one fourth of publications assess countries’ preparedness and response to the pandemic, and another twenty percent describe indirect health impacts of the pandemic. Empirical methodology of CRR is mostly based on prospective observational studies and case series (clinical series), whereas non-CRR papers are mainly composed of randomized controlled trials; at least in the top medical journals (based on 402 papers published in 2019-2020, Gai et al., 2021).

International scientific collaboration and cooperation to address the global pandemic has been emphasized immediately after the outbreak of the global COVID-19 pandemic in January 2020 by the scientific community, policy and the public at large. The WHO actively coordinated international efforts, and various consortia leveraged their international networks for CRR (Kinsella et al., 2020).

Timely and widespread diffusion of relevant research results has been one pillar in IRC. The public release of the 2019-nCoV viral sequence by Lu et al. (2020), a Chinese research group, in January 2020 is one example. The upsurge of preprint papers on COVID-19 related papers marks the shift to (more) open science by the scientific community as a whole (Aviv-Reuven and Rosenfeld, 2020; Homolak et al., 2020). Due to travel restrictions, the health science community met at large

digital conferences. Also publishers contributed to fast dissemination of research by considerably shortening the time from submission to publication; something already seen in prior pandemics (Aviv-Reuven and Rosenfeld, 2020; Palayew et al., 2020).

The other pillar has been joint research on an international level. Scientific breakthroughs on CRR, as documented in highly cited scientific papers, have been often achieved by international teams (Aviv-Reuven and Rosenfeld, 2020). Travel restrictions made physical exchange more difficult, and perhaps the need to reduce transaction costs due to urgency led to lower team size in IRC on CRR compared to pre-pandemic levels (Cai et al., 2020; Fry et al., 2020; Lee and Haupt, 2020). Aviv-Reuven and Rosenfeld (2020) compare coronavirus related research papers with other health science papers published during the pandemic, finding that coronavirus research teams tend to be somewhat smaller and less international. Despite that, international research teams account for around one third of the overall publication output on coronavirus related research (Aviv-Reuven and Rosenfeld, 2020; Cai et al., 2020; Lee and Haupt, 2020). That share remained fairly stable during exponential upscaling of coronavirus related research in 2020, despite travel restrictions and other impediments to international collaboration during the pandemic (Cai et al., 2020). This has been made possible by a high influx of scientists new to coronavirus related research. Around 80 percent of all co-authors on relevant papers did not cooperate before on CRR (but potentially on related topics) (Liu et al., 2021).

Zhang et al. (2020); Haghani and Bliemer (2020); Fry et al. (2020) provide an account — up to April 2020 — of the transformation of the IRC network on CRR during the global Covid-19 pandemic. A pertinent finding is that the network expanded rapidly (within the first two to three months) worldwide. Another common finding is that countries are extremely heterogeneous not only in terms of individual science production but also their network centrality. Fry et al. (2020); Cai et al. (2020) note that the IRC network on CRR has become more ‘elitist’ with the pandemic. The central role of two countries, USA and China, and their strong IRC interaction has been emphasized (e.g Fry et al., 2020). Interestingly, clinical medicine has been a long-standing core subject of US-Chinese IRC (shown for the period 2000 to 2010 by Niu and Qiu, 2014). Cai et al. (2020) document for the second half of 2020 network position dynamics of (some) central countries, and in particular a relative weakening in the USA-China interaction on CRR.

In sum, previous empirical research on the development of international CRR during the COVID-19 pandemic mostly focused on CRR in isolation. Contextual factors put forward are rarely accounted for in systematic empirical analyses of (inter-)national CRR dynamics. Ideas on the role of CRR capacity built prior the pandemic, global virus spreading dynamics, travel restrictions, and ICT solutions for science dissemination are clearly part of the discussion. The relevance of the global health science system — in which CRR is embedded — is clearly visible across all studies, but somewhat remains ‘the elephant in the room’. Therefore we ask, how does the structure of global health sciences relate to the scientific response to the pandemic?

3 Data and methods

3.1 Overview

The empirical analysis consists of three interrelated parts. The first and the second analysis investigate factors driving national CRR output and international collaboration on CRR respectively. The third analysis investigates how the development of national and international CRR relates to the worldwide distribution of health science. This section continues with a description of the data before we turn to the empirical methods used in the analyses.

3.2 Data

We measure scientific output and international research collaboration on scientific articles.⁸ Our main dataset is a collection of peer-reviewed articles in journals indexed by MEDLINE. The restriction to MEDLINE indexed journals ensures that papers in the sample fall into our scope of biomedical research and are of (minimum) scientific quality. We collect papers from the pre-COVID-19 period (Jan.–Dec.2019) and from the COVID-19 period (Jan.–Dec.2020). Furthermore, we distinguish CRR papers from non-CRR papers. The analysis is based on the papers' submission dates to stay close to the actual research activity.

The dataset has been constructed as follows. We downloaded papers appearing in MEDLINE journals from the PubMed database as of June 2021. Coronavirus related papers are identified through a text search query suggested by PubMed Central Europe on the papers' title, abstract, and MESH terms.⁹ Countries are identified in paper affiliations through regular expressions and subsequent manual cleaning.

In order to assess the data, one has to keep in mind the existence of several timelags. First, there is an (unknown) timelag from research to submission. Then there is a timelag from submission to acceptance by the journal. And finally a timelag from journal acceptance to entry on PubMed. For the papers in our dataset timelags from submission to acceptance to entry in PubMed are known. Appendix A provides respective distributions for CRR and non-CRR papers. Looking at the timelag from journal acceptance to PubMed entry, some underreporting becomes likely from October 2020 on. From January 2021 on, data cut offs become clearly visible in our sample. Therefore, our main analysis is on 2020, the first year of the pandemic. Appendix D provides an outlook up to May 2021.

Our bibliometric sample (2019 and 2020) consists of 1,681,636 papers. We distinguish CRR from non-CRR, and pre-COVID-19 period (Jan.–Dec.2019) from COVID-19 period (Jan.–Dec.2020). This yields four categories: 775,738 non-CRR, pre-COVID-19 papers, 734 CRR, pre-COVID-19 papers, 830,914 non-CRR, COVID-19 papers, and 74,250 CRR, COVID-19 papers.

Research output in CRR is extremely dynamic, in particular in spring 2020, while it is relatively stable in non-CRR over the analysis period (see Fig 1). In the pre-COVID-19 period, CRR output is relatively stable, at roughly 50 papers per month. Starting with the January 2020 outbreak, CRR grows exponentially up to about 11,700 submissions in May 2020 and then decreases again to about 5,000 submissions in December 2020. Non-CRR output is stable throughout, at about 67,000 papers, and even increases slightly with the pandemic.¹⁰

The distribution of national health science output is highly skewed, with a Gini coefficient of around 0.9 before and during the pandemic in non-CRR and CRR.¹¹ The ten most prolific countries in CRR generate 65 percent of output during COVID-19. In the order of CRR output the top ten

⁸ This is common practice in the literature. The pros and cons are discussed elsewhere (Katz and Martin, 1997, is the seminal reference)

⁹ In detail, the search query is: ("2019-nCoV" OR "2019nCoV" OR "COVID-19" OR "SARS-CoV-2" OR "COVID19" OR "COVID" OR "SARS-nCoV" OR ("wuhan" AND "coronavirus") OR "Coronavirus" OR "Corona virus" OR "corona-virus" OR "corona viruses" OR "coronaviruses" OR "SARS-CoV" OR "Orthocoronavirinae" OR "MERS-CoV" OR "Severe Acute Respiratory Syndrome" OR "Middle East Respiratory Syndrome" OR ("SARS" AND "virus") OR "soluble ACE2" OR ("ACE2" AND "virus") OR ("ARDS" AND "virus") or ("angiotensin-converting enzyme 2" AND "virus")).

¹⁰ Looking closely, one sees non-CRR output decreasing slightly by the end of 2020. For interpretation recall that from October 2020 on the PubMed sample underreports increasingly submissions due to a time lag from submission to publication in PubMed (see Appendix A). Therefore, the slight drop of submissions in autumn 2020 (in our sample) does not necessarily indicate potentially negative effects of the pandemic due to frictions in the research machinery, or crowding out of non-coronavirus research. Quantification of such effects must be left to future studies.

¹¹ Gini coefficients are for CRR pre-pandemic 0.90, CRR pandemic 0.87, non-CRR pre-pandemic 0.88, non-CRR pandemic 0.88.

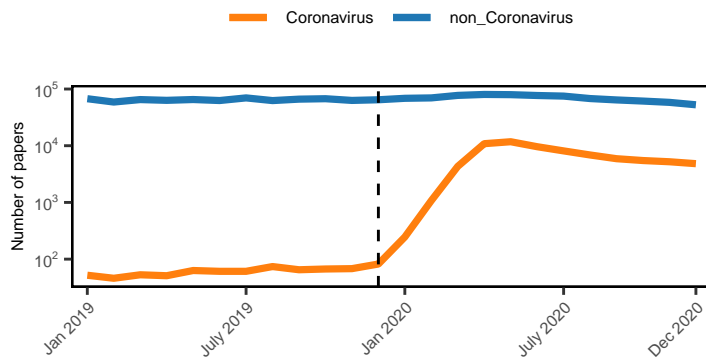


Fig. 1 Coronavirus and non-coronavirus papers by submission month (log-scale).

countries are (CRR papers in 2019; CRR papers in 2020): USA (225; 21,902), China (192; 9,942), Italy (17; 7,730), UK (54; 6,289), India (12; 5,260), Spain (14; 3,122), Canada (34; 3,037), Germany (38; 2,931), France (23; 2,923), Australia (29; 2474). All these countries contributed considerably to CRR during the pandemic, but not consistently to CRR before the pandemic. Some scholars explained the (early) high CRR output of China and Italy by the global infection process. On the other hand, rankings based on CRR during the pandemic (in Fig 1B) correspond surprisingly well to rankings based on overall scientific production in the last two decades (see Table 1); supporting the idea of the relevance of (slowly-accumulating) science capacity discussed in the ‘Background’ section.

National scientific output is closely related to IRC also in our sample. Around one fourth of the papers in our sample are internationally co-authored. This holds true for CRR as well as non-CRR papers, before as well as during the pandemic (see Appendix A, Table 7). Thus, CRR papers during the pandemic are, on average, as international as other papers.

Looking at the country-level, our data confirms a strong positive relationship between national scientific output and national IRC intensity. In a simple OLS regression of number of internationally co-authored papers on total number of papers (both in logs), we obtain an elasticity of 0.9 for (any) health science papers in 2019, (any) health science papers in 2020, and also for CRR papers in 2020. This exceeds well the estimate of elasticity of 0.7 by (Frame and Carpenter, 1979; Luukkonen et al., 1992), but seems reasonable if one takes into account the positive trends in IRC observed by Adams (2013). Elasticity for CRR papers in 2019 is slightly lower; around 0.8. Thus IRC may have become somewhat more relevant to national output of CRR during the pandemic, but neither more or less as would be expected on IRC in non-CRR.

We create international science networks for CRR as well as non-CRR research for each month, from January 2019 to December 2020. Networks are weighted following a full-count assignment, i.e. we increment the weight of the edge between two countries for each paper where both countries appear in the affiliations. Networks are accumulated over time by adding up all edge weights from the beginning of the analysis, January 2019, up to the focal month. Table 2 provides basic statistics on the accumulating networks.

In total, we identify 203 countries in our paper affiliations. We aggregate the 2019 period to indicate international scientific activity before the pandemic. At the end of 2019, the (accumulated) CRR network includes 68 countries, with about 500 international co-author ties (edge weights). The accumulated non-CRR network includes in Dec. 2019 nearly all countries, 201, connected through

585k ties. The non-CRR network is stable in 2020. Its decreasing growth rate of IRC papers (Weight % growth) is due to the addition of a relatively constant number of around 67k IRC papers each month.

Looking at CRR network growth in Table 2, one may distinguish four phases during which the CRR network expands in 2020. The first month, January 2020, may be considered a first phase in which the CRR network grows mostly in terms of joint papers (edge weights) and less in terms of collaborating countries entering the network (nodes). The second phase, from February to April 2020, is characterized by high growth in terms of both network entrants, around 30 percent growth rate, and joint collaborations, 130 to 300 percent. In the third phase, from May to July 2020, network entry slows down considerably. The growth rate of joint collaborations is very high at the beginning of the phase but slows down during this period from 68 to 22 percent. In a fourth phase, from August to December 2020, few latecomers enter the network and collaboration growth rates stabilize at around ten percent growth per month.

Table 2 International science networks, CRR and non-CRR, accumulated over months.

<i>CRR network (accumulated), 2019</i>												
	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
Countries	18	26	29	34	40	45	54	54	59	61	66	68
Country % growth		44	12	17	18	12	20	0	9	3	8	3
Edge weights	23	39	75	97	123	146	240	266	307	381	427	494
Weight % growth		70	92	29	27	19	64	11	15	24	12	16
<i>non-CRR network (accumulated), 2019</i>												
	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
Countries	188	194	194	198	198	200	201	201	201	201	201	201
Country % growth		3	0	2	0	1	0	0	0	0	0	0
Edge weights	49.8k	93.6k	147k	189k	240k	286k	342k	389k	438k	491k	539k	585k
Weight % growth		88	57	29	27	19	20	14	13	12	10	9
<i>CRR network (accumulated), 2020</i>												
	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
Countries	72	96	116	151	165	171	174	176	179	180	181	182
Country % growth		6	33	21	30	9	4	2	1	2	1	1
Edge weights	673	1.5k	6.1k	17.5k	29.4k	40.2k	48.9k	55.0k	60.9k	67.5k	73.9k	79.8k
Weight % growth		36	131	295	185	68	37	22	12	11	11	8
<i>non-CRR network (accumulated), 2020</i>												
	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
Countries	201	201	201	201	202	203	203	203	203	203	203	203
Country % growth		0	0	0	0	0	0	0	0	0	0	0
Edge weights	635k	684k	739k	795k	853k	910k	966k	1016k	1061k	1109k	1149k	1184k
Weight % growth		9	8	8	8	7	7	6	5	4	4	3

The first two analyses, Analysis 1 and Analysis 2, aim to shed some light on the role of different factors in driving national and inter-national CRR output over the course of the pandemic. Contextual factors are obtained from several supplementary datasets: The national situation during the pandemic is captured through COVID-19 incidences and governmental measures. The COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University provides COVID-19 cases and deaths (Dong et al., 2020).¹² Social restrictions and international travel restrictions are obtained from Hale et al. (2021).¹³ National economic data is from the ‘Penn World Table’ (Feenstra et al., 2015). Finally, national socio-economic development is proxied by the human development index obtained from the Human Development Report Office of the United Nations (see hdr.undp.org). Not all countries observed in the publication data are included in the supplementary datasets. Due to missing observations, the number of countries

¹² The data has been downloaded from <https://github.com/owid/covid-19-data/> in March 2021.

