O EPJ Data Science
a SpringerOpen Journal

REGULAR ARTICLE   **Open Access**

Attention dynamics on the Chinese social media Sina Weibo during the COVID-19 pandemic

Hao Cui[1] and János Kertész[1][*]   (iD)

[*]Correspondence: kerteszj@ceu.edu
[1]Department of Network and Data Science, Central European University, Quellenstrasse 51, A-1100 Vienna, Austria

Abstract

Understanding attention dynamics on social media during pandemics could help governments minimize the effects. We focus on how COVID-19 has influenced the attention dynamics on the biggest Chinese microblogging website Sina Weibo during the first four months of the pandemic. We study the real-time Hot Search List (HSL), which provides the ranking of the most popular 50 hashtags based on the amount of Sina Weibo searches. We show how the specific events, measures and developments during the epidemic affected the emergence of different kinds of hashtags and the ranking on the HSL. A significant increase of COVID-19 related hashtags started to occur on HSL around January 20, 2020, when the transmission of the disease between humans was announced. Then very rapidly a situation was reached where COVID-related hashtags occupied 30–70% of the HSL, however, with changing content. We give an analysis of how the hashtag topics changed during the investigated time span and conclude that there are three periods separated by February 12 and March 12. In period 1, we see strong topical correlations and clustering of hashtags; in period 2, the correlations are weakened, without clustering pattern; in period 3, we see a potential of clustering while not as strong as in period 1. We further explore the dynamics of HSL by measuring the ranking dynamics and the lifetimes of hashtags on the list. This way we can obtain information about the decay of attention, which is important for decisions about the temporal placement of governmental measures to achieve permanent awareness. Furthermore, our observations indicate abnormally higher rank diversity in the top 15 ranks on HSL due to the COVID-19 related hashtags, revealing the possibility of algorithmic intervention from the platform provider.

Keywords: Public attention dynamics; COVID-19; Social media; Ranking

1 Introduction

In our times of information deluge the dynamics of public attention is of eminent importance from many aspects, including education, politics, marketing and governance. In the new media the flow of information has dramatically accelerated, leading often to rapidly changing public attention. At the same time these media provide unprecedented possibilities to study attention dynamics [1, 2] as they produce Big Data open for inves-

🖄 Springer

tigation. The microblogging service Twitter [3] is particularly suited to provide the basis for quantitative studies on the dynamics of public attention as the content of the messages is available [4]. Accordingly, Twitter data have been used to identify classes of dynamic collective attention [5], investigate party-related activity and to predict election outcomes [6]. Furthermore, Twitter data have served as the basis of modelling attention dynamics during pre-election time [7] or studying the relationship between public attention and social emotions [8].

Public attention becomes a focal issue in times of crises like pandemics. As early as 2010, four years after it was launched, Twitter was shown to be an adequate, real-time content, sentiment, and public attention trend-tracking tool [9] and was used to study rapidly-evolving public sentiment with respect to the epidemic H1N1 [10]. The analysis of tweets enabled to quantify the difference between attention and fear and their distance-dependence in the case of the Ebola epidemic [11]. For the present pandemic COVID-19, Twitter studies on public attention have occurred, focusing on the perception of policies by the public [12, 13], fighting COVID-19 misinformation [14], and the psychological impacts associated with social media exposure during the pandemic [15, 16].

The first COVID-19 epicenter was in China where the service of Twitter is blocked, whereas its local substitute Sina Weibo, is very popular [17]. Therefore, it is natural to use data from Weibo for similar purposes as was introduced earlier for Twitter in other countries. Scientists have already recognized that this microblogging service provides important insight into the function of the Chinese society [18, 19]. Recently, some studies have appeared dealing with the public attention towards COVID-19 on Sina Weibo. The attention of Chinese netizens to COVID-19 was measured by analyzing keyword frequency in retrieved COVID-related posts from randomly sampled Weibo users [20] or hashtags on Sina Weibo Hot Search List (HSL) [21], and was studied for identifying the sentiment and emotion trends of public attention [21, 22].

Social media search indices have shown correlations with the epidemic curve [23–27] and have been used for prediction of the transmission of infectious diseases, such as cholera [28], Ebola [29], influenza [24] and Middle East Respiratory Syndrome (MERS) [23]. For the recent COVID-19, online search trends such as Google Trends and Baidu Index have shown strong correlations with real-world cases and deaths [25–27] and were used to predict the number of new suspected or confirmed cases [26, 27]. In these studies, the searched keywords on social media are mainly the symptoms and the names of the disease. Nevertheless, there are various aspects of the pandemic that influence the society, including the measures and regulations given by the government, the scientific knowledge provided by healthcare experts, aspects related to frontline doctors and nurses, the donation behaviors between countries, the remote working new norm and so on. In this study, we also pose the so far less studied question about the temporal correlation patterns between different COVID-related topic categories with the changing world situations of COVID-19.

The Sina Weibo Hot Search List (HSL) depicts the popularity ranking of the top 50 hashtags in real-time, according to an algorithm in which the number of searches on the hashtags is dominant. The ranking dynamics on Weibo enables us to investigate the dynamics of popular attention in such details which is not available on other platforms. We will focus on the temporal evolution of this ranking in order to study the attention dynamics.

Recently, ranking dynamics has been studied widely from sports [30–32] to scientists, journals or companies [30]. There are stable rankings (word frequencies) with little or no changes in the ranks [33] and there are volatile ones with vivid dynamics (mentions of Twitter hashtags) [30]. Clearly, Weibo HSL belongs to the latter with rich dynamic properties. This ranking provides a proxy for the attention preferences of Weibo users enabling the quantification of the dynamics thereof, which reflects the changes in the attention due to events and measures. Instead of capturing the public attention using segmented key words from retrieved posts, we adopted the "rank diversity" [32] measure on Weibo HSL to describe the rank dynamics of real-time popular hashtags which can serve as an indication of public attention in the Sina Weibo system. To the best of our knowledge, we are the first in relating the ranking dynamics of social media hot topics to the public attention dynamics. To see the effect of the pandemic on the public attention we compare the situation before the break-out of COVID-19 with the period during the pandemic up to the time when consolidation sets in.

Collective attention towards online news decays with time due to the fade of novelty and attractiveness in the competition with the other news [1]. Recent studies show a peak of collective attention towards COVID-19 in late January 2020 and a subsequent collapse, in terms of the dynamic behavior of words used on Twitter [34, 35]. To further understand the less studied attention decay mechanism and the decay pattern with time, we describe the public attention decay towards COVID-19 based on the dynamics of the ranks and the durations of COVID-related hashtags on the Weibo HSL.

In this study, we focus on the HSL of Sina Weibo during the period of COVID-19, which provides rich data about public attention and its dynamics in China. Based on that data, we have been able to track the evolution of public attention in different periods during the pandemic, follow how the attention of the population shifted from one group of topics to another and study the changing correlation patterns of different COVID-related topic categories with the evolving COVID-19 situation in the world. During our studies we have discovered signatures of the possible algorithmic control on this social media platform. Understanding the dynamics of public attention on social media promotes instant and effective communications among governments, health experts and the public, helping the government to monitor public opinion, maintain the stability of the society as well as develop and deliver more effective measures to minimize the effects of the pandemic.

