

Assessing the Impacts of COVID-19 on Reddit

COVID-19, online communities, social platform change, language evolution, Reddit

Introduction. When the COVID-19 pandemic hit, much of the world moved online. Platforms of all types reported surges of activity, and many commented on various functions that online platforms suddenly fulfilled. However, we lack a rigorous understanding of the pandemic’s impacts on social platforms—and crucially, whether they were temporary or long-lasting. We present a conceptual framework for studying the large-scale evolution of social platforms and apply it to the study of Reddit’s history, with a particular focus on the COVID-19 pandemic.

Data and Methodology. Our data consists of all comments and submissions on Reddit from its inception in December 2005 to the present day in June 2022 downloaded from the Pushshift dumps [1]. This includes almost 13 billion comments by 89 million users and 5 billion submissions by 48 million users. Reddit as a platform is composed of separate communities (known as subreddits) which act as forums for users to discuss different topics.

We propose an analysis framework for capturing platform changes over time using two axes: structure versus content (major axis) and macro- versus micro-level (minor axis). Structural analysis examines the meta-level factors associated with *how* Reddit changes, including the size of the platform and the composition of its user base. On the other hand, content analysis examines *which* topical areas users are active and the language they use. Macro-level (platform-level) versus micro-level (user-level) is a complementary axis that instead focuses on changes occurring at different resolutions on the platform. Macro changes examine aggregated metrics across various dimensions, whereas the micro level change captures the user-level mechanisms that may be driving some of the macro-level movements.

Results. We showcase our general approach alongside the questions we ask in Figure 1 (a), and provide the answers to the questions in Figure 1 (b). Next we will showcase a few of the key findings below.

Structural-macro. The first variable of change we study is the structure of Reddit. To gauge how Reddit changed as a whole, we track the evolution of several platform-level metrics (activity, users, subreddits). To account for natural changes in the metrics we use Prophet, Facebook’s time series prediction tool which captures non-linear trends, holiday effects, and ignores outliers. For each metric we fit Prophet on Reddit’s historic pre-COVID data and then stop incorporating new data at the start of COVID-19 (February 2020) and predict into the future. The results from this setup are shown in Figure 2 where we track the actual change in these metrics (dark blue line) and the predicted change (yellow line). Overall, we observe large jumps in each of the metrics with the most extreme changes occurring in the number of new users and comments per month. Across all metrics, the start of COVID-19 is featured as one of the five largest shocks to the platform; and occupies the top position for the number of comments, active users, and new users.

Structural-micro. The new users that joined Reddit during the pandemic were also less likely to be active for a longer period of time than users who joined before the pandemic. We define a user as active for a longer stretch of time if they comment in a window of three to twelve months after creating their account. This window technique allows us to equally compare users across different periods of time. Figure 3 shows how the proportion of active users changes

over time. Here we observe a reversal in a previous trend which saw users being active for longer periods of time. Instead, the users that joined during COVID-19 users were less likely to stay active for longer.

Content-macro. In Figure 4 we illustrate collections of subreddits broken down by their topic. The topic classification was designed using a similar approach to Waller and Anderson through a clustering on the neural-embeddings of communities [2]. We observe that following the pandemic, the topical landscape of where discussions took place changed. To begin, the International cluster saw a major growth in activity. This cluster was composed of a series of global communities that saw a dramatic rise in COVID-19 language and political discussion. As the COVID-19 pandemic continued, this topic saw no sign of dying down, and indicates that COVID-19 led to a rapid increase in Reddit’s global adoption. Another two topic that experienced an immediate and sharp increases were (1) Advice: Dating and Life, and (2) Hobbies: 2. These topics were composed of subreddits like (1) r/NoFap, r/datingoverthirty, and r/selfimprovement, and (2) r/getdisciplined, and r/selfimprovement. The biggest winner, however, was likely NSFW content on Reddit. This content in general had little mention of COVID-19 related language, but experienced large gains in activity. In fact, during the pandemic NSFW content came to represent almost a two-fold increase in total Reddit activity going up to 13% from 6.7%.

