A Tale of Two Countries: A Longitudinal Cross-Country Study of Mobile Users’ Reactions to the COVID-19 Pandemic Through the Lens of App Popularity

Liu Wang  
Beijing University of Posts and Telecommunications  
Beijing, China  
w_liu@bupt.edu.cn

Haoyu Wang  
Huazhong University of Science and Technology  
Beijing, China  
haoyuwang@hust.edu.cn

Yi Wang  
Beijing University of Posts and Telecommunications  
Beijing, China  
iwang@bupt.edu.cn

Gareth Tyson  
The Hong Kong University of Science and Technology  
Hong Kong, China  
g.tyson@qmul.ac.uk

Fei Lyu  
Beijing University of Posts and Telecommunications  
Beijing, China  
feilyu@bupt.edu.cn

ABSTRACT
The ongoing COVID-19 pandemic has profoundly impacted people’s lives around the world, including how they interact with mobile technologies. In this paper, we seek to develop an understanding of how the dynamic trajectory of a pandemic shapes mobile phone users’ experiences. Through the lens of app popularity, we approach this goal from a cross-country perspective. We compile a dataset consisting of six-month daily snapshots of the most popular apps in the iOS App Store in China and the US, where the pandemic has exhibited distinct trajectories. Using this longitudinal dataset, our analysis provides detailed patterns of app ranking during the pandemic at both category and individual app levels. We reveal that app categories’ rankings are correlated with the pandemic, contingent upon country-specific development trajectories. Our work offers rich insights into how the COVID-19, a typical global public health crisis, has influenced people’s day-to-day interaction with the Internet and mobile technologies.

CCS CONCEPTS
• Human-centered computing → Ubiquitous and mobile computing.

KEYWORDS
COVID-19, app store, app popularity, cross country

1 INTRODUCTION
The ongoing COVID-19 pandemic has impacted almost every aspect of human life. The pandemic introduced changes in people’s perceptions and attitudes, as well as how they work, pursue leisure, interact with others, and so on [12, 15, 20]. Among these changes, interacting with internet and information technologies could be very significant. Compared with prior years, technologies have played a critical role in the pandemic, as an essential infrastructure for supporting the response to the pandemic. For instance, to help curb its spread, governments and public health authorities around the world have launched contact-tracing apps [9, 14, 19, 23]. Given that mobile phones have been the most popular and pervasive interfaces for people to interact with the Internet, most technological experience has occurred via mobile apps. We therefore posit that the popularity of mobile apps, as well as their dynamics, may provide a lens for understanding people’s reactions to the pandemic.

Due to the different strategies employed by countries in fighting COVID-19, as well as the degree of public compliance, the regional development of the pandemic has exhibited distinct patterns. For example, China introduced a strict national lockdown during the initial outbreak and achieved strong public cooperation, making the pandemic largely under control in most of areas from April, 2020 [11]. In contrast, the United States, although implementing multiple measures at different levels, failed to contain the pandemic, resulting in chaos spanning over an entire year [6]. We conjecture that such differences may be reflected in the app popularity in different countries. Thus, understanding people’s experiences with mobile apps must consider the country-specific patterns of the pandemic.

This Work. In this paper, we seek to understand, characterize, and compare the dynamics of app popularity during the pandemic, under the circumstance of different country-specific pandemic development patterns. We are particularly keen to explore the relationship between app utilization and the infection trajectories experienced across countries. We first create the daily snapshots of the ranking of popular apps in the iOS App Store in both China and the United States (US), from the January 2020 to the end of June 2020, the time period of initial rapid outbreak of the virus (See Section 3). This longitudinal dataset enables us to study app usage behavior evolution across billions of mobile users in both China and US. We next characterize the dynamics of app popularity from both a Macro- (i.e., app category) and Micro- (i.e., individual app) lens.
We investigate which app categories are positively or negatively correlated with the coronavirus pandemic and try to formulate some trends in their ranking evolution (See Section 4), especially when comparing the differing situations and cultures in China and the US. We further perform fine-grained per-app analysis (See Section 5), to summarize the dynamics of app popularity. Finally, we investigate whether the COVID-19 pandemic has introduced side-effects on the app maintenance behaviors (e.g., performance) by analyzing millions of app reviews (See Section 6). We highlight a number of key findings:

- **COVID-19 has played a key role in changing the popularity of some categories of app.** During the outbreak, several app categories experienced significant ups and downs in popularity in both China and the US, particularly Business, Education, Navigation, and Travel. This change reflects the reactions and efforts people made in response to the outbreak, and highlights differences between China and the US.

- **The impact of COVID-19 on app rankings are diverse and can vary from country to country.** There are many apps in our dataset whose rankings are strongly correlated, either positively or negatively, with the pandemic case rates. Their ranking variations can be classified into several groups. Most apps within the same group share some similarities in their adaptation to the pandemic situation.

- **The rapid popularity of some apps may be accompanied by the side effect of declining ratings, especially for Chinese apps.** We explore the reasons for this from the app reviews and found a massive increase in the number of negative reviews compared to the pre-pandemic period. We reveal some potential challenges for apps.