The paper is organized as follows: in Sect. 2, we provide background information on Sina Weibo real-time Hot Search List (HSL) and methodologies on quantifying attention dynamics. In Sect. 3, we present our results on attention dynamics, correlations between different types of hashtags, and attention decay. In Sect. 4, we discuss and summarize the results.

2 Materials and methods

2.1 Sina Weibo real-time hot search list (HSL)

Sina Weibo is the biggest Chinese microblogging website, with MAU (monthly active users) reaching 550 millon and DAU (daily active users) 241 million in March 2020 [36]. Instead of using one hashtag at the beginning like Twitter, topics on Weibo are confined in double hashtags one at the beginning and one at the end of the topic description, for example, #Pneumonia of unknown cause detected in Wuhan#. In this paper, we use hashtag as the synonym of topic, we refer a hashtag as the content contained within the double
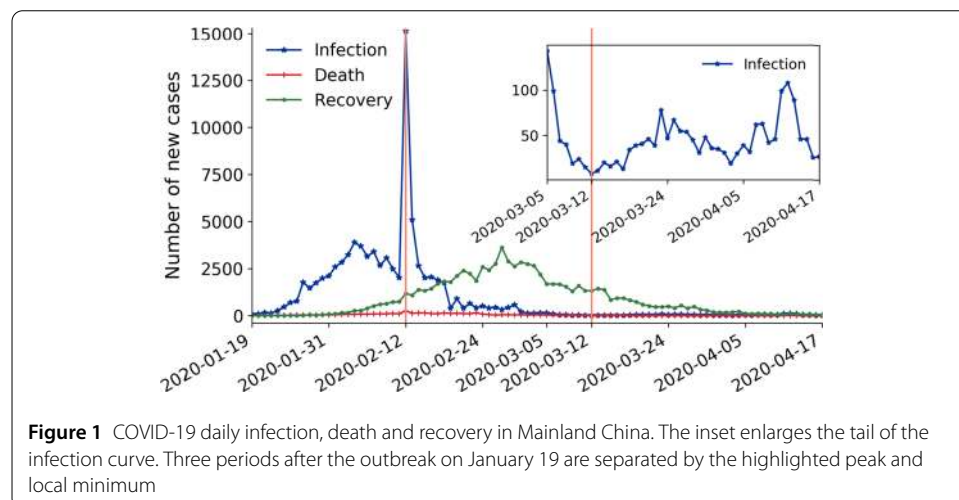
#s. A hashtag becomes popular at a given time as it is used in many tweets, gains a large number of searches, likes and discussions by the users. The Hot Search List on Weibo is a section that displays the 50 most popular hashtags in real-time. The hashtags on the HSL, together with their ranks and search volume indices are updated every minute [37] and new popular hashtags may emerge and others vanish. The search volume index is a comprehensive measure which takes into account multiple dimensions such as the number of searches in Sina Weibo and the quality of the user accounts involved in the search, for the aim of preventing manipulated fake popularity [38]. The third and sixth ranks on the HSL are sometimes occupied by promoted advertisements labeled with the character "荐" [39] (meaning recommendation).

2.2 Data

We took data from Weibo HSL to study attention dynamics as it captures vibrant real-time change of public attention. Due to the random existence of one or two commercial advertisements at the third and the sixth ranks, in order to get a constant length of non-advertisement hashtags on the HSL at each timestamp, we removed all the hashtags labeled with "荐", re-ranked the original HSL and took the top 48 hashtags for each timestamp. All the HSL we mentioned later in this paper mean the re-ranked HSL with 48 ranks. We directly downloaded the data from the HSL with a frequency of every 5 minutes from December 16, 2019 to April 17, 2020. There are in total 26,022 hashtags and 9120 of them are related to the aspects of COVID-19. To relate social media contents with real-life pandemic situation in Mainland China, we collected the daily number of infections, deaths, and recoveries from the official website of National Health Commission of China [40]. In the following subsection we explain how we identified the different categories of hashtags.

2.3 Topical classification and correlations

Figure 1 shows the number of daily infections, deaths and recoveries in Mainland China. The number of daily infections and deaths have a sharp peak on February 12 due to the adoption of new diagnostic criteria [41]. The decreasing trend of daily infections since the peak turned to increasing after March 13, as a result of the rising number of imported coronavirus cases from abroad [42]. We will argue that there are three periods to be dis-

Figure 1 COVID-19 daily infection, death and recovery in Mainland China. The inset enlarges the tail of the infection curve. Three periods after the outbreak on January 19 are separated by the highlighted peak and local minimum

tinguished after the outburst of COVID-19 around January 19, separated by the maximum and local minimum of the daily number of infections on February 12 and March 12, respectively.

To study the public attention towards COVID-related information, we first extracted hashtags which encompass all aspects of COVID-19 and classified them based on geographic regions and the exposure order under the pandemic into three categories: Mainland China, East Asia outside of Mainland China and Other Countries outside of East Asia. With a focus on COVID-hashtags related to Mainland China, we manually classified them based on semantic meanings into the following seven disjoint sub-categories. The Bad News category comprises hashtags on confirmed infections and deaths in different regions of Mainland China as well as shortages of essential supplies. The Good News category consists of news on cases of recovery, sufficiency of supplies, and decrease in daily infections or deaths. The Regulations category consists of authority announcements of national, regional, institutional laws, rules and regulations associated with public behaviors and concerns during the pandemic. The Life Influence category contains hashtags that reflect the pandemic influence on the aspects of citizen lives. The Front Lines category includes hashtags related to the lives of front line workers (mainly doctors and nurses) and their interactions with patients in hospitals. The Science category incorporates scientific understandings of the virus properties, vaccine development, and ways for public protection given by authoritative doctors. The Supports category takes into account hashtags on worldwide donations and emotional supports. For ambiguous cases which contain information of more than one category, our classifications were based on the focus of the main subject. Due to the syntactic-semantic complexity of Chinese language, the classifications were made by two independent annotators. Final consensus was reached in case of disagreement. The Mainland China sub-categories are summarized in Table 1 together with examples. The full list of COVID-related hashtags is available in the dataset, which we have made public (see declaration at the end of the paper).

To further understand how the Mainland China related COVID-hashtags are correlated with each other and with the daily number of infections/deaths/recoveries in the three separated time periods, we measured the Pearson's correlations between the seven series of daily number of new hashtags in each of the sub-categories defined above, together with the three series of daily number of infections/deaths/recoveries. The correlation of these ten time series are calculated using the percentage change between the current and the prior element instead of the actual value in order to reduce the effect of the trend which can cause spurious correlations. For time series category $X = \{X_{t_i} : t_i \in T, i = 1, 2, \ldots, n\}$ and category $Y = \{Y_{t_i} : t_i \in T, i = 1, 2, \ldots, n\}$, where T is the time index set, the Pearson's correlation is calculated using the percentage change series $\tilde{X} = \{\frac{X_{t_{i+1}} - X_{t_i}}{X_{t_i}}, t_i \in T, i = 1, 2, \ldots, n\}$ and $\tilde{Y} = \{\frac{Y_{t_{i+1}} - Y_{t_i}}{Y_{t_i}}, t_i \in T, i = 1, 2, \ldots, n\}$.

