Labor reallocation and remote work during COVID-19: Real-time evidence from GitHub

Grant McDermott[1*] and Benjamin Hansen[1,2,3]

[1]Department of Economics, University of Oregon, Eugene, OR, USA
[2]National Bureau of Economic Research, Cambridge, MA, USA
[3]IZA Institute of Labor Economics, University of Bonn, Bonn, Germany

[*]To whom correspondence should be addressed; E-mail: grantmcd@uoregon.edu.

We investigate the effect of the COVID-19 pandemic on labor activity using real-time data from millions of GitHub users around the world. We show that the pandemic triggered a sharp pattern of labor reallocation at both the global and regional level. Users were more likely to work on weekends and outside of traditional 9 am to 6 pm hours, especially during the early phase of the pandemic. We also document considerable heterogeneity between different user groups and locations. Some locations show a steady reversion back to historical work patterns, while others have experienced persistent trend deviations in the wake of COVID-19. The pattern of labor reallocation is slightly more pronounced among males in our sample, suggesting that men may have benefited more from the increased flexibility provided by remote work than women. Finally, we show that the pattern of reallocation was accompanied by a simultaneous increase in overall activity, though this effect is more transient. We discuss several potential mechanisms and draw tentative conclusions for

broader workplace trends given our study population.

Introduction

The novel Sars-Cov2 virus (and the disease it causes, COVID-19) has profoundly changed the global economy. Recovery from the initial pandemic is ongoing amid continued outbreaks and the arrival of new variants. However, real-time data suggest that unemployment peaked at 20 percent in the United States, levels not seen since the Great Depression (*1*). Even those workers who retained employment status throughout the pandemic have had to confront dramatic changes in work conditions and habits.

Perhaps the most notable of these changes is the shift to remote work, or "working from home" (WFH). Cities, states, and countries across the world have implemented stay-at-home orders of varying forms to combat the virus. At the same time, private companies and public entities have responded by implementing large-scale WFH policies of their own. In some cases, this has been to comply with government orders and temporary lockdowns. Yet many employers issued internal allowances or WFH mandates in advance of any official requirements and have since approved permanent remote working terms for a large fraction of their employees.[1] There appears to be a growing consensus that WFH will become an enduring legacy of the pandemic (*4, 5*).

Despite these large-scale social changes, the extent to which remote work changes when and how people work remains open question. Experimental evidence collected before the pandemic suggests that some types of workers are more productive at home (*6*). But WFH during COVID-19 may be uniquely challenging due to the presence of contemporaneous shocks such

---

[1]A case in point is Microsoft, which issued a WFH recommendation to employees at its main Seattle-Redmond campus (and other nearby locations) on March 4, 2020 (*2*). This was several weeks before the official mayoral order to remain at home on March 23rd. The company later adopted guidelines for enabling a large portion of its employees to remote work on a permanent basis, mirroring similar moves by other tech giants such as Google and Facebook (*3*).

as stress or illness from the virus itself. In turn, policy responses like the widespread school closures in the early phase of the pandemic may have inadvertently triggered (or exacerbated) a disproportionate burden among certain population groups (*7, 8*). Other individuals may have conversely selected into doing *more* work during the pandemic, simply because their traditional leisure options were enjoined by lockdown measures. Further complicating matters is the fact that WFH may portend changes to how we work and collaborate in teams. At least two recent studies, each drawing on internal data from a different tech company, document several negative outcomes following the switch to WFH (*9, 10*). Collaboration was stymied as communication networks became siloed and productivity fell even as work hours increased.

It is within this context that the present study seeks to further our understanding of COVID-19's labor impacts, specifically, and the longer-run effects of WFH, generally. We do so by drawing on a rich and underused resource: real-time activity data from millions of GitHub users across the world. The increased provision of (near) real-time data has become a defining feature of the COVID-19 pandemic. Policymakers and citizens alike have been informed by continuous updates to local case numbers and public health dashboards. For their part, researchers have turned to a variety of novel data sources and methods to better understand the effects of COVID-19. While real-time data come with their own challenges (*11, 12*), they have helped to overcome the limitations of more traditional data sources and provided a direct window into the pandemic response as it has unfolded.[2] Examples include company emails, internal communications and analytics (*9, 10, 14*), mobility data from cellphones and social media accounts (*15, 16, 17*), online vacancy postings and unemployment insurance claims (*18*), time-clock data from Homebase

---

[2]Contemporaneous data on socioeconomic outcomes, such as employment or weekly hours work, have traditionally been based on comprehensive surveys administered by federal entities like the US Census Bureau. However, the delays associated with these surveys make them ill-suited for capturing rapidly evolving situations as in a global pandemic. For example, the Current Population Survey (CPS) and American Community Survey (ACS), both of which are administered by the US Census Bureau, respectively, have 1- and 18-month lags. Survey responses also decreased dramatically following the start of the pandemic and, in some cases, temporarily ceased altogether (*13*).

and other scheduling software (*11*), Google Trends (*19*), and rapid online market surveys (*1, 8, 20*). One prominent group of researchers has even combined several high-frequency private data sources—including anonymized information on credit card use and business revenues—into a publicly available database that tracks real-time economic activity (*12*). All told, real-time data have helped measure COVID-19's impact on a diverse range of socioeconomic outcomes, from employment and energy use to mental health and family strife.

The GitHub data that we use in this study complement these and other novel data sources. However, it also offers some unique benefits in terms of scale and coverage. Our primary finding is that the pandemic triggered a sharp pattern of labor reallocation that persisted through much of 2020. Users were significantly more active on the platform during weekends and outside of traditional 9 am to 6 pm hours. While this reallocation is evident at both the global and regional level, we also document considerable heterogeneity between different user groups and locations. In addition to documenting changes in how GitHub users allocate their time (intensive margin), we also consider changes to total productivity (extensive margin). We observe a large, positive deviation in total user activity relative to the predicted counterfactual. This deviation coincides neatly with the onset of the pandemic, yet generally subsides after a few months. Finally, we also find compelling evidence of spatial and temporal spillovers. The pattern of labor reallocation that we observe in many locations occurs well in advance of any local lockdown orders. Instead, we observe a global synchronization of shifting activity around the start of March 2020. This observation not only has narrow implications for our own research design, but also underscores the importance of adopting a wider perspective for estimating causal relationships in an interconnected world (*15*).

Background

GitHub data

GitHub (www.github.com) is the world's largest platform for software development and scientific code, currently serving over 50 million users.[3] The platform was originally targeted at professional software developers, but now claims a userbase across a broad swathe of industries and academic institutions (*21*). GitHub's core purpose is to facilitate version control of code and data.[4] Specifically, it enables tracking of individual files and lines of code so that earlier versions of a project's codebase can be recalled at any point in time. Individuals, teams, and organizations that host their code or data on GitHub can manage contributions from collaborators, field issues and requests, and generally control any changes made to a project.

An example may help to clarify some terms as well as the scope of our data. Consider an app developer who has identified a small bug in her code. She fixes the bug on her local computer and saves ("commits") these changes, thereby adding it to the project's version control history. She then uploads ("pushes") the fix to the corresponding GitHub repository in the cloud. This GitHub repository essentially provides a clone of her local project—all of the files, directories, etc.—that is accessible to others online. At this point, there are several possible ways to proceed. For example, users of the app might be free to download the update immediately. Or, a collaborator might be assigned to review the proposed fix on GitHub first, only merging it into the main code branch once they are satisfied. Regardless of how things do proceed, the most salient features from our perspective are the following: 1) any changes to the GitHub repository are automatically timestamped and logged as part of the version control history, and 2) this version control history can be viewed and analyzed by anyone with read access to the reposi-

---

[3]https://github.com/search?q=type:user&type=Users

[4]The name *GitHub* reflects tight integration with the *Git* version control system. Originally written by Linus Torvalds to facilitate collaborative development of the Linux kernel (*22*), Git provides the underlying protocol and functionality that GitHub's services are built upon.

tory. The features of our hypothetical example extend to all other projects and repositories on GitHub. The platform automatically records any changes made to a project's codebase—who made them, when they were made, and what they consisted of—with perfect fidelity. More-over, it logs these changes for posterity. To put a fine point on it, GitHub data provide a direct measurement of real-time labor activity for tens of millions of knowledge workers across the world.

In this study, we use GitHub data to investigate how user activity and work habits have changed as a result of COVID-19. Our focus is on activity recorded on the public GitHub timeline — which is to say, activity on any repository that is not access-limited to a group of private collaborators. All of this information is provisioned by GitHub's official API (Applica-tion Protocol Interface).[5] However, we primarily draw our data from two third-party sources, *GHTorrent* (*23*) and *GH Archive* (*24*), which were established to collate information across the entire GitHub platform and make it accessible to researchers. Despite the exclusion of private GitHub activity from our study, we emphasize that the public timeline still encompasses a vast number of users and activity. For example, we observe activity from over 15 million unique users during 2020 alone. Fig. 1 depicts the timeline of daily activity from 2015 to 2020 (our designated study period). Toward the latter part of this period, we observe as many as two million individual events per day.[6]

[Fig. 1 about here.]

For a subset of approximately 2.5 million registered GitHub users, we also observe their self-reported geographic location. Of these users, just over a million were active on public

---

[5]For example, the recent public activity of the corresponding author may be found at the following API end-point: https://api.github.com/users/grantmcdermott/events.

[6]Note that "event" is a formal designation given by the GitHub API to a variety of user actions. This includes committing and pushing changes to a repository (the most common event type), forking a repository, leaving a comment on an issue thread, etc. We do not distinguish between event types in our primary analysis since we are interested in tracking activity in the aggregate. Limiting the analysis to individual events types (e.g. push events) yields very similar results to the ones that we present here. A full list of the recognized event types is available here: https://docs.github.com/en/developers/webhooks-and-events/events.

repositories during 2020. Table S1 and Fig. 2 depict the geographic distribution of daily GitHub activity over our study period. Drilling down yet further, a large fraction of GitHub users resides in traditional tech hubs. Table S2 highlights the top ten cities in our dataset by identifiable user geography. As part of our later analysis, we focus on six of these markets—London, New York, San Francisco, Beijing, Bengaluru (Bangalore), and Seattle—and examine the response of user activity to both local and global COVID-19 shocks. Finally, while GitHub does not report gender, many users report their first and last names. We use this information to statistically impute gender among users of the five Anglo-Indian cities within our sample, based on tabulated frequencies drawn from relevant census data. For simplicity, we exclude uncertain matches (e.g., unisex names). The resulting subsample populations are heavily skewed toward males. For every female user that we impute, we impute nine male users. This pattern holds consistently across cities, corroborating existing evidence about the lack of gender diversity in the tech industry (*25*).

[Fig. 2 about here.]

Timing and response to the COVID-19 pandemic

The exact origin and timing of human-to-human SarsCov2 transmission remains unknown. China began locking down several cities in Hubei province in December 2019 following the initial outbreak in the city of Wuhan. The World Health Organization declared an international public health emergency on January 30, 2020 and officially declared it a global pandemic on March 11, 2020. While localized outbreaks varied in their precise timing and intensity, global awareness of COVID-19 increased throughout February 2020, with widespread awareness emerging in March 2020 ahead of the many local, regional, and even national lockdown measures that were temporarily introduced.

With that timeline in mind, we first focus on weekly activity, considering the week 10 of

7

2020 (the first week of March) as a baseline measure of the pandemic onset. This abstraction may fail to account for localized or company specific policies, which may have begun sooner in some cases. However, it will also limit bias in the case of geographic and network spillovers. To the extent that we do mismeasure company or local shifts to WFH, our approach will serve to understate the effects of WFH on labor reallocation. We also emphasize that our study population is not necessarily representative of the modal worker in the economy. Tech professionals and other knowledge workers experienced very different work stresses and constraints during COVID-19 compared to, say, restaurant owners and their employees. Yet this population is ideally suited for measuring the effect of WFH on labor timing and intensity. While we do not observe remote work directly, (26) find that workers in digital knowledge spaces like computers are ex ante easily able to do their job remotely, and (27) find that ex post the vast majority of these workers were working remotely in the pandemic.