Content-micro. We finally examined the emergence of a new COVID-19 vocabulary during the time of COVID and compared how different cohorts of users (based on the year they joined) invoked COVID-related language differently. Figure 5 shows that the use of COVID-19 related text was driven by the older cohorts turning to the platform to discuss the pandemic whereas newer members spoke less about COVID. This shows that older users were more likely to use Reddit to discuss COVID, whereas newer users used Reddit as a way to engage with content that’s not necessarily COVID-19 related.

Discussion. Our work provides an empirical foundation for future analyses of social media trends. By answering the question of “what changed,” we establish a necessary precursor to exploring the “why” behind these changes. Through our comprehensive analysis, we highlight the transformations that Reddit underwent during this period and encourage further investigations into the underlying reasons for these shifts. At the beginning we defined a series of questions that our framework was able to solve. Now in Figure 1 (b) we summarize the answers to each of these questions. In summary, Reddit experienced a truly historic shock to the system, and two years on, mostly the content-level changes remained. The initial COVID growth rate of the activity metrics was not sustained, and two years on into the pandemic a lot of Reddit’s features returned to their pre-COVID growth levels. However, COVID-19 did mark a turning point in the interests of Reddit users, as we observed a surge in NSFW content, investment discussions, and international content.

References

- [1] Jason Baumgartner, Savvas Zannettou, Brian Keegan, Megan Squire, and Jeremy Blackburn. The pushshift reddit dataset. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14, pages 830–839, 2020.
- [2] Isaac Waller and Ashton Anderson. Quantifying social organization and political polarization in online platforms. *Nature*, 600(7888):264–268, 2021.

	Structure	Content		Structure	Content
Macro	<ul style="list-style-type: none"> Q1: Was the COVID-19 increase in online activity unprecedented? Q2: Did this increase persist or was it a temporary anomaly? 	<ul style="list-style-type: none"> Q5: Did new language emerge and was it predictable? Q6: Did COVID-19 focus or broaden our interests and language? 	Macro	<ul style="list-style-type: none"> A1: COVID-19 brought about a historic change. A2: Across many of the metrics, the change was momentary. 	<ul style="list-style-type: none"> A5: A new set of words emerged that was unpredictable. A6: COVID-19 broadened both the topics we spoke about and the communities we participated in.
Micro	<ul style="list-style-type: none"> Q3: Was the activity increase driven by new or existing users? Q4: Did the COVID-19 cohort stay on the platform? 	<ul style="list-style-type: none"> Q7: Did all cohorts users experience the disruption equally? Q8: How likely were all cohorts to talk about COVID-19? 	Micro	<ul style="list-style-type: none"> A3: Newer users tended to be less active on average, but older cohorts became more active. A4: The COVID-era cohort was less likely to stay active on Reddit for long. 	<ul style="list-style-type: none"> A7: Everyone was disrupted equally, but older cohorts were quicker to shift their interests back. A8: Older cohorts were more likely to invoke COVID-related language than newer cohorts.

(a) (b)

Figure 1: An illustration of the 2x2 (structure v.s. content, macro v.s. micro).(a) Illustrates the questions we pose using our 2x2 approach, and (b) provides the answers to these questions.