¹³ We use the two data files ‘international-travel-covid.csv’ and ‘stay-at-home-covid.csv’

reduces from 201 to 156 countries. All the countries dropped from the sample are small, except for Taiwan that is dropped because the UN does not provide statistics for Taiwan separately from China.

The third analysis, Analysis 3, focuses on how the development of global CRR during the pandemic relates to global non-CRR. That analysis is based only on publication data, and therefore includes all 201 countries.

3.3 Analysis 1: National scientific output

3.3.1 Variables

CRR. Our dependent variable is the number of CRR papers accumulated from January 2020 to a given month t' , which we shorthand by $(c_{i,t'})$. Accumulating output over months seems reasonable because we are ultimately interested in the dynamics of scientific activity — not submissions. For example, a paper submitted in April 2020 may well rely on research conducted from January 2020 to March 2020, and may have been influenced by events during that period (e.g. by the infection process for which we control; see below). Alternatively, one may consider the outcome variable as a proxy for the formation of national scientific capacity in CRR during the pandemic.

Initial scientific capacity. Initial scientific capacity is captured by the accumulated number of CRR (non-CRR) papers up to December 2019, denoted $c_{i,t0}$ ($n_{i,t0}$).

Pandemic context. The accumulated number of confirmed deaths associated with the coronavirus, $deaths_{i,t'}$, may be interpreted as a proxy for research needs and opportunities. The number of hospitalized cases would be closer to what we have in mind, but is a less reliable statistic. Pandemic policy restrictions may also be constraining but also inciting research. We include the accumulated number of days with requirement not to leave the house (with minimal exceptions), $lock-down_{i,t'}$, and the number of days with total border closure, $border-closed_{i,t'}$.

Socio-economic context. Population size in 2019 (in Mio.), $pop_{i,t0}$, GDP per capita in 2019, $gdp_{i,t0}$, (expenditure based PPP in 2017 US Dollars), and the human development index in 2019, $hdi_{i,t0}$, serve to control for country size, economic wealth, and state of development respectively.

We take logs (i.e. $\log(x+1)$) of all variables except for hdi. The pragmatic argument is that our variables are highly skewed to the right. In theory, one may think of CRR paper production akin to a Cobb-Douglas production process, where factors combine multiplicatively. In this case taking logs of all variables generates a linear model.¹⁴

Table 3 provides basic descriptive statistics of the variables used in the analysis on national scientific output. In addition we add a time indicator running from one in January 2020 to twelve in December 2020. Respective correlation coefficients take up the time trend for all accumulated variables, while initial conditions are fixed over time (zero correlation).

3.3.2 Empirical model

The estimating equation is the following linear model:

$$c_{i,t'} = \beta_{0,t'} + \beta_{1,t'} n_{i,t0} + \beta_{2,t'} c_{i,t0} + \beta_{3,t'} deaths_{i,t'} + \beta_{4,t'} lock-down_{i,t'} + \beta_{5,t'} border-closed_{i,t'} + \beta_{6,t'} pop_{i,t0} + \beta_{7,t'} gdp_{i,t0} + \beta_{8,t'} hdi_{i,t0} + \epsilon_{i,t'}$$

¹⁴ One may think of hdi in this case as the exponent of an exponential in the original multiplicative equation.

Table 3 Descriptive statistics (Analysis 1)

	Mean	Std.Dev.	Correlation coefficients									
			1	2	3	4	5	6	7	8	9	10
(1) $c_{i,t'}$	3.09	2.34	1.00	0.78	0.69	0.79	0.19	0.14	0.50	0.38	0.42	0.49
(2) $n_{i,t0}$	6.36	2.28	0.78	1.00	0.84	0.46	-0.00	-0.23	0.60	0.52	0.57	0.00
(3) $c_{i,t0}$	0.82	1.17	0.69	0.84	1.00	0.38	-0.00	-0.22	0.55	0.43	0.46	0.00
(4) $death_{i,t'}$	4.31	3.31	0.79	0.46	0.38	1.00	0.29	0.35	0.40	0.22	0.26	0.64
(5) $lock-down_{i,t'}$	0.75	1.49	0.19	-0.00	-0.00	0.29	1.00	0.31	0.10	-0.00	-0.00	0.24
(6) $border-closed_{i,t'}$	2.80	2.17	0.14	-0.23	-0.22	0.35	0.31	1.00	-0.07	-0.16	-0.17	0.55
(7) $pop_{i,t0}$	2.65	1.42	0.50	0.60	0.55	0.40	0.10	-0.07	1.00	-0.18	-0.14	0.00
(8) $gdp_{i,t0}$	9.39	1.23	0.38	0.52	0.43	0.22	-0.00	-0.16	-0.18	1.00	0.93	0.00
(9) $hdi_{i,t0}$	0.73	0.15	0.42	0.57	0.46	0.26	-0.00	-0.17	-0.14	0.93	1.00	0.00
(10) t	6.50	3.45	0.49	0.00	0.00	0.64	0.24	0.55	0.00	0.00	0.00	1.00

where the $\epsilon_{i,t'}$ denotes the error term. The model is estimated with simple OLS separately for each month t' in 2020. This allows for varying coefficients over time to uncover dynamics. Time effects common to all countries, as for example augmenting number of papers on open source platforms, are then naturally accounted for. For ease of interpretation, all variables are standardized (after taking logs) separately for each month t' to zero mean and standard deviation of one.

3.4 Analysis 2: International research collaboration

3.4.1 Variables

Analysis 2 focusses on drivers of bi-national collaboration on CRR. Some variables describe a single country i (or j) that is part of the dyad, others characterize the dyad ij . Variables measured before the shock form the initial conditions and are indicated by a time subscript $t0$. Variables that vary over the period of the pandemic are all accumulated over the pandemic (indicated by t'). The reason, again, is that a mapping from research context to research output within each month is less likely than a mapping of past and current research context to past and current research output.

Joint CRR. The dependent variable is the number of CRR papers signed by two countries i and j between January 2020 up to a given month t' , denoted $c_{ij,t'}$.

Initial scientific capacity. The number of joint publications in 2019 between both countries on CRR ($c_{ij,t0}$) and non-CRR ($n_{ij,t0}$) are part of initial conditions. For each country, say i , we indicate scientific capacity by the total number of CRR papers ($c_{i,t0}$) and non-CRR papers ($n_{i,t0}$). In the link formation context these measures may be interpreted as factors that help to attract or initiate collaborations. In addition, the combination of health science competence is indicated by the sum ($n_{i,t0} + n_{j,t0}$) and absolute difference ($|n_{i,t0} - n_{j,t0}|$). These two factors indicate whether pooled health science capacity in the dyad is high ($n_{i,t0} + n_{j,t0}$) and to what extent scientific capacity is equally distributed among the partners ($|n_{i,t0} - n_{j,t0}|$).

Pandemic context. The relative pandemic situation of both countries is indicated by taking into account the (accumulated) number of COVID-19 related deaths ($death_{i,t'}$), number of days under lock-down ($locked_{i,t'}$), and number of days with closed borders ($closed_{i,t'}$). Lock-down and closed borders of a country may hinder initiation or maintenance of international collaboration. Therefore, we keep these variables as individual factors. The number of deaths approximate the number of COVID-19 cases in the country on which research may be conducted. Therefore, we sum cases over the dyad ($death_i + death_j$), and take into account their absolute difference $|death_i - death_j|$. If for example the sum is positively correlated with joint CRR, while the absolute difference is negatively

correlated with CRR, then countries with few cases would interact with countries having many cases. Potential explanations for such an observation could be international solidarity but also resource interdependence.

Socio-economic context. We control for country size by the number of inhabitants in 2019 pop_{i,t_0} . Economic wealth and development is indicated by gdp per capita (gdp_{i,t_0}) and the Human Development Index (hdi_{i,t_0}) respectively. The discussion on core-periphery processes creates an interest in understanding to what extent developed (developing) countries interact among and with each other. Again, the sum and absolute difference of the partners' characteristics provide some indication; leading to the four composite variables $gdp_i + gdp_j$, $|gdp_i - gdp_j|$, $hdi_i + hdi_j$, $|hdi_i - hdi_j|$.

Geographical space. The literature is clear in that there is a geographical bias in IRC link formation (see Section Background above). Therefore we control for the geographic distance between two countries, $distance_{ij}$, measured by the distance (in km) of the countries' geographic centers. Another control $same_region_{ij}$ indicates whether two countries are located in the same world region, i.e. continent as defined by the World Bank Development Indicators.

3.4.2 Empirical model

The empirical model is essentially a gravity model, where the interaction intensity between two countries (here number of joint CRR papers) depends on the countries' distance and further (relational) factors. The model is common in spatial scientometrics (Frenken et al., 2009) and has been applied in the same form e.g. by Hoekman et al. (2009).

In detail, we estimate a zero-inflated negative binomial model in order to take into account the fact that many (mostly small and distant) countries do not have any joint CRR paper. The model combines the negative binomial count density¹⁵ with a binary process¹⁶ to model excess zeros in the outcome (see e.g. Cameron and Trivedi, 2005, p.681).

The first part of the model, termed the zero-model, models the binary process of two countries i and j having *no* joint CRR papers. The linear predictor of the zero-model includes the typical ingredients of a gravity model:

$$p = Pr [c_{ij,t} = 0 | \gamma] = \text{logit}(\gamma_0 + \gamma_1 \log(n_{i,t0} + n_{j,t0}) + \gamma_2 \log(|n_{i,t0} - n_{j,t0}|) + \gamma_3 \log(\text{distance}_{ij}) + \gamma_4 \text{same_region}_{ij}) ,$$

where $(n_{i,t0} + n_{j,t0})$ captures the joint weight of the two countries in 'health sciences', and $(|n_{i,t0} - n_{j,t0}|)$ their absolute difference. Geographical proximity is captured by distance_{ij} and the *same region* $_{ij}$ dummy.

With probability $(1 - p)$ the count density applies with expectation

$$E [c_{ij,t} | \beta] = \exp(\beta_0) c_{ij,t0}^{\beta_1} n_{ij,t0}^{\beta_2} (c_{i,t0} c_{j,t0})^{\beta_3} \times (d_{i,t} + d_{j,t})^{\beta_4} (|d_{i,t} - d_{j,t}|)^{\beta_5} (gdp_{i,t0} + gdp_{j,t0})^{\beta_6} \times (|gdp_{i,t0} - gdp_{j,t0}|)^{\beta_7} (hdi_{i,t0} + hdi_{j,t0})^{\beta_8} (|hdi_{i,t0} - hdi_{j,t0}|)^{\beta_9} \text{distance}_{ij}^{\beta_{10}} \text{same_region}_{ij}^{\beta_{11}} ,$$

where $\exp(\beta_0)$ is a scaling factor (the intercept). Individual variables are introduced above but some notes on the model structure are in order. First, the model allows for a lasting impact of an established relation in non-CRR and CRR through $(n_{ij,t0}, c_{ij,t0})$. Second, countries may 'attract' IRC through their research competences $(n_{i,t0}, c_{i,t0})$. Because this effect is symmetric (the relationship ij is the same as ji), we enforce the same coefficient for i and j . Control variables are entered such that they capture the joint 'mass' of the partners as well as their absolute difference. For example scientific interaction may be driven by countries having many COVID-19 cases in sum $(d_{i,t} + d_{j,t})$, but also if one country has many more cases than the other $(|d_{i,t} - d_{j,t}|)$. The former can be interpreted as a common pool of resources (or incentives), the second captures dyadic interdependence due to inequality in resources.¹⁷ Third, we introduce geographic distance and same region indicators in the count model because geographic proximity is susceptible of driving not only whether but also how much research is conducted jointly.

The zero-model is kept relatively light, in form of a basic gravity model, for the following reasoning. First, one may interpret the zero-model as the potential to collaborate in the health sciences as such. This is mostly an issue of mutual (or one-sided) awareness driven by a combination of global visibility and geographic distance. Given awareness, the intensity of joint CRR may then be determined by various factors capturing the needs and benefits of collaboration. There is also a more pragmatic argument. The two-step process is an artificial interpretation of the model. In fact, we deal here with one convolution of densities that is determined by all the factors that we

¹⁵ The choice of the negative binomial density over a Poisson density is supported by a strong and significant estimate of the variance related parameter.

¹⁶ We chose a logit model.

¹⁷ We tried various ways of introducing country specific factors into the model, and found this formulation to be the most compelling (in terms of fit and reasoning).

consider at once. In this sense, separating factors in the different parts helps to clarify the overall contribution of individual factors to the outcome.

The link formation model is estimated on dyadic data which by construction may result in network correlation of errors. In general, correlated errors maintain unbiasedness and consistency of coefficient estimates, but create a downward bias of the estimated standard deviations of coefficient estimates. The reason is essentially that (network) dependence of observations reduces the information content compared to independent observations (which is assumed by standard estimators). Most of the dependence across dyad observations can be expected from repeated observations for the same individual countries. The Multiple Regression Quadratic Assignment Procedure (MRQAP) therefore creates a Null distribution through random permutation of rows and columns of the adjacency matrix (in effect a random re-labeling of nodes in the network). In principle, there are different ways to implement the idea of MRQAP (Decker et al., 2007). In our case, the preferred choice is to permute one right-hand-side factor, keeping everything else fixed, to generate the distribution of the z-statistic (coefficient estimate divided by its standard deviation) under the Null hypothesis that the right-hand-side factor is not systematically related to the outcome controlling for all other factors.¹⁸ In the results, we report the estimated z-value together with significance levels of one-sided tests based on the Null distribution of z-values from permutation.

3.5 Analysis 3: Convergence to global health science

The COVID-19 shock changed dramatically the global distribution of CRR. Because the global health science system responded in very short time, CRR during the pandemic must have built essentially on resources developed before the outbreak. Therefore it seems reasonable that the global structure of CRR during the pandemic converges to the (pre-pandemic) global structure of health sciences. We investigate this hypothesis by looking at i) national scientific output, ii) countries' network centrality in the IRC network, and iii) IRC network structure.

3.5.1 National scientific output

For each month in 2019 and 2020, we create a ranking of countries based on CRR output ($c_{i,t}$) and non-CRR output ($n_{i,t}$) respectively. We speak of convergence in national scientific output when the two rankings become more similar over time. Similarity is appropriately captured by the rank correlation coefficient τ_X , as described in Emond and Mason (2002). This statistic is similar to Kendall's τ , except that it handles ties the same as dominant relationships (entering 1 and not 0 in the dominance matrix). In principle, this is favorable in case of many ties in the rankings, as we have in corona pre-Covid-19 research; but does not really effect the results. A 90 percent confidence interval around τ_X is then obtained by a traditional jackknife, or leave-one-out, approach as described for example in Abdi and Williams (2010).