Methods

In observational data settings such as ours, researchers commonly exploit variation in treatment timing across units. For example, one might be tempted to use variation in the onset of COVID-19 (and lockdowns) across countries to construct treatment and comparison units. This research design could then be used to infer the causal effect on GitHub activity in one location, given the valid counterfactual provided by another location that had not yet been treated. Unfortunately, such a strategy is difficult to justify in the context of a globalized shock like the COVID-19 pandemic. Temporal and geographic spillovers—whether due to coordinated political responses, multinational firm policies, or network effects—would lead to biased estimates. Indeed, (15) document considerable spillover effects during the early phase of the pandemic, which cut across large geographic distances and even served to undermine local policy responses. A lesser, but also important, complication involves the presence of compositional

effects. The number of GitHub users grew continuously during the study period. While it is possible to control for this growth directly, correctly attributing changes along the extensive and intensive margins of activity requires some care.

We adopt two separate empirical strategies to avoid these inferential problems. Our first approach compares proportional year-on-year GitHub activity within the same geography. For example, did GitHub users in San Francisco work more on weekends during March 2020 (after the COVID-19 lockdown) than they did in previous years? This allows us to detect whether activity during our 2020 treatment year deviates from the patterns that we typically observe during a specific month or calendar week. The year-on-year comparative approach has the virtue of being easy to visualize and intuit. Indeed, we present direct visual evidence of labor reallocation patterns by plotting the raw GitHub data in the next section. Yet we also complement these visualizations by estimating event-study regression models of the form

$$Y_{i,t} = \alpha_i + \gamma_t + X_{i,t}\Gamma + \Sigma_{w=\underline{W}}^{\overline{W}}\beta_w L_{i,t-w} + \epsilon_{i,t}. \tag{1}$$

Our treatment group in the above equation is defined as the pandemic year (i.e., 2020), while the control group consists of the years immediately preceding it (e.g., 2015–2019). Thus α_i denotes a year fixed effect, γ_t a week-of-year time effect, and $X_{i,t}$ a vector of controls such as location fixed effects. The outcome variable $Y_{i,t}$ is a measure of interest, such as a proportion of weekend or out-of-hours GitHub activity. The term $\Sigma_{w=\underline{W}}^{\overline{W}}\beta_w L_{i,t-w}$ allows us to track the evolution of this outcome variable over time in response to COVID-19 lockdown(s) L. Our default specification considers the evolution of GitHub activity from $\underline{W} = 10$ weeks before the lockdown until $\overline{W} = 20$ weeks after it. We extend this basic event-study framework to various locations and cuts of the data (e.g., by gender).

The focus on proportionality measures in Eq. (1) allows us to detect reallocation patterns

even if the number of GitHub users is changing (increasing) over time.[7] However, it does not allow us to take a definitive stance on whether COVID-19 induced changes in the overall level of activity. For example, did San Francisco GitHub users work longer hours in total after they went into lockdown? Or were they just reallocating a set time budget from regular work hours to evenings and weekends? To address this question, our second empirical approach adopts a forecasting counterfactual strategy ([28, 29]). In brief, we train a time-series model on historical data and forecast a counterfactual for 2020 against which the observed values can be compared. The "Prophet" model developed by researchers at Facebook is ideally suited to our use case ([30]).[8] As per Fig. 1, the daily GitHub activity data exhibit clear trends and patterns of layered seasonality (day of week, month of year, etc.). We similarly need to account for large holiday effects that may be location-specific and bleed into surrounding dates (e.g., around Christmas and New Year's Day). Prophet takes the form of a nonlinear regression model, where different separate time-series components are combined in additive fashion:

$$y_t = g(t) + s(t) + h(t) + \varepsilon_t. \tag{2}$$

Here $g(t)$ is a piecewise-linear trend function (or "growth" term), $s(t)$ captures various seasonal patterns, $h(t)$ represents (potentially irregular) holiday effects, and ε_t is an idiosyncratic error term. We use 2017–2019 activity data for model training and only evaluate forecasts on a set cohort of existing users to avoid spurious compositional effects.

---

[7] As described in the supplementary text, we also take steps to compare consistent cohorts over time to avoid picking up spurious compositional trends.

[8] Use case fit aside, we also cross-validated a variety of forecast models. Prophet performed as well or better than the other options while being considerably faster and requiring minimal hyperparameter tuning for good performance.

Results

We begin our results section by presenting the raw year-on-year global GitHub activity data. Fig. 3 highlights a sharp reallocation towards weekend activity following the onset of COVID-19. This proportional increase, from a baseline of 20 percent to over 24 percent, represents a sizable shift in individuals' time use. If we assume a constant 40-hour work week, then it translates to an additional 2 hours of work every weekend. The gap between our treatment and control groups does narrow as we move toward the later calendar weeks of the year. However, the proportional shift is long-lasting and remains above trend throughout 2020.

[Fig. 3 about here.]

In Fig. 4 we present similar proportionality plots for a group of major tech cities from different parts of the world. This time, we include the proportion of out-of-hours work in addition to weekends, and highlight local lockdown orders. Several cities demonstrate a clear shift toward both weekend and out-of-hours work. However, the effects are more heterogeneous and nuanced in others. Residing in the initial treatment country of China, Beijing experienced a clear and early structural change as it came out of Chinese New Year in early February. Yet it is interesting to note that the reallocation pattern in the other cities generally precedes—by several weeks—their own, local lockdown orders. Instead, we observe a general synchronization around a "global" treatment date at the beginning of March (week 10). There is also a fair degree of heterogeneity in terms of persistence. For example, London had reverted to its historical pattern by somewhere between weeks 30 and 40, coinciding with local lockdown orders being lifted. In contrast, the out-of-hours proportion for Bengaluru remained stubbornly above trend for the entirety of 2020.

[Fig. 4 about here.]

The descriptive analyses presented thus far reveal large shifts in the times that GitHub users

choose to work. We formalize these insights by estimating several regression models per Eq. (1). Informed by the visual patterns evident in the preceding figures, we fix a global treatment date at the beginning of March—as opposed to relying on local lockdown orders—to avoid the inferential problems associated with geo-temporal spillovers in treatment. By and large, our regression results simply confirm the patterns that are readily apparent in the raw data and can be viewed in supplementary material (e.g Tables S3 and S4). We limit our discussion here to the event-study plots in Fig. 5, which break out the results by gender.

A consideration of gender differences in our dataset is salient for several reasons. Female workers have generally experienced larger employment declines and slower recoveries throughout the pandemic (7). Moreover, these effects cannot be fully explained by differences in occupation or industry (31). Yet it is not a priori obvious whether these gendered effects will be fully evident among the GitHub users in our sample. Many knowledge workers were shielded from COVID-19's (un)employment impacts due to their ability to work remotely. These individuals were also more likely to work longer and flexible hours before the pandemic (32, 33). At the same time, women typically bare a disproportionate parenting burden during unforeseen events like school closures, regardless of industry (8, 34). To interrogate these potential outcomes, we split our city-level data by imputed gender and estimate separate regression models to allow for heterogeneous effects. Because of sample size differences across this split—male users greatly outnumber females users—we pool the gender data rather than estimate separate regressions for each city. Doing so allows us to maximize statistical power across geography while capturing overall uncertainty within gender groups.

Looking at Fig. 5, both male and female GitHub users show an increased tendency to work outside of traditional hours following the initial lockdown period. Despite tapering toward the end of the study period, these treatment effects persist for most of 2020. Yet we also observe some differences between gender groups. The treatment response is more immediate and con-

centrated among men than women, which suggests that men may have benefited more from the initial flexibility provided by remote work. At least in the aggregate, male users in our sample appear to have been in a position that allowed them to adapt nearly immediately to COVID-19. In contrast, female users reallocated work effort more slowly and inconsistently. This latter claim is supported by inspecting the raw data trends in gender activity across cities in Figs. S7 and S8. Again, such a finding would be consistent with prior research on gender roles and the disproportionate burden that women bear in home production, particularly given the widespread school closures starting in early March 2020.

[Fig. 5 about here.]

Having established a consistent pattern of proportional labor reallocation—we observe comparatively more activity during traditional leisure periods—we next investigate the net effect of COVID-19 on overall productivity. The results from this analysis are summarized in Fig. 6, where we zoom out to focus on country-level data rather than individual cities.[9] As described earlier, our approach here adopts a forecasting counterfactual strategy. For each country, we train our model on 2017–2019 activity data, before forecasting into 2020 and then comparing against observed GitHub activity.[10] Following the onset of the pandemic, most countries in our sample record a large (15–20 percent) increase in overall activity relative to the forecast counterfactual. Assuming that these changes are reflected proportionally against a regular 40-hour work week, this translates to approximately eight hours of additional activity each week. In some countries like India (30–40 percent increase), this jump was even larger. Interestingly, the exceptions to this trend are concentrated among the Asian nations that had to confront COVID-19 first. China and South Korea do not exhibit the large and significant increase in user activity

---

[9]Our city-level focus was motivated by the need to investigate the effect of hyperlocalized lockdown orders versus a more globalized effect with treatment spillovers.

[10]Note in passing that our emphasis on proportionality measures thus far allowed us to abstract from compositional changes in the underlying GitHub userbase. But for measuring total productivity, any changes to the number of users would clearly distort year-on-year comparisons on the extensive margin. We therefore limit our forecasting exercises to a set cohort of users who already had active GitHub accounts at the start of the training period.

of other countries. If anything, Chinese GitHub activity appears to have fallen below trend in the wake of COVID-19. However, the sudden structural reallocation of time towards weekly activity is clearly apparent, as evidenced by the increased amplitude of the Chinese series.

Other notable features of Fig. 6 include the fact that the elevated activity levels across our sample countries had generally returned to trend by the start of July. This reversion coincides with the reopening of the global economy, as the first wave of the pandemic subsided and vaccines were increasingly made available. We also observe—in some countries, at least— evidence of slight dip leading into the pandemic before the sharp uptick. This dip possibly reflects anticipation effects or difficulties associated with the initial coordination problem in moving to remote work and education (*9*). To help validate these results, we replicate our forecasting exercise for user activity in 2019, i.e., the year before the pandemic. The results from this placebo test can be seen in Fig. S10. While the observed activity differed widely from what we observed in 2020, no such difference emerges for 2019. We interpret this as supporting evidence that our forecasting approach credibly identifies deviations specific to the pandemic rather than stemming from systematic forecasting errors.

[Fig. 6 about here.]

Stepping back, one way we might summarize the results across Figs. 3–6 is as follows. The pattern of *proportional* labor reallocation shows a good deal of persistence, especially out-of-hours activity. Despite some regional and gender-based heterogeneity, we observe a sustained shift in work patterns that essentially lasted through the end of 2020. In contrast, the jump in *total* activity appears to have been more transient. Total GitHub activity across most regions in our sample had reverted to trend by July 2020. This divergence highlights the differential impacts of the pandemic itself, the imposition of lockdowns and switch to WFH, and the lasting cultural changes that were seeded. As economies around the world have reopened and leisure opportunities have returned, so too have workers return to a regular 40- or 50-hour week. But

they have not necessarily returned to the same 9 am to 6 pm, Monday to Friday schedule.

We conduct a number of supplementary analyses beyond those presented here. For example, thus far we have generally shied away from making claims about "productivity", which economists would define in strict terms of output per unit effort. However, we can still glean some insight into productivity by tracing the evolution of GitHub events per number of active users. Figs. S4 and S5 illustrate the resulting time-series plots for this (narrow) measure of productivity.[11] While we observe a moderate increase in productivity over 2019–2020, COVID-19 does not appear to trigger any sustained changes beyond that. In a related but separate exercise, we confirm that our results do not hinge on our decision to group all event activity together. Indeed, the relative frequency of different GitHub event types—pushes, pull requests, etc.—is remarkably stable over time (generally) and around the COVID-19 treatment period (specifically). See Fig. S6 for an illustration. Finally we consider a set of organisation case studies, which offer additional insights because of their stable employment environments and long-standing norms of public-facing code. See Figs. S11–13 and further discussion in the supplementary text, particularly with regard to generalising our findings to activity in private GitHub repositories.

Conclusions

COVID-19 continues to reshape much of human life. In this paper, we have focused on labor (re)allocation and how individuals with flexible work conditions have adapted their schedules in response to the pandemic. We rely on real-time activity data from GitHub, covering millions of knowledge workers from across the world. Despite regional heterogeneity, our primary finding is that the onset of the pandemic triggered a sharp and sustained pattern of labor reallocation

---

[11]Here we highlight the same group of countries our forecasting analysis, since we again wish to hold a fixed cohort of users.

at the global level. GitHub users were demonstrably more active on weekends and outside of regular work hours than they were in previous years. Concomitant with this proportional reallocation of time, COVID-19 also precipitated a jump in overall activity. GitHub users were not simply reshuffling a set budget of work hours during the early lockdown period, but actually working more in general. However, this latter effect is largely transient and aggregate activity had trended back to expected levels within a few months. By the end of the 2020 we observe a fairly steady pattern that is tempting to call a "new normal". People appear to be working more during traditional leisure times—but not necessarily more in the aggregate—as schedules adjust to a world where more and more of us are WFH.