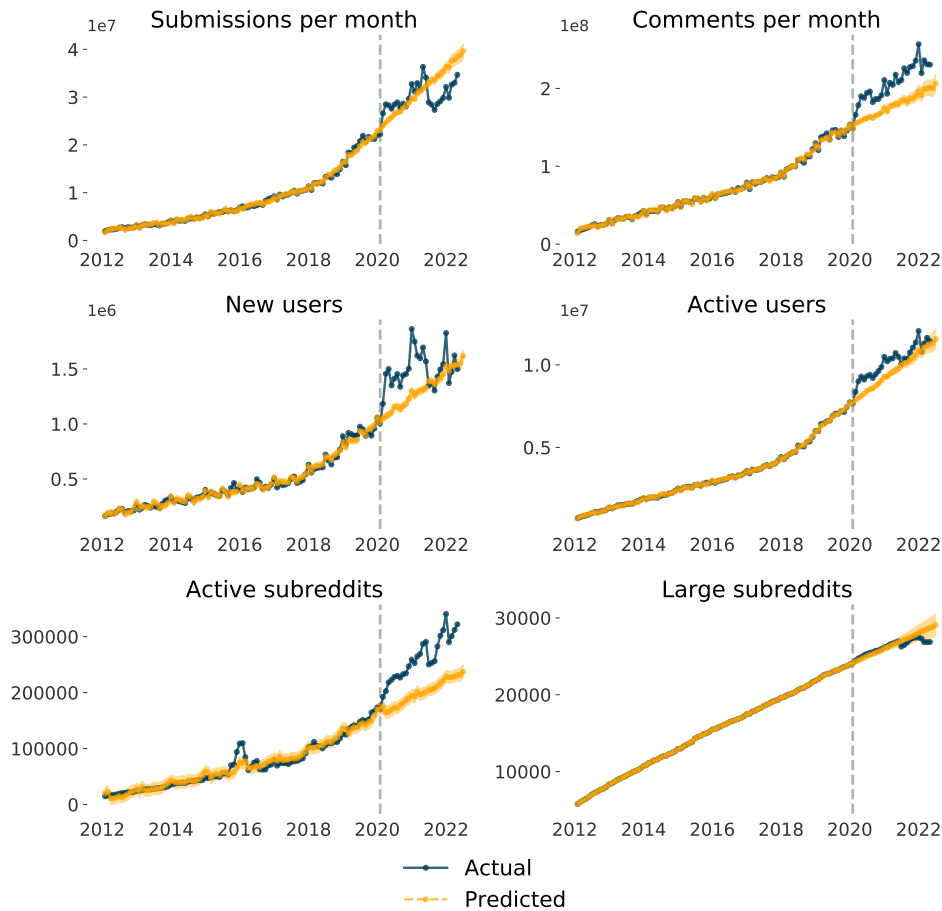


Figure 2: Predict every month until COVID-19 and then predict into the future.

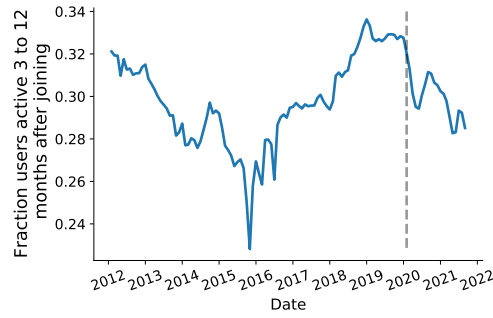


Figure 3: Proportion of users active between three and twelve months of making their first post.