2.4 Attention dynamics

One natural measure of social media attention towards a topic category is the quantity of the related hashtags. The growing pattern of the cumulative number of hashtags on the HSL with time reflects the dynamics of the public attention. We separately measured the growth of the cumulative number of all hashtags and all COVID-related hashtags that ever appeared on the HSL in our observation period. To understand how much

Table 1 Mainland China COVID-hashtag details. A summary of the example hashtags in each sub-category of Mainland China category and the number of hashtags in different time periods

Category	Examples	Period 1	Period 2	Period 3
Bad News	#全国累计确诊新冠肺炎66492例# (#National cumulative confirmed COVID-19 cases reach 66492#) #黑龙江聚集性疫情共48起发病194人# (#Heilongjiang in total 48 clustered epidemic 194 infected cases#) #武汉多家医院物资紧张# (#Wuhan many hospitals supplies in shortage#)	451	193	272
Good News	#火神山医院累计治愈患者破千# (#Huoshenshan Hospital has cured over a thousand patients#) #7省区现有确诊病例清零# (#7 provinces current infected cases down to zero#) #疫情形势出现3个积极变化# (#Epidemic situation shows 3 positive changes#)	145	257	121
Regulations	#上海地铁不戴口罩不得进站# (#Shanghai metro station not allowed to enter without wearing a mask#) #疫情影响严重的地区可增发生活补助# (#Additional living allowances can be issued in areas severely affected by the epidemic#) #非疫情严重国家进京者居家观察14天# (#Home observation for 14 days for visitors to enter Beijing from non-severe epidemic countries#)	318	325	633
Life Influence	#武汉市民江滩唱起国歌# (#Wuhan citizens sing national anthem at the River Beach#) #疫情期间点外卖指南# (#Guide to ordering takeout during the epidemic#) #一季度民航业亏损398亿# (#Civil aviation industry suffered a loss of 39.8 billion in the first quarter#)	310	371	649
Front Lines	#钟南山等专家连线武汉ICU团队# (#Zhong Nanshan and other experts connected to the Wuhan ICU team#) #护士握手呼唤79岁新冠患者# (#Nurse shakes hands and calls 79-year-old COVID-19 patient#) #方舱医院收治第一批患者现场# (#Site of Fangcang shelter hospital taking the first batch of patients#)	251	329	347
Science	#各年龄段人群普遍易感新冠病毒# (#People of all ages are generally susceptible to coronavirus#) #口罩的正确使用方法# (#The correct use of masks#) #如何区分感冒流感和新冠肺炎# (#How to distinguish between flu and COVID-19#)	180	170	123
Supports	#汶川村民自发支援武汉100吨蔬菜# (#Wenchuan villagers spontaneously support Wuhan 100 tons of vegetables#) #欧盟对华运送12吨急需物资# (#EU sends 12 tons of urgently needed supplies to China#) #武汉给援汉医疗队全员的感谢信# (#Thank you letter from Wuhan to all members of the medical aid team#)	151	144	116

COVID-information occupies the HSL at each timestamp, we constructed the historical ratio trajectory of the COVID-related hashtags on the HSL since the first COVID-hashtag #武汉发现不明原因肺炎# (#Pneumonia of unknown cause detected in Wuhan#) appeared on December 31, 2019.

2.4.1 Lifetime duration

The lifetime duration of a hashtag on the HSL indicates the ability of obtaining persistent attention from the public. We quantified the duration (continuous existence on the HSL) of a hashtag with τ:

$$\tau = \tau_1 - \tau_0,$$

where τ_0 is the timestamp of the first and τ_1 is the timestamp of the last appearance of a hashtag on the HSL.

We compared the duration of the hashtags across various categories and different time scopes. We compared the duration of the hashtags before the outbreak on January 19, all COVID-related hashtags, and non-COVID hashtags after the outbreak. To ensure complete life cycles of the hashtags, we took all hashtags whose first arrivals on the HSL are between December 19, 2019 and January 18, 2020 as the sample for hashtags before the pandemic, which includes 6161 in total. Similarly, we took all COVID-hashtags whose first arrivals are no later than April 14, with a total number of 8808. For the non-COVID hashtags after the outbreak, we took a random sample of all non-COVID hashtags with the same size as the COVID sample. Hashtags that reappeared after disappearing from the HSL were excluded from our calculation. To understand the overall attention variation towards COVID-hashtags with time, we investigated the daily value of their cumulative average duration. We denote D_j as the cumulative average of duration from December 31, 2019 (day 0) until day j. D_j is calculated as follows:

$$D_j = \frac{1}{|S(j)|} \sum_{i=0}^{j} \sum_{\alpha \in S(j)} d_i^{\alpha}, \tag{1}$$

where d_i^{α} is the duration of hashtag α whose first appearance was on day i. $S(j)$ is the set of all the hashtags whose first appearance is in the interval $[0,j]$.

2.4.2 Ranking

The changes in the ranking patterns of the hashtags at different time periods reflect the general public attention dynamics. Rank diversity $d(k)$ [32] is defined as the number of distinct elements in a complex system that occupy the rank k at some point during a given length of time. Rank diversity is known to give characteristic profiles for different types of systems; e.g., in open systems (where only the top part of the competing items is ranked) behaves differently from closed systems (where all the items are ranked) [32, 33]. In this paper, we use rank diversity to measure the number of different hashtags occupying a given rank on the HSL over a given length of time, and thus obtain overall information on the total dynamical trend of the hashtags on the HSL. We normalize the rank diversity value by the total number of unique hashtags that have appeared on the HSL in a given time interval. We compared the rank diversity in the 48 ranks on the HSL before the outbreak and during the different periods after the outbreak, with and without COVID-19 hashtags.

The public attention towards a hashtag can also be indicated by its highest rank during the lifetime on the HSL. The highest rank of a hashtag reveals its highest ability and achievement when competing for attention with the other hashtags. We studied the highest rank distribution of the classified COVID-hashtags and compared the results with the