Many questions remain for future work. While we have provided evidence of spatial spillovers, the exact network propagation remains unclear. The GitHub data that we have used here would readily lend itself to such an analysis, since the full network tree of collaborators and joint projects is encapsulated in the event logs. Similarly, we have assumed an implicit constant quality to event activity. A follow-up study could evaluate the content of the individual commits or frequency of corrections; again using a similar dataset to ours. Users could also be matched to additional demographic information such as age and workplace experience, to see how these factors interact with gender and school closures. As these and other questions are investigated, we believe that real-time data sources—such as the GitHub data that we document here—will continue to offer unique insights into the ways the firms and workers have adapted in the wake of COVID-19.

References and Notes

1. A. Bick, A. Blandin, *Available at SSRN 3692425* (2020).

2. Microsoft News Center, Kurt DelBene's March 4 guidance to King County employees, https://blogs.microsoft.com/blog/2020/10/09/embracing-a-flexible-workplace [Published: 2020-03-04. Last accessed: 2021-10-08].

3. Microsoft Official Blog, *Embracing a flexible workplace* (by Kathleen Hogan), https://blogs.microsoft.com/blog/2020/10/09/embracing-a-flexible-workplace [Published: 2020-10-09. Last accessed: 2021-10-08].

4. J. M. Barrero, N. Bloom, S. J. Davis, Why working from home will stick, *Working Paper 28731*, National Bureau of Economic Research (2021).

5. D. Thompson, Superstar Cities Are in Trouble, *The Atlantic*, https://www.theatlantic.com/ideas/archive/2021/02/remote-work-revolution/617842/ [Published: 2020-02-01. Last accessed: 2021-10-08].

6. N. Bloom, J. Liang, J. Roberts, Z. J. Ying, *The Quarterly Journal of Economics* **130**, 165 (2015).

7. S. Albanesi, J. Kim, The Gendered Impact of the COVID-19 Recession on the US Labor Market, *Tech. rep.*, National Bureau of Economic Research (2021).

8. T. Deryugina, O. Shurchkov, J. E. Stearns, COVID-19 disruptions disproportionately affect female academics, *Tech. rep.*, National Bureau of Economic Research (2021).

9. L. Yang, *et al.*, *Nature Human Behaviour* pp. 1–12 (2021).

10. M. Gibbs, F. Mengel, C. Siemroth, Work from Home & Productivity: Evidence from Personnel & Analytics Data on IT Professionals, *Working Paper 2021-56*, University of Chicago, Becker Friedman Institute for Economics (2021).

11. A. W. Bartik, M. Bertrand, F. Lin, J. Rothstein, M. Unrath, Measuring the labor market at the onset of the covid-19 crisis, *Tech. rep.*, National Bureau of Economic Research (2020).

12. R. Chetty, J. N. Friedman, N. Hendren, M. Stepner, T. O. I. Team, The economic impacts of COVID-19: Evidence from a new public database built using private sector data, *Tech. Rep. w27431*, National Bureau of Economic Research (2020).

13. ATUS and Covid19 Data Collection, https://www.bls.gov/tus/covid19.htm. Accessed: 2021-10-21.

14. E. DeFilippis, S. M. Impink, M. Singell, J. T. Polzer, R. Sadunand, Collaborating During Coronavirus: The Impact of COVID-19 on the Nature of Work, *Tech. Rep. 21-006*, Harvard Business School working paper (2020).

15. D. Holtz, *et al.*, *Proceedings of the National Academy of Sciences* **117**, 19837 (2020).

16. S. Cicala, S. P. Holland, E. T. Mansur, N. Z. Muller, A. J. Yates, Expected health effects of reduced air pollution from covid-19 social distancing, *Tech. rep.*, National Bureau of Economic Research (2020).

17. A. Goolsbee, C. Syverson, *Journal of Public Economics* **193**, 104311 (2021).

18. E. Forsythe, L. B. Kahn, F. Lange, D. Wiczer, *Journal of Public Economics* **189**, 104238 (2020).

19. A. Brodeur, A. E. Clark, S. Fleche, N. Powdthavee, *Journal of Public Economics* **193**, 104346 (2021).

20. A. W. Bartik, *et al.*, *Proceedings of the National Academy of Sciences* **117**, 17656 (2020).

21. J. Perkel, *Nature* pp. 127–128 (2016).

22. S. Chacon, B. Straub, *Pro Git* (Apress, 2014), second edn.

23. G. Gousios, *Proceedings of the 10th Working Conference on Mining Software Repositories*, MSR '13 (IEEE Press, Piscataway, NJ, USA, 2013), pp. 233–236.

24. I. Grigorik, *GH Archive*. https://www.gharchive.org [Last accessed: 2021-10-08].

25. R. M. Fernandez, S. Campero, *Ilr Review* **70**, 73 (2017).

26. J. I. Dingel, B. Neiman, *Journal of Public Economics* **189**, 104235 (2020).

27. E. Brynjolfsson, *et al.*, Covid-19 and remote work: an early look at us data, *Tech. rep.*, National Bureau of Economic Research (2020).

28. H. R. Varian, *Journal of Economic Perspectives* **28**, 3 (2014).

29. R. J. Hyndman, G. Athanasopoulos, *Forecasting: principles and practice* (OTexts, 2021), third edn.

30. S. J. Taylor, B. Letham, *The American Statistician* **72**, 37 (2018).

31. G. M. Cortes, E. C. Forsythe, The heterogeneous labor market impacts of the covid-19 pandemic, *Tech. rep.*, Upjohn Institute Working Paper (2020).

32. M. Bertrand, C. Goldin, L. F. Katz, *American economic journal: applied economics* **2**, 228 (2010).

33. C. Goldin, *American Economic Review* **104**, 1091 (2014).

34. K. Houghton, Childcare and the new part-time: Gender gaps in long-hour professions, *Tech. rep.*, University of Oregon Working Paper (2021).

35. Acknowledgments, sources of funding, and the like, per *Science* style...
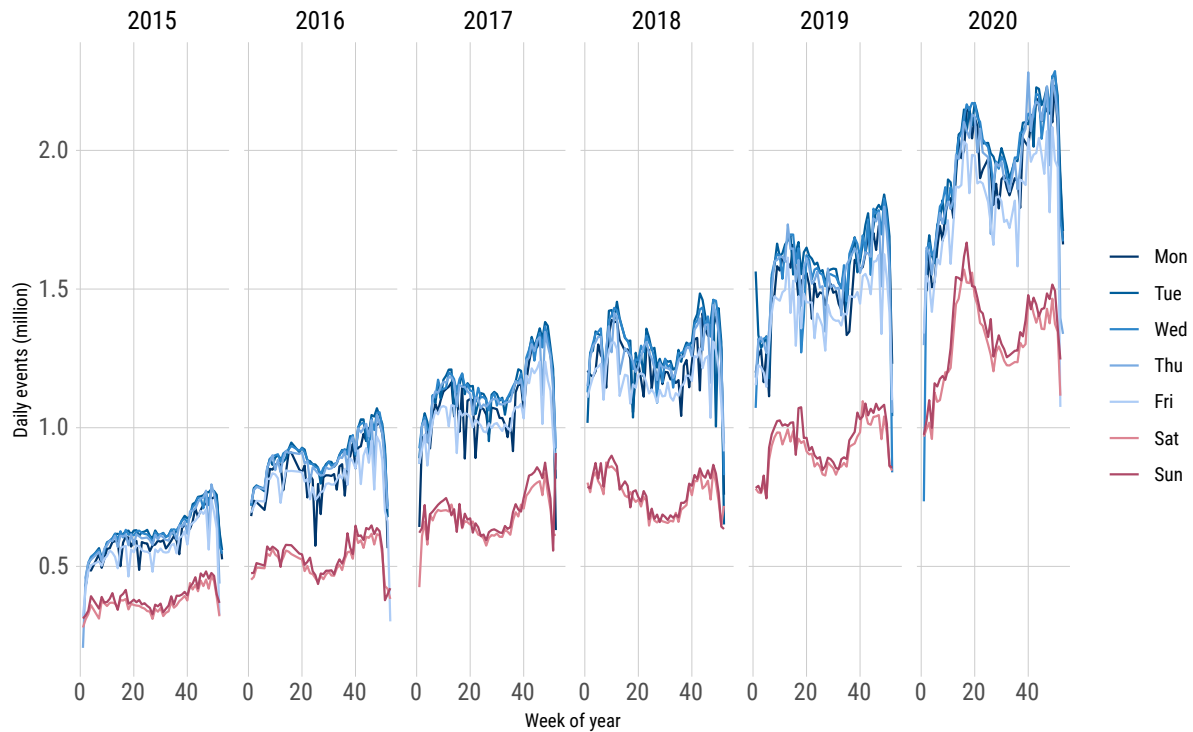
Tables and figures

Figure 1: **Global GitHub activity.** The figure shows the total daily events recorded on the public GitHub timeline from 2015 to 2020. The time series is broken out by day of week and week of year to underscore the strong seasonality in the data (weekdays versus weekends, public holidays, etc.). Peak daily activity increases from approximately half a million events at the start of 2015 to over two million daily events by the end of 2020. This is largely driven by a proportional increase in active daily users—from 150,000 to 500,000—over the same time period. Timestamps are relative to UTC (Coordinated Universal Time).

Average daily events (2015−2019)
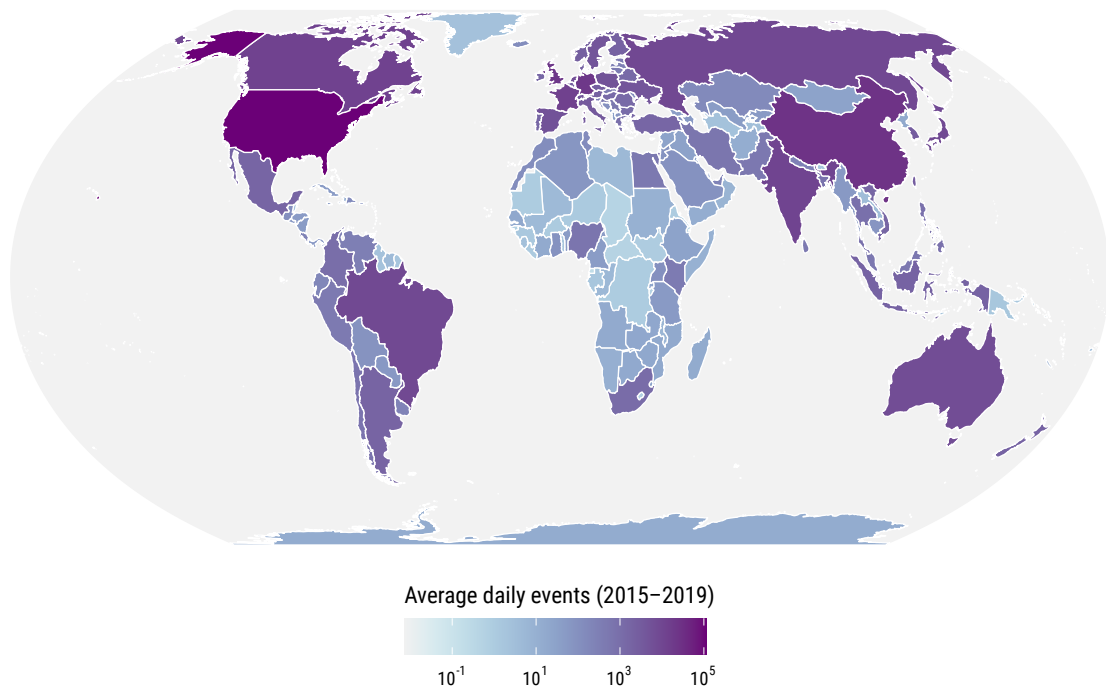
10^{-1} 10^{1} 10^{3} 10^{5}