3.5.2 Network centrality

The international science network is highly hierarchical, and network centrality captures the position of a country within that hierarchy. Therefore, we ask 'How does network centrality of countries in the CRR network align with their centrality in the overall health science network?'

¹⁸ There are pros and cons to the different ways of implementing MRQAP. The simplest way would be to permute the outcome but that has the disadvantage to create a distribution under the null that no factor is relevant. Furthermore, our (non-linear) model fails to converge if the model does not fit the data, which is almost always the case when the outcome is permuted. Permutation of error terms after partialing is a common alternative but is only valid under relatively strict assumptions on the error term which are unlikely to be met in our model as it is a convolution of two different distributions. Our preferred alternative of permuting individual right hand side factors has the disadvantage to break existing correlation with other right hand side factors. The effect seems however to be bearable if a pivotal statistic is used, as we do (Decker et al., 2007).

Although one may argue on the details, the international science network can be safely said to exhibit a core-periphery structure. Therefore, network centrality is appropriately captured through (normalized) s-core decomposition (Eidsaa and Almaas, 2013). The s-core decomposition identifies sets of nodes that are heavily connected among each other. Together they form the network core(s). Roughly, the algorithm starts from the complete network and proceeds by iterative removal of the least connected node in the remaining network. The s-core is the strength of the node in the (remaining) network at the time of removal. In order to compare different networks over time, we normalize the s-core by the maximum s-core in the network. The s-core ranges from 0 for isolates in the network, to 1 for (highest) core members. S-cores are measured for each country on a monthly bases on the CRR network and the non-CRR network respectively. As the difference between the s-cores decreases, hierarchies converge.

3.5.3 Network structure

Convergence in network structure is first investigated by correlating the CRR adjacency matrix with the non-CRR adjacency matrix on a monthly basis. The weights in the adjacency matrix are in logs because they are highly skewed — many countries do not collaborate (zero entry) and some do heavily (many joint papers). Statistical significance of correlations is obtained through the Quadratic Assignment Procedure (QAP). The QAP test creates a null distribution through re-labeling of nodes in one network; maintaining the structure of the networks. The correlation analysis tells us whether the CRR network and the non-CRR network become more similar over time.

In a second step, we investigate in which aspects the CRR network differs from the non-CRR network. The background section highlights network hierarchy and communities as two salient features in ICR networks. This is what we focus on.

Hierarchy is captured by the largest (absolute) eigenvalue of the adjacency matrix, say λ . More hierarchical networks tend to have a higher largest eigenvalue. In fact, it has been shown that the largest eigenvalue is maximal for nested-split graphs. A nested-split graph is a specific type of a hierarchical network, in which the most central node connects to all other nodes, and less central nodes connect to subsets of alters of more central nodes. Interestingly, nested-split graphs emerge in network games where payoffs are strategic complements in effort levels (König et al., 2014), which is a reasonable assumption for science networks.

We measure λ on the (accumulating) CRR network for each month from January 2019 to December 2020. The value of λ in itself is not very telling as it depends on various features of the network; most importantly network size. For interpretation it is therefore useful to compare this statistic to a network null model in order to see whether the network of interest is more or less hierarchical as the null network. Most studies create a null distribution of networks by fixing some aspects of the focal network, e.g. network size and degree distribution, and randomize the rest. Our interest is in how the CRR network structurally deviates from the non-CRR network. Therefore we use the non-CRR network formed in 2019 to generate our null distribution. In detail, one realization s is obtained by drawing links uniformly at random with replacement from the non-CRR network until a network is created of the same size (same number of links, and hence same total strength) as the CRR network. Each realization yields a statistic λ_s . We draw 100 realizations to obtain the null distribution for our statistics. Based on the null distribution, we calculate the z -value for the largest eigenvalue λ , as $z_\lambda = \frac{\lambda_{crr} - \hat{E}[\lambda_s]}{sd(\lambda_s)}$. A statistic z_λ below (above) zero indicates that the CRR network is more (less) hierarchical than would be expected by the structure of the (prior) non-CRR network.

Network communities are commonly thought of as subsets of nodes with relatively strong interaction. IRC networks feature communities due to scientists' tendency to collaborate with other scientists that are somewhat close in space; with space broadly defined as scientific, geographical, cultural etc. (see Section Background). The hypothesis we wish to test is that during the pandemic

the (accumulating) CRR network converges towards the same community structure as the (prior) non-CRR network. The outline of our empirical approach is as follows: In a first step, we detect communities in the (prior) non-CRR network. This becomes our reference community structure, say benchmark. Then, for each month in the observation period, we measure how well the (accumulating) CRR network ‘fits’ that benchmark. As in the hierarchy analysis, we account for varying network sizes over time and across networks by creating a null distribution through resampling from the non-CRR network. We discuss the details in the following.

Community detection in the (prior) non-CRR network follows closely the procedure of Fitzgerald et al. (2021). The main idea is to find a network partition that maximizes some network modularity statistic, say Q . Newman (2004) proposed a statistic that measures to what extent nodes belonging to the same community form (weighted) ties beyond what would be expected based on their link strength alone.¹⁹ Reichardt and Bornholdt (2006) proposed multiple generalizations known as spin-glass algorithm. Among others they introduce a parameter γ for tuning the resolution of the network partition. A smaller (higher) γ tends to yield less (more) clusters. For $\gamma = 1$ the objective function coincides with Newman’s network modularity measure (Newman, 2004). We use that extension. The optimization is done through simulated annealing (as implemented in the R-package ‘igraph’). Simulated annealing is a stochastic optimization algorithm and hence may provide different partitions for the same data and parameter settings. This is actually useful, because the robustness of communities for a given parameter γ across multiple optimizations signals whether communities are well identified. We search a robust and informative community partitioning through a grid search on γ . Appendix XY details how.

Figure 2 displays a partition of the (2019) non-CRR network for $\gamma = 1.2$. It is remarkable how similar that partitioning is to the partitioning obtained by Fitzgerald et al. (2021, Fig.5b) based on all Scopus publications in 2015. The map displays seven communities with numbers ordered by the average network strength of their members (Appendix C lists the countries belonging to each community). Community 1 is a global community that includes most importantly the US, China, Great Britain, and Japan. All other communities cover world regions. Roughly, Community 2 covers Central Europe, Community 3 Northern Europe, Community 4 includes only Israel, Jersey, and Montserrat, Community 5 spans East Europe and Russia, Community 6 corresponds to South America, and Community 7 covers Africa and the Middle East.

This partition is used in the analysis as the benchmark for the (accumulating) CRR network. For each month in the observation period, we calculate the modularity statistic on the (accumulating) CRR network (Q_{crr}) on the benchmark partitioning obtained in the previous step on the non-CRR network. In addition, multiple resamples are obtained from the non-CRR network such that each resample s has the same size (i.t.o. total edge weights) as the (current) CRR network. As for the CRR network, we calculate for each resample of the non-CRR network a modularity statistic, Q_s . This provides us a z -value, i.e. $z_Q = \frac{Q_{crr} - \hat{E}[Q_s]}{sd(Q_s)}$. If the z -value is high (low) we can say that the CRR network fits well (badly) the community structure of the (2019) non-CRR network.

4 Results

4.1 National scientific output

Table 5 shows regressions of the cumulative number of national CRR papers for each month in 2020. In the first column, January 2020, pre-pandemic CRR (c_{t0}) and the number of COVID-19 related deaths ($deaths_{t'}$) are estimated to be significantly and positively correlated. Other factors

¹⁹ In some sense, the null model here is a simple gravity model based on only node size. In the prior IRC literature a similar normalization has been applied by calculating Salton’s measure, i.e. observed strength of interaction between two nodes divided by the product of the nodes network strength. Salton’s measure revealed in particular country preferences of interaction, revealing the role of distance in IRC.

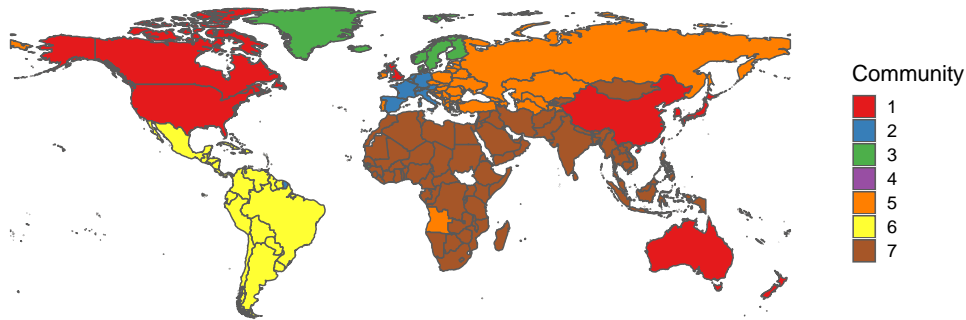


Fig. 2 Communities in the 2019 non-CRR network from spin-glass algorithm with tuning parameter $\gamma = 1.2$; the reference community structure for Analysis 3.3.

appear to be weakly correlated to CRR in January 2020. However, within the first months of the pandemic — as CRR takes off — this pattern reverses. More general health science capacity (n_{t0}) becomes the dominating factor of CRR output. The relevance of CRR specific experience diminishes gradually, and the impact of (national) COVID-19 cases immediately goes down on a stable and low level. Experiencing a lock down ($locked_{t'}$) tends to be also positively related to output but not very significantly. The discussion above already hints to a potential ambiguity in the effect from lock down experience in that it may signal research needs and opportunities (certainly not fully captured by the number of COVID-19 related deaths), as well as actual constraints to research. Closed international borders (as we measure by $border_{t'}$) seem to have no effect on scientific output. Also population size (pop_{t0}) and GDP per capita (gdp_{t0}) take over little explanatory power once (prior) research output is controlled for. Finally, the state of economic and social development (hdi_{t0}) is negatively correlated with output controlling for confounding factors. Although the estimated coefficient is (mostly) insignificant, there is a clear negative trend, suggesting that in particular in the second half of 2020 less developed countries engaged more in CRR research. Finally, note that model fit in terms of R^2 increases from around 0.8 in January to 0.95 in December. This suggests that the factors taken into account in the regression ‘fit’ increasingly well with variations of national CRR.

Table 5 Accumulated number of CRR papers (c_{it}) in 2020 (all variables in logs, standardized to zero mean and one std.dev., 156 countries).

	Jan. '20	Feb. '20	Mar. '20	Apr. '20	May '20	June '20	July '20	Aug. '20	Sept. '20	Oct. '20	Nov. '20	Dec. '20
n_{t0}	0.048 (0.095)	0.219 * (0.087)	0.62 *** (0.07)	0.786 *** (0.059)	0.851 *** (0.058)	0.929 *** (0.054)	0.928 *** (0.053)	0.931 *** (0.05)	0.94 *** (0.049)	0.948 *** (0.048)	0.943 *** (0.049)	0.942 *** (0.05)
c_{t0}	0.764 *** (0.071)	0.577 *** (0.066)	0.308 *** (0.051)	0.144 *** (0.041)	0.099 * (0.041)	0.063 (0.038)	0.066 (0.038)	0.05 (0.036)	0.044 (0.035)	0.035 (0.035)	0.029 (0.036)	0.025 (0.036)
$deaths_{it}$	0.134 *** (0.04)	0.089 (0.059)	0.086 (0.047)	0.1 * (0.04)	0.084 * (0.037)	0.08 * (0.032)	0.075 * (0.032)	0.091 ** (0.03)	0.095 ** (0.029)	0.088 ** (0.029)	0.072 * (0.029)	0.062 * (0.029)
$locked_{it}$	—	0.023 (0.055)	0.022 (0.028)	0.049 * (0.023)	0.058 * (0.023)	0.053 * (0.021)	0.06 ** (0.021)	0.057 ** (0.02)	0.053 ** (0.02)	0.052 ** (0.02)	0.053 ** (0.02)	0.055 ** (0.02)
$borders_{it}$	—	—	0.013 (0.028)	0.01 (0.023)	-0.013 (0.024)	-0.006 (0.022)	-0.004 (0.022)	-0.006 (0.022)	-0.003 (0.021)	-0.004 (0.021)	-0.007 (0.021)	-0.01 (0.021)
pop_{t0}	0.06 (0.07)	0.13 * (0.064)	0.021 (0.053)	-0.019 (0.045)	-0.026 (0.046)	-0.048 (0.044)	-0.044 (0.045)	-0.041 (0.042)	-0.039 (0.041)	-0.033 (0.041)	-0.011 (0.041)	0.001 (0.041)
gdp_{t0}	0.002 (0.103)	0.079 (0.095)	0.017 (0.075)	0.042 (0.061)	-0.005 (0.062)	0.018 (0.058)	0.05 (0.057)	0.076 (0.054)	0.097 (0.052)	0.101 (0.052)	0.113 * (0.052)	0.115 * (0.053)
hdi_{t0}	0.001 (0.11)	-0.056 (0.101)	-0.07 (0.081)	-0.064 (0.067)	-0.043 (0.066)	-0.096 (0.061)	-0.123 * (0.061)	-0.143 * (0.058)	-0.167 ** (0.056)	-0.17 ** (0.056)	-0.171 ** (0.057)	-0.168 ** (0.058)
R^2	0.795	0.829	0.895	0.929	0.928	0.938	0.938	0.944	0.948	0.949	0.949	0.947

Notes: Recall from Section 3.3 that c_{t0} and n_{t0} proxy initial scientific capacity in CRR and non-CRR respectively. The main finding is that correlation of c_{t0} with CRR output starts off high but decreases as n_{t0} takes over the main explanatory power. There is some support that national pandemic severity ($deaths_{it}$, $locked_{it}$) are positively, and economic development (hdi_{t0}) is negatively associated with CRR by the second half of 2020.

4.2 International research collaboration

Estimation results presented in Table 6 provide information on the dynamics of joint CRR over the pandemic, taking into account various factors. Regressions of accumulated number of joint CRR papers start in February 2020 and are estimated separately for each month up to December 2020.²⁰ The discussion below focusses on coefficient estimates. Note that coefficient estimates are not to be interpreted as elasticities. Average marginal effects are therefore provided in the Appendix B, Table 8.