hashtags before the outbreak as well as the non-COVID hashtags after the outbreak (Additional file 1). To understand the overall highest rank variation towards COVID-hashtags with time, we investigated the daily value trajectory of their cumulative average highest rank. We denote H_j as the cumulative average of highest rank from December 31 (day 0), 2019 until day j. H_j is calculated as follows:
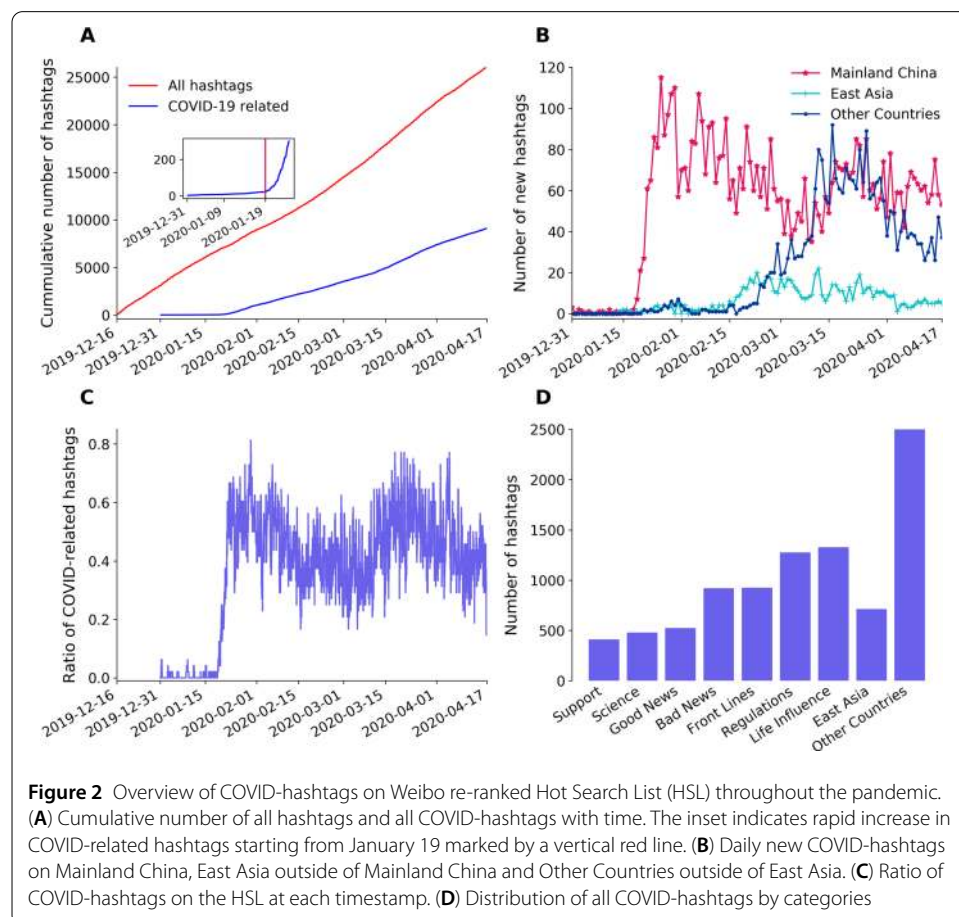
$$H_j = \frac{1}{|S(j)|} \sum_{i=0}^{j} \sum_{\alpha \in S(j)} h_i^{\alpha}, \tag{2}$$

where h_i^{α} is the highest rank of hashtag α whose first appearance was on day i. $S(j)$ is the set of all the hashtags whose first appearance was in the interval $[0, j]$.

3 Results

3.1 Statistics and categorization of hashtags

The cumulative number of new hashtags on HSL grows approximately linearly (see Fig. 2(A)), indicating a nearly constant attention capacity and need for news of the users. Closer inspection tells, however, that the rate of new hashtags decreases between January 10 and February 12 followed by an increased rate until March 28 after which the original slope of 225 ± 4 new hashtags/day sets in. We attribute this change in the slope to the effect of COVID-related hashtags.

Figure 2 Overview of COVID-hashtags on Weibo re-ranked Hot Search List (HSL) throughout the pandemic. (**A**) Cumulative number of all hashtags and all COVID-hashtags with time. The inset indicates rapid increase in COVID-related hashtags starting from January 19 marked by a vertical red line. (**B**) Daily new COVID-hashtags on Mainland China, East Asia outside of Mainland China and Other Countries outside of East Asia. (**C**) Ratio of COVID-hashtags on the HSL at each timestamp. (**D**) Distribution of all COVID-hashtags by categories

The first COVID-related hashtag appeared on the HSL on December 31, 2019, followed by only a few ones in the following week. As the first death case occurred on January 11, second one occurred on HSL on January 16 and more infected cases detected in other cities in China as well as in the surrounding Asian countries, rumours and scared emotions about the unknown pneumonia were permeating in the society and the number of daily COVID-related hashtags started to increase rapidly on January 19. On January 20, Chinese authorities announced to the public that the new coronavirus is transmissible between humans.
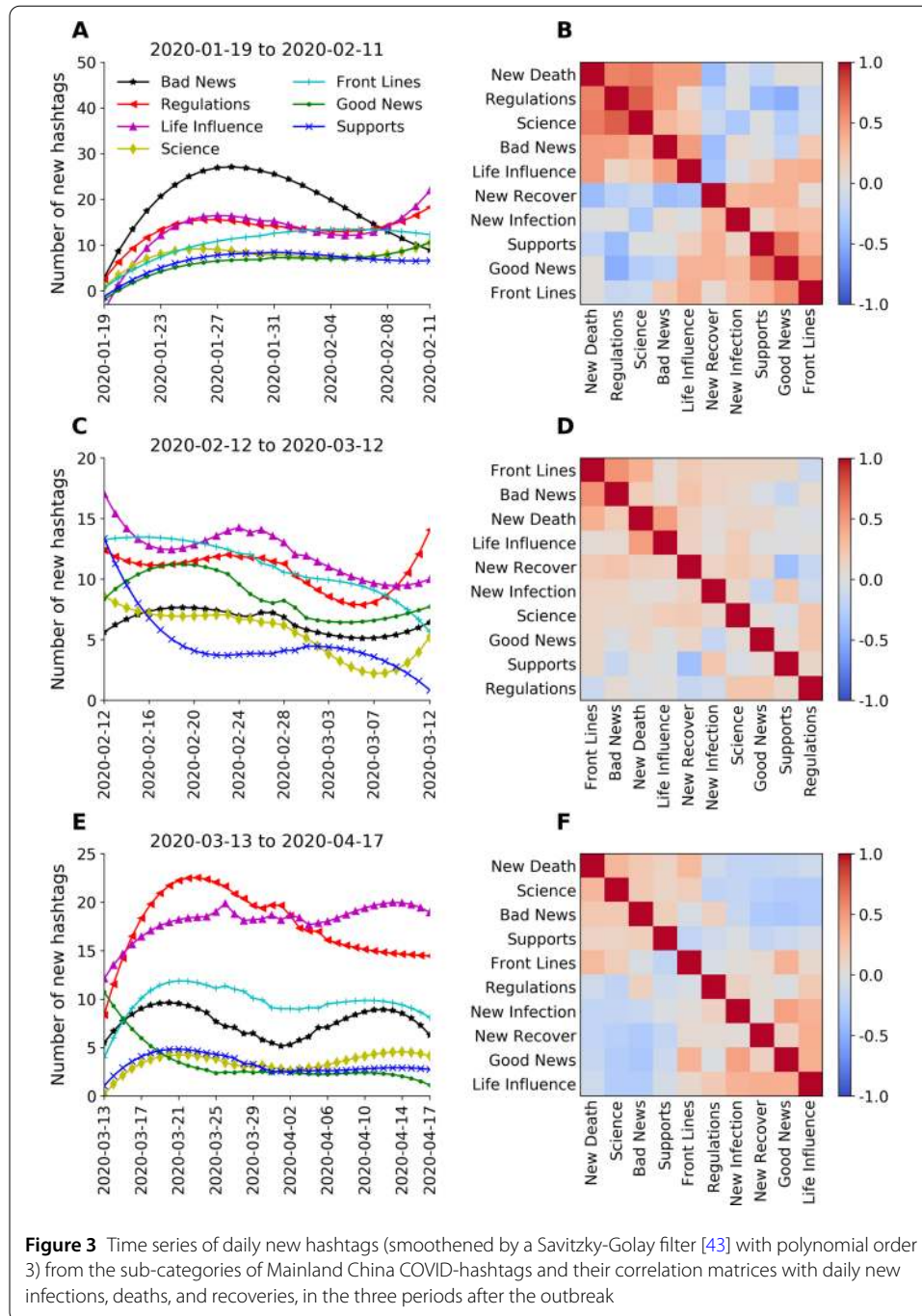
From our point of view the period until January 19 can be considered as pre-COVID. During that time at most three COVID-related hashtags per day have occurred on the HSL and the cumulative number of different hashtags on HSL has grown approximately linearly with an unaltered slope (see Fig. 2(A)). Around January 19 the number of COVID-related hashtags started growing and, at the same time, the overall growth of the total number of hashtags slightly decreased, indicating that the new COVID-related hashtags stay longer on HSL as compared with those before the outbreak. This results in a decrease of the total number of new hashtags per unit time on HSL. After January 19, a rapid increase can be observed in the number of COVID-hashtags (see the inset of Fig. 2(A)). This has, finally, also an effect on the total cumulative number of hashtags resulting in an increased slope in Fig. 2(A).

