9^{th} International Conference on Computational Social Science IC^2S^2
July 17-20, 2023, Copenhagen, Denmark

The COVID-19 research outbreak: how the pandemic culminated in a surge of new researchers

Keywords: COVID-19, pandemic, science of science, newcomers, team dynamics

Extended Abstract

The urgency of the COVID-19 pandemic forced scientific research to operate on an accelerated timeline, as both policymakers and the public relied on the most current evidence to guide their behaviors and decisions [1]. With a lower barrier to entry to academic publishing, partly aided by the advent of preprint servers, scientists with long-standing experience in epidemic research have been joined by *newcomers*—scientists from other fields such as computer science, physics, and economics, and younger researchers still in training—with the common goal of advancing the frontiers of science and informing policy decisions [2, 3]. Given this sudden surge in research output starting in early 2020, we investigate the large influx of newcomers and unravel their team dynamics over the first two years of the pandemic. We emphasize that we do not establish any causal relationship between credibility and the relative experience of researchers. We contrast our profiling of COVID-19 research newcomers with that from a similarly burgeoning field of computer science, *deep learning*.

Dataset & Methodology

Dataset. We use the Feb 2023 snapshot of the OpenAlex database and use the provided concept tags to filter articles. To ensure reproducibility and consistency, we select only the published papers and preprints (abbreviated as works or papers) with a digital object identifier (DOI). Because we are primarily interested in studying collaboration and team dynamics, we also discard single-authored works. For each topic, we identify an *observation window* (OW) spanning 2 years, preceded by a 5-year non-overlapping *training window* (TW). Furthermore, the OW is split into four consecutive, disjoint 6-month long *phases*.

Concept tags. For COVID-19 papers, we select works tagged with any of the three following concepts: COVID-19 (`C3008058167`), 2019-20 coronavirus outbreak (`C3006700255`), and SARS-CoV-2 (`C3007834351`). Experience in epidemic research is ascertained through works covering the concepts: infectious disease (`C524204448`), pandemic (`C89623803`), epidemiology (`C107130276`), and outbreak (`C116675565`). Similarly, for deep learning (DL), we use the concepts: Deep learning (`C108583219`), RNN (`C147168706`), CNN (`C81363708`), and GAN (`C2988773926`). Prior experience in deep learning research is judged based on ANN (`C50644808`), MLP (`179717631`), supervised (`C136389625`), and unsupervised learning (`C8038995`).

Author classification. We define *experienced* researchers as those who have published at least one paper during the TW with the corresponding prior experience concepts. In contrast, *newcomers* do not have any papers during that window with these concepts. We note that newcomers comprise existing researchers from other fields and new authors who publish for the first time during the OW. The status of an author remains static over the OW, as defined at the end of the TW.

Paper classification and team dynamics. We quantify the proportion of authors in a paper who are newcomers, classifying papers with strictly over 50% of newcomers as *majority*

newcomers and the rest as *minority newcomers*. To evaluate the temporal dynamics of researcher representation, we record the monthly evolution of the imbalance between these two subgroups and compare the two topics across different phases of the OW in Fig. 1B. To better understand the interaction between team composition, prior expertise, and potential differences across emerging research topics, we consider the distribution of the number of authors stratified by paper type in Fig. 1C.

Findings and Conclusions

For COVID-19, we select Mar 1, 2020 – Feb 28, 2022, as the OW, which captures the rapid growth followed by a period of sustained production, and also roughly corresponds to the pandemic response timeline in the western hemisphere. We analyze a total of 314,801 works by 907,476 authors during the OW. While being admittedly smaller, DL has also proliferated in popularity in recent years. We pick Jan 1, 2018, and Dec 31, 2019, as the OW when the field consistently exceeds 1,000 works a month (Fig. 1A top right). Across the OW, we process 49,442 works by 119,506 authors for DL. The two observation windows are kept disjoint to avoid confounding effects of the pandemic on DL.

In Fig. 1A, we track the monthly count of COVID-19 (left) and DL (right) papers and their corresponding prior concepts (in green and purple) through their respective training and observation windows. Because we rely on OpenAlex's automated concept tagging process, our data is susceptible to misclassification errors, as evidenced by COVID-19 articles before 2020.

The proportion of articles with the majority of newcomers in Fig. 1B (bottom) for both topics increases over the OWs. In Phase I, the fraction tracks the overall growth of the topics. Starting in Phase II, it is remarkable that over 75% of all COVID-19 articles, compared to slightly over 50% for DL, were authored by a team comprising a majority of newcomers. This highlights the diversity of research interests of researchers involved, likely due to the broad, sweeping effects of the pandemic on society. Moreover, it suggests that there may be notions of *viral* topics in the scientific community that attract a disproportionate volume of out-of-domain researchers in a short time, and COVID-19 was an example of such.

The two topics' team composition violin plots (Fig. 1C) also feature a few key differences. The split between newcomers and experts for DL is more balanced. DL papers authored by teams with a minority of newcomers, on average, are larger. This may result from increased collaborations between academic and industrial research labs. COVID researchers, particularly newcomers, tend to work in smaller teams than their DL counterparts. This may also partly be due to various restrictions, *e.g.*, remote work and local lockdowns, that were in effect during the pandemic.

This initial study and its findings lay the groundwork for a more thorough analysis of the scientific community's response to the pandemic. Using more sophisticated mathematical models, we can go forward to account for temporal and size effects and closely inspect the behavior of specific groups of newcomers across different fields.