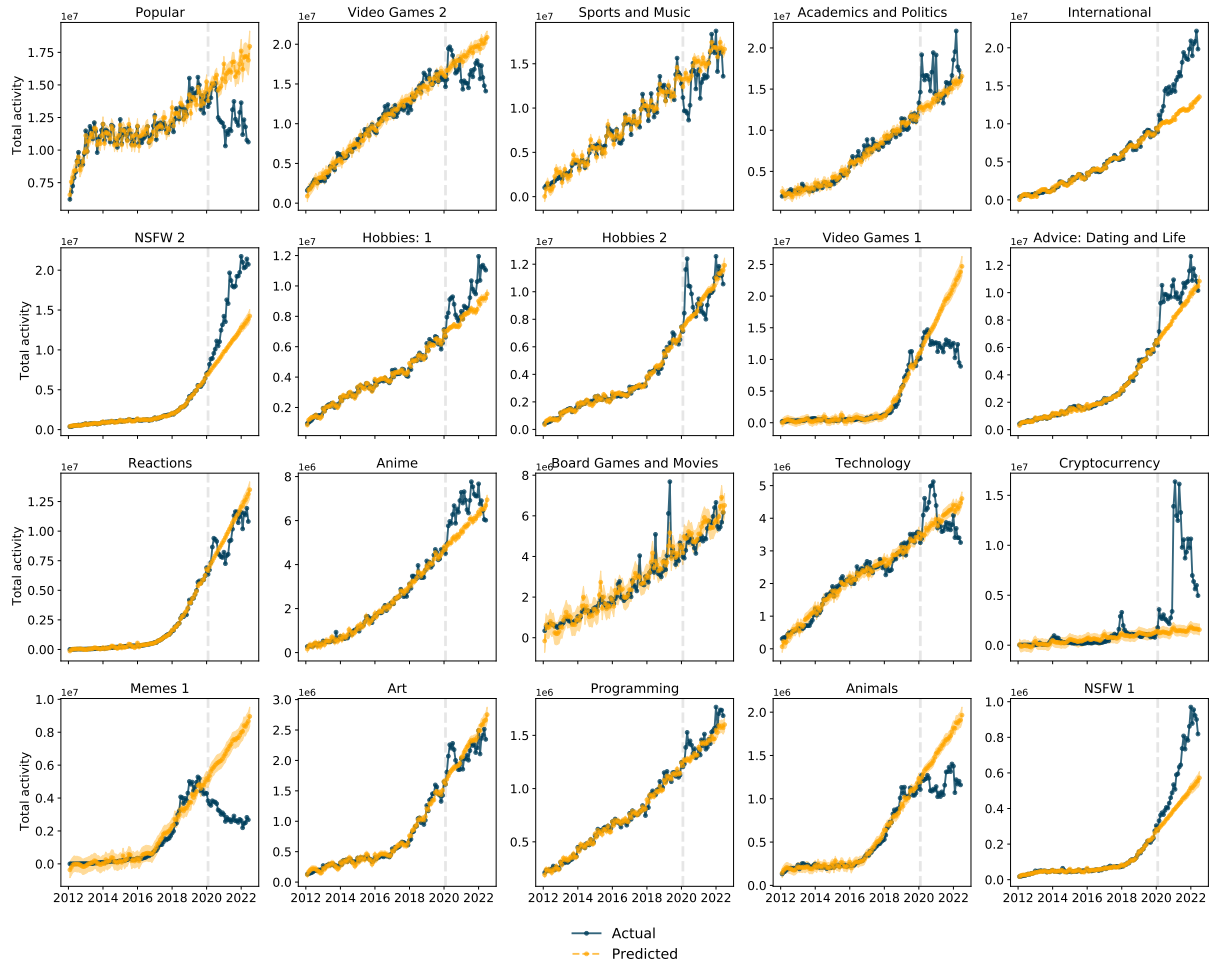


Figure 4: Evolution of topic interests over time. The blue line shows the actual change in activity in a topic whereas the yellow line shows the predicted change. The yellow area surrounding the yellow plot is the 95 confidence intervals.

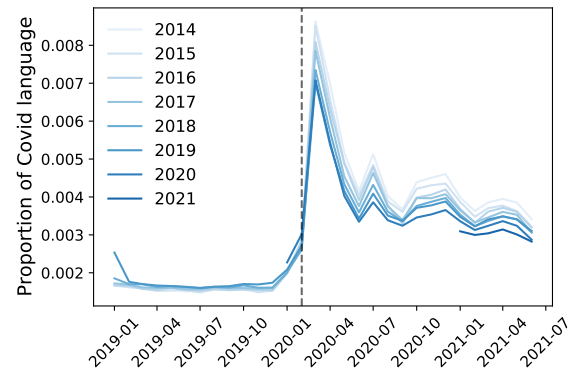


Figure 5: Differences in fraction of language in a given month that is COVID-19 related across cohorts.