Figure 2: **Geographic distribution.** The map shows the average number of daily GitHub events over the 2015–2020 study period, aggregated at the country level. The data are limited to active GitHub users for whom a defined geographic location is available. Note that the color scale has been logged to preserve aesthetic attributes.
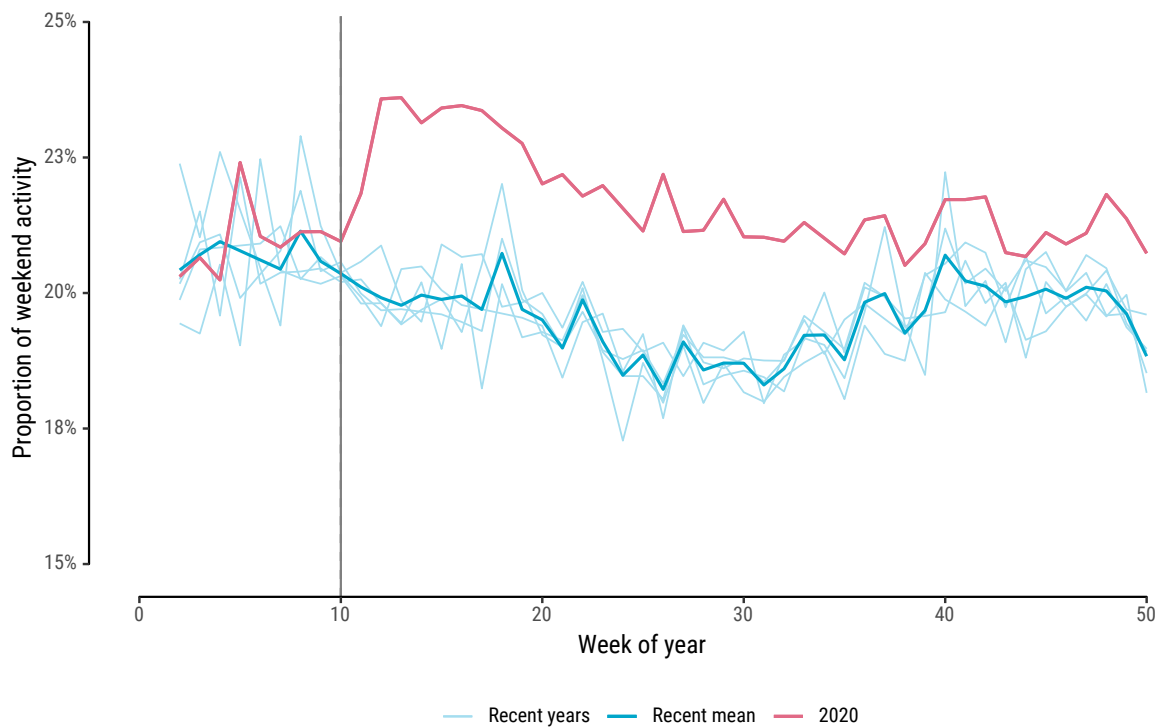
Figure 3: **Working more on weekends.** The figure shows the proportion of global GitHub event activity taking place on weekends, by week of year. The year of the pandemic (2020, in magenta) is contrasted against the five years immediately before that (2015–2019, in cyan). The vertical line denotes the 10th week of the year, corresponding to the beginning of March 2020 and the onset of the pandemic in most industrialized nations. Timestamps are relative to UTC.

Figure 4: **Labor reallocation across cities.** As per Figure 3, except broken out by major global cities and including the proportion of out-of-hours activity (i.e., outside 9 am to 6 pm). As with Figure 3, the solid vertical lines in each panel denote the beginning of March 2020. The additional dashed vertical lines correspond to the official lockdown orders or WFH mandates associated with each city. Recent comparison years are limited to 2017–2019 to limit confounding migration and relocation patterns. Timestamps are adjusted to local time zones.

Figure 5: **Gender event-study plot.** The figure shows the dynamic evolution of lockdown treatment effects in accordance with regression Eq. (1). The data have been subset by, and stratified according to, imputed gender. This subsample consists of approximately 72,000 active GitHub users across five major Anglo-Indian cities (London, New York, San Francisco, Bengaluru, and Seattle-Redmond), of which nearly 90 percent are male. The inner lines depict the coefficient point estimates, while the outer ribbons depict the 95 percent confidence intervals associated with standard errors that have been clustered at the city-year level.
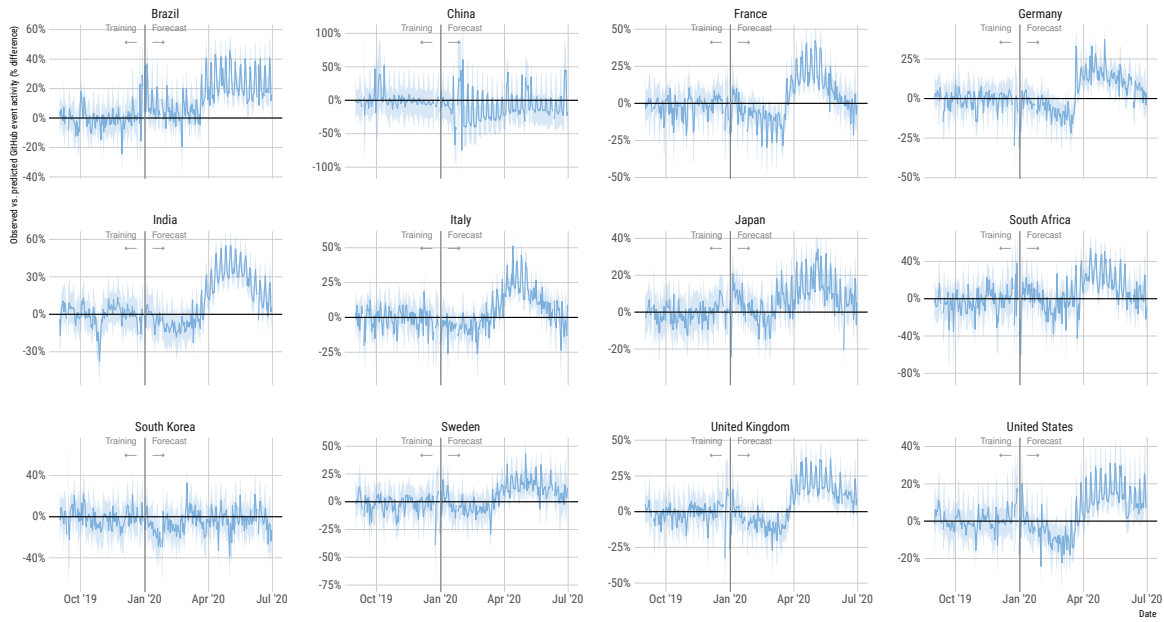
Figure 6: **Observed vs. predicted GitHub activity.** The figure shows the percentage difference between observed and predicted daily GitHub event activity across a sample of countries, zoomed in around the COVID-19 onset period. Predictions are obtained using the Prophet forecasting model (*30*), with the full training sample extending over 2017–2019. Shading denotes the 90 percent forecast interval. For each country, activity is limited to a set cohort of users who had established GitHub accounts before the start of the training period. Values above zero indicate that observed activity was higher than the counterfactual level predicted by the model, and vice versa.

Supplementary Materials for
Remote work and labour reallocation during COVID-19: Real-time evidence from GitHub

Grant McDermott* and Benjamin Hansen

*Correspondence to: grantmcd@uoregon.edu.

This PDF file includes:

Methods

Supplementary text

Tables S1 to S4

Figs. S1 to S13

Methods

GitHub data — additional information

We obtain the bulk of our GitHub data from *GH Archive* (*1*) and *GHTorrent* (*2*). The former catalogs all of the public event activity recorded by the GitHub API and is updated daily. The latter catalogs similar information, but is updated less regularly. However, it also provides periodic snapshots of user information—including location and organization affiliation—which we use for subsetting and covariate matching. Both (*1*) and (*2*) host their tables on Google's *BigQuery* database warehouse system.[1] We query and join these tables programatically via the BigQuery API. The raw data require several cleaning steps to exclude undesirable features and outliers. For example, we ignore user entries with more than 30 events in an hour (or 150 events in a day) to exclude bots and fake accounts. We similarly ignore and remove all activity on days with notable GitHub outages.

An important feature of the data is that the underlying GitHub API logs all event activity in terms of Coordinated Universal Time (UTC). This has ramifications for our research design, insofar as we cannot simply infer the time of day when a particular user is active—i.e., relative to their local time zone—based on the observed timestamps of the data. Instead, we must first match users to their self-reported location and then convert to the appropriate timezone. At least, we must do this for queries that require precise estimation of time-based effects, such as the proportion of out-of-hours activity. The UTC artefact in part motivates our decision to query and analyse representative geographic samples of the data (e.g., a detailed snapshot of certain countries, rather than a panel of activity data for every country in the world).

While (*1*) and (*2*) provide the bulk of our GitHub data, neither provides information that would allow us to impute gender. Nevertheless, the underlying GitHub GraphQL API records

---

[1]https://cloud.google.com/bigquery

first and last names for all users that self-report this information.[2] We therefore query the GraphQL API directly, matching on the subset of users that we have identified as located within one of the highlighted cities in Table S2. We extract all available first names and then impute gender based on the tabulated frequencies of relevant census data. We exclude Beijing because our imputation procedure is not reliable for Chinese names.[3] We also exclude ambiguous gender imputation matches, such as common unisex names. In total, we impute gender for about 40 percent of users from the five remaining highlighted cities in Table S2.

---

[2]https://docs.github.com/en/graphql
[3]This is primarily due to inconsistent use of native Chinese and Anglicized names, as well as the prevalence of different Chinese character (hanzi) standards.

Supplementary text

Organisation case studies

A potential concern is that our results may, at least in part, reflect compositional effects stemming from the types of repositories that GitHub users choose to work on. This concern relates to the fact that our data are all drawn from the public GitHub timeline and we don't observe activity within any private repositories. For example, some users might have switched over to doing more work on public repositories following the pandemic to better advertise their skills to potential employers and thereby mitigate job uncertainty. Others might have switched over to working on "hobby" projects, rather than renumerated ones, during traditional leisure times. We believe there are good reasons *a priori* to be sanguine about the former, since employment opportunities for knowledge workers during the pandemic were very stable compared to workers in other sectors. For the latter, this is largely a matter of framing, but also seems unlikely given the persistence of out-of-hours work that we observe throughout 2020 (and certainly well after most economies reopened). Nevertheless, we further explore this issue with the help of several organizational case studies in Figs. S11–S13.

These case studies consider event activity within repositories belonging to three organisations: Microsoft (https://github.com/microsoft), Alibaba (https://github.com/alibaba), and the UK Government Digital Service division (https://github.com/alphagov). All three are among the most active and well-represented organisations on GitHub; certainly within their respective categories.[4] More importantly, each in their own way provides a useful test case for analysing how GitHub activity responded to the pandemic in an environment where: i) job security among this subset of users was high, and ii) working

---

[4]As GitHub's parent firm, Microsoft is the largest organisation on the entire platform by some distance. To avoid confounding geographic and time-zone effects, we further limit our analysis of Microsoft activity to members from Seattle-Redmond (i.e. where the company is headquartered).

within public repositories and on open-source code was part of organisational norms.

Looking at Figs. S11–S13, there is a clear treatment effect across all three organisations. As with the city- and country-level results presented in the main text, the pattern of labour reallocation appears to be more persistent for out-of-hours activity than weekend activity. In the case of Microsoft's Seattle-Redmond employees, for example, the proportional jump in out-of-hours activity is much larger than the same metropolitan region as a whole. We observe a sustained 6–8 percent increase in activity outside of traditional 9 am to 6 pm work hours. Our takeaway from these organization case studies is that they support our our main findings, and enhance our understanding of the underlying treatment effects. We are left to conclude that our focus on the public GitHub timeline does not unduly distort the results that we would observe if we also had access to activity data from private repositories.

At the same time, there are interesting divergences between the organization trends that underscore the heterogeneity in our data. To highlight an obvious example, we observe a large anticipatory spike in the UK Government Digital Service activity immediately prior to the country's lockdown. Drilling down into the actual GitHub logs, this spike was driven by a rush to provide ancillary services to the British public that needed to be available once the lockdown went into effect.[5] Yet we note that these strong anticipatory effects do not explain (or undermine) the persistent pattern of labor reallocation that follows in the weeks and months thereafter.

---

[5]For example, forms that small businesses would use when requesting government support to help defray financial losses.