References

[1] S. P. Horbach, "Pandemic publishing: Medical journals strongly speed up their publication process for covid-19," *Quantitative Science Studies*, vol. 1, no. 3, pp. 1056–1067, 2020.

[2] N. Fraser, L. Brierley, G. Dey, J. K. Polka, M. Pálfy, F. Nanni, and J. A. Coates, "The evolving role of preprints in the dissemination of covid-19 research and their impact on the science communication landscape," *PLoS biology*, vol. 19, no. 4, p. e3000959, 2021.

[3] M. S. Majumder and K. D. Mandl, "Early in the epidemic: impact of preprints on global discourse about covid-19 transmissibility," *The Lancet Global Health*, vol. 8, no. 5, pp. e627–e630, 2020.

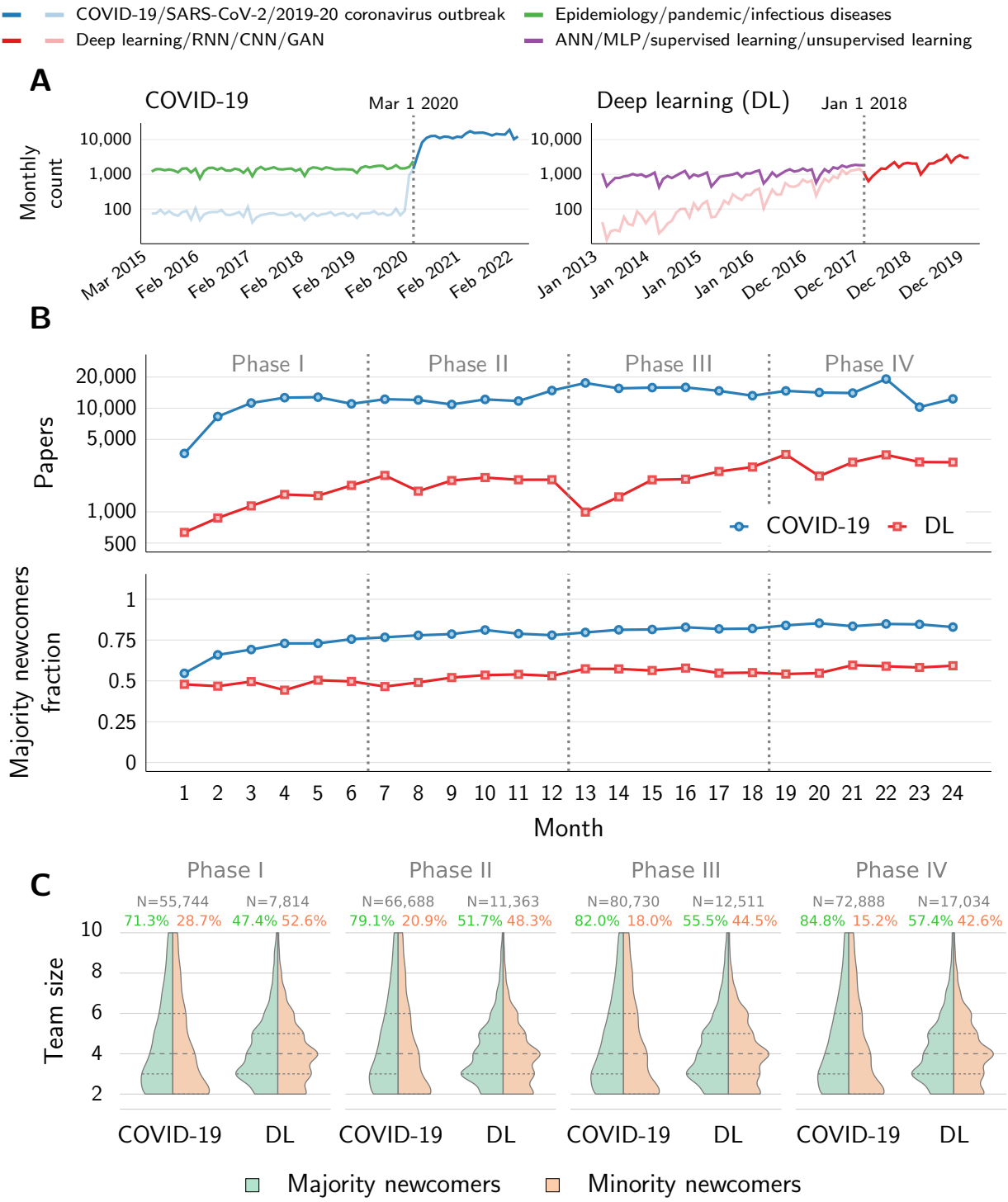

Figure 1: (A) Monthly count of works tagged with COVID-19 and deep learning (DL) with more than 2 authors (in blue and red resp). The start of the respective observation windows (OW) are marked with dotted lines. The paper counts for COVID-19/DL during the training window (TW) are in a lighter shade. Work counts of the prior concepts in the TW are in green and purple resp. (B) Number of papers and a fraction of papers written by the majority of newcomers (resp. above and below) tagged with COVID-19 and DL in the OW (resp. Mar 1, 2020 – Feb 28, 2022, and Jan 1, 2018 – Dec 31, 2019). The timeline is divided into four 6-month-long phases. (C) Violin plots concerning the number of authors per paper (between 2 and 10 authors) tagged respectively with COVID-19 and DL. The dotted lines inside mark the quartiles. The green left (resp. orange right) part concerns the papers with a majority of (resp. minority) newcomers. Above, the total number and the percentage of works in the two classes.