Focus on the first column (February 2020). The zero model (lower part of the table) estimates the probability of having *no* joint CRR. The sum of non-CRR papers in 2019 ($n_{i,t0} + n_{j,t0}$) is negative and highly significant, meaning that chances increase to have (at least one) joint CRR paper in 2020. Absolute difference in non-CRR papers make little difference; if anything (as a positive coefficient is estimated) inequality makes it less likely to work together. Geographic distance plays no (significant) role as such, but being in the same region is positively related to joint CRR in February 2020. Looking at the count model of the intensity of dyadic CRR, we estimate a coefficient of one for joint non-CRR in the prior year 2019. Joint CRR papers are positively but not significantly correlated. Once past dyadic interaction is taken into account, countries with large health science capacity ($n_{i,t0}$) are not sought initially. On the other hand, countries with specific CRR capacity attract collaborators ($c_{i,t0}$). The pandemic state of the country, i.e. COVID-19 related deaths, lock-downs, and closed borders, do not play a role initially; probably because too few countries in the sample are hit. Country size (pop) and economic wealth (gdp per capita) play a role initially. There is some sign that developing countries collaborate with each other (negative $hdi_i + hdi_j$), but not particularly with more advanced countries (negative $|hdi_i - hdi_j|$). One percent increases reduce the outcome by 1.9 and 0.5 percent respectively. Geographic distance and being located in the same region are not significant in the count model. The significant technical parameter τ indicates that the negative binomial model actually does capture over-dispersion.

²⁰ For January 2020, the first month of the pandemic, the outcome variable has been very sparse. Therefore the model did not fit the data and the estimation failed to converge.

Table 6 Zero-inflated negative binomial model of (accumulated) joint coronavirus related papers ($c_{i,j,t}$) during the pandemic, est. coefficient (z -value, p -value)

	Feb. '20	Mar. '20	Apr. '20	May '20	June '20	July '20	Aug. '20	Sept. '20	Oct. '20	Nov. '20	Dec. '20
<i>Count-model</i>											
Intercept	-4.809*** (-3.088, 0)	-8.207*** (-10.428, 0)	-5.198*** (-3.735, 0)	-3.032*** (-1.858, 0)	-3.617*** (-2.316, 0)	-3.919*** (-2.698, 0)	-4.286*** (-3.162, 0)	-4.346*** (-3.255, 0)	-4.178*** (-3.222, 0)	-4.287*** (-3.211, 0)	-4.198*** (-3.231, 0)
$n_{i,t,0}$	1.048*** (0.816, 0)	0.871*** (0.686, 0)	0.809*** (0.621, 0)	0.841*** (0.656, 0)	0.762*** (0.577, 0)	0.722*** (0.537, 0)	0.717*** (0.522, 0)	0.722*** (0.522, 0)	0.728*** (0.537, 0)	0.744*** (0.552, 0)	0.753*** (0.567, 0)
$c_{i,t,0}$	0.174 (0.000, 0)	0.044 (-0.122, 0)	0.005 (-0.168, 0)	0.005 (-0.168, 0)	0.053 (-0.117, 0)	0.059 (-0.111, 0)	0.057 (-0.109, 0)	0.034 (-0.148, 0)	0.025 (-0.164, 0)	0.028 (-0.171, 0)	0.025 (-0.171, 0)
$n_{i,t,0}$	1.69, 0.088 (1.312, 0.178)	0.872, 0.228 (0.488, 0.093)	0.105, 0.419 (-0.125, 0.464)	0.125, 0.464 (-0.125, 0.464)	1.312, 0.178 (0.872, 0.228)	1.438, 0.177 (0.929, 0.418)	1.376, 0.186 (0.862, 0.266)	1.438, 0.177 (0.929, 0.418)	1.488, 0.186 (0.948, 0.266)	1.488, 0.186 (0.948, 0.266)	1.488, 0.186 (0.948, 0.266)
$c_{i,t,0}$	0.301** (0.151, 0.009)	0.019 (-0.028, 0)	0.049 (0.009, 0.059)	0.049 (0.009, 0.059)	0.003 (-0.003, 0)	0.005 (-0.003, 0)	0.007 (-0.003, 0)	0.022 (-0.003, 0)	0.032 (-0.003, 0)	0.015 (-0.003, 0)	0.017 (-0.003, 0)
$death_i + death_j$	-0.943 (-1.386, 0.266)	0.023 (0.741, 0.36)	0.067 (0.165, 0.196)	0.059 (0.161, 0.142)	0.062 (0.203, 0.486)	0.096* (0.419, 0.418)	0.108* (0.577, 0.431)	0.131*** (1.814, 0.265)	0.166*** (2.583, 0.186)	0.136*** (1.25, 0.337)	0.134*** (0.738, 0)
$death_i - death_j$	1.036 (-1.168, 0.313)	-0.038 (0.574, 0.411)	-0.054 (-2.592, 0.119)	-0.034 (-2.513, 0.145)	-0.025 (-2.978, 0.115)	-0.05 (-3.199, 0.093)	-0.057 (-4.914, 0.012)	-0.079* (-4.914, 0.012)	-0.092** (-6.45, 0.002)	-0.07* (-4.931, 0.023)	-0.075* (-5.44, 0.02)
$locked_{i,t}$	-0.059 (-0.586, 0.315)	0.035 (2.252, 0.175)	0.029 (3.635, 0.1)	0.024 (3.705, 0.082)	0.019 (3.405, 0.121)	0.03* (5.842, 0.026)	0.024 (4.771, 0.059)	0.022 (4.665, 0.051)	0.024* (5.096, 0.043)	0.019 (4.176, 0.087)	0.02 (4.475, 0.075)
$closed_{i,t}$	0.06* (4.846, 0.014)	0.018 (2.815, 0.105)	0.018 (2.815, 0.105)	0.007 (1.301, 0.295)	0.012 (2.737, 0.128)	0.021* (5.074, 0.017)	0.03** (7.135, 0.006)	0.031** (7.816, 0.003)	0.032*** (8.057, 0)	0.031** (7.693, 0.002)	0.03*** (7.681, 0)
$pop_{i,t}$	0.239*** (5.136, 0)	0.008 (-0.323, 0.433)	0.037 (2.081, 0.177)	0.035 (2.26, 0.155)	-0.013 (-0.91, 0.351)	0.01 (0.752, 0.404)	0.019 (1.451, 0.31)	0.005 (0.383, 0.458)	0.006 (0.47, 0.411)	0.019 (1.566, 0.268)	0.024 (2.013, 0.221)
$gdp_i + gdp_j$	0.671* (3.119, 0.014)	0.415* (3.827, 0.025)	0.164 (2.152, 0.141)	0.074 (1.164, 0.302)	0.225* (4.073, 0.031)	0.229* (4.42, 0.031)	0.309** (6.123, 0.003)	0.327** (6.707, 0.001)	0.341** (6.384, 0.005)	0.345** (7.243, 0.002)	0.341** (7.262, 0.001)
$gdp_i - gdp_j$	0.059 (1.17, 0.207)	0.029 (3.31, 0.257)	0.049 (2.972, 0.084)	0.05 (3.361, 0.057)	0.047 (3.43, 0.075)	0.043 (3.592, 0.058)	0.04 (-2.058***)	0.042 (-2.303***)	0.038 (-2.403***)	0.034 (-2.33***)	0.036 (-2.322***)
$hdi_i + hdi_j$	-2.703* (-3.17, 0.016)	-2.78** (-5.925, 0.001)	-1.216* (-3.723, 0.032)	-1.16* (-4.364, 0.019)	-1.896*** (-9.438, 0)	-1.835*** (-9.459, 0)	-1.95*** (-10.862, 0)	-1.95*** (-10.862, 0)	-1.95*** (-10.862, 0)	-1.898*** (-11.547, 0)	-1.866*** (-11.646, 0)
$hdi_i - hdi_j$	-2.23*** (-3.217, 0.009)	-1.627*** (-4.633, 0.005)	-1.852*** (-7.404, 0.001)	-1.585*** (-7.509, 0)	-1.896*** (-10.41, 0)	-1.835*** (-9.459, 0)	-1.95*** (-10.862, 0)	-1.95*** (-10.862, 0)	-1.95*** (-10.862, 0)	-1.898*** (-11.547, 0)	-1.866*** (-11.646, 0)
distance	0.17 (2.094, 0.075)	0.042 (1.214, 0.261)	0.036 (1.444, 0.264)	0.031 (1.442, 0.246)	0.011 (0.383, 0.419)	-0.015 (-0.85, 0.371)	-0.039 (-2.167, 0.197)	-0.042 (-2.378, 0.159)	-0.038 (-2.187, 0.18)	-0.029 (-1.633, 0.244)	-0.022 (-1.245, 0.318)
same region	0.137 (0.888, 0.206)	0.166* (2.541, 0.038)	0.062 (1.267, 0.139)	0.092 (2.225, 0.05)	0.095* (3.295, 0.012)	0.124** (3.604, 0.009)	0.12** (3.524, 0.004)	0.146*** (4.421, 0)	0.185*** (5.659, 0)	0.179*** (5.451, 0)	0.19*** (5.836, 0)
$\log(\theta)$	0.785*** (3.563, 0)	1.779*** (13.876, 0)	1.533*** (19.336, 0)	1.487*** (21.71, 0)	1.425*** (24.631, 0)	1.425*** (26.305, 0)	1.425*** (26.816, 0)	1.441*** (28.767, 0)	1.357*** (28.25, 0)	1.288*** (27.91, 0)	1.258*** (28.142, 0)
<i>Zero-model</i>											
Intercept	9.385** (2.271, 0.005)	22.461*** (3.59, 0)	7.433*** (3.248, 0)	5.323*** (3.716, 0)	6.008*** (4.713, 0)	6.051*** (4.489, 0)	5.666*** (4.305, 0)	6.498*** (4.896, 0)	7.372*** (5.666, 0)	7.078*** (5.684, 0)	7.55*** (6.151, 0)
$n_{i,t,0} + n_{j,t,0}$	-2.055*** (-4.084, 0.006)	-10.23*** (-2.837, 0)	-3.237** (-4.034, 0.002)	-2.454** (-9.751, 0.004)	-2.373*** (-8.829, 0)	-3.092*** (-7.1, 0)	-3.12*** (-7.14, 0)	-3.049*** (-7.329, 0)	-3.68*** (-7.391, 0)	-3.152*** (-8.236, 0)	-3*** (-8.991, 0)
$ n_{i,t,0} - n_{j,t,0} $	0.674 (1.864, 0.101)	8.523 (2.517, 0.088)	1.788 (2.355, 0.073)	1.088** (4.916, 0.004)	1.129*** (4.648, 0)	1.755* (4.266, 0.0122)	1.755* (4.31, 0.009)	1.651** (4.245, 0.004)	2.196** (4.717, 0.009)	1.656** (4.753, 0.004)	1.499** (4.992, 0.005)
distance	0.293 (0.72, 0.314)	-0.812 (-1.739, 0.079)	0.923 (2.067, 0.057)	0.699** (4.189, 0.004)	0.456* (3.295, 0.012)	0.699** (3.604, 0.009)	0.357** (4.057, 0.001)	0.357** (4.057, 0.001)	0.388** (3.641, 0.008)	0.523** (4.168, 0.002)	0.64** (3.795, 0.00702)
same region	-2.257*** (-5.194, 0)	-3.649*** (-4.654, 0)	-4.549*** (-6.51, 0)	-3.531*** (-6.51, 0)	-2.958*** (-6.429, 0)	-3.047*** (-6.693, 0)	-2.844*** (-6.863, 0)	-2.999*** (-7.369, 0)	-2.836*** (-7.663, 0)	-2.684*** (-7.615, 0)	-2.617*** (-7.683, 0)
obs.	12090	12090	12090	12090	12090	12090	12090	12090	12090	12090	12090
loglik	-1746	-4274	-7216	-9635	-11503	-12540	-13063	-13393	-14169	-14704	-15226

— P-values are based on MRQAP (1000 permutations) and one-sided (because null distributions are not symmetric). One, two, and three stars signal significance values below 5%, 1%, and 0.1% respectively. Marginal effects are in Appendix B, Table 8.

— Regressions start in Feb. 2020 because in Jan. 2020 collaboration is sparse and not very systematic.

— Recall from Section 3.4 that n and c stands for non-CRR and CRR respectively. Indices i indicate a country, and ij a country dyad. Country-level variables are joined to capture the country dyad's sum ($x_i + x_j$) and their absolute difference ($|x_i - x_j|$). $\log(\theta)$ captures over-dispersion in the count model.

— The main finding is that in February initial CRR competence ($c_{i,t,0}$) is positively and non-CRR ($n_{i,t,0}$) is negatively associated with collaboration. Prior collaborations in non-CRR ($n_{i,t,0}$) are relevant throughout. In the second half of the year, countries with strong non-CRR competence ($n_{i,t,0}$) attract collaborations, and less developed countries tend to collaborate more among each other ($hdi_i + hdi_j$ and $|hdi_i - hdi_j|$). Same region effect is present in all months.

In March 2020 interaction appears to have changed. Countries characterized by high scientific capacity ($n_{i,t0}$) and/or high gdp per capita (gdp_i) engage now extensively in international collaboration. And countries with high pre-pandemic CRR capacity ($c_{i,t0}$) do not anymore attract international partners in a significant way. The negative effect of development ($hdi_i + hdi_j$ and $|hdi_i - hdi_j|$) also gets stronger and more significant, meaning that less developed countries collaborate systematically with each other.

In April and May 2020 in particular past non-CRR collaboration (in the count model) and being located in the same region (in the zero model) seems to count for CRR, but not much more.

From June 2020 until December 2021 correlation patterns remain rather stable. Countries that are scientifically strong ($n_{i,t0}$) and have a strong track record of joint research in the past ($n_{ij,t0}$) are likely to have many joint CRR papers. A noteworthy trend from June 2020 to December 2021 is the increasing tendency of country pairs with high GDP per capita on average ($gdp_i + gdp_j$) to join in CRR. On the other hand, less developed economies continue to collaborate among each other ($hdi_i + hdi_j$ and $|hdi_i - hdi_j|$ both negative). Looking at pandemic specific national circumstances, we find a positive effect of (sum of) COVID-19 deaths as well as border closures on international CRR collaborations. The positive correlation of border closures suggests that it signals rather research needs than impediments to international research.

We further note a systematically increasing coefficient estimate of ‘same region’ on the number of joint CRR in the count regression. In the zero-model, being located in the same (world) region is also estimated to have a significant positive effect on the outcome. The overall effect of ‘same region’ is difficult to see because both equations combine. Doing the calculation shows that being in the same region has a strong effect overall as it scales expected joint CRR by a factor between two and six; with a peak of six in April and stabilizing at around 2.3 by December 2020 (see marginal effects in Appendix B, Table 8).