To facilitate further study along this direction, we have released our dataset and the experimental results to the research community at Github: https://app-popularity-covid19.github.io/

## 2 BACKGROUND AND RELATED WORK

### 2.1 App Ranking in iOS App Store

App ranking refers to the position of an app in a store or marketplace. In general, app rankings can reference not only how an app ranks in a search, but also the overall app rankings from the store. These rankings are country specific, and can be viewed at an overall or category level. The iOS App Store, one of the most common app marketplaces, provides the real-time app ranking lists both overall and by category. In the world of apps, ranking is of utmost importance. A minor change in rank can make a significant difference in traffic and revenue. AppTentive’s recent mobile consumer survey [1] reveals that nearly half of all mobile app users identified browsing the app store charts and search results (the placement on either of which depends on rankings) as a preferred method for finding new apps. Simply put, higher rankings mean more downloads, and to some degree, more money for app developers.

The iOS App Store has a complex and highly guarded algorithm for determining rankings for both keyword-based app store searches and composite top charts. Although the exact ranking algorithms are not publicly available, there are several known factors that influence the rankings, including downloads (how many people have downloaded the app), rating (how many stars people give the app), rating count (how many ratings the app has) and trends (how quickly the app is growing), etc [5]. Since it is the combination of several other factors (average rating, number of ratings, installs, etc.), this ranking speaks to the overall awareness, usefulness, and satisfaction of an app. Consequently, app ranking is a particularly useful metric for mobile app analysis, particularly, its popularity.

### 3 STUDY DESIGN

#### 3.1 Research Questions

Our study is driven by the following research questions:

**RQ1 General Trends.** Across different regions (i.e., China and US), what are the characteristics of app popularity during COVID-19, and how are they impacted by the pandemic trajectory? We examine the popularity of app categories and the correlation between app ranking and the pandemic.

**RQ2 Behavior Patterns.** What are the different patterns in the change of app popularity during the pandemic? We attempt to classify the app ranking variations based on relevance to the pandemic, then examine the distribution of the number of apps in each class and explore their characteristics.

**RQ3 Side-effect.** Considering that app popularity may change greatly during the COVID-19 pandemic, does it introduce any side-effect on app maintenance behaviors? Specifically, we characterize the side-effect from the perspective of app ratings and reviews.

#### 3.2 Data Collection

We first harvest a dataset of app popularity. Given that the most straightforward impact of COVID-19 on the mobile app ecosystem will be mirrored in a number of popular apps, we leverage the App Store’s ranking list and focus on the top-ranked apps. However, it is non-trivial to obtain such a dataset as it requires effort to continuously monitor apps in the app market to check their rankings. To
this end, we implement a crawler to retrieve the daily app ranking list from the App Store. Considering that the pandemic in different countries can evolve in different ways, we take two representative countries (China and the US), as case studies for comparison. To be specific, we crawl the ranking list with top 1500 apps (total available apps in the App Store) each day from January 1 to June 30, 2020, to keep track of the apps and their daily rankings, containing over 543K records in total for the two countries. This was the time when the abrupt COVID-19 unexpectedly erupted, during which people’s lives underwent drastic changes. We then take the top 100 apps per day as the target of this study as they better reflect the direct impact of COVID-19. Overall, there are 586 unique apps in China and 590 in the US covering 22 categories\(^1\), each of which ranked in the top 100 for at least one day during a given half year. Besides, we collect ratings for each app on a daily basis, as well as the reviews for some of these apps. Table 1 summarizes our dataset.

| Country | # Categories | # Unique Apps | # Ranking Records | # Ratings | # Reviews |
|---------|--------------|---------------|-------------------|-----------|-----------|
| China   | 22           | 590           | 271,765           | 106,653   | 980,573   |
| US      | 22           | 590           | 271,522           | 107,381   | 756,617   |

### 3.3 Analysis Preliminaries

We briefly present the analysis methods used in this paper.

#### 3.3.1 Popularity of app categories

To understand which app categories are more competitive and explore some trends in ranking evolution, we begin our analysis by investigating the most intuitive metric of app-category popularity, i.e., the number of apps ranked in the top 100 for each category. Specifically, we measure the popularity of app by app volume, i.e., the number of apps belonging to that app category in the top-100. The larger the number, the more popular the category is. We examine the number of apps in each category on a daily basis to gain insight into how their popularity trends are changing. In more detail, we zoom in on each category to check the ranking variations of all apps in that category.

#### 3.3.2 Correlation analysis

To understand if any apps have been significantly affected by the pandemic, and whether the impact has been positive or negative, we also study the correlation between the pandemic evolution and the app ranking changes, which could reflect the popularity changes of apps, for each app. We first calculated the Cohen’s d [2] to measure the effect size. Cohen’s d is a standard score that summarizes the difference in terms of the number of standard deviations. Typically if \(d > 0.8\) can be considered as having a large effect size. In our experiment, the calculated Cohen’s d ranges from 0.7 to 1.3, most of which are interpreted as a large effect size. Moreover, we also performed Tukey’s test and we can see a statistically significant difference (\(p\)-value \(< 0.05\)) among the results. In our case, we measure the development of the pandemic situation over the half year through the metric of daily active cases retrieved from Google Statistics [3]. As for the correlation analysis, we considered two different commonly used correlation coefficients, i.e., Pearson correlation coefficient [4] and Spearman correlation coefficient [8]. We calculated and compared both them and found that there is minimal difference between them. For apps with really strong correlation, the correlation coefficient calculated with either correlation method is high. For instance, for app examples listed in Section 5 (e.g., Dianping, TikTak Travel), the correlation coefficients calculated by both methods can reach above 0.9. Similarly, for most apps with insignificant correlation, the values calculated by both correlation methods are also very low. This makes the choice between the two correlation coefficients become less problematic. Finally, we choose the Pearson correlation and present a correlation comparison strategy based on it.