Figure 2(B) shows the cumulative number of geographically categorized COVID-hashtags with Mainland China, East Asia outside of Mainland China, and Other Countries outside of East Asia as categories. The Mainland China category starts to rise rapidly from January 19, reaches a peak in the following week, and then gradually drops with a few rebounds. The second peak and the decline of Mainland China category is intertwined with the trajectory of the Other Countries category in mid-March. The East Asia category remains at a relatively low level throughout the pandemic.

COVID-19 was first observed in east Asia, with Mainland China being the hardest-stricken region, followed by places with growing infections such as South Korea, Diamond Princess cruise ship and Japan. The epicenter of COVID-19 later shifted to Europe and the rest of the world as the situation mitigated in east Asia. The results depicted in Fig. 2(B) follow these events closely, confirming the role of the real-time HSL on Weibo as a reflection of the real world. Unsurprisingly, the upward and downward trend periods of Mainland China and Other Countries coincide with Fig. 2(C), where the ratio of COVID-related hashtags on the HSL at each timestamp is displayed. The swift third peak on April 4 in Fig. 2(C) is due to the national Qingming Festival (also known as the Tomb-Sweeping Day), where the victims who died in the COVID-19 pandemic were mourned. The dynamics of the COVID-related hashtags on the HSL demonstrates vibrant generations of newly created COVID-19 hashtags about the relevant up-to-date events around the world. Figure 2(D) shows the distribution of the hashtags in the sub-categories of Mainland China category along with East Asia and Other Countries. Among the seven sub-categories that belong to Mainland China, Support, Science, and Good News have relatively fewer hashtags, compared with Front lines, Life Influence, Bad News, and Regulations.

3.2 Periodization and correlations

Figure 3 illustrates the attention dynamics of the sub-categories of Mainland China by showing the quantity variations in Fig. 3(A), (C), (E), paired with their correlation matrices

Figure 3 Time series of daily new hashtags (smoothened by a Savitzky-Golay filter [43] with polynomial order 3) from the sub-categories of Mainland China COVID-hashtags and their correlation matrices with daily new infections, deaths, and recoveries, in the three periods after the outbreak

with daily infections, deaths, and recoveries in Fig. 3(B), (D), (F). As noted above, we have identified three periods in the investigated time interval: The first period is January 19–February 11, separated by the huge peak in Fig. 1 from the second one (February 12–March 12). The third period (March 13–April 17) is separated from period 2 by the second vertical line where the number of new infections has a local minimum (Fig. 1 inset).

In Table 1, we show the number of hashtags related to Mainland China in the different categories for the three periods. In Fig. 3, we show that the daily emergence of the categorized COVID-hashtags is dominated in the first two periods by Bad News, with increasing
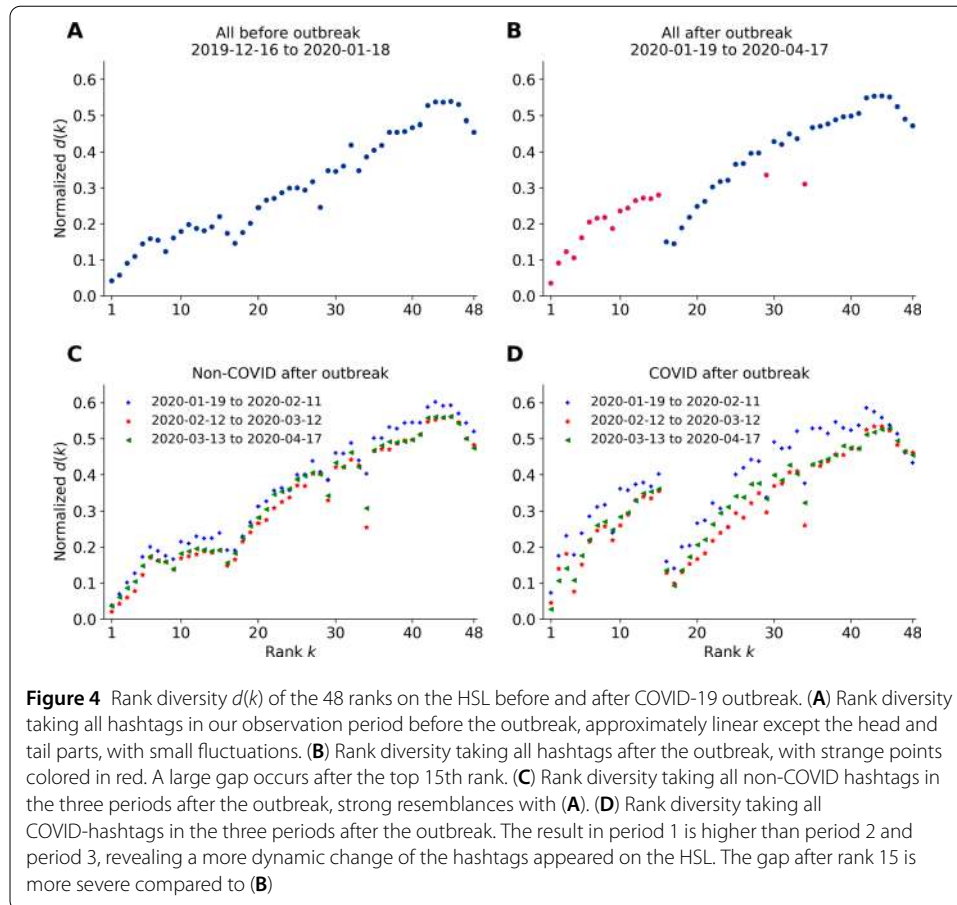
and decreasing trends in period 1 and period 2, respectively. In period 3, the categories Regulations, Life Influence, and Front Lines receive more attention as compared to the rest of the categories. Here the consistently high values in Regulations and Life Influence could result from the worsening world pandemic situation along with the rise of the imported infected cases in Mainland China, necessitating the establishment of measures to handle it. The categories of the Mainland China COVID-hashtags move with the number of infections and deaths in the world.