As we are estimating the *accumulated* number of two countries’ joint CRR papers, one way of interpreting the results is to consider it as a process of capital formation, i.e. formation of (joint) CRR capacity. As a very first reaction, in February 2020, prior social capital and prior CRR capacity have both been leveraged. More generic health science capacity as such became relevant immediately in March 2020. From then on generic health science capacity remained the dominant factor, be it in established interaction patterns or in forming new connections. We also find an increasing tendency of rich countries to collaborate, and a strong tendency of less developed countries to collaborate among each other. Throughout the period, being located in the same world region has a strong positive impact on joint collaboration.

4.3 Convergence to global health sciences

4.3.1 National scientific output

Countries take rapidly very similar positions in rankings on coronavirus papers as they do in rankings on other health papers. This result is evident in Fig 3A that provides the rank correlation coefficient of country rankings in CRR and non-CRR research by month. Until the outbreak in January 2020 (vertical dashed line in Fig 3A) rank correlations are rather low at around 0.2. After the outbreak, the (monthly) scientific output of countries in corona aligns with non-CRR output until a (high) correlation of 0.80 in April 2020. From then on, correlations decrease slightly to around 0.75 in December 2020.

This trend is consistent with estimation results on the (accumulated) national CRR output (Table 5), where pre-pandemic health science output ($n_{i,t0}$) becomes the dominant factor from March 2020 on.

4.3.2 Network centrality

Similarly, countries’ network centrality in the coronavirus research network aligns with their centrality in the overall health science network.

Fig 3B provides the monthly s-cores on the non-CRR network (left panel), CRR network (middle panel), and the difference of s-cores in CRR and non-CRR networks (right panel). The figure shows the 60 most central countries in the non-CRR network and applies that same ordering across all three panels. The remaining countries are highly peripheral in either network.

The left panel of Fig 3B shows that the global network hierarchy is very stable. The core is formed by developed countries and China. Centrality in the coronavirus network is more dynamic

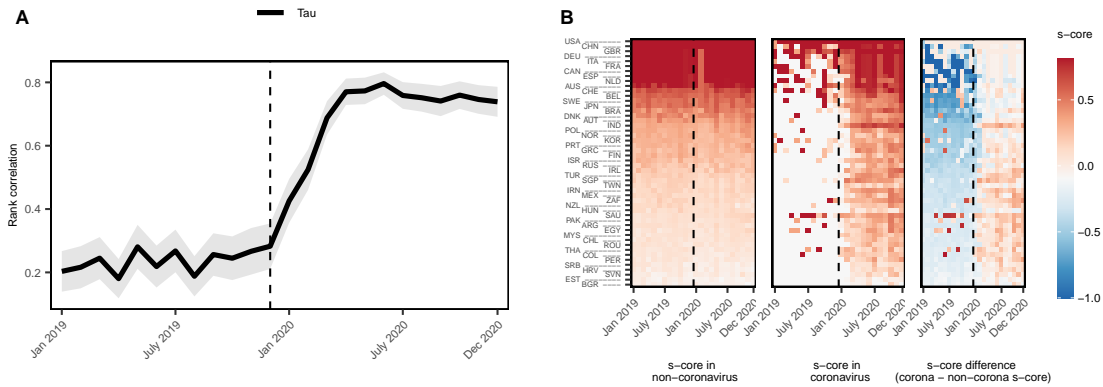


Fig. 3 Countries take on the same role in coronavirus research as in the global health sciences. (A) Correlation of country rankings by coronavirus and non-coronavirus research by month. (B) Country centrality based on s-core decomposition of the coronavirus and non-coronavirus network by month.

(middle panel). Pre-COVID-19, most countries are not involved in CRR collaborations and, hence, are found in the extreme periphery (white). The core of the CRR network includes only a few of the leading countries in health sciences. On the other hand, Saudi Arabia is part of the core in the CRR network, but peripheral in the overall health science network. The presence of Saudi Arabia in the pre-pandemic CRR network may be explained by previous regional MERS-CoV outbreaks. (Variations in core membership over time may be explained by lower research activity overall which leads to more erratic signals.) After the shock, the structure of the CRR network shifts rapidly towards the hierarchy in health science at large. This is easily seen in the right panel of Fig 3B that shows the difference between the s-core centrality in the CRR and non-CRR network. Prior to the shock, s-core differences range from -1 in dark blue (for countries at the extreme periphery in the coronavirus network and in the core in the other network), over 0 in white (same s-core in both networks), up to 1 in red (for countries in the CRR network core and peripheral in the non-CRR network). After the shock, the global core rapidly takes its role in CRR, and so does the global periphery (all countries appear in light colors with an s-core difference of around zero from April 2020).

Closer inspection further reveals that the normalized s-core of more peripheral countries tends to be somewhat higher in the CRR network as in the non-CRR network (Panel ‘s-core difference’ turns rather red than blue in 2020). This means that the distance to the highest core in the network is somewhat reduced. Thus, while the ranking by network centrality of countries in the CRR network is maintained as in the non-CRR network, hierarchy has become less steep (at least among the top 60 countries).

4.3.3 Network structure

Fig. 4 shows how CRR network structure compares to non-CRR network structure over the course of the pandemic.

Fig. 4A provides the correlation coefficient of the (accumulated) CRR and 2019 non-CRR adjacency matrix (in logs) over the observation period. During the pre-pandemic period (2019) the two adjacency matrices are increasingly but only weakly correlated with a coefficient of around 0.25. After the shock in January 2020, within three months (April 2020), the correlation coefficient jumps to 0.75, and increases further to around 0.9 until the end of the analysis period, December 2020. All correlations (pre- and post-pandemic) are highly significant based on a QAP test. Thus,

with the COVID-19 pandemic the CRR network structure rapidly approached the prior non-CRR network structure.

However, the correlation coefficient is not approaching one either, which implies that some structural difference remains between the two networks. So, what's the difference?

This can be seen in the middle panel (Fig. 4B), which provides statistics on network hierarchy (z_λ) and network modularity (z_Q) of the (accumulated) CRR network relative to the prior non-CRR network.

First consider the hierarchy measure z_λ , the line in red. We start with a z_λ -value of 1.3 in January 2019 and increase as more CRR collaborations are accumulated up to 7.7 by December 2019. Thus, before the shock the CRR network is more hierarchical than the non-CRR network. This is because only very few countries actually do have some CRR activity making worldwide CRR collaboration highly unequal. In other words, in the land of the blind, the one-eyed man is king. Immediately after the shock, in January 2020 hierarchy increases to $z_\lambda = 11.7$. That level of hierarchy is mostly kept in February 2020 but then decreases constantly to attain a value of $z_\lambda = -22.6$ in December 2020, indicating that by the end of 2020 the CRR network is much less hierarchical than the non-CRR network.

The development of network modularity (z_Q) is different. During the pre-pandemic period, 2019, modularity in the CRR network is comparable to modularity in the non-CRR network with a z_Q around zero. This means that pre-pandemic international collaborations on CRR tend to follow the same community structure as pre-pandemic international collaborations on non-CRR.²¹ The first reaction of the community statistic is observed in February 2020 where it jumps on a high level of 6.3 and immediately drops thereafter. Then, from March to December 2020, network modularity increases constantly.

For interpretation we considered z_Q for each community individually (numbers not reported here). The bump of z_Q in February 2020 originates in particular from within cluster interaction in community 1 (in particular USA, China) and to some extent also in community 6 (South America) and community 7 (Middle East and Africa). In contrast, the subsequent more continuous rise of z_Q is driven only by regional communities, i.e. all communities except community 1. In particular community 2 (Central Europe), community 3 (Norther Europe), 6 (South America) and 7 (Middle East and Africa) contribute to this trend.

The panels in Fig. 4C show the adjacency matrices of the accumulated CRR network for the three time periods, i.e. the network before the shock in Dec. 2019, immediately after the shock in Feb. 2020, and at the end of the observation period in Dec. 2020. The upper row of panels focusses on the communities. The community panels display the edge weight between two countries above expectations based on their network strength ($w - E(w)$), as used in the community detection algorithm. Countries are ordered by community membership and network centrality, such that squares along the diagonal correspond to communities (the square in the upper left corner corresponds to community 1, the square in the lower right corner to community 7). As the network accumulates, we see that communities become more densely connected. The high z_Q value indicates that the CRR network complies with the modularity structure even more than the (2019) non-CRR network from which these communities have been obtained.

The lower row of panels gives an impression on the nested hierarchy in the network. Hierarchy plots display the edge weight (w) between countries and orders countries from top to down and left to right by their eigenvalue centrality. Nested hierarchy in February 2020 is displayed in the middle panel of the lower row. Given this plot, it is apparent that hierarchy is relatively high mostly because there are very few actors on which IRC activity concentrates. The network of Dec. 2020 is apparently also highly nested, which is a special form of a hierarchical network. However, the low z_λ value indicates that the CRR network is in fact less hierarchical (or nested) than the (2019) non-CRR network.

²¹ Recall that by construction the modularity statistic takes only into account countries that have (any) CRR collaboration.

In sum, by the end of the observation period, the accumulated CRR network has become very similar to the non-CRR network, but less hierarchical and more modular (only) along regional communities.

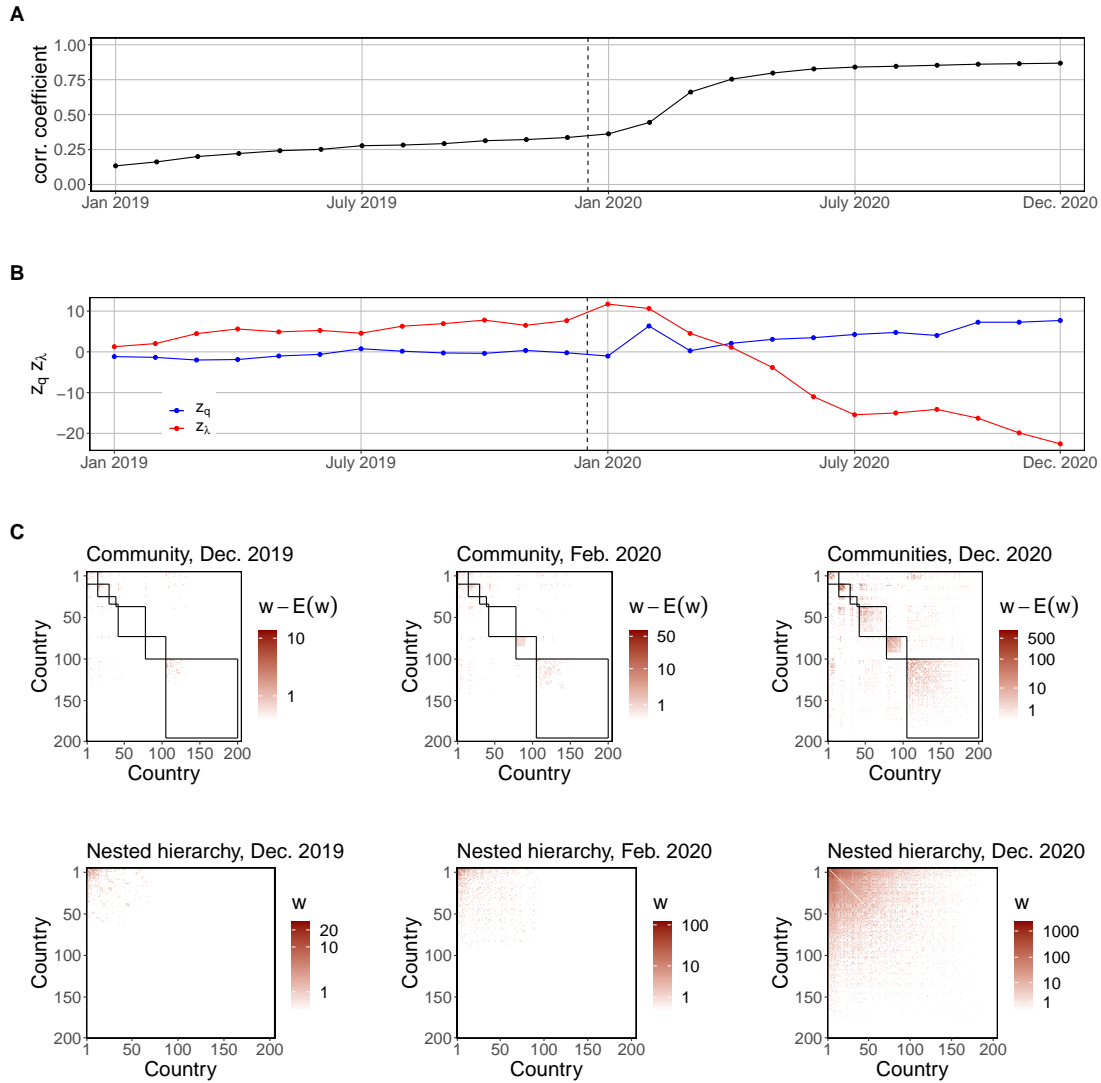


Fig. 4 Convergence of CRR network to non-CRR network. In all plots, CRR network is accumulated from Jan. 2019 up to a given month, and the non-CRR network, accumulated from Jan. 2019 to Dec. 2019, serves as the reference network. (A) Pearson’s correlation coefficient of (log of) CRR adjacency matrix and non-CRR adjacency matrix. All correlations highly significant according to QAP. (B) Development of z_q (z-value of modularity statistic Q of CRR network with null from non-CRR) and z_λ (z-value of largest eigenvalue of CRR network with null from non-CRR network). (C) CRR adjacency matrices arranged by (fixed) communities (upper panels) and nested hierarchy (lower panels). Values in community plots correspond to joint papers in excess of expectations. Values in nested hierarchy plots correspond to the number of joint papers observed. Values mapped to color in log-scale.

5 Discussion and Conclusion

Based on the population of Medline indexed journal articles, we investigated the global scientific response to the COVID-19 pandemic. Taking an (inter-)national perspective, our analyzes focused on three interrelated aspects: national scientific output, international research collaboration, and network formation. The aim has been to describe the dynamics of CRR during the pandemic and to put them into context. The following first summarizes our main results. Then we turn to the limitations of our study in order to provide policy implications and directions for future research.

Our main results may be summarized in three points: First, our three analyses on three different levels yield highly consistent results. Regressions of national CRR output (Analysis 1), and regressions of international collaboration on CRR (Analysis 2), show that national and inter-national scientific activity are closely coupled. Both follow similar dynamics and are driven by essentially the same factors. Analysis 3 on the international CRR network describes the order of global CRR that emerges as a consequence. The close coupling of national and international positioning in CRR is in strong agreement with the broader literature on (inter-)national science systems (see e.g. Adams, 2013; Frame et al., 1977; Luukkonen et al., 1992; Pan et al., 2012; Gui et al., 2018, in combination).