Supplementary Tables

Table S1: Sample of top GitHub countries by identifiable user geography

Rank	Country	Users (2019)	Users (2017)
1	United States	706,038	695,185
2	India	178,689	173,582
3	China	164,613	161,818
4	United Kingdom	122,862	120,867
5	Germany	103,662	101,935
6	Brazil	85,660	83,027
8	France	71,902	70,637
9	Japan	42,680	41,983
16	Sweden	28,868	28,521
18	Italy	26,330	25,701
22	South Korea	18,506	17,978
33	South Africa	11,709	11,435

This sample of 12 countries was selected to yield good geographic coverage and a mix of COVID-19 response policies. We favour countries with single time zones to enable accurate weekend and out-of-hours proportionality. Exceptions for the latter criteria are made for the United States (owing to its status as the top-ranked country and where we use CST) and Brazil (where >90% of the population reside in the BRT time zone). For the forecasting exercises described in the main text, we limit the sample to users who had established a GitHub account before 2017.

Table S2: Top 10 GitHub cities by identifiable user geography

Rank	City	Users (2019)	Gender imputed	% Male
1	**London, UK**	**44,759**	**18,266**	**92**
2	**New York, US**	**44,413**	**15,920**	**87**
3	**San Francisco, US**	**40,713**	**16,327**	**89**
4	**Beijing, CN**	**38,901**	**–**	**–**
5	**Bengaluru, IN**	**35,706**	**11,110**	**91**
6	Shanghai, CN	25,921	–	–
7	**Seattle-Redmond, US**	**24,205**	**10,256**	**89**
8	Paris, FR	22,792	–	–
9	Moscow, RU	18,910	–	–
10	Chicago, US	18,487	–	–

Cities in bold are highlighted in this paper. For a subset of cities, we impute gender based on the user's first name. Gender imputation is limited to GitHub users who were active during 2019 and excludes common unisex name matches.

Table S3: Event-study regressions: Proportion of weekend activity

Location	Global	Full city sample	Beijing	Bengaluru	London	New York	San Francisco	Seattle
Dependent Variable:				Events				
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variables								
Treated × TimeToTreatment=-8	-0.0067	0.0010				0.0039	0.0122	0.0088
	(0.0042)	(0.0133)				(0.0064)	(0.0105)	(0.0075)
Treated × TimeToTreatment=-7	-0.0060	0.0185		-0.0256	0.0056	0.0232***	0.0234	0.0129
	(0.0040)	(0.0165)		(0.0167)	(0.0084)	(0.0039)	(0.0184)	(0.0101)
Treated × TimeToTreatment=-6	-0.0070	-0.0042		-0.0289	-0.0002	0.0167***	0.0172	0.0050
	(0.0075)	(0.0124)		(0.0170)	(0.0062)	(0.0023)	(0.0154)	(0.0047)
Treated × TimeToTreatment=-5	0.0276	0.0174	0.0010	0.0036	0.0126	0.0336*	0.0192	0.0152
	(0.0142)	(0.0163)	(0.0575)	(0.0176)	(0.0081)	(0.0128)	(0.0112)	(0.0067)
Treated × TimeToTreatment=-4	-0.0024	0.0118	0.0296	-0.0068	0.0104**	0.0118	0.0129	-0.0060
	(0.0082)	(0.0126)	(0.0657)	(0.0195)	(0.0026)	(0.0146)	(0.0076)	(0.0153)
Treated × TimeToTreatment=-3	-0.0014	0.0032	-0.0070	-0.0229	0.0097	0.0170	0.0112	-0.0041
	(0.0052)	(0.0182)	(0.0438)	(0.0199)	(0.0044)	(0.0118)	(0.0115)	(0.0033)
Treated × TimeToTreatment=-2	-0.0056	0.0233	-0.0049	0.0117	0.0090	0.0260***	0.0140	-0.0021
	(0.0051)	(0.0168)	(0.1348)	(0.0215)	(0.0066)	(0.0025)	(0.0155)	(0.0049)
Treated × TimeToTreatment=0	0.0004	0.0219*	0.0536	-0.0021	0.0118***	0.0227*	0.0196	0.0076**
	(0.0023)	(0.0110)	(0.0532)	(0.0060)	(0.0016)	(0.0077)	(0.0103)	(0.0020)
Treated × TimeToTreatment=1	0.0118***	0.0214	0.0081	-0.0037	0.0240**	0.0069	0.0538*	0.0207*
	(0.0012)	(0.0169)	(0.1146)	(0.0088)	(0.0060)	(0.0114)	(0.0178)	(0.0076)
Treated × TimeToTreatment=2	0.0312***	0.0494***	0.0669	0.0308	0.0497***	0.0443***	0.0589**	0.0272**
	(0.0020)	(0.0125)	(0.0568)	(0.0203)	(0.0083)	(0.0064)	(0.0126)	(0.0065)
Treated × TimeToTreatment=3	0.0328***	0.0593***	0.0560	0.0619**	0.0684***	0.0650**	0.0476**	0.0352**
	(0.0023)	(0.0137)	(0.0599)	(0.0174)	(0.0049)	(0.0141)	(0.0107)	(0.0065)
Treated × TimeToTreatment=4	0.0147	0.0507***	0.0527	0.0348	0.0369***	0.0598***	0.0400**	0.0337***
	(0.0102)	(0.0122)	(0.0527)	(0.0268)	(0.0062)	(0.0057)	(0.0103)	(0.0039)
Treated × TimeToTreatment=5	0.0298***	0.0505***	0.0525	0.0711**	0.0436***	0.0442***	0.0562**	0.0231
	(0.0026)	(0.0149)	(0.0572)	(0.0190)	(0.0060)	(0.0073)	(0.0098)	(0.0103)
Treated × TimeToTreatment=6	0.0296***	0.0495***	0.0513	0.0419	0.0530*	0.0478**	0.0469**	0.0435***
	(0.0023)	(0.0134)	(0.0565)	(0.0233)	(0.0197)	(0.0083)	(0.0142)	(0.0050)
Treated × TimeToTreatment=7	0.0336***	0.0488**	0.0145	0.0795***	0.0492**	0.0547***	0.0413***	0.0396***
	(0.0054)	(0.0202)	(0.0614)	(0.0114)	(0.0085)	(0.0023)	(0.0056)	(0.0029)
Treated × TimeToTreatment=8	0.0234**	0.0439**	0.0209	0.0544***	0.0554**	0.0295**	0.0367**	0.0415***
	(0.0079)	(0.0163)	(0.0534)	(0.0062)	(0.0056)	(0.0056)	(0.0080)	(0.0028)
Treated × TimeToTreatment=9	0.0251***	0.0402***	0.0368	0.0372*	0.0444***	0.0345**	0.0445**	0.0280**
	(0.0030)	(0.0128)	(0.0559)	(0.0134)	(0.0046)	(0.0060)	(0.0084)	(0.0074)
Fixed-effects								
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
TimeToTreatment	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Location		Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics								
Observations	174	724	112	124	124	124	124	116
R^2	0.76274	0.71156	0.66461	0.90867	0.88779	0.89589	0.89356	0.88778
Within R^2	0.56894	0.36484	0.46067	0.77718	0.78345	0.79217	0.77946	0.70980

Clustered (Location-Year) standard-errors in parentheses

Signif. Codes: ***: 0.01, **: 0.05, *: 0.1

Table S4: Event-study regressions: Proportion of out-of-hours activity

Location	Full city sample	Beijing	Bengaluru	London	New York	San Francisco	Seattle
Dependent Variable:				Events			
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Variables							
Treated × TimeToTreatment=-8	-0.0303***				-0.0177***	-0.0207	0.0040
	(0.0099)				(0.0027)	(0.0105)	(0.0081)
Treated × TimeToTreatment=-7	-0.0084		-0.0237**	-0.0082	-0.0111	-0.0061	0.0170
	(0.0101)		(0.0054)	(0.0077)	(0.0049)	(0.0047)	(0.0116)
Treated × TimeToTreatment=-6	-0.0236**		-0.0165	-0.0101	-0.0032	-0.0016	-0.0031
	(0.0088)		(0.0218)	(0.0085)	(0.0055)	(0.0032)	(0.0158)
Treated × TimeToTreatment=-5	-0.0183	-0.0423	-0.0315	-0.0077	0.0074	0.0141	0.0227**
	(0.0129)	(0.0535)	(0.0199)	(0.0063)	(0.0096)	(0.0078)	(0.0039)
Treated × TimeToTreatment=-4	-0.0303**	-0.0566	-0.0192	-0.0131	-0.0117	0.0099*	-0.0126***
	(0.0118)	(0.0611)	(0.0191)	(0.0104)	(0.0160)	(0.0036)	(0.0015)
Treated × TimeToTreatment=-3	-0.0144	-0.0582	0.0168	-0.0129	0.0152	0.0007	-0.0111***
	(0.0130)	(0.0587)	(0.0097)	(0.0098)	(0.0119)	(0.0031)	(0.0016)
Treated × TimeToTreatment=-2	0.0059	-0.0211	0.0312***	-0.0070	0.0045	-0.0097	-0.0164*
	(0.0086)	(0.0878)	(0.0015)	(0.0045)	(0.0062)	(0.0052)	(0.0068)
Treated × TimeToTreatment=0	-0.0028	0.0361	-0.0076	-0.0226***	0.0090	0.0016	-0.0019
	(0.0060)	(0.0767)	(0.0093)	(0.0019)	(0.0144)	(0.0051)	(0.0040)
Treated × TimeToTreatment=1	0.0031	0.0202	-0.0025	-0.0063*	-0.0014	0.0134***	0.0128
	(0.0061)	(0.0770)	(0.0067)	(0.0023)	(0.0168)	(0.0017)	(0.0113)
Treated × TimeToTreatment=2	0.0093*	0.0340	0.0266**	5.64×10^{-5}	0.0063	0.0246**	0.0040
	(0.0050)	(0.0535)	(0.0051)	(0.0062)	(0.0039)	(0.0058)	(0.0120)
Treated × TimeToTreatment=3	0.0124	0.0199	0.0529***	0.0069	-0.0055	0.0025	0.0144
	(0.0086)	(0.0560)	(0.0061)	(0.0060)	(0.0064)	(0.0093)	(0.0106)
Treated × TimeToTreatment=4	0.0210**	0.0181	0.0476**	0.0034	0.0252**	0.0078*	0.0109
	(0.0090)	(0.0626)	(0.0109)	(0.0026)	(0.0054)	(0.0028)	(0.0070)
Treated × TimeToTreatment=5	0.0160*	0.0102	0.0552***	-0.0086	0.0185*	0.0175	0.0181**
	(0.0087)	(0.0583)	(0.0045)	(0.0099)	(0.0077)	(0.0077)	(0.0042)
Treated × TimeToTreatment=6	0.0162	0.0271	0.0711***	-0.0021	0.0218	0.0080	0.0112**
	(0.0108)	(0.0582)	(0.0084)	(0.0022)	(0.0109)	(0.0063)	(0.0027)
Treated × TimeToTreatment=7	0.0132	-0.0349	0.0653***	0.0214***	0.0164**	0.0082	0.0122*
	(0.0142)	(0.0605)	(0.0040)	(0.0012)	(0.0049)	(0.0073)	(0.0044)
Treated × TimeToTreatment=8	0.0250**	-0.0035	0.0558**	0.0348**	0.0390***	0.0104	0.0232***
	(0.0103)	(0.0642)	(0.0096)	(0.0079)	(0.0009)	(0.0085)	(0.0029)
Treated × TimeToTreatment=9	0.0252***	0.0131	0.0735***	0.0024	0.0290**	0.0252**	0.0265***
	(0.0089)	(0.0573)	(0.0080)	(0.0088)	(0.0063)	(0.0049)	(0.0025)
Fixed-effects							
Location	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes
TimeToTreatment	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics							
Observations	724	112	124	124	124	124	116
R^2	0.86056	0.76375	0.94462	0.87963	0.86721	0.80745	0.86325
Within R^2	0.30276	0.53828	0.77708	0.71258	0.67445	0.50719	0.64801