The patterns of the Pearson's correlation matrix of the ten time series reflect temporal structure with the three periods. Figure 3(B) shows a positive correlation block structure. There are strong correlations between New Death, Regulations, Science, and Bad News (upper left block) as well as between Supports, Good news and Front Lines (lower right block) and there is considerable anti-correlation between the two blocks. Figure 3(D) (period 2) exhibits much weaker correlations, in fact, very few elements of the matrix reach values beyond the noise level (see Additional file 1). Exceptions are new strong correlations between New Death and Front Lines, as well as Bad News and Front Lines. In the third period (Fig. 3(F)) the block structure gets again more pronounced, though not as pronounced as in the first period. Note that the categories had to be rearranged in order to achieve this structure. The major change is that Supports/Front Lines and Life influence/Regulations have exchanged positions. In period 1, the Bad News (mainly infections and deaths) of domestic cases in Mainland China were flooding, this lead to the urgent establishment of regulations, which caused life influences. In period 3, the domestic situation was under control, therefore, the Bad News in Mainland China were mainly caused by the worsening international situation (infections/deaths and Chinese coming back from abroad). Then the Regulations and corresponding Life Influences towards these issues were not anymore strongly associated with domestic deaths.

3.3 Rank diversity and hashtag dynamics

What is the effect of COVID-19 on the ranking dynamics? Fig. 4 shows a comparison of the rank diversity in the top 48 ranks taking non-COVID and COVID hashtags in different periods. Striking differences are observed between the rank diversity plots before and after the outbreak. As Fig. 4(A) suggests, the rank diversity plot before the outbreak was approximately linear with moderate fluctuations. A clear gap emerges in the rank diversity after rank 15 in Fig. 4(B) during the COVID period. We recognize resemblances in the rank diversity plots before the outbreak and after the outbreak considering only non-COVID hashtags, except for the strange drops at ranks 29 and 34 in Fig. 4(C). Comparing Fig. 4(D) with Fig. 4(B), the gap after rank 15 is larger in the rank diversity plot considering only COVID-related hashtags. The rank diversity plots for hashtags in period 1 surpass period 2 and period 3 with both non-COVID and COVID hashtags as depicted in Fig. 4(C) and (D), while the difference is much higher in the latter case.
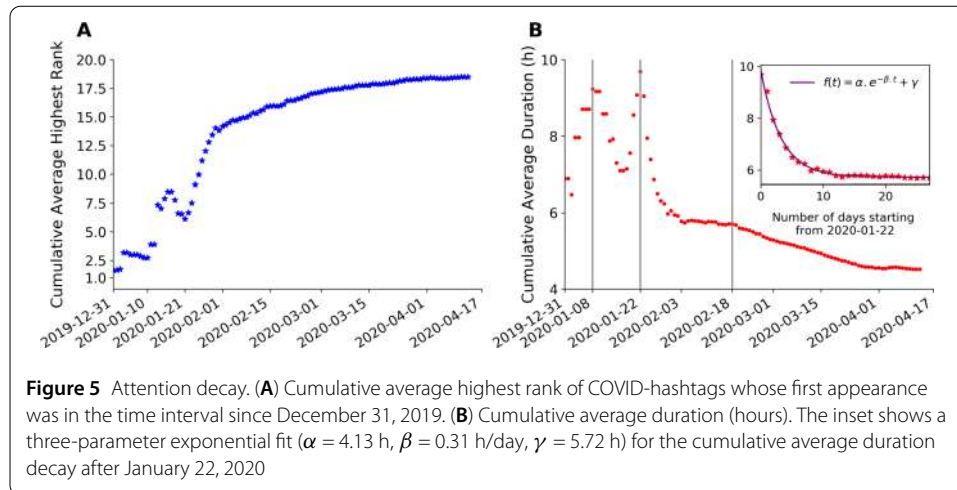
Figure 4 gives evidence that the COVID-hashtags cause the gap in the rank diversity plot after the outbreak. Taking the normalized rank diversity plot before the outbreak as a reference, a higher normalized rank diversity at a certain rank position represents a higher number of unique occurrences within the observation period, so that the COVID-related hashtags in the top 15 ranks change faster (with higher frequency) than normal. One possible explanation is that the COVID-hashtags kept emerging with higher frequency than before the outbreak and people payed much attention to these new hashtags. Additionally, when the flooding hashtags contained similar information such as the new infections

Figure 4 Rank diversity *d(k)* of the 48 ranks on the HSL before and after COVID-19 outbreak. (**A**) Rank diversity taking all hashtags in our observation period before the outbreak, approximately linear except the head and tail parts, with small fluctuations. (**B**) Rank diversity taking all hashtags after the outbreak, with strange points colored in red. A large gap occurs after the top 15th rank. (**C**) Rank diversity taking all non-COVID hashtags in the three periods after the outbreak, strong resemblances with (**A**). (**D**) Rank diversity taking all COVID-hashtags in the three periods after the outbreak. The result in period 1 is higher than period 2 and period 3, revealing a more dynamic change of the hashtags appeared on the HSL. The gap after rank 15 is more severe compared to (**B**)

and deaths in different cities or provinces of China, the public interest towards individual hashtags could drop quickly, resulting in a higher number of unique hashtags at certain ranks in unit time on HSL. This effect of higher rank diversity for higher ranks seems to be amplified by the algorithm leading to the observed gap.

Strange drops of rank diversity at ranks 29 and 34 can also be seen on our plots in Fig. 4. As provided in Additional file 1, there are hashtags that stay at the ranks 29 and 34 for an unusually long time and then disappear from the HSL, indicating algorithmic intervention from Weibo. As one of the most popular and influential social media in China, Weibo might shoulder the responsibility during the global public health emergency to keep people informed about related news in China and around the globe, by means of changing the algorithm towards COVID-hashtags to promote crucial news and keep them updating in the top 15 positions and leave the list at rank 29 or 34. Our methods are sensitive enough to demonstrate this type of interventions. Therefore our observations reflect a combination of both spontaneous attention dynamics from the public and the controlled effects from Sina Weibo.

Rank diversity captures attention dynamics from the point of view of the overall dynamical rank movements of the hashtags on the HSL. It is interesting to follow the dynamics also from the aspect of the individual hashtags. The average highest rank of a category of hashtags on a given day is characteristic to the attention paid to that category. (Note, of course, that getting to the HSL expresses already considerable attention.) Similarly, the average duration is another measure of attention. However, in the latter case it should be

Figure 5 Attention decay. (**A**) Cumulative average highest rank of COVID-hashtags whose first appearance was in the time interval since December 31, 2019. (**B**) Cumulative average duration (hours). The inset shows a three-parameter exponential fit (α = 4.13 h, β = 0.31 h/day, γ = 5.72 h) for the cumulative average duration decay after January 22, 2020

mentioned that short duration can be caused by decaying attention to the general topic (in this case the hashtag is likely to be replaced by another from a different topic) or because of the heavy stream of new hashtags of the same topic.