Second, regression analyses 1 and 2 show that CRR specific science capacity accumulated before the pandemic has been influential for national CRR output and international CRR collaboration in particular during the first three months of the pandemic. Broader health science capacity has become the dominant factor for (inter-)national CRR at least two months after the beginning of the pandemic. Pandemic related contextual factors, COVID-19 related deaths, lockdowns, and border closure, have been estimated to be systematically and positively correlated with (inter-)national CRR output mostly in the second half of 2020. These effects remain small however compared to broader health science capacity. Similarly, socio-economic conditions, i.e. population size, GDP per capita, and human development (HDI), do take some but limited explanatory power over the course of the pandemic. Regression analysis 2 further shows a strong effect of being located in the same region on international CRR collaboration.

We thus complement prior, more qualitative, studies. Haghani and Bliemer (2020) and Zhang et al. (2020) show that pre-pandemic CRR and non-CRR both provide relevant knowledge for the COVID-19 pandemic. Our regressions show a clear order in time of how coronavirus specific knowledge and more general health science capacity has been leveraged. The role of pandemic related factors has been put forward to explain observations on (inter-)national CRR in several studies (Haghani and Bliemer, 2020; Fry et al., 2020; Zhang et al., 2020). For example the dynamics of the infection process has been related to dynamics in CRR output in particular for China and Italy (e.g. Fry et al., 2020). While pandemic related explanations may well hold for individual countries, we find that such effects are *on average* relatively small. This can be explained with the fact that the global scientific community rapidly felt some urgency. Similarly, the relatively small effect of GDP and HDI suggest that, in the short run, science capacity is largely fixed. Our estimates on the relevance of regional collaboration complies with the general literature on IRC (Frenken et al., 2009) as well as observations on regional epidemics (Haghani and Bliemer, 2020; Zhang et al., 2020).

Third, consistent with regression results, Analysis 3 shows that global CRR during the pandemic rapidly converges towards the global order of broader health science capacity. Within three months, country rankings by national CRR output approach rankings by pre-pandemic non-CRR output (Analysis 3.1), countries take the same (centrality) position in the international CRR network as in the pre-pandemic broader health science network (Analysis 3.2), and the CRR network converges to the structure of the pre-pandemic health science network (Analysis 3.3). However, the alignment is not perfect. Global CRR deviates systematically from the attracting global distribution of broader health science before the pandemic in that the global CRR network is significantly less hierarchic and more regional than global health science (Analysis 3.3).

These results are all consistent with prior empirical studies on the expansion of the international CRR network (Cai et al., 2020; Fry et al., 2020; Haghani and Bliemer, 2020; Zhang et al., 2020). Our study clarifies however that the pre-existing global health science system systematically structures the expansion of global CRR, and add that global CRR systematically deviates in relevant dimensions from the inherited global structure.

One major limitation of this study is the arguably rough division of research papers into two categories; CRR and non-CRR. The following discussion takes this limitation into account to delineate policy implications and to outline potential avenues for future research.

The COVID-19 shock opened a scientific race. In order to start off, most scientists had to re-orient or adapt ongoing research. Our observation that countries with prior experience in CRR occupied pole positions but countries of the global health science core rapidly took the lead, suggests that a broad science base provides the ability to re-orient research effectively. This is an argument for a more autonomous development of the science system, because it deliberates policy and society from the burden (and impossibility) to specify a detailed scientific agenda to meet future challenges.

However, having no notion of distance between non-CRR and CRR, we can say nothing about the degree of flexibility in research orientation, nor what exactly underlies such flexibility. In particular we would like to know to what extent CRR research during the pandemic benefited from very generic scientific capacity (obtained through basic science), from closely related scientific capacity (broad array of applied sciences somehow related to CRR), or their interaction. This would give some indication for science policy on how to balance basic science against applied research to guard against future challenges.

Immediately after the shock the need for international collaboration has been emphasized by all stakeholders. Our results show that national CRR output and international CRR collaboration have been so closely aligned that they should be considered as two dimensions of the same activity. Furthermore, we found that the international collaboration network on CRR largely followed the prior international health science network. The implication is that any (national) strategy to develop science capacity must take a (global) systems perspective — the embedding of the national science system in the international collaboration network is part of the (slowly accumulating and path dependent) national science base.

It is hard to spell out that insight without raising two follow-up questions: What would be an ‘appropriate’ network structure (and for whom)? How could such an ‘appropriate’ network structure be obtained? Our observation that the international CRR network systematically deviates from the broader health science network may help to scratch on these issues — but with major limitations.

According to our analysis, the international CRR network exhibits a less steep hierarchy and stronger regionalization than the international broader health science network. Yet, what exactly that signifies remains unclear. Recall that we do not differentiate papers by quality, we only measure quantities. COVID-19 has been a hot science topic in 2020, and it may be that many papers have been hastily crafted but add little knowledge. A higher propensity of low quality papers in countries with lower scientific capacity could contribute to our finding of a less steep global hierarchy as measured on output quantity, while global hierarchy in terms of research excellence may be in fact maintained. Similarly, increased regionalization, as we observe it, is consistent with more independent science in the different world regions and hence more independent accumulation of science capacity on a regional level. But it is also consistent with maintained but unresolved resource (inter-)dependence and increasing vertical stratification (by scientific excellence) as well as horizontal stratification (by research topics). Thus, the paper at hand clearly describes structural changes, but leaves the question of individual network gains and global network efficiency for future research.

Finally, the underlying cause of structural changes in global CRR remains unclear. On the one hand, stronger regionalization may be largely due to the pandemic context. Our regressions include relevant parameters such as national pandemic policy measures and virus spreading dynamics, but these are certainly no perfect controls. Given those controls bite, our results are consistent with the

idea that we observe the accentuation of a long-term trend towards regionalization; a trend already observed by Fitzgerald et al. (2021) for the sciences in general. In case the pandemic context is the underlying cause, the organization of global science during the pandemic may be as transitory as the pandemic itself. If the organization of global CRR during the pandemic accentuates a general global trend, it may not only cast the shadow of a future multi-polar world but even accelerate that trend.

References

- Abdi, H., Williams, L.J. “Jackknife” in *Encyclopedia of Research Design*, NJ Salkind, Ed., pp. 655-660 (Thousand Oaks, CA, SAGE Publications Inc., 2010).
- Adams, J. The fourth age of research. *Nature* **497**, 557-560 (2013).
- Allik, J., Lauk, K., Realo, A. Factors predicting the scientific wealth of nations. *Cross-Cultural Research* 54(4):364-397 (2020).
- Aviv-Reuven, S., Rosenfeld, A. Publication patterns’ changes due to the Covid-19 pandemic: A longitudinal and short-term scientometric analysis, *arXiv* **2010.02594**, (2020)
- Beaver, D.deB., Rosen, R. (1978) Studies in scientific collaboration: Part I. The professional origins of scientific co-authorship. *Scientometrics* 1:65-84.
- Beaver, D.deB., Rosen, R. (1978) Studies in scientific collaboration: Part III. Professionalization and the natural history of modern scientific co-authorship. *Scientometrics* 1:65-84.
- Bell, M.L., Fong, K.C. Gender differences in first and corresponding authorship in public health research submissions during the Covid-19 pandemic. *AJPH* **111**, 159-163, <https://doi.org/10.2105/AJPH.2020.305975> (2021).
- Ben-David, J. *The scientist’s role in society*, Prentice-Hall, Inc., New Jersey, USA (1971).
- Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech.* **P10008**, (2008).
- Braun, T., Glänzel, W. International collaboration: Will it be keeping alive East European research? *Scientometrics* 36(2):247-254 (1996).
- X Cai, CV Fry, C Wagner, International Collaboration During the COVID-19 Crisis: Autumn 2020 Developments. Available at SSRN: <https://ssrn.com/abstract=3729672>, <http://dx.doi.org/10.2139/ssrn.3729672> (2020).
- Cameron, A.C., Trivedi, P.K. *Microeconometrics: Methods and Applications* Cambridge University Press, New York (2005).
- Cantner, U., Rake, B. International research networks in pharmaceuticals: structure and dynamics. *Research Policy* 43:333-348 (2014).
- Chahrour, M., Assi, S., Bejjani, M., Nasrallah, A.A., Salhab, H., Fares, M., Khachfe, H.H. A bibliometric analysis of COVID-19 research activity: a call for increased output. *Cureus* **12(3)**, e7357 (2020).
- Dekker, D., Krackhardt, D., Snijders, T.A.B. Sensitivity of MRQAP tests to collinearity and autocorrelation conditions. *Psychometrika* 72(4):563-581, (2007).
- Dong E., Du, H., Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Inf Dis.* 20(5):533-534. doi: 10.1016/S1473-3099(20)30120-1 (2020).
- EC-OECD. STIP Compass: International Database on Science, Technology and Innovation Policy (STIP), edition May 28, 2021, <https://stip.oecd.org> (2021).
- Eidsaa, M., Almaas, E. s-core network decomposition: A generalization of k-core analysis to weighted networks. *Phys. Rev. E* **88**, 062819 (2013).
- Emond, E.J., Mason, D.W. A new rank correlation coefficient with application to the consensus ranking problem. *J. Multi-Crit. Decis. Anal.* **11**, 17-28 (2002).
- Feenstra, R.C., Inklaar, R., Timmer, M.P. The Next Generation of the Penn World Table, *American Economic Review*, 105(10): 3150-3182, available for download at www.gdpc.net/pwt, (2015).

- Fitzgerald, J., Ojanerä, S., O'Clery, N. Is academia becoming more localised? The growth of regional knowledge networks within international research collaboration. *Applied Network Science* 6:38, <https://doi.org/10.1007/s41109-021-00371-w>, (2021).
- Frame, J.D., Carpenter, M.P. International Research Collaboration. *Social Studies of Science*, 9:481-497 (1979).
- Frame, J.D., Narin, F., Carpenter, M.P. The distribution of world science. *Social Studies of Science* 7:501-516 (1977).
- Frenken, K., Hardeman, S., Hoekman, J. Spatial scientometrics: Towards a cumulative research program. *Journal of informetrics* 3:222-232 (2009).
- Friedmann, J. A general theory of polarized development. *The Ford Foundation Urban and Regional Advisory Program in Chile* Int-2296 (1967).
- Fry, C.V., Cai, X., Zhang, Y., Wagner, C.S. Consolidation in a crisis: Patterns of international collaboration in early COVID-19 research. *PLoS ONE* **15(7)**, e0236307 (2020).
- Gai N, Aoyama K, Faraoni D, Goldenberg NM, Levin DN, Maynes JT, et al. General medical publications during COVID-19 show increased dissemination despite lower validation. *PLoS ONE* 16(2): e0246427. <https://doi.org/10.1371/journal.pone.0246427> (2021).
- Gazni, A., Sugimoto, C.R., Didegah, F. Mapping world scientific collaboration: Authors, Institutions, and Countries. *Journal of the American Society for Information Science and Technology* 63(2):323-335 (2012).
- Glänzel, W. National characteristics in international scientific co-authorship relations. *Scientometrics* 51(1):169-115 (2001).
- Gui, Q., Liu, C., Du, D. International knowledge flows and the role of proximity. *Growth and Change*, 49(3):532-547 (2018).
- Guleid FH, Oyando R, Kabia E, et al. A bibliometric analysis of COVID-19 research in Africa. *BMJ Global Health*; 6:e005690. doi:10.1136/bmjgh-2021-005690 (2021).
- Haghani, M., Bliemer, M. C. Covid-19 pandemic and the unprecedented mobilisation of scholarly efforts prompted by a health crisis: Scientometric comparisons across SARS, MERS and 2019-nCov literature. *Scientometrics* **125(3)**, 2695-2726 (2020).
- Hale, T.?, Angrist, N.?, Goldszmidt, R. , Kira, B., Petherick, A.?, Phillips, T., Webster, S., Cameron-Blake, E.?, Hallas, L., Majumdar, S., Tatlow, H. A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). *Nature Human Behaviour* <https://doi.org/10.1038/s41562-021-01079-8> (2021).
- Háncean, M.-G., Perc, M., Lerner, J. The coauthorship networks of the most productive European researchers. *Scientometrics* 126:201-224 (2021).
- Hoekman, J., Frenken, K., van Oort, F. The geography of collaborative knowledge production in Europe. *The Annals of Regional Science* 43:721-738 (2009).
- Homolak, J., Kodvanj, I., Virag, D. Preliminary analysis of COVID-19 academic information patterns: a call for open science in the times of closed borders. *Scientometrics* 124:2687-2701 (2020).
- Horlings, E., van den Besselaar, P. Convergence in science growth and structure of worldwide scientific output, 1993-2008. *Working Paper. The Hague: Rathenau Instituut* **6064471**, (2013). Available at <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6064471>. Accessed 19.01.2021.
- Jones, B.F., Wuchty, S., Uzzi, B. Multi-university research teams: Shifting impact, geography, and stratification in science. *Science* **322**, 1259-1262 (2008).
- Jonkers, K., Tijssen, R. Chinese researchers returning home: Impacts of international mobility on research collaboration and scientific productivity. *Scientometrics*, 77(2):309-333 (2008).
- Kato, M., Ando, A. National ties of international scientific collaboration and researcher mobility found in Nature and Science. *Scientometrics* 110:673-694 (2017).
- Katz, J.S. Scale-independent indicators and research evaluation. *Science and Public Policy* 27(1):23-36 (2000).
- Katz, J.S., Martin, B.R. What is research collaboration? *Research Policy* 26(1):1-18 (1997).