For each app, the daily ranking is viewed as a vector in Pearson correlation coefficient formula, then the correlation degree is obtained from the daily number of active cases in the corresponding countries. Note that we also take \(P\)-value into account because it tells us whether the result of an experiment is statistically significant. Typically if the \(P\)-value is lower than the conventional 5% (\(P < 0.05\)), the correlation coefficient is called statistically significant. Thus, results with \(P\)-value less than 0.05 are of concern to us, otherwise are ignored as insignificantly correlated. Finally, we calculate the correlation coefficient for each app and obtained 425 (72.5%) apps in China and 502 (85.1%) apps in the US with statistically significant correlation with daily active cases. The results illustrate that the majority of apps have a statistically significant correlation between their rankings and the COVID-19 case rates. This encourages and motivates the following analysis, i.e., the distributions of these correlation coefficients at the overall and category level. And note that all the correlation coefficients provided below are statistically significant.

### 4 POPULARITY OF APP CATEGORIES

#### 4.1 The dynamics of the popularity

##### 4.1.1 Coarse-grained inspection

Figure 1 visualizes the dynamics of the popularity of each app category. We use stacked bar charts to show the number of apps in each category among each day’s top 100 apps in the app stores of China and the US, respectively. As aforementioned, there are 22 categories (see the legend in Figure 1). Apparently, the number of apps in each category over time is represented by the fluctuations of the vertical lengths of the corresponding color. In general, most categories’ popularity does not exhibit significant fluctuations in either China or the US. Some categories have been constantly popular over the long run, such as Games, Social Networking, Photo & Video and Entertainment, as well as some being consistently less popular, such as Reference, Medical and Weather. However, there are four categories that are worth noting in both countries, i.e., Business, Education, Travel, and Navigation, whose popularity exhibits remarkable dynamics along with the outbreak in the respective countries. We further examine the details of the dynamics of their popularity.

##### 4.1.2 Fine-grained inspection

For each of the four categories, we plot the overlay of scatter and line charts for both countries in Figure 2. The scatter represents the ranking of each app in that category and the lines show the number of newly confirmed and active COVID-19 cases per day. First, we observe a number of Business apps (e.g., Zoom Cloud Meetings) and Education apps (e.g., Tencent Classroom) jumped to the top of the list. This started in early February in China and mid-March in the US, just as the COVID-19 began to spread widely in the corresponding countries. This observation is quite straightforward since the outbreak had quickly

\(^1\)The categories are defined by the iOS App Store.
led to a large-scale shift to remote work for employees and online learning for students. Besides, the Travel apps (e.g., Ctrip) and Navigation apps (e.g., Baidu Maps) show the opposite trend during the same time, with almost all of them falling out of the top 100. This is also reasonable because the mandatory social distancing and quarantines are required for containing virus transmission, which left such apps less desired. It is important to note that the ranking changes in these four categories starts almost simultaneously with the outbreak, which reflects the pandemic’s quick impact. In sum, the findings indicate that mobile app ecosystem has had rapid reactions to the COVID-19 outbreak. Regarding app categories, COVID-19 has an effect of enhancing or decreasing the popularity of certain categories of apps.

Comparing the two countries, it is also interesting to note that Navigation and Travel apps in the US rebounded in mid-April and mid-May, respectively, while the US was still in the midst of an increasingly severe pandemic situation. In contrast, the rankings for these two categories in China improved only after the conditions began to improve. To some extent, this may reflect the different attitudes of the two countries in dealing with COVID-19.

4.1.3 Compare to the previous year. Considering that there are also possible seasonal trends in app usage and popularity that would not be accounted for in the data for only one year in 2020. For more reliable results, we then collected the relevant data for the same period in 2019 and compared them to see if there are seasonal trends that appear to line up with the pandemic trajectory. As a result, the dynamics of the popularity of those app categories (i.e., Business, Education, Navigation, Travel) in 2019 have been distributed relatively even in the first half of 2019 without reflecting a pandemic trajectory at all. (For the sake of saving space, we did not paste the figure.) This comparison indicates that there is no clear seasonal trend between 2019 and 2020 in both countries, and thus we can safely ignore the influence of seasonal factors on the popularity of app categories. Undeniably, there are other confounding factors that may influence the ranking and popularity of apps, such as release of new apps. But these effects are usually reflected in individual apps and have a minimal impact on the whole category. COVID-19, however, as a global public crisis event of tremendous damage and extensive scope during this period, has undoubtedly made a huge impact on people’s lives, and we believe it is the main cause leading to such changes in popularity of app categories.

4.2 Correlation Analysis
In this section, we provide correlation analysis results at both the overall and category levels.
4.2.1 Overall distribution. Figure 3 presents the Cumulative Distribution Function (CDF) of the correlation coefficients between app ranking and the number of active confirmed cases in the two countries. The trends are very similar in the US and China, with a roughly 50:50 number of positively and negatively correlated apps. Besides, there are indeed a number of apps whose ranking changes have a clear correlation with the pandemic evolution in both countries. Specifically, over 47% of the apps in China and 60% in the US have correlation coefficients greater than 0.4 in absolute value. In addition, 33% of the apps in the US and 20% in China have correlation coefficients greater than 0.6 in absolute value, which can be considered as strong or very strong correlations (see Table 2). This observation suggests that many apps’ popularity does have a significant correlation with the development of the pandemic, positively or negatively.