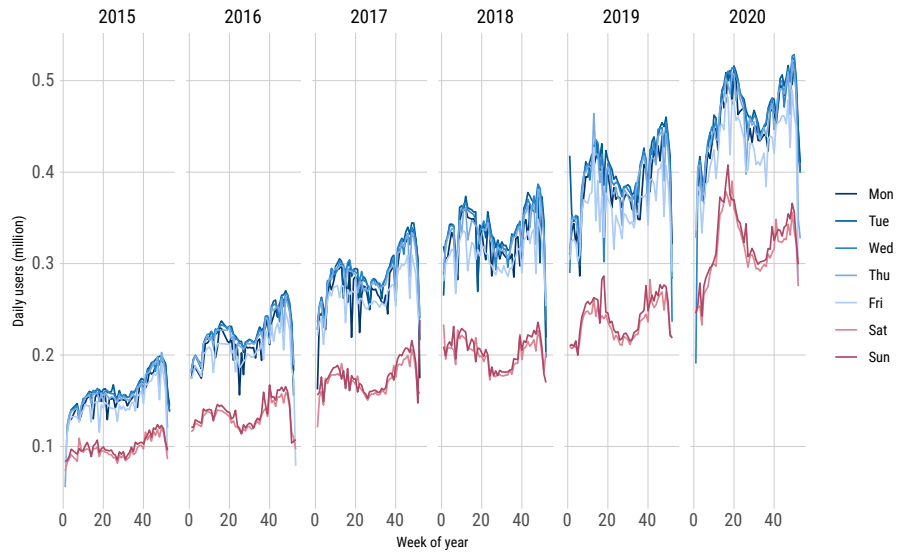*Clustered (Location-Year) standard-errors in parentheses*
Signif. Codes: ***: 0.01, **: 0.05, *: 0.1

Supplementary Figures

Additional time series plots
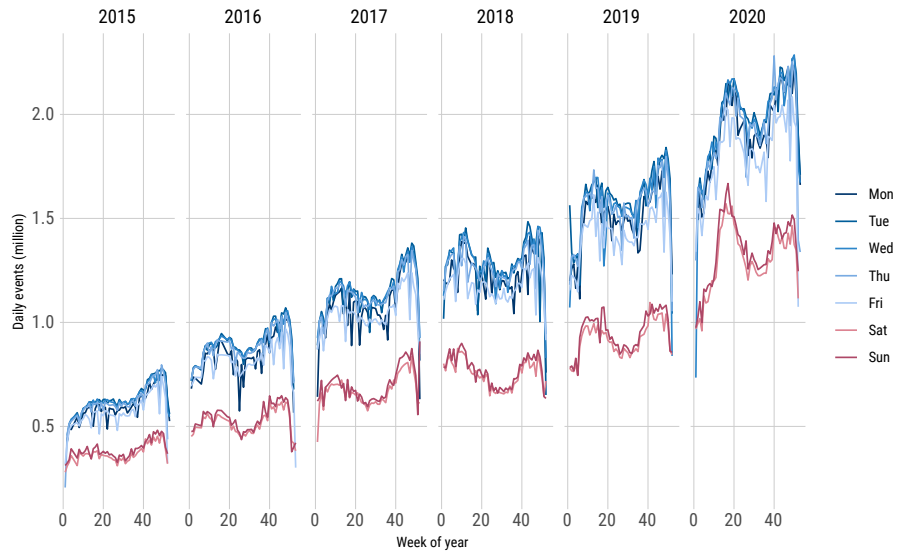
(a) Events

(b) Users

Figure S1: **Daily global GitHub activity (2015–2020).** Panel (a) recapitulates Fig. 1 from the main text, while panel (b) records active GitHub users rather than events. We include both panels here to demonstrate the direct proportionality between users and events.
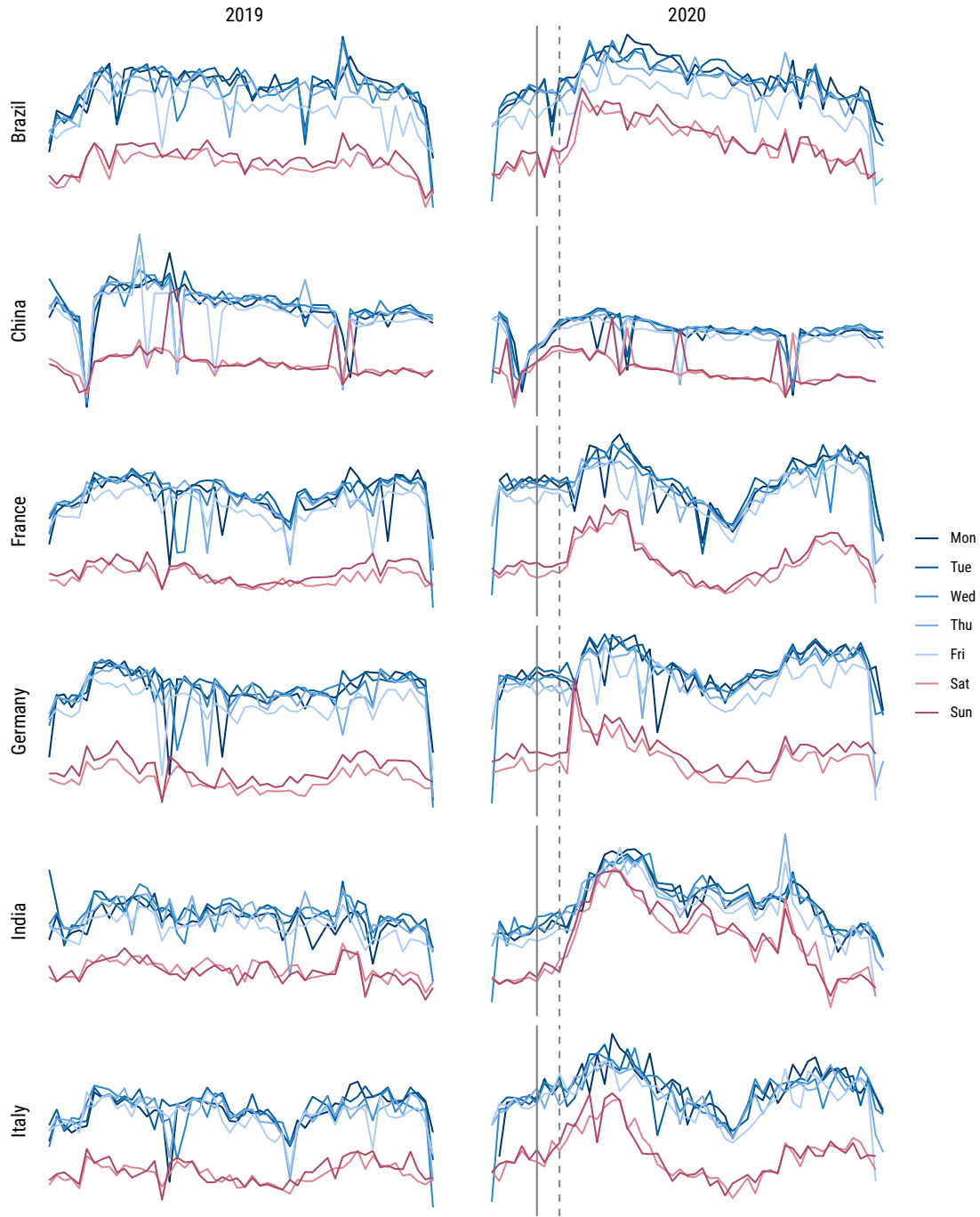
Figure S2: **Daily country GitHub activity (2019–2020), part I.** As per Fig. 1 in the main text, but highlighting event activity within the sample of countries from Table S1 over 2019–2020. The vertical lines represent the first official lockdown of any country (China) and the start of March 2020, respectively. Note that the y-axis is free-scaled to emphasise within-country trends; absolute activity levels should not be compared across countries. *Continues in the next figure.*
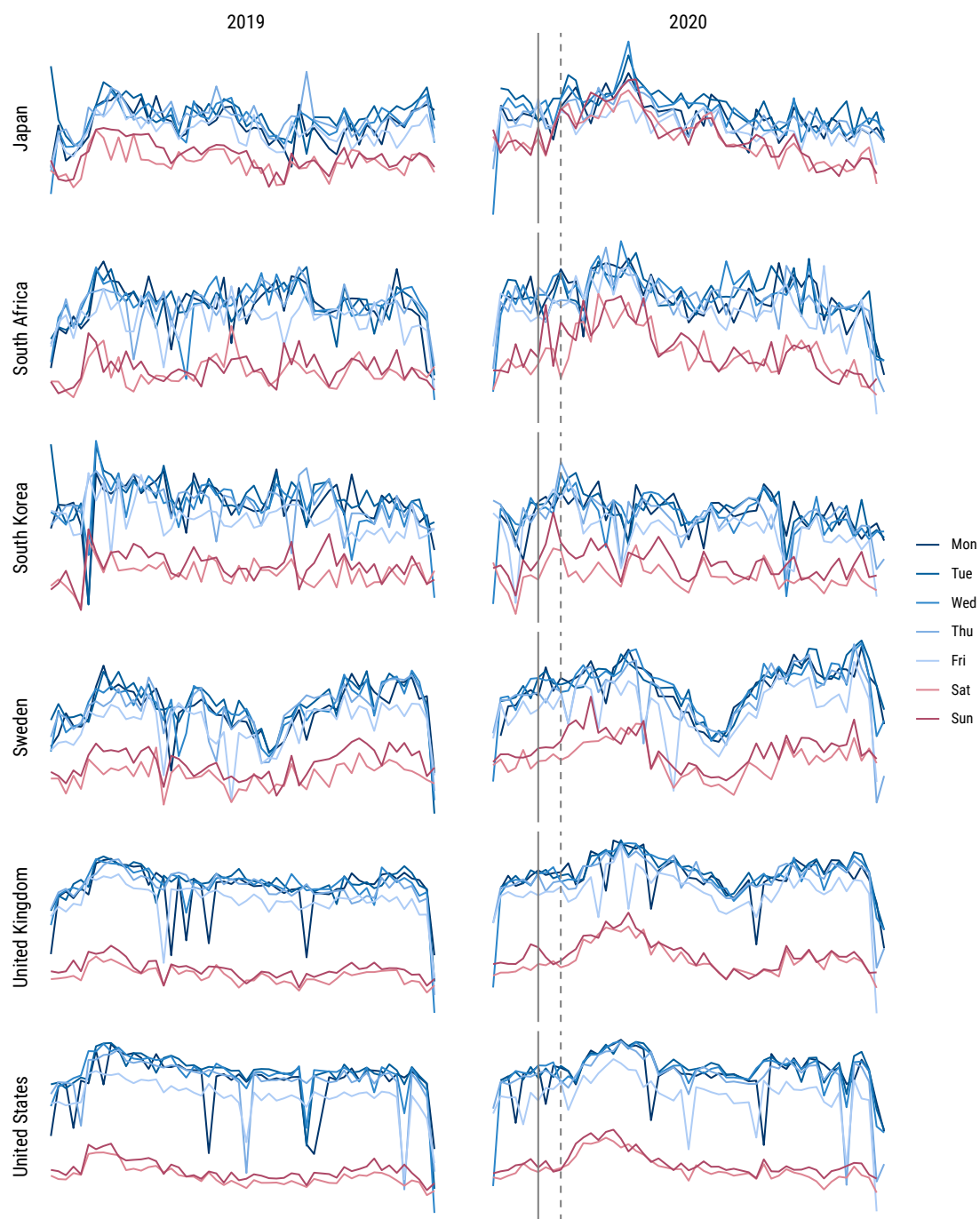
Figure S3: **Daily country GitHub activity (2019–2020), part II.** *Continues from Fig. S2.*
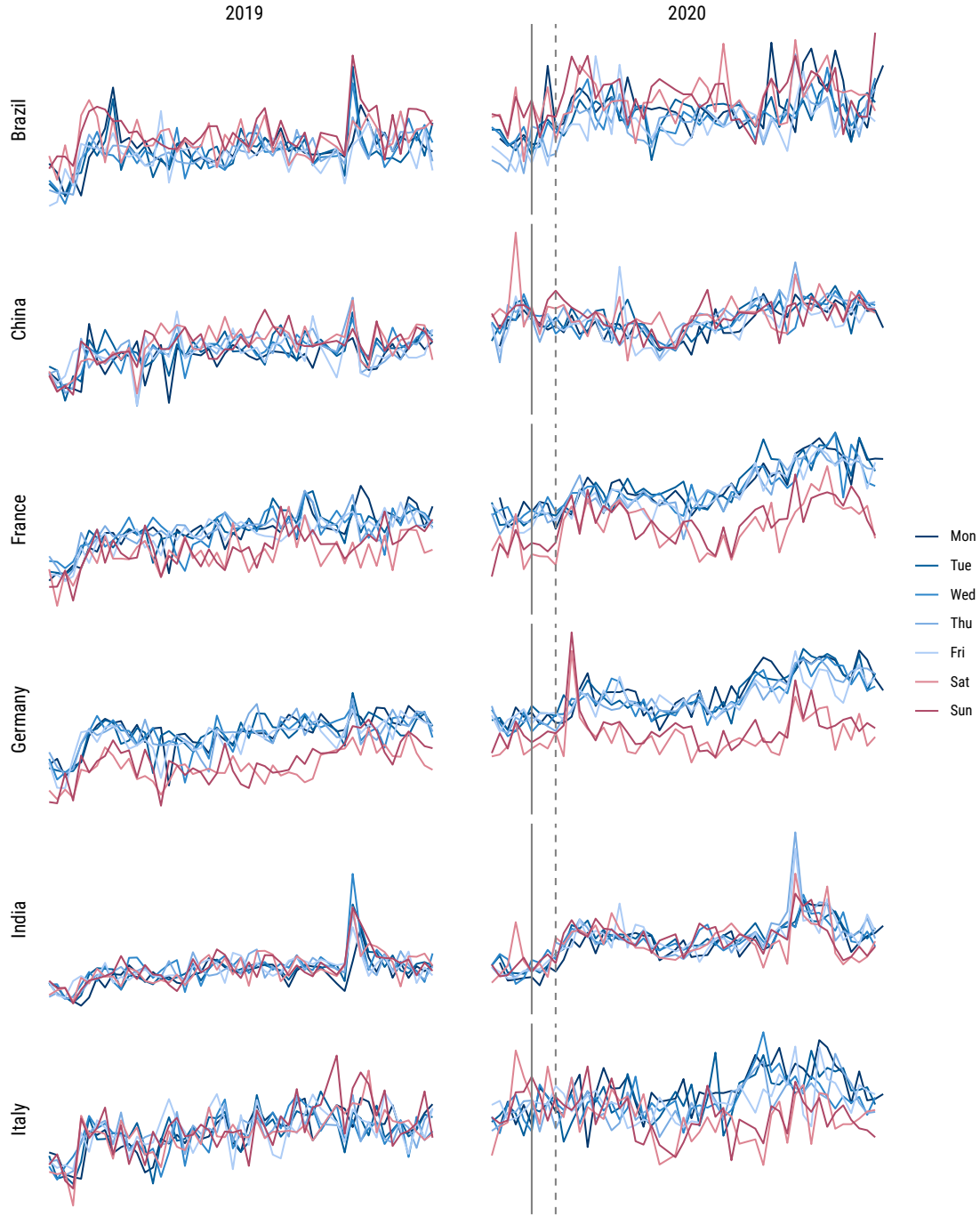
Figure S4: **Daily country GitHub productivity (2019–2020), part I.** As per Fig. S2, but now measuring productivity, which we defined here as: $Productivity_t = Events_t/ActiveUsers_t$. Again, note that the y-axis is free-scaled to emphasise within-country trends; absolute activity levels should not be compared across countries. *Continues in the next figure.*
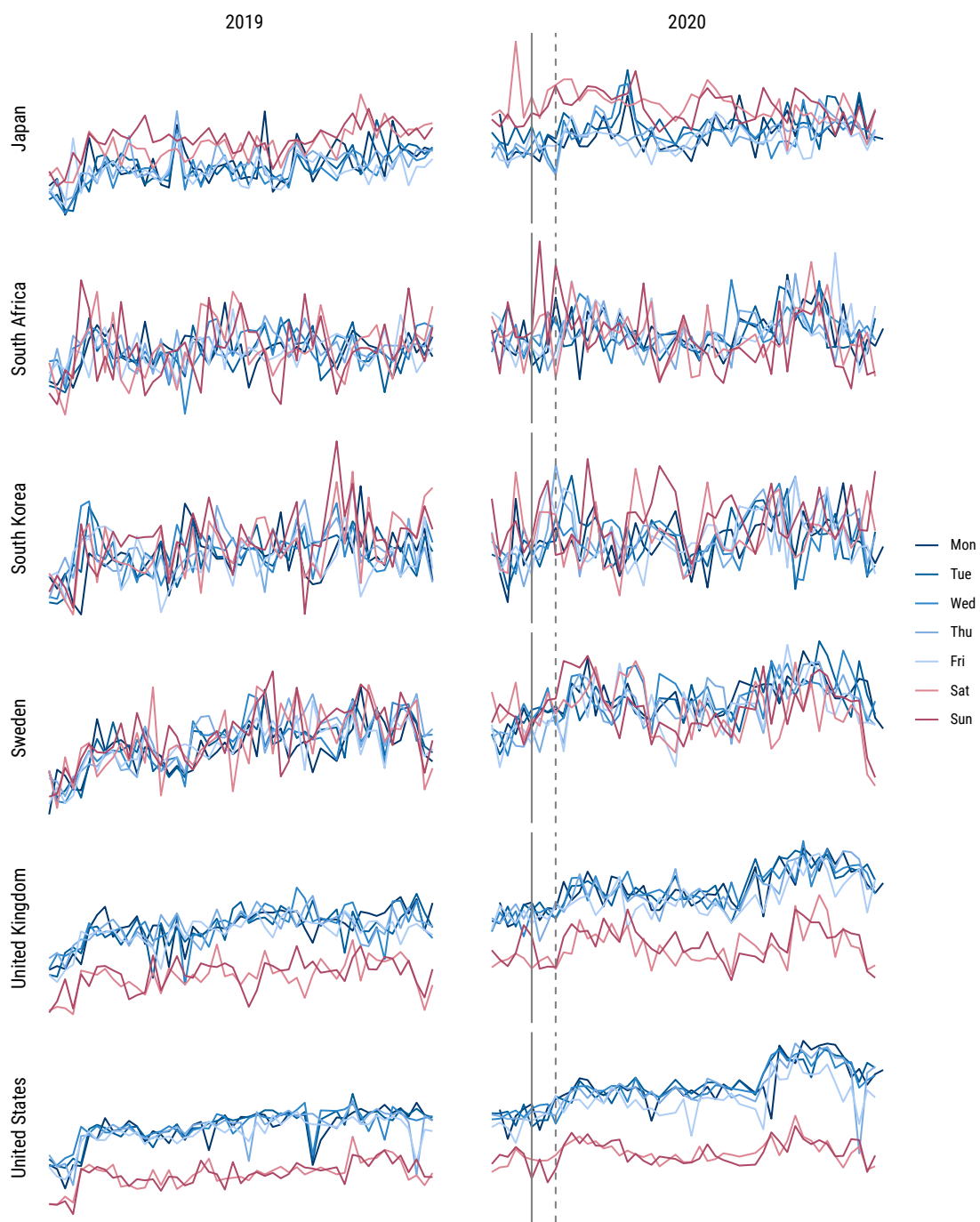
Figure S5: **Daily country GitHub productivity (2019–2020), part II.** *Continues from Fig. S4.*