How do the average highest rank and average duration accumulate with time? As Fig. 5(A) shows, the cumulative average highest rank, H_j is initially at a top rank, indicating that the first few hashtags about the unknown pneumonia received a huge amount of attention from the public. As more COVID-related hashtags occurred, H_j becomes lower, with a rapid change at the beginning and a slower change later, separated by around January 30. This is due to the rapidly increasing number of COVID-related hashtags and the limited number of ranks on HSL. In Fig. 5(B), the first peak of the cumulative average duration, D_j is on January 8, when the hashtag was posted that eight patients infected by the unknown pneumonia recovered from hospital. Then the D_j decreases first and then increases again, reaching the second peak on January 22, after which the increasing daily new hashtags with short durations start to play a greater role than the few hashtags with long durations.

The fast decay of D_j in the period between January 22 and February 18 (see the inset in Fig. 5(B)) was fitted by an exponential function:

$$f(t) = \alpha e^{-\beta \cdot t} + \gamma, \tag{3}$$

with α = 4.13 h, β = 0.31 h/day, γ = 5.72 h. On February 18, hashtags of positive changes in the COVID situation started to appear on HSL. After that, the D_j exhibits a slower and longer decay.

4 Discussion and summary

In this work, we have studied the public attention dynamics on the real-time Hot Search List (HSL) of the biggest Chinese microblogging website Sina Weibo under the influence of the COVID-19 pandemic. On the one hand, such study contributes to the understanding of the dynamics of public attention on social media and how it reflects the dynamics of the public thoughts and behaviors. On the other hand, identifying the online attention dynamics patterns and their relationship to events and measures during pandemic may contribute to its efficient management.

In order to follow the dynamics of public attention we have introduced sub-categories of COVID-19-related hashtags. Our results show diversification of the public attention after the outbreak on January 19, 2020 as indicated by changing frequencies of such hashtags in the different sub-categories. Moreover, the pattern of correlations with the real-world events and measures vary in three identified periods during the investigated time span. We conclude that at the beginning the dominant driving force of the public attention was the infection and death situation in Mainland China, with mainly domestic cases, while the international situation and the imported cases influenced the attention later. Our observations point towards the complexity of the attention patterns indicating that several components should be taken into account if such data are used to the prediction of the epidemic curve [26, 27].

Furthermore, we have shown that the cumulative average duration follows exponential decay immediately after the attention peak in the pandemic, but a slower decay for a longer time. The exponential decay suggests that the speed of governmental response is crucial in the early pandemic phase. This exponential decay sets a scale for the governments within which it should take quick actions and publish crucial measures and regulations to control the pandemic, healthcare experts should deliver scientific knowledge to inform the public how to protect themselves efficiently.

The attention towards COVID-hashtags decayed as the circumstances in China got better. Nevertheless, the attention was influenced by the world pandemic situation which kept changing, hence the decay of public attention on the Chinese social media Weibo has become less clear cut. In any case, targeted and timely stimulus should be given to keep the attention and awareness of the public throughout the pandemic to prevent future waves of COVID-19.

In this paper we have made the first step to relate the ranking dynamics of hot topics on social media with the public attention dynamics. We have provided a novel approach to study and quantify the attention dynamics taking advantage of the real-time Hot Search List (HSL) on Weibo. The rank diversity in the top 15 ranks containing COVID-hashtags are higher than normal. This could result from the spontaneous preference from the public towards COVID-related information. More likely, it is due to an algorithmic intervention towards COVID-hashtags from the platform provider. Sina Weibo may intentionally promote the COVID-related important information to make sure people will get aware of them. In this sense, such an algorithmic intervention can be useful for the public. The empirical rank diversity could be a combined influence of both the Weibo algorithm and the spontaneous public preference. This observation shows the possibility that rank diversity could be an adequate tool to investigate further the important aspect of algorithmic intervention in social media data.

Besides exploring the attention dynamics on the Chinese social media Sina Weibo, we also studied the cumulative growth of all topics and all the COVID-topics on Twitter trending list in the United States. As is shown in the supplementary material Fig. SI1(A), the cumulative number of all the Twitter trending topics in the United States was almost perfectly linear from January 1, 2020 to April 16, 2020. The time period that the cumulative number of all COVID-topics on Twitter trending list increases is in accordance with the rising period of the number of hashtags in the Other Countries category in Fig. 2(B). The similarity of results on Sina Weibo HSL and Twitter trending list is a reflection that both platforms are influenced similarly by the major events worldwide during the COVID-19

pandemic. Though having more daily new topics on Twitter trending list than Weibo HSL, the number of COVID-topics on Twitter is much fewer. The topics on Twitter are generally shorter and have broader meaning, for example, #QuarantineLife, while the hashtags on Weibo are more detailed, for example, #小区窗台演唱会庆祝解除隔离# (#Community window concert to celebrate the lifting of quarantine#), contributing to the rich number of diverse hashtags on Weibo. It should be emphasized that both for Sina Weibo and Twitter the lists are produced by unknown algorithms and in the case of Sina Weibo we have been able to pinpoint direct interventions from the side of the provider into the ranking. However, the detailedness of Weibo HSL, its fixed length and the fact that HSL is the same for all users seem to make Weibo HSL more suitable to study attention dynamics through ranking than Twitter, as Twitter trending lists are without fixed length and can be personalized.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1140/epjds/s13688-021-00263-0.

> **Additional file 1.** Supplementary information (PDF 2.3 MB)

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Wu F, Huberman BA (2007) Novelty and collective attention. Proc Natl Acad Sci USA 104:17599–17601
2. Russell Neuman W, Guggenheim L, Mo Jang S, Bae SY (2014) The dynamics of public attention: agenda-setting theory meets big data. J Commun 64:193–214
3. Twitter micoroblog and social network service. https://about.twitter.com/. Accessed 2 Dec 2020
4. Twitter: research and experiments. https://help.twitter.com/en/rules-and-policies#research-and-experiments. Accessed 2 Dec 2020
5. Lehmann J, Gonçalves B, Ramasco JJ, Cattuto C (2007) Dynamical classes of collective attention in Twitter. In: Proceedings of the 21st international conference on world wide web (WWW), pp 251–260
6. Eom Y-H, Puliga M, Smailovič J, Mozetič I, Caldarelli G (2015) Twitter-based analysis of the dynamics of collective attention to political parties. PLoS ONE 10:0131184
7. Ko J, Kwon HW, Kim HS, Lee K, Choi MY (2014) Model for Twitter dynamics: public attention and time series of tweeting. Physica A 404:141–149
8. Pen T-Q, Sun G, Wu Y (2017) Interplay between public attention and public emotion toward multiple social issues on Twitter. PLoS ONE 12:0167896
9. Chew C, Eysenbach G (2010) Pandemics in the age of Twitter: content analysis of tweets during the 2009 H1N1 outbreak. PLoS ONE 5:14118
10. Signorini A, Segre AM, Polgreen PM (2011) The use of Twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic. PLoS ONE 6:19467