4.2.2 Category-level distribution. We further study the distribution of correlations at the category level. Figure 4 shows the boxplot of correlation coefficients for each category in China. The red line in the box indicates the median, while the blue line represents the mean. First, we observe that the two most prominent categories are Travel and Navigation, with all correlation coefficients being positive and a median greater than 0.75. This indicates that there is a significant positive correlation between pandemic situation and the ranking of apps in these two categories, i.e., the ranking decreases as the pandemic situation worsens and increases when the situation improves (as measured by case rate).

Moreover, most apps in Education category have negative correlation coefficients with a median of about -0.6, suggesting that Education apps are in general negatively correlated with the pandemic situation. The findings are consistent with the evolution of the popularity of these three categories, as previously discussed. However, for the Business category, there seems to be no significant negative correlation discernible. A possible explanation is the long-term popularity of most Business apps after the outbreak, as we can observe in Figure 2(a). Some Business apps jumped up the ranking during the outbreak and have been leading ever since, even though the number of active cases has dwindled to almost nothing. As a result, there is no significant positive or negative correlation in

Figure 2: Evolution of the daily ranking of apps and the number of active cases in four categories (Business, Education, Travel and Navigation) in two countries.

Figure 3: The overall distribution of correlation coefficients.

Figure 4: Correlation coefficients across app categories in China.
the later phases, thereby undermining the overall correlation coefficient. We posit that we can examine it in more detail by splitting it into two stages, which we will discuss later in Section 5. In general, although most of the categories do not show a significant positive or negative correlation, there are certain categories that do, such as Travel and Navigation showing a strong positive correlation and Education presenting a strong negative correlation.

Besides, since a larger correlation coefficient shows a more significant correlation, we consider introducing a threshold \( n \), and regard an app as strongly positively or negatively correlated with the outbreak (overall or in a stage) whose absolute correlation coefficient is greater than \( n \). As such, the key challenge is to select an appropriate \( n \). Nevertheless, the correlation coefficient represents an effect size and so we can verbally describe the strength of the correlation using the guide that Evans [16] suggests for the absolute value of \( r \) in Table 2. Thus, we set the threshold to 0.6. Specifically, we take \( r > 0.6 \) as strong positive correlation (SPC) and \( r < -0.6 \) as strong negative correlation (SNC). At last, we create a taxonomy including eight groups for China and two groups for the US.

**Table 2: A commonly used interpretation[16] of the \( r \).**

| Magnitude of Correlation | Description of Strength |
|--------------------------|-------------------------|
| 0.01 - 0.19              | Very Weak               |
| 0.20 - 0.39              | Weak                    |
| 0.40 - 0.59              | Moderate                |
| 0.60 - 0.79              | Strong                  |
| 0.80 - 1.00              | Very Strong             |

5.1 Taxonomy

5.1.1 China. In China, we consider the situation from both an overall and a phased perspective.

(1) Overall-level. We start with checking the correlation coefficient \( r \) between the app ranking and the overall period of the pandemic. If \( r > 0.6 \) we classify the app as “SPC with the overall.” Similarly, we classify the app as “SNC with the overall” if its \( r < -0.6 \).

(2) Stage-level. For the apps without a strong correlation between its ranking and the overall pandemic period, we then examine their correlations with the two stages respectively, i.e., the growth stage of the pandemic and the decline stage of the pandemic. If an app’s rank is strongly correlated to only one of the stages, we label it as a one-stage correlation. Concretely, there are four groups: “SPC with growth stage”, “SNC with growth stage”, “SPC with decline stage” and “SNC with decline stage.”

If an app’s ranking is strongly correlated with both stages but in opposite directions, we deem it as two-stages correlation. Specifically, there are two groups: “SPC with growth stage, SNC with decline stage” and “SNC with growth stage, SPC with decline stage.”

5.1.2 US. In the US, the situation is relatively straightforward due to the single uptrend of the pandemic, and we can simply focus on two groups, i.e., “SPC with the overall” and “SNC with the overall.”

5.1.3 Results. Based on the above, we classify the apps in the two countries separately and the results are shown in Table 3. For China, we have eight groups with the number of apps ranging from 2 to 44. In the US, there are two groups with app quantities of 72 and 116, respectively.

5.2 Per Group Analysis

5.2.1 SPC with the overall in China. There are 41 apps in China whose app rankings are strongly positively related to the pandemic, i.e., they suffered a regression in ranking when the pandemic got worse. Subsequently, their rankings rebounded as the situation recovered. The app with the strongest correlation (0.911) is Dianping.
a local lifestyle information and trading platform in China, providing users with information services such as consumer reviews and transaction services such as group buying. Followed by TikTak Travel (0.885), Ctrip Travel (0.882) and Tencent Maps (0.876), all of which cater to services like travel and navigation. Unsurprisingly, most of the apps in this group are closely tied to outdoor activities. Given the strict quarantine policy during the outbreak in China, it is no wonder that the rankings of this kind of apps undergo such a shift. The trend comparisons for two representative apps are visualized in Figure 6. The red line shows the ranking of the app each day and the green line presents the number of active confirmed cases per day. We can observe that their trends are almost identical.