- King, D.A. The scientific impact of nations. *Nature* 430:311-316 (2004).
- Kinsella, C.M., Santos, P.D., Postigo-Hidalgo, I., Folguez-González, A., Passchier, T.C., Szil-lat, K.P., et al. (2020) Preparedness needs research: How fundamental science and international collaboration accelerated the response to COVID-19. *PLoS Pathog* 16(10): e1008902. <https://doi.org/10.1371/journal.ppat.1008902> (2020).
- König, M.D., Tessone, C.J., Zenou, Y. Nestedness in networks: A theoretical model and some applications. *Theor. Econ.* **9**, 695-752 (2014).
- Krackhardt, D. QAP partialling as a test of spuriousness. *Social Networks* **9**, 171-186 (1987).
- Kuznets, S. *Modern economic growth*, Oxford & IBH Publishing Co., New Delhi, India.
- Lancichinetti, A., Radicchi, F., Ramasco, J.J., Fortunato, S. Finding Statistically Significant Communities in Networks. *PLoS ONE* **6(4)**, e18961 (2011).
- Lee, J.J., Haupt, J.P. Scientific globalism during a global crisis: research collaboration and open access publications on COVID-19. *Higher Education*, <https://doi.org/10.1007/s10734-020-00589-0>, (2020).
- Lee, D., Heo, Y., Kim, K. A strategy for international cooperation in the COVID-19 pandemic era: Focusing on national scientific funding data. *Healthcare* 8, 204, 204; doi:10.3390/healthcare8030204, (2020).
- Liu, M. et al. Can pandemics transform scientific novelty? Evidence from COVID-19. *arXiv* 2009.12500 (2021).
- Lu, R. et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet* 395(10224):565-574 (2020).
- Luukkonen, T., Persson, O., Sivertsen, G. Understanding Patterns of International Scientific Collaboration. *Science, Technology, and Human Values* 17(1):101-126 (1992).
- May, R.M. The scientific wealth of nations. *Science* 275(5301):793-796 (1997).
- Mendes, T., Carvalho, L. Shifting geographies of knowledge production: the coronavirus effect. *Tijdschrift voor Economische en Sociale Geografie* 111(3):205-210 (2020).
- The social capital of global ties in science: The added value of international collaboration. *Review of Policy Research* 27(4):389-414 (2010).
- Mervis, J. (2021) Biden, Congress roll out big plans to expand National Science Foundation, *Sciencemag*, Science and Policy, doi:10.1126/science.abi8778 .
- Newman, M.E.J. Analysis of weighted networks. *Physical Review E*, 70, 056131 (2004).
- Niu, F., Qiu, J. Network structure, distribution and growth of Chinese international research collaboration. *Scientometrics* 98:1221-1233 (2014).
- Palayew, A., Norgaard, O., Safreed-Harmon, K., Andersen, T. H., Rasmussen, L. N., & Lazarus, J. V. (2020). Pandemic publishing poses a new COVID-19 challenge. *Nature Human Behaviour*, 4(7), **666-669**.
- Pan, R.K., Kaski, K., Fortunato, S. World citation and collaboration networks: uncovering the role of geography in science. *Sci. Rep.* **2**, 902 (2012).
- Pavitt, K. What makes basic research economically useful? *Research Policy* 20:109-119 (1991)
- The limits to internationalization of scientific research collaboration. *J Technol Transf* 34:76-94 (2009).
- Quan, W., Mongeon, P., Larivière, V. China's increasing leadership in scientific collaboration. *Science, Technology and Innovation indicators in transition STI 2018 Conference Proceedings* <https://hdl.handle.net/1887/64521> (2018).
- Radanliev, P., De Roure, D., Walton, R., Van Kleek, M., Santos, O., Mantilla Montalvo, F., La Treall, M. What country, university or research institute, performed the best on COVID-19? Bibliometric analysis of scientific literature. *arXiv* **2005.10082**, (2020).
- Reichardt, J., Bornholdt, S. Statistical mechanics of community detection. *Physical Review E*, 74, 10.1103/PhysRevE.74.016110, (2006).
- Rommel, A. US National Science Foundation set for a funding boom. *Nature*, 593:19:20, (2021).

- Schot, J., Steinmueller, W.E. Three frames for innovation policy: R&D, systems of innovation and transformative change. *Research Policy*, 47:1554-1567, (2018).
- Stephan, P.E. The economics of science funding for research. International Centre for Economic Research Working Paper Series, N.12/2010 (2010).
- Stephan, P.E., Levin, S.G. Exceptional contributions to US science by the foreign-born and foreign-educated. *Population Research and Policy Review*, 20:59-79 (2001).
- Wagner, C.S., Whetsell, T.A., Leydesdorff, L. Growth of international collaboration in science: revisiting six specialties. *Scientometrics* 110:1633-1652 (2017).
- Wang, L., Wang, X. Who sets up the bridge? Tracking scientific collaborations between China and the European Union. *Research Evaluation*, 26(2):124-131 (2017).
- Network structure, self-organization, and the growth of international collaboration in science. *Research Policy* 34:1608-1618 (2005).
- Wagner, C.S., Whetsell, T.A., Mukherjee, S. International research collaboration: Novelty, conventionality, and atypicality in knowledge recombination. *Research Policy* 48:1260-1270 (2019).
- Yonezawa, A. The internationalization of Japanese higher education: policy debates and realities, in S. Marginson et al. (eds.), *Higher Education in the Asia-Pacific*, Higher Education Dynamics 36, Springer Science+Business Media B.V. (2011).
- Zhang, L., Zhao, W., Sun, B., Huang, Y., Glänzel, W. How scientific research reacts to international public health emergencies: a global analysis of response patterns. *Scientometrics* **124**, 747-773 (2020).
- Zhu, J., Liu, W. Comparing like with like: China ranks first in SCI-Indexed research articles since 2018. *Scientometrics* 124:1691-1700 (2020).
- Zitt, M., Bassecoulard, E., Okubo, Y. Shadows of the past in international cooperation: Collaboration profiles of the top five producers of science. *Scientometrics* 47(3):627-657 (2000).
- Zucker, L.G., Darby, M.R., Brewer, M.B. Intellectual human capital and the birth of US biotechnology enterprises. *The American Economic Review*, 88(1):290-306 (1998).

A Sample

A.1 Timelags from research over acceptance to entry into the PubMed dataset

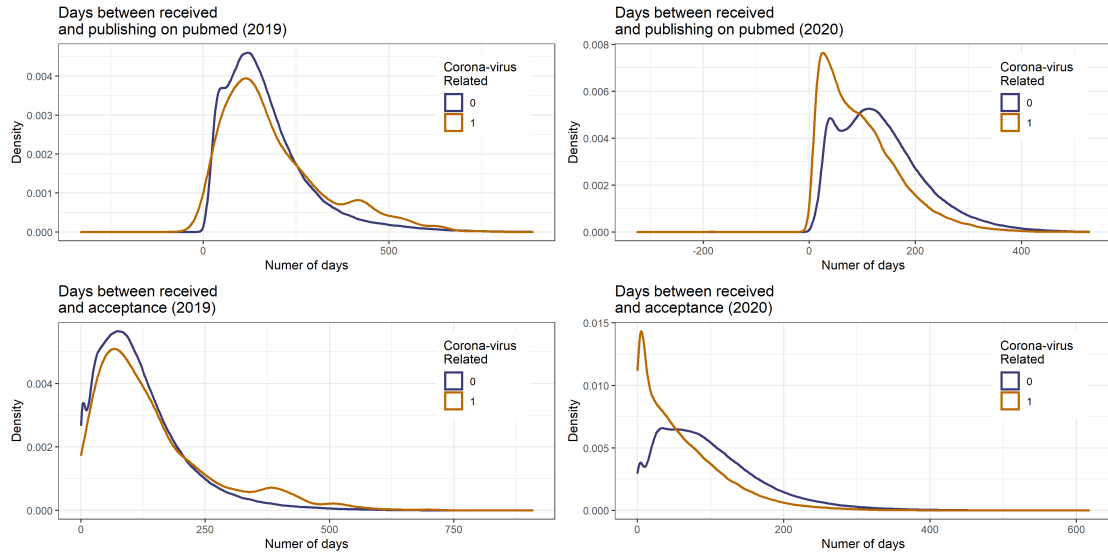


Fig. 5 Timelags from submission date (data received) to (clockwise, starting with the upper left panel): i) publishing on PubMed in 2019, ii) publishing on PubMed in 2020, iii) journal acceptance in 2020, iv) journal acceptance in 2019. Our sample is from PubMed. Therefore, the analysis period is based on the timelag distribution of the upper right panel.

Table 7 Descriptive statistics on scientific production (at submission date)

	<i>Coronavirus related</i>		<i>Others Documents</i>		<i>All Documents</i>	
	2019	2020	2019	2020	2019	2020
Nb of authors	6/7.36 (4.86)	5/6.86 (7.05)	6/6.66 (17.51)	6/6.65 (14.67)	6/6.66 (17.5)	6/6.66 (14.2)
Solo authors	2.83	6.72	3.38	3.22	3.38	3.5
International collab.	25.44	23.06	23.57	23.63	23.57	23.58)
Nb of country	1/1.36 (0.92)	1/1.35 (1.29)	1/1.31 (1.05)	1/1.32 (1.01)	1/1.31 (1.05)	1/1.33 (1.03)
Days received-accept	110.04/143.31 (119.52)	45/60.87 (58.53)	99.96/118.97 (93.82)	84.96/97.96 (73.24)	99.96/119 (93.85)	81.96/94.88 (72.86)
Share aff. captured	1/0.95 (0.2)	1/0.92 (0.26)	1/0.93 (0.24)	1/0.95 (0.21)	1/0.93 (0.24)	1/0.94 (0.22)
# Document	743(0.1%)	74250(8.2%)	775738(99.9%)	830914(91.8%)	776481	905164

Notes: Binary indicators in [%], for continuous measures [median/mean (s.d.)].

B Regression analysis: Marginal effects

Marginal effects of the zero-inflated negative binomial regression of (accumulated) joint CRR papers are obtained from the estimated coefficient in Table 6. To keep it simple, we calculate numerically to what extent a ‘small’ increase in the value of one variable affects on average the expected outcome.²² For all variables, except for the ‘same region’ dummy, a small increase is considered to be a one percent increase of the observed value (before taking logs). Observations with value zero are shifted to one. The marginal effect for one observation is then the ratio of the prediction based on the increased value over the prediction based on the originally observed value. The dummy variable ‘same region’ is the exception in that we compare the expected outcome based on a value of zero (not in same region) with the expected outcome based on a value of one (same region) for each observation. The average marginal effect is the average over all individual marginal effects.

Table 8 provides the average marginal effects obtained for each regression. As a reading example, we discuss the marginal effects of factors with ‘significant’ coefficients in the February 2020 regression in Table 6: In terms of average marginal effects, a one percent increase of $(n_{i,t0} + n_{j,t0})$ translates into a 1.1 percent increase of expected joint CRR papers. Average marginal effect of ‘same region’ appears to be large: On average, being located in the same region, rather than not, increases the expected number of joint CRR papers by about four times (i.e. 384 percent).²³

Looking at the count model of the intensity of dyadic CRR, we estimate a significant coefficient of one for joint non-CRR in the prior year 2019. A one percent increase of that variable translates into an expected increase of the outcome by around 36 percent. On the other hand, a one percent increase of (sole) non-CRR papers reduces the expected outcome by half a percent. Countries with specific CRR capacity attract collaborators ($c_{i,t0}$) in that a one percent increase is expected to increase the outcome by seven percent. A one percent increase of country size (pop) and economic wealth (gdp per capita) shift average expected outcomes by 0.25 percent and 0.65 percent respectively. One percent increases of $(hdi_i + hdi_j)$ and $(|hdi_i - hdi_j|)$ reduce the outcome by 1.9 and 0.5 percent respectively.

Considering the table as a whole, we see that being located in the same region and having many prior non-CRR ties are associated with the largest marginal effects.

²² Note that we cannot calculate significance values of the marginal effects by using the ‘Delta’ method, because standard deviations based on asymptotics are not reliable (see Methods).

²³ The high marginal effect of the ‘same region’ dummy variable compared to the marginal effects of other variables with comparable coefficient estimates is due to the binary nature of the dummy. Marginal effects of the dummy are based on a comparison of zero value to a value of one. For all other variables, average marginal effects are calculated on a one percent increase of the underlying variable, before taking logs for estimation. This corresponds to adding one percent (0.01) after taking logs. Thus, we consider a step size for the dummy variable ‘same region’ that is 100 times larger as for all other variables.

Table 8 Marginal effects of zero-inflated negative binomial regression of (accumulated) joint CRR papers (in percent).

	Feb. '20	Mar. '20	Apr. '20	May '20	June '20	July '20	Aug. '20	Sept. '20	Oct. '20	Nov. '20	Dec. '20
$n_{i,t0}$	39.02	21.66	27.53	28.96	25.45	23.77	23.55	23.80	24.04	24.54	25.07
$c_{ij,t0}$	12.53	3.03	0.33	0.37	3.68	4.06	3.92	2.36	1.72	1.91	1.74
$n_{i,t0}$	-0.53	0.26	0.09	0.04	0.15	0.17	0.17	0.17	0.15	0.15	0.14
$c_{i,t0}$	7.01	0.39	0.86	1.05	-0.06	0.11	0.16	0.48	0.67	0.33	0.36
$death_i + death_j$	-43.25	0.13	0.12	0.08	0.08	0.10	0.12	0.14	0.18	0.15	0.14
$ death_i - death_j $	94.60	-0.28	-0.13	-0.05	-0.04	-0.06	-0.06	-0.08	-0.10	-0.08	-0.08
$locked_{i,t'}$	-3.95	1.75	1.27	1.00	0.74	1.14	0.84	0.77	0.80	0.63	0.65
$closed_{i,t'}$	0.46	0.46	0.10	0.03	0.05	0.08	0.11	0.11	0.11	0.10	0.10
$pop_{i,t0}$	0.23	-0.01	0.04	0.03	-0.01	0.01	0.02	0.00	0.01	0.02	0.02
$gdp_i + gdp_j$	0.67	0.41	0.16	0.07	0.22	0.23	0.31	0.33	0.30	0.34	0.34
$ gdp_i - gdp_j $	0.06	0.03	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.04
$hdi_i + hdi_j$	-2.03	-2.09	-0.92	-0.88	-1.64	-1.55	-1.73	-1.81	-1.64	-1.75	-1.75
$ hdi_i - hdi_j $	-0.49	-0.37	-0.42	-0.36	-0.42	-0.41	-0.43	-0.43	-0.41	-0.42	-0.42
$n_{i,t0} + n_{j,t0}$	1.08	3.27	1.21	0.95	0.84	0.99	1.00	0.97	1.18	1.04	0.99
$ n_{i,t0} - n_{j,t0} $	-0.38	-2.62	-0.66	-0.43	-0.41	-0.55	-0.56	-0.52	-0.69	-0.55	-0.50
distance	0.02	0.30	-0.16	-0.22	-0.14	-0.18	-0.22	-0.20	-0.20	-0.20	-0.17
same region	339.05	464.83	615.06	422.27	235.25	225.90	206.43	237.96	245.15	231.95	234.30

Notes: Average marginal effects have been obtained by comparing for each variable predictions based on observations with predictions on observed increased by one percent. In case observed value is zero, we compare it with having one. Marginal effects of the dummy 'same region' are obtained by comparing zero with one outcomes.

C Community detection in the 2019 non-CRR network

The procedure is the same as in Fitzgerald et al. (2021), but relevant details are repeated here for readers' convenience.