Additional proportionality plots
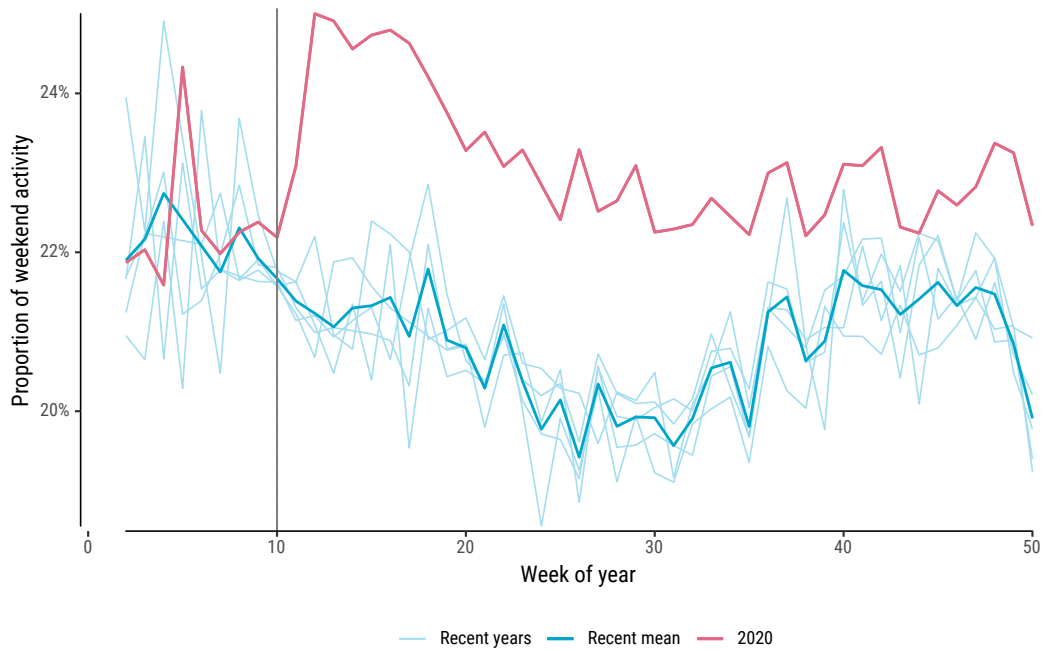
Figure S6: **Working more on weekends (push events only).** As per Fig. 3 in the main text, but limited to push events only.
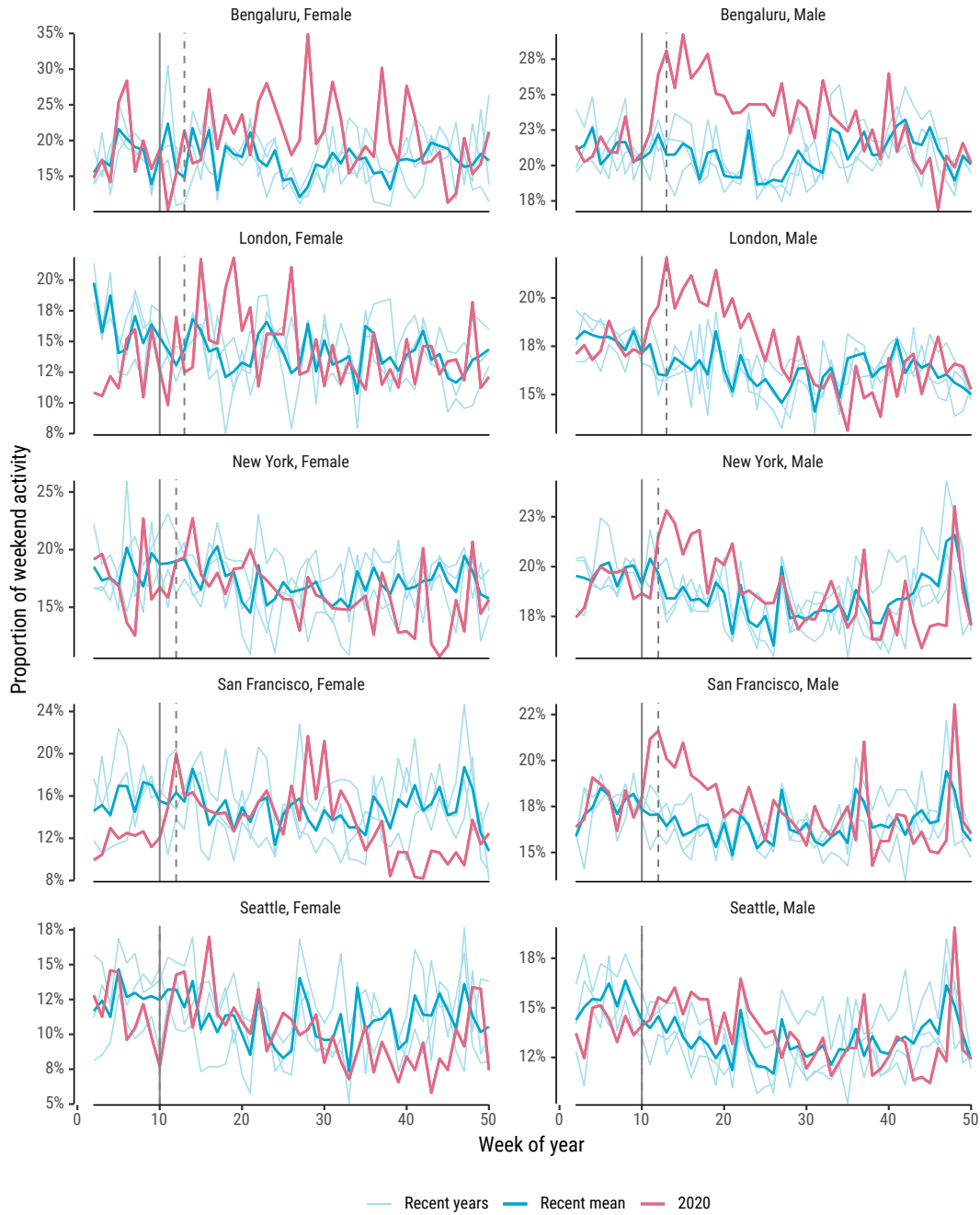
Figure S7: **Proportion of weekend activity by city and gender.** As per Fig. 4 in the main text, except (i) limited to weekend activity and (ii) broken out by imputed gender.