![Figure 6: Two examples for SPC with the overall in China.](image)

5.2.2 **SNC with the overall in China.** The situation in this group is opposite to the previous one. There is a strong negative correlation between the app ranking and the pandemic case rate for 44 apps in this group. This means that the outbreak has contributed to the growth of their rankings. And the subsequent recovery has resulted in a drop in the rankings. For context, Figure 7 presents two examples with these properties. We see opposite trends in the situation where app rankings are strongly negatively correlated with the growth stage of the pandemic. In other words, as the situation gets worse, the app ranking drops; yet it does not show a definitive upward or downward trend as the situation recovers. For example, there are two typical apps shown in Figure 8. One is called Eastern Airlines that aims to provide safe and convenient ticketing and travel experiences. The other is Damek, a comprehensive live entertainment ticket marketing platform in China, covering concerts, dramas, musicals, sporting events, etc. Both of them have a high correlation coefficient between app ranking and active cases in the pandemic growth stage (0.933 and 0.772, respectively). However, such correlations are not that strong during the decline stage. This means there are some apps that are adversely affected by the pandemic but do not recover in tandem with the pandemic.

![Figure 7: Two examples for SNC with the overall in China.](image)

5.2.3 **SPC with the growth stage in China.** In this group, the app ranking is strongly positively correlated only with the growth stage of the pandemic. In other words, as the situation gets worse, the app ranking drops; yet it does not show a definitive upward or downward trend as the situation recovers. For example, there are two typical apps shown in Figure 8. One is called Eastern Airlines that aims to provide safe and convenient ticketing and travel experiences. The other is Damek, a comprehensive live entertainment ticket marketing platform in China, covering concerts, dramas, musicals, sporting events, etc. Both of them have a high correlation coefficient between app ranking and active cases in the pandemic growth stage (0.933 and 0.772, respectively). However, such correlations are not that strong during the decline stage. This means there are some apps that are adversely affected by the pandemic but do not recover in tandem with the pandemic.

5.2.4 **SNC with the growth stage in China.** Similarly, we look at the situation where app rankings are strongly negatively correlated with only the growth stage of the pandemic, i.e., as the pandemic worsened, the app ranking improved, but there was no clear trend of a growing or declining thereafter. We highlight the following examples: 1) Tencent Classroom (-0.943) is a professional online education platform that provides teachers and students with interactive
5.2.5 SPC with decline stage in China. In this group, the correlation between app ranking and the pandemic case rate is not significant in the first stage of the pandemic, but as the situation turns better, the app ranking rises. For example, the app with the highest correlation coefficient (0.924) in the decline stage is Shell Finder, a home searching platform featuring real property information as well as industry innovations such as VR viewings. In addition, Traffic Management 12123 (0.881) is the official client of the traffic safety management platform, providing a full range of traffic services and reservation services. As shown in Figure 10, we can observe that the rankings of these apps were only slightly affected in the early days of the pandemic. This reflects the fact that as the situation improves and people’s lives gradually return to normal, some apps are coming back to the forefront as soon as possible.

5.2.6 SNC with decline stage in China. This group contains apps whose rankings have a strong negative correlation with the active cases only in the decline stage, i.e., as the pandemic improves, the ranking declines. For this case, some apps may seem less important in people’s normal work and life. For example, the app with the highest correlation coefficient (-0.752) is FastView, an information app that provides users with different forms of content including graphics, videos, etc. Also, there is an online gaming app called Doudizhu (-0.762), which is a very popular card game in China. It rose to the top of the ranking at the beginning of the outbreak, but quickly declined for a long time as shown in Figure 11(b). These apps seem to become popular for a short period of the outbreak as a means for people to keep abreast of events or to pass the time. However, they cannot last very long, especially when people return to their normal life routines.

5.2.7 SPC with growth stage, SNC with decline stage & SNC with growth stage, SPC with decline stage. Besides the above apps, there are also a few apps that have strong correlations with both stages but in different directions. Their rankings can be roughly treated as constantly declining or increasing over the outbreak. For example, there is a puzzle elimination game called Cube Battle that has fallen steadily out of the top 1,500 since the outbreak began, as shown in Figure 12(a). It has a high correlation coefficient of 0.803 within the growth stage and -0.878 within the decline stage. In contrast, there is a beauty and cosmetic app called SoYong that has a generally fluctuating upward trend in the ranking from February to May, as displayed in Figure 12(b). It has the correlation coefficient of -0.86 within the growth stage and 0.717 within the decline stage. Note that both types are few in number, indicating that the two evolving patterns may be very uncommon.

5.2.8 SPC and SNC in the US. Due to the single trend in the number of the active confirmed cases in the US, i.e., consistently increasing since mid-March, we construct two groups: SPC and SNC with the overall, with 72 and 116 apps, respectively. Since the pandemic situation in the US has been increasingly serious, the apps with an SPC to the pandemic case rate should show a largely declining ranking. For example, the app with the highest correlation coefficient (0.965) is a gaming app called Trivia Crack, which has its ranking change in parallel with the number of active cases, as shown in Figure 13(a). In contrast, the apps’ ranking with an SNC to the pandemic display a generally rising trend. An example is displayed in Figure 13(b), which is an app called Videoleap Video Editor & Maker (-0.88) for producing videos.