11. van Lent LGG, Sungur H, Kunneman FA, van de Velde B, Das E (2017) Too far to care? Measuring public attention and fear for ebola using Twitter. J Med Internet Res 19:193

12. Zavarrone E, Grassia MG, Marino M, Cataldo R, Mazza R, Canestrari N CO.ME.T.A.—COVID-19 media textual analysis. A dashboard for media monitoring. https://arxiv.org/pdf/2004.07742.pdf. Accessed 2 Dec 2020

13. Lopez CE, Vasu1 M, Gallemore C Understanding the perception of COVID-19 policies by mining a multilanguage Twitter dataset. https://arxiv.org/ftp/arxiv/papers/2003/2003.10359.pdf. Accessed 2 Dec 2020

14. Pennycook G, McPhetres J, Zhang Y, Lu JG, Rand DG (2020) Fighting COVID-19 misinformation on social media: experimental evidence for a scalable accuracy-nudge intervention. Psychol Sci 31(7):770–780

15. Gao J, Zheng P, Jia Y, Chen H, Mao Y, Chen S, Wang Y, Fu H, Dai J (2020) Mental health problems and social media exposure during COVID-19 outbreak. PLoS ONE 15(4):0231924

16. Dubey S, Biswas P, Ghosh R, Chatterjee S, Dubey MJ, Chatterjee S, Lahiri D, Lavie CJ (2020) Psychosocial impact of COVID-19. Diabetes Metab Syndr Clin Res Rev 14(5):779–788

17. An introduction to Sina Weibo: background and status quo. https://www.whatsonweibo.com/sinaweibo/. Accessed 2 Dec 2020

18. Tong J, Zuo L (2014) Weibo communication and government legitimacy in China: a computer-assisted analysis of Weibo messages on two 'mass incidents'. Inf Commun Soc 17:66–85

19. Nip JYM, Fu K-W (2016) Networked framing between source posts and their reposts: an analysis of public opinion on China's microblogs. Inf Commun Soc 19:1127–1149

20. Zhu Y, Fu K, Grépin K, Liang H, Fung I (2020) Limited early warnings and public attention to coronavirus disease 2019 in China, January–February, 2020: a longitudinal cohort of randomly sampled weibo users. Disaster Med Public Health Prep 14(5):E24–E27

21. Zhao Y, Cheng S, Yu X, Xu H (2020) Chinese public's attention to the COVID-19 epidemic on social media: observational descriptive study. J Med Internet Res 22:18825

22. Li X, Zhou M, Wu J, Yuan A, Wu F, Li J Analyzing COVID-19 on online social media: trends, sentiments and emotions. https://arxiv.org/pdf/2005.14464.pdf. Accessed 2 Dec 2020

23. Seo D-W, Shin S-Y (2017) Methods using social media and search queries to predict infectious disease outbreaks. Healthc Inf Res 23(4):343–348

24. Alessa A, Faezipour M (2018) A review of influenza detection and prediction through social networking sites. Theor Biol Med Model 15(1):2

25. Higgins TS, Wu AW, Sharma D, Illing EA, Rubel K, Ting JY (2020) Correlations of online search engine trends with coronavirus disease (COVID-19) incidence: infodemiology study. JMIR Public Health Surveill 6(2):19702

26. Qin L, Sun Q, Wang Y, Wu K-F, Chen M, Shia B-C, Wu S-Y (2020) Prediction of number of cases of 2019 novel coronavirus (COVID-19) using social media search index. Int J Environ Res Public Health 17(7):2365

27. Li C, Chen LJ, Chen X, Zhang M, Pang CP, Chen H (2020) Retrospective analysis of the possibility of predicting the COVID-19 outbreak from Internet searches and social media data, China, 2020. Euro Surveill 25(10):2000199

28. Chunara R, Andrews JR, Brownstein JS (2012) Social and news media enable estimation of epidemiological patterns early in the 2010 Haitian cholera outbreak. Am Soc Trop Med Hyg 86(1):39–45

29. Liu K, Li L, Jiang T, Chen B, Jiang Z, Wang Z, Chen Y, Jiang J, Gu H (2016) Chinese public attention to the outbreak of ebola in West Africa: evidence from the online big data platform. Int J Environ Res Public Health 13(8):780

30. Blumm N, Ghoshal G, Forró Z, Schich M, Bianconi G, Bouchaud J-P, Barabási A-L (2012) Dynamics of ranking processes in complex systems. Phys Rev Lett 109:128701

31. Criado R, Garcia E, Pedroche F, Romance M (2013) A new method for comparing rankings through complex networks: model and analysis of competitiveness of major European soccer leagues. Chaos 23:043114

32. Morales JA, Sánchez S, Flores J, Pineda C, Gershenson C, Cocho G, Zizumbo J, Rodríguez RF, Iñiguez G (2016) Generic temporal features of performance rankings in sports and games. EPJ Data Sci 5:33

33. Morales JA, Colman E, Sánchez S, Sánchez-Puig F, Pineda C, Iñiguez G, Cocho G, Flores J, Gershenson C (2018) Rank dynamics of word usage at multiple scales. Front Phys 6:45

34. Alshaabi T, Minot JR, Arnold MV, Adams JL, Dewhurst DR, Reagan AJ, Muhamad R, Danforth CM, Dodds PS How the world's collective attention is being paid to a pandemic: COVID-19 related n-gram time series for 24 languages on Twitter. https://arxiv.org/pdf/2003.12614.pdf. Accessed 2 Dec 2020

35. Dewhurst DR, Alshaabi T, Arnold MV, Minot JR, Danforth CM, Dodds PS Divergent modes of online collective attention to the COVID-19 pandemic are associated with future caseload variance. https://arxiv.org/pdf/2004.03516.pdf. Accessed 2 Dec 2020

36. Weibo reports first quarter 2020 unaudited financial results. http://ir.weibo.com/news-releases/news-release-details/weibo-reports-first-quarter-2020-unaudited-financial-results/. Accessed 2 Dec 2020

37. Wang Y An introduction to Sina Weibo for journalists. https://www.interhacktives.com/2018/02/22/how-to-use-sina-weibo-as-a-journalist/. Accessed 2 Dec 2020

38. Service WC Common questions on the rules of real-time hot-search-list, hot-message-list and hot-topic-list. https://www.weibo.com/ttarticle/p/show?id=2309404007731978739654. Accessed 2 Dec 2020

39. Weibo Advertising. https://www.marketingtochina.com/weibo-advertising/. Accessed 2 Dec 2020

40. National Health Commission of People's Republic of China. http://www.nhc.gov.cn/xcs/xxgzbd/gzbd_index.shtml. Accessed 2 Dec 2020

41. China confirms 15,152 new coronavirus cases, 254 additional deaths. https://www.cnbc.com/2020/02/13/coronavirus-latest-updates-china-hubei.html. Accessed 2 Dec 2020

42. Sajid I China reports 99 new virus cases majority imported. https://www.aa.com.tr/en/asia-pacific/china-reports-99-new-virus-cases-majority-imported/1801667. Accessed 2 Dec 2020

43. Savitzky–Golay filter. https://en.wikipedia.org/wiki/Savitzky-Golay_filter. Accessed 2 Dec 2020