Figure S8: **Proportion of out-of-hours activity by city and gender.** As per Fig. 4 in the main text, except (i) limited to out-of-hours activity and (ii) broken out by imputed gender.

Additional forecasting plots
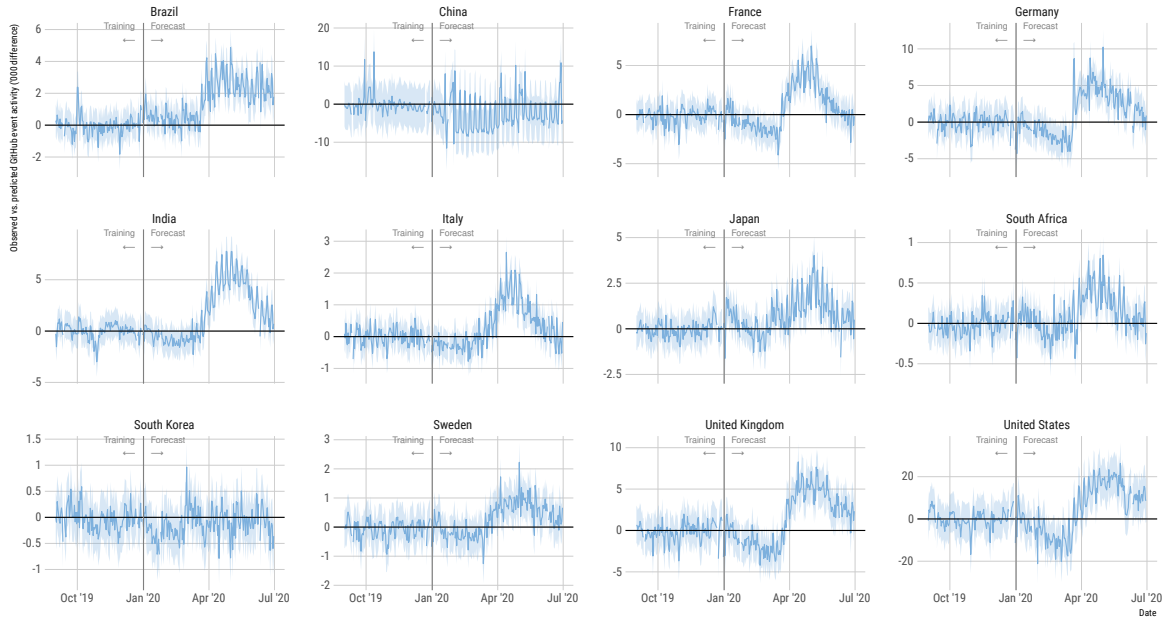
Figure S9: **Observed vs. predicted GitHub activity (absolute differences).** As per Fig. 6 in the main text, except using absolute deviations in the number of events (in thousands) rather than percentages.
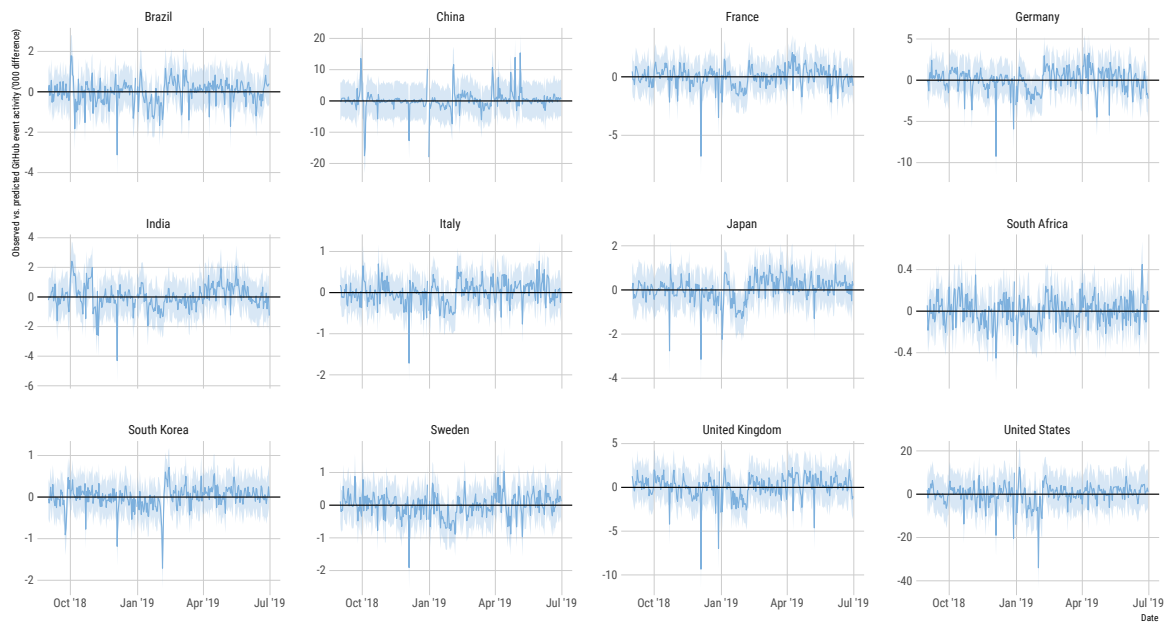
Figure S10: **Observed vs. predicted GitHub activity (placebo).** As per Fig. S9 and similar to Fig. 6 in the main text, except shifted backwards by one year (i.e. October 2018 through June 2019). The absence of any treatment effect around the equivalent COVID-19 onset — approximately the start of March — underscores that we are not picking up a spurious seasonal effect in our primary analysis.
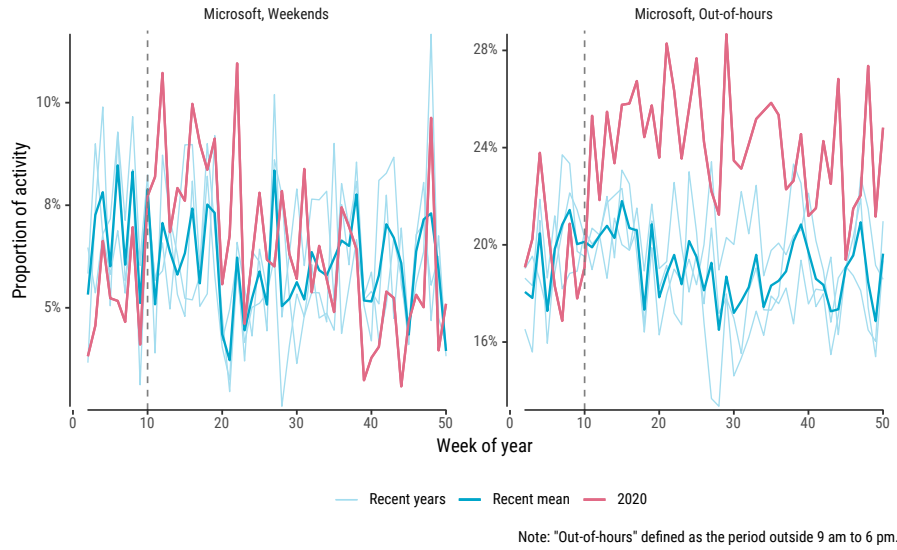
Organisation case studies

Figure S11: **Microsoft activity.** As per Fig. 4 in the main text, except limited to (i) activity on repositories belonging to Microsoft, and (ii) users located in the Seattle-Redmond area.
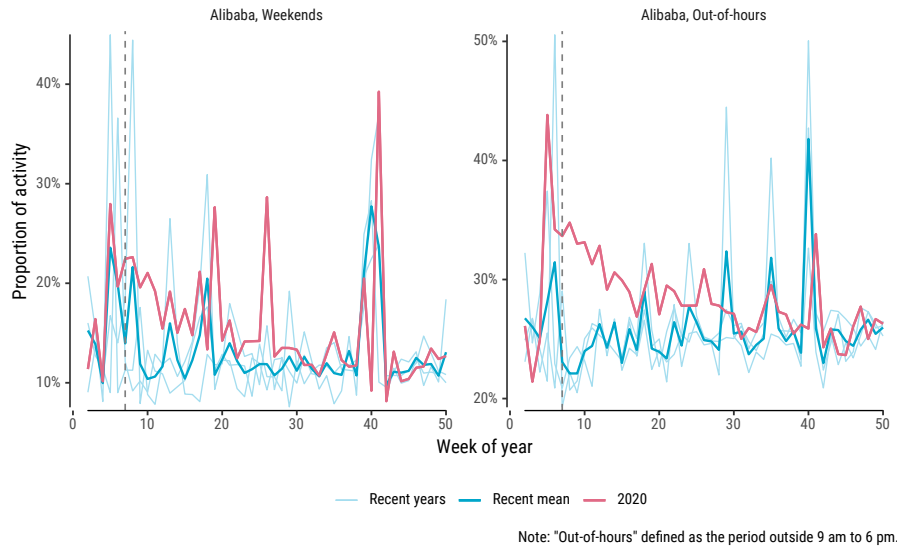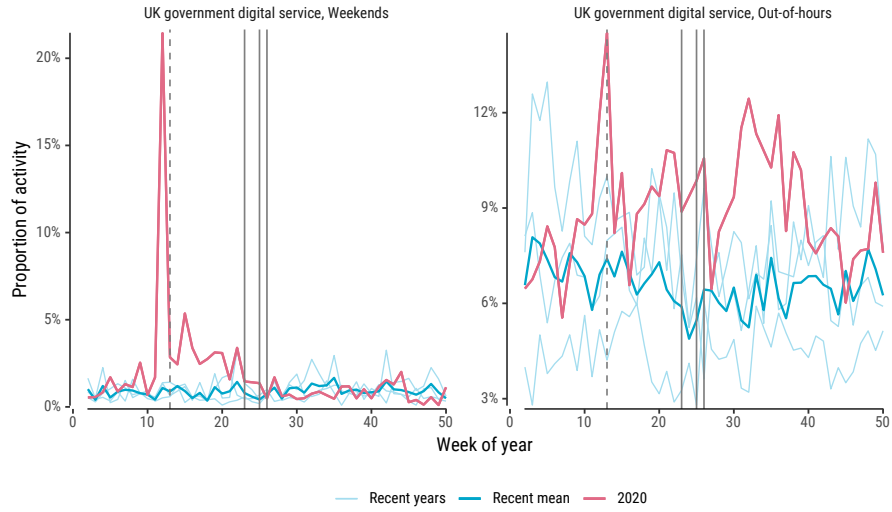

Figure S12: **Alibaba activity.** As per Fig. 4 in the main text, except limited to activity on repositories belonging to Alibaba. The timestamp adjustment assumes that all contributors are located in China.

Note: "Out-of-hours" defined as the period outside 9 am to 6 pm.

Figure S13: **UK Government Digital Service.** As per Fig. 4 in the main text, except limited to activity on repositories belonging to the (UK) Government Digital Service division. The large spikes immediately prior to lockdown (dotted vertical line) are driven by activity related to the provision of COVID-19 web services (e.g. government assistance applications). The three solid vertical lines demarcate successive reopening dates: 1) schools, 2) shops, 3) all other.

References and Notes

1. I. Grigorik, *GH Archive*. https://www.gharchive.org [Last accessed: 2021-10-08].

2. G. Gousios, *Proceedings of the 10th Working Conference on Mining Software Repositories*, MSR '13 (IEEE Press, Piscataway, NJ, USA, 2013), pp. 233–236.