6 CHARACTERIZING SIDE EFFECTS

We conclude by briefly inspecting side effects that may be caused by these rapid fluctuations in app popularity. Particularly, an interesting phenomenon we observe is that the ratings of some apps are likely to decrease in varying degrees as their rankings go up. For
We have shown that some apps that gain popularity while experiencing a decline in their ratings, its rating is experiencing a decline. This prompts us to investigate whether the rise in app ranking and popularity may have some side effects on app maintenance behavior.

Specifically, for China, we choose the period when the ranking exhibits strong negative correlation with the number of active confirmed cases during the worsening of the COVID-19 outbreak, i.e., from January to mid-February. For the US, we pay attention to the two months just after the outbreak, i.e., from March to May.

We then calculate the correlation coefficients between the ratings and rankings of the focused apps. The results show that there were 95 apps in China and 88 apps in the US that improved their rankings significantly at the beginning of the COVID-19 outbreak. Of these, 63 apps in China and 24 apps in the US have a correlation coefficient $r>0$ between rating and ranking. This means that their ratings decreased as their ranking rises, i.e., introducing side effects.

### 6.1 Ratings Analysis

We briefly focus on those apps that gained popularity for a short or long time as COVID-19 hit, exploring the relationship between their ratings and rankings. Considering that there is a wide diversity in the evolution of many app rankings throughout the time span, we shorten the time frame to the period when COVID-19 was first spreading, and target only those apps whose rankings rose rapidly during this period.

For some apps, there are some problems with opening video or voice, such as delay, not smooth, etc. Usually such complaints arise in apps featuring multi-person online real-time sessions, such as online meetings and online lecture apps.

### 6.2 Reviews Analysis

We have shown that some apps that gain popularity while experiencing a decline in their ratings. We next dive into the reasons behind this phenomenon and investigate the challenges for app developers. Previous studies [13, 22] showed that reviews represent a rich source of information for app developers, such as user requirements, bug reports, feature requests, and documentation of user experiences. Thus, we look at user reviews to analyze the possible reasons for the drop in app ratings. It is important to note that while the ratings of all of these apps have dropped, the magnitude of the drop varies greatly, with many showing only slight drops. For the sake of representativeness, we target apps where the correlation coefficient between app rankings and ratings is greater than 0.6 and the absolute value of the range of rating decline is greater than 0.1. This leaves 37 apps for review analysis.

We use AppBot [7] to extract the app reviews. This tool provides a large number of data-mining and sentiment analysis features, and is used in many studies, e.g., [18]. For each app, we use AppBot for automated text mining and sentiment analysis of reviews. Beyond this, we use reviews from the second half of 2019 as the benchmark for comparison.

Overall, we find a significant increase in the number of reviews for these apps in the first half of 2020 compared to the second half of 2019. This is intuitive considering their increased demand. A total of 1,355,460 reviews occur in the first half of 2020, which is 4.3 times more than the 313,209 reviews in the previous six months. More importantly, sentiment analysis reveals that the number and percentage of negative reviews among them has also increased considerably, with a total of 716,129 negative reviews in the first half of 2020 (53%) compared to 72,345 in the second half of 2019 (23%), a visible number about 10 times higher. This is also the case for individual apps. For example, Tencent Meeting had only 33 user reviews during the second half of 2019, of which only 5 (15%) were negative, while the number of reviews during the first half of 2020 was astonishingly high at 10,507, of which 4,583 (44%) were negative. This may mean that such apps have struggled to fulfill user expectations as their user base has grown. Further, we focus on the negative reviews and seek to get a sense of the topics of user complaints. We therefore rely on AppBot’s topic grouping feature and identify six core topics raised within these reviews:

(1) **Bugs.** There are lots of negative reviews describing problems with the app that should be corrected, such as crashes, misbehavior, or performance issues. This is likely a product of more users experimenting with the apps’ features.

(2) **Device.** There are many negative comments regarding the device, including device incompatibility, interface mismatch, lack of matching function and device features, unfriendly to iPad devices, triggering device heat and lag, etc.

(3) **Connectivity.** This mainly refers to network connection problems, such as difficult WiFi connections, slow loading, long waiting times, etc.

(4) **Design & UX.** There are some negative reviews about the user interface, specifically that the user interface is not attractive, and the UI design and interaction are not friendly enough, making the user experience poor.

(5) **Video/Audio.** For some apps, there are some problems with opening video or voice, such as delay, not smooth, etc. Usually such complaints arise in apps featuring multi-person online real-time sessions, such as online meetings and online lecture apps.

(6) **Privacy.** There are also some concerns on privacy and security, such as invasion of privacy, sharing information with third-parties, stealing edits and tracking location, etc.

We further measure the distribution of different kinds of topics for each app, as shown in Figure 15. As can be seen, the distribution of user concerns varies greatly across apps, where bugs and device issues are prevalent. We also notice that some of the issues are related to the increase in the number of users. For example, Play It, a multiplayer interactive trivia game app in China, suffered several
crashes due to a surge in the number of users during the initial outbreak of COVID-19 in China as people were quarantined at home. Generally speaking, as more people use the app, the number of complaints grows as well. Some issues that used to be obscure seem to be magnified, such as the aesthetics of the user interface. This sudden popularity presents a great challenge to the apps in many aspects including functionality, compatibility, stability, fault tolerance and UI design, etc. In a way, it serves as a wake-up call for developers to learn from the experience in dealing with unexpected situations in the future.