We identify communities with the spin glass community detection algorithm of Reichardt and Bornholdt (2006). The objective criterion to be maximized is

$$Q = \frac{1}{m} \left(\sum_{ij} c_{ij} - \gamma \frac{k_i k_j}{m} \right) d(\sigma_i, \sigma_j),$$

where m is the sum over all edges, c_{ij} is the number of joint papers between country i and j , k_i (k_j) is the strength of country i (j) (i.e. the sum over all weighted edges of the country), σ_i (σ_j) indicates the community of i (j) in a given network partition, with the function $d()$ evaluating to one if both countries belong to the same community (i.e. $\sigma_i = \sigma_j$) and zero else. Finally, the tuning parameter γ trades-off the two objectives of having high edge weights within clusters against having few edge weights between clusters. A smaller (higher) γ tends to yield less (more) clusters.

The optimization is done through simulated annealing (as implemented in the R-package 'igraph'). Simulated annealing is a stochastic optimization algorithm and hence may provide different partitions for the same data and parameter settings. In order to find robust communities, we run many (i.e. 100) times the optimization algorithm over a grid of γ , starting from 0.8 to 1.6. The resulting community partitions for a given γ are then compared, each partition against all others. Similarity of two partitions is measured by the 'Variation of Information' (VI). The following definition is paraphrased from (Fitzgerald et al., 2021, Appendix 4):

"Consider two partitions X and Y of a set A into disjoint subsets, $X = X_1, X_2, \dots, X_k$ and $Y = Y_1, Y_2, \dots, Y_l$, VI is defined as follows. Let $n = \sum_i |X_i| = \sum_j |Y_j| = |A|$, $p_i = |X_i|/n$, $q_j = |Y_j|/n$, $r_{ij} = |X_i \cap Y_j|/n$. Then the normalised variation of information between the two partitions is:

$$VI(X, Y) = -1 \log N \sum_{i,j} r_{ij} [\log(r_{ij}/p_i) + \log(r_{ij}/q_j)]."$$

Table 9 provides the VIs for 100 optimizations for each γ on the grid. The most robust partitioning is obtained for $\gamma = 0.8$. The resolution of that partitioning however is very low, dividing the world into one cluster of less developed economies (consisting of countries in Africa, Middle East, South Asia, and India, Mongolia, and Kazakhstan) and another cluster including all other countries. The partitioning with $\gamma = 1.2$ is slightly less robust but at much higher resolution. Figure 2 shows the partitioning with lowest average VI under $\gamma = 1.2$. We choose that partition as our benchmark. Interestingly, the partitioning that we obtain on the 2019 non-CRR network is not identical but very similar to the partitioning obtained by Fitzgerald et al. (2021, Fig.5b) based on all Scopus publications in 2015.

Table 9 Variation of Information for varying γ over 100 optimizations.

γ	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5	1.6
VI	0.099	0.302	0.465	0.161	0.102	0.349	0.255	0.158	0.232

The partitioning displayed in Figure 2 consists out of the following communities (ordered by average degree of their members):

- Cluster 1: Australia, Canada, China, French Polynesia, Gibraltar, Grenada, Japan, Macao SAR China, New Zealand, Singapore, Solomon Islands, South Korea, Taiwan, United Kingdom, United States
- Cluster 2: Austria, Belgium, Curacao, France, Germany, Italy, Liechtenstein, Luxembourg, Martinique, Monaco, Netherlands, Reunion, Spain, Switzerland, Vatican City
- Cluster 3: Denmark, Faroe Islands, Finland, Greenland, Guinea-Bissau, Iceland, Isle of Man, Norway, Sweden
- Cluster 4: Israel, Jersey, Montserrat
- Cluster 5: Albania, Angola, Armenia, Azerbaijan, Belarus, Bosnia & Herzegovina, Bulgaria, Croatia, Cyprus, Czechia, Estonia, Georgia, Greece, Hungary, Ireland, Kazakhstan, Latvia, Lithuania, Malta, Moldova, Montenegro, North Macedonia, Poland, Portugal, Romania, Russia, San Marino, Serbia, Sint Maarten, Slovakia, Slovenia, Tajikistan, Turkey, Turkmenistan, Ukraine, Uzbekistan
- Cluster 6: Andorra, Antigua & Barbuda, Argentina, Belize, Bolivia, Brazil, Chile, Colombia, Costa Rica, Cuba, Dominica, Dominican Republic, Ecuador, El Salvador, French Guiana, Guatemala, Guyana, Honduras, Mexico, Nicaragua, Panama, Paraguay, Peru, St. Kitts & Nevis, Suriname, Uruguay, Venezuela
- Cluster 7: Afghanistan, Algeria, Aruba, Bahamas, Bahrain, Bangladesh, Barbados, Benin, Bermuda, Bhutan, Botswana, Brunei, Burkina Faso, Burundi, Cape Verde, Cambodia, Cameroon, Cayman Islands, Central African Republic, Chad, Congo - Kinshasa, Egypt, Eritrea, Ethiopia, Fiji, Gabon, Gambia, Ghana, Guadeloupe, Guinea,

Haiti, India, Indonesia, Iran, Iraq, Côte d'Ivoire, Jamaica, Jordan, Kenya, Kuwait, Kyrgyzstan, Laos, Lebanon, Lesotho, Liberia, Libya, Madagascar, Malawi, Malaysia, Maldives, Mali, Mauritania, Mauritius, Mongolia, Morocco, Mozambique, Myanmar (Burma), Namibia, Nepal, New Caledonia, Niger, Nigeria, Oman, Pakistan, Palau, Palestinian Territories, Philippines, Qatar, Congo - Brazzaville, Rwanda, St. Lucia, St. Vincent & Grenadines, Samoa, Saudi Arabia, Senegal, Seychelles, Sierra Leone, Somalia, South Africa, Sri Lanka, Sudan, Syria, Tanzania, Thailand, Togo, Tonga, Trinidad & Tobago, Tunisia, Turks & Caicos Islands, Uganda, United Arab Emirates, Vietnam, Yemen, Zambia, Zimbabwe

No cluster: Djibouti, Åland Islands, British Virgin Islands, Sao Tome and Principe, Tuvalu

D Extended analysis up to April 2021

Table 10 National output regression (Jan.-Apr. 2021).

	Jan. '21	Feb. '21	Mar. '21	Apr. '21
n_{t0}	0.946 *** (0.05)	0.946 *** (0.05)	0.948 *** (0.048)	0.952 *** (0.049)
c_{t0}	0.016 (0.036)	0.015 (0.036)	0.01 (0.035)	0.007 (0.036)
$death_{t'}$	0.052 (0.029)	0.049 (0.029)	0.043 (0.028)	0.035 (0.029)
$locked_{t'}$	0.06 ** (0.02)	0.064 ** (0.02)	0.066 *** (0.019)	0.066 ** (0.02)
$border_{t'}$	-0.013 (0.021)	-0.015 (0.021)	-0.016 (0.02)	-0.015 (0.02)
pop_{t0}	0.007 (0.04)	0.004 (0.04)	0.011 (0.039)	0.015 (0.04)
gdp_{t0}	0.114 * (0.053)	0.11 * (0.052)	0.12 * (0.051)	0.123 * (0.052)
hdi_{t0}	-0.159 ** (0.058)	-0.151 ** (0.057)	-0.156 ** (0.056)	-0.155 ** (0.057)
R^2	0.947	0.948	0.951	0.95

Notes: Recall from Section 3.3 that c_{t0} and n_{t0} proxy initial scientific capacity in CRR and non-CRR respectively.

Table 11 Zero-inflated negative binomial model of (accumulated) joint coronavirus related papers (Jan.-Apr. 2021), est. coefficient (z-value, p-value)

	Jan. '20	Feb. '20	Mar. '20	Apr. '20
<i>Count-model</i>				
Intercept	-4.023*** (-12.164, 0)	-3.885*** (-12.163, 0)	-3.99*** (-12.544, 0)	-4.062*** (-12.772, 0)
$n_{ij,t0}$	0.75*** (50.115, 0)	0.751*** (51.896, 0)	0.752*** (52.291, 0)	0.744*** (51.752, 0)
$c_{ij,t0}$	0.032 (0.76, 0.248)	0.025 (0.622, 0.321)	0.022 (0.535, 0.31)	0.024 (0.581, 0.343)
$n_{i,t0}$	0.149*** (9.859, 0)	0.135** (9.315, 0.001)	0.139*** (9.594, 0)	0.149*** (10.317, 0)
$c_{i,t0}$	0.006 (0.492, 0.458)	0.01 (0.857, 0.393)	0.006 (0.482, 0.47)	0.006 (0.467, 0.434)
$death_i + death_j$	0.112** (6.177, 0.004)	0.106** (6.165, 0.006)	0.101** (6.027, 0.009)	0.101** (5.867, 0.009)
$ death_i - death_j $	-0.059* (-4.181, 0.038)	-0.056* (-4.214, 0.036)	-0.055* (-4.327, 0.035)	-0.056* (-4.247, 0.034)
$locked_{i,t'}$	0.028* (6.364, 0.014)	0.03** (7.323, 0.005)	0.031** (7.779, 0.001)	0.03** (7.627, 0.009)
$closed_{i,t'}$	0.025** (6.596, 0.005)	0.022** (5.95, 0.009)	0.021** (5.766, 0.006)	0.022** (6.092, 0.007)
$pop_{i,t0}$	0.023 (1.953, 0.224)	0.027 (2.359, 0.186)	0.031 (2.653, 0.133)	0.029 (2.484, 0.172)
$gdp_i + gdp_j$	0.358*** (7.735, 0)	0.359*** (8.007, 0)	0.385** (8.609, 0.001)	0.396** (8.83, 0.001)
$ gdp_i - gdp_j $	0.033 (2.789, 0.121)	0.037 (3.183, 0.069)	0.037 (3.194, 0.087)	0.036 (3.103, 0.078)
$hdi_i + hdi_j$	-2.363*** (-12.007, 0)	-2.21*** (-11.588, 0)	-2.285*** (-12.033, 0)	-2.348*** (-12.348, 0)
$ hdi_i - hdi_j $	-1.835*** (-11.816, 0)	-1.834*** (-12.17, 0)	-1.867*** (-12.443, 0)	-1.848*** (-12.325, 0)
distance	-0.044 (-2.525, 0.16)	-0.064* (-3.848, 0.034)	-0.071* (-4.233, 0.024)	-0.077* (-4.576, 0.023)
same region	0.165*** (5.152, 0)	0.175*** (5.687, 0)	0.168*** (5.446, 0)	0.163*** (5.276, 0)
$\log(\theta)$	1.251*** (28.604, 0)	1.322*** (30.667, 0)	1.295*** (30.729, 0)	1.275*** (30.595, 0)
<i>Zero-model</i>				
Intercept	7.56*** (6.181, 0)	7.091*** (5.87, 0)	7.533*** (6.235, 0)	7.323*** (6.045, 0)
$n_{i,t0} + n_{j,t0}$	-2.972** (-9.464, 0.002)	-2.998*** (-9.845, 0)	-3.031** (-10.001, 0.002)	-3.064** (-9.753, 0.009)
$ n_{i,t0} - n_{j,t0} $	1.448** (5.18, 0.004)	1.45** (5.35, 0.002)	1.457** (5.412, 0.001)	1.502** (5.354, 0.001)
distance	0.471** (3.894, 0.003)	0.547** (4.551, 0.002)	0.518** (4.299, 0.002)	0.532** (4.375, 0.005)
same region	-2.492*** (-7.476, 0)	-2.57*** (-7.988, 0)	-2.704*** (-8.462, 0)	-2.797*** (-8.274, 0)
obs.	12090	12090	12090	12090
loglik	-15655	-15885	-16075	-16199

— P-values are based on MRQAP (1000 permutations) and one-sided (because null distributions are not symmetric). One, two, and three stars signal significance values below 5%, 1%, and 0.1% respectively. Marginal effects are in Appendix B, Table 8.

— Recall from Section 3.4 that n and c stands for non-CRR and CRR respectively. Indices i indicate a country, and ij a country dyad. Country-level variables are joined to capture the country dyad's sum ($x_i + x_j$) and their absolute difference ($|x_i - x_j|$). $\log(\theta)$ captures over-dispersion in the count model.

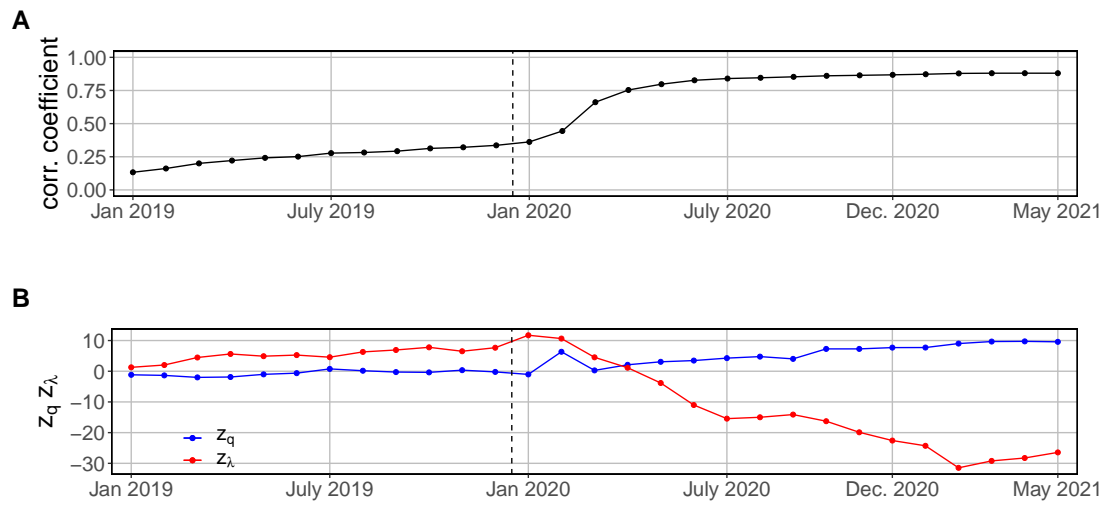


Fig. 6 Convergence of CRR network to non-CRR network. In all plots, CRR network is accumulated from Jan. 2019 up to a given month, and the non-CRR network, accumulated from Jan. 2019 to Dec. 2019, serves as the reference network. (A) Pearson's correlation coefficient of (log of) CRR adjacency matrix and non-CRR adjacency matrix. All correlations highly significant according to QAP. (B) Development of z_q (z -value of modularity index Q of CRR network with sampling from non-CRR network for null distribution and communities) and z_λ (z -value of largest eigenvalue of CRR network with sampling from non-CRR network for null distribution).