7 DISCUSSION
7.1 Implications

The most significant result in our paper is that COVID-19 has had a significant impact on the app rankings in app store. This impact is reflected in both the popularity of several categories and the rankings of individual apps. Moreover, the evolution of app rankings are diverse and have a lot to do with whether the features and functionalities of the apps are compatible with the current environment. Besides, we notice that some apps that rapidly rise in popularity may also experience the side effect of rating drops. This reflects the challenges that rapid changes in popularity can cause app developers. We also note that findings are different between China and the United States.

The relevant stakeholders should take note of these observations. Our findings have further implications for understanding the behavior of the mobile app ecosystem during public health crises, and help developers to make better decisions on app development and management. As our first finding indicates, the popularity of apps may experience disturbances resulting from public events. Such disturbances demonstrate the importance of having a contingency plan. When some sudden public events such as the COVID-19 pandemic happens, a contingency plan could help app developers, operators, as well as managers better cope with them. For example, one may develop a quick server deployment plan when expecting high demand. App developers may diversify their portfolios to avoid potential loss caused by popularity decreases. The second finding suggests that the impact may be complicated, depending on how the event is going on. Thus, app developers may need to develop the ability to develop the short-term and long-term insights, and plan accordingly. For example, for an app that gains popularity at the early phase of the pandemic but gradually return to normal later, app developers need to decide if it is necessary to invest resources for the short-term spikes. We also notice that increases of app popularity often coincide with decreases in ratings, and growing number of reviews containing users’ insights. App developers thus need to be vigilant to users’ feedback, and take this opportunities to improve the app quality. Moreover, our results and findings are not restrict to a specific event; it could be helpful in other public emergencies at the global scale.

7.2 Limitations

We recognize that our study carries several limitations and threats to validity. First, since the main impact of the pandemic was reflected on only a small number of apps, we selected the top 100 apps each day, yet there was no accepted standard for selecting this number. Second, our entire correlation analysis relied only on the Pearson correlation coefficient and did not use other correlation coefficients. This is mainly because we feel that the results of the Pearson correlation coefficient have yielded good results and basically satisfy our experimental requirements and purposes. Although the correlation coefficients calculated from a few of the app ranking data with outliers may be affected. However, the number of such data is small and acceptable. Third, in determining the strength of the correlation, we set a threshold value of 0.6. However, there is no standard definition for the selection of this threshold. Forth, only a small number of apps have been used for our review analysis. The major reason is that many apps have insignificant side effects. Hence, we are concerned that reviews of such apps are not representative and reliable enough.

8 CONCLUSION

This paper presents the first longitudinal study of the evolution of app rankings during COVID-19. Our analysis covers 586 apps in China and 590 apps in the US with records of being ranked in the top 100 from January 1 to June 30, 2020, across 22 categories. We performed analysis from the perspectives of category popularity, the evolution patterns of app rankings, and the side effect of declining ratings accompanied by the rising popularity of some apps. Our observations reveal mobile users’ reactions in the mobile ecosystem in the face of unexpected public crises, and provide insights for app developers to make better decisions on developing apps.

ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China (grants No.62072046, 61873069 and 62102042). Haoyu Wang (haoyuwang@hust.edu.cn) is the corresponding author.
REFERENCES

[1] 2020. AppTentive. https://www.apptentive.com/blog/2020/02/04/mobile-app-ratings-and-reviews.

[2] 2020. Cohen’s d. https://en.wikipedia.org/wiki/Cohen%27s_d.

[3] 2020. Google Statistics. https://news.google.com/covid19/map?hl=en-US&mid%3D%2Fg%2F11j99onlq&gl=US&ceid=US%3Aen.

[4] 2020. Pearson correlation coefficient. https://en.wikipedia.org/wiki/Pearson_correlation_coefficient.

[5] 2020. Ranking Factors. https://moz.com/blog/app-store-rankings-formula-deconstructed-in-5-mad-science-experiments?_ga=2.18395884.2057695595.1604471809-802527328.1604471809.

[6] 2020. Spearman’s rank correlation coefficient. https://en.wikipedia.org/wiki/Spearman%27s_rank_correlation_coefficient.

[7] 2020. App review & ratings tools to level up your whole customer experience. https://appbot.co/.

[8] 2020. Spearman correlation coefficient. https://en.wikipedia.org/wiki/Pearson_correlation_coefficient.

[9] Nadeem Ahmed, Regio A Michelin, Wanli Xue, Sushmita Ruj, Robert Malaney, Sahil S Kanhere, Aruna Seneviratne, Wen Hu, Helge Janicke, and Sanjay K Jha. 2020. A survey of covid-19 contact tracing apps. IEEE Access 8 (2020), 134577–134601.

[10] N. Ahmed, R. A. Michelin, W. Xue, S. Ruj, R. Malaney, S. S. Kanhere, A. Seneviratne, W. Hu, H. Janicke, and S. K. Jha. 2020. A Survey of COVID-19 Contact Tracing Apps. IEEE Access 8 (2020), 134577–134601. https://doi.org/10.1109/ACCESS.2020.3010226

[11] Nourah S AlTakarli. 2020. China’s response to the COVID-19 outbreak: a model for epidemic preparedness and management. Dubai Medical Journal 3, 2 (2020), 44–49.

[12] Arndt Büssing, Jutta Hübner, Stefanie Walter, Wolfgang Gießler, and Jens Bünztrkl. 2020. Tumor Patients’ Perceived Changes of Specific Attitudes, Perceptions, and Behaviors Due to the COVID-19 Pandemic and Its Relation to Reduced Wellbeing. Frontiers in psychiatry 11 (2020).

[13] Laura V Galva Correio and Kristina Winbladh. 2013. Analysis of user comments: an approach for software requirements evolution. In 2013 35th international conference on software engineering (ICSE). IEEE, 582–592.

[14] HyungHoon Cho, Daphne Ippolito, and Yun William Yu. 2020. Contact tracing mobile apps for COVID-19: Privacy considerations and related trade-offs. arXiv preprint arXiv:2003.11511 (2020).

[15] Yogesh K Dwivedi, D Laurie Hughes, Crispin Coombs, Joanna Constantiou, Yanqing Duan, John S Edwards, Babita Gupta, Banita Lal, Santosh Mastra, Prakhra Prashant, et al. 2020. Impact of COVID-19 pandemic on information management research and practice: Transforming education, work and life. International Journal of Information Management 55 (2020), 102211.

[16] James D Evans. 1996. Straightforward statistics for the behavioral sciences. Thomson Brooks/Cole Publishing Co.

[17] Denae Ford, Margaret-Anne Storey, Thomas Zimmermann, Christian Bird, Sonia Jaffe, Chandra Maddila, Jenna L. Butler, Brian Housk, and Nachia Meganan. 2020. A Tale of Two Cities: Software Developers Working from Home During the COVID-19 Pandemic. arXiv:arXiv:2008.11147 [cs.SE]

[18] Valéid Garoussi, David Cutting, and Michael Felderer. 2020. Mining user reviews of COVID contact-tracing apps: An exploratory analysis of nine European apps. arXiv preprint arXiv:2012.15589 (2020).

[19] Robert A Kleiman and Colin Mertel. 2020. Digital contact tracing for COVID-19. CMAJ 192, 24 (2020), E653–E656.

[20] Amit Kramer and Karen Z Kramer. 2020. The potential impact of the COVID-19 pandemic on occupational status, work from home, and occupational mobility.

[21] Nuria Oliver, Emmanuel Letouzé, Sébastien Delattalde, Marco De Nadai, Bruno Lepri, Renaud Lambiotte, Richard Benjamins, Ciro Cattuto, Vittoria Colizza, Nicolas de Cordes, Samuel P. Fraiberger, Till Koene, Sune Lehmann, Juan Murillo, Alex Pentland, Phuong N Pham, Frédéric Pixetta, Albert Ali Salah, Jari Saramäki, Samuel V. Scarpino, Michele Tizzoni, Stefan Verhulst, and Patrick Vinck. 2020. Mobile phone data and COVID-19: Missing an opportunity? arXiv:2003.12347 [cs.CY]

[22] Dennis Pagano and Walid Maalej. 2013. User feedback in the appstore: An empirical study. In 2013 21st IEEE international requirements engineering conference (RE). IEEE, 125–134.

[23] Frantz Rowe, Ojelanki Ngwenyama, and Jean-Loup Richet. 2020. Contact-tracing apps and alienation in the age of COVID-19. European Journal of Information Systems 29, 5 (2020), 545–562.

[24] Tansuree Sharma, Tian Wang, and Masooda Bashir. 2020. Advocating for Users’ Privacy Protections: A Case study of COVID-19 apps. In 22nd International Conference on Human-Computer Interaction with Mobile Devices and Services. 1–4.

[25] Rusxi Sun, Wei Wang, Minhui Xue, Gareth Tyson, Seyes Catmerpe, and Damith C Ranasinghe. 2021. An empirical assessment of global COVID-19 contact tracing applications. In 2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE). IEEE, 1085–1097.

[26] Liu Wang, Ren He, Haoyu Wang, Pengcheng Xia, Yuanchun Li, Lei Wu, Yajin Ran. 2020. A survey of covid-19 contact tracing apps. arXiv preprint arXiv:2003.12347 [cs.CY]

[27] Linda Wang and Alexander Wong. 2020. COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-Ray Images. arXiv preprint arXiv:2003.09871 (2020).

[29] Jianpeng Zhang, Yutong Xie, Yi Li, Chunhua Shen, and Yong Xia. 2020. Covid-19 screening on chest x-ray images using deep learning based anomaly detection. arXiv preprint arXiv:2003.12538 (2020).

[30] Danraj Ranasinghe, Christian Bird, N. Ahmed, R. A. Michelin, Wanli Xue, Sushmita Ruj, Robert Malaney, Sahil S Kanhere, Aruna Seneviratne, Wen Hu, Helge Janicke, and Sanjay K Jha. 2020. A survey of covid-19 contact tracing apps: An exploratory analysis of nine European apps. arXiv preprint arXiv:2012.15589 (2020).

[31] Robert A Kleiman and Colin Mertel. 2020. Digital contact tracing for COVID-19.