Competition for popularity and interventions on a Chinese microblogging site

Keywords: social media, public attention, popularity dynamics, rank dynamics, interventions

Extended Abstract

Microblogging sites are important vehicles for the users to obtain information and shape public opinion thus they are arenas of continuous competition for popularity. Most popular topics are usually indicated on ranking lists. In this study, we investigate the public attention dynamics through the Hot Search List (HSL) of the Chinese microblog Sina Weibo, where trending hashtags are ranked based on a multi-dimensional search volume index. We characterize the rank dynamics by the time spent by hashtags on the list, the time of the day they appear there, the rank diversity, and by the ranking trajectories. We show how the circadian rhythm affects the popularity of hashtags, and observe categories of their rank trajectories by a machine learning clustering algorithm. By analyzing patterns of ranking dynamics using various measures, we identify anomalies that are likely to result from the platform provider's intervention into the ranking, including the anchoring of hashtags to certain ranks on the HSL. We propose a simple model of ranking that explains the mechanism of this anchoring effect.

Findings: The duration of the hashtags on the HSL in relation to the time of the day they enter the list shows trimodality (Fig. 3). This is related to the fact that the appearance of hashtags on the HSL have circadian patterns (Fig. 1A). On the one hand, the pattern is caused by the circadian rhythm of the users whose activities depend on the time of the day, on the other hand it is imposed by the apparent working mode of Sina Weibo, which reduces the night-time flow of new hashtags to the HSL almost to zero level. The night break is reflected in the very low number of points in the stripe separating the two triangles in Fig. 3A and in the particularly sharp upper boundary of this stripe. A further observation indicating intervention is that some hashtags on the HSL appear at high ranks and disappear in short time (Fig. 4C), we found these hashtags are mostly from the Star category. Similarly, there are many hashtags that just stay on the HSL for short time which is shown in the first peak in Fig. 3B. The fact that the peak is separated from the rest of the distribution is also likely be related to intervention.

More importantly, we show an anchoring effect at some rank positions on the HSL, where rank diversity is suppressed as compared to the expected smooth behavior of this quantity. Using a simple ranking model, we show how anchoring at some rank positions changes rank diversity. The intervention into the rank produces a deep valley at the anchoring position, very similar to those observed in the measured curves in Fig. 5 which shows the comparison between the real data and our model with anchoring. At certain positions (ranks 8, 16, 28, and 33) there are large drops in the values of the function, indicating intervention by "anchoring" hashtags at these specific ranks. With the simple model, reproducing qualitatively the effect, we support the assumption that the observed anomalies in the ranking functions are due to intervention. To understand the background why some hashtags get anchored, we classified the hashtags that have stayed at the anchoring ranks for longer than 2 hours into four categories based on semantic meaning. Figure 6 (A)(B)(C)(D) show the proportion of such hashtags by category at each of the anchoring ranks 8, 16, 28, and 33, respectively. Comparing with Fig. 6E, where the percentages are the average of each category at six non-anchoring ranks (5, 12, 21, 25, 30, 37), ranks 8 (Fig. 6A), 28 (Fig. 6C), and 33 (Fig. 6D) clearly have a large promoted proportion of International hashtags where the majority are related to international politics. Social hashtags also have a larger proportion at anchoring ranks except for rank 33.

Impact: Sina Weibo is the microblogging site with world-wide the largest number of active users, who are overwhelmingly Chinese speakers. While we believe that alone the size of Sina Weibo justifies focused study, we know that most of our results are idiosyncratic. However, this is true only in a narrow sense as our results provide general lessons. We demonstrated that studying the ranking dynamics in popularity lists is worth for several reasons. First, we uncovered relationships between ranking dynamics and the circadian pattern of user activity. Moreover, we identified different trajectory categories on the list, which characterize different dynamic patterns of popularity. Finally, and most importantly, we showed, how pinpointing anomalies in ranking statistics can be used to identify interventions by the service provider. As service providers have financial interests and may be under political pressure, objectivity of the ranking lists and its truth content can be questioned.

As the platform algorithms may change from time to time, it is challenging to keep track of the interventions, as they can be detrimental being a possible tool of online mass manipulation. Thus, similar to the fight against fake news, the fight against manipulation of public attention is in the interest of the society and it also needs the tools of detecting interventions. Our studies give important reference not only in terms of intervention detection on social media, but also for other research disciplines, such as communication science, journalism, political science, to investigate in further details the specific messages and different aspects of online political contents and learn more about the motivations of such interventions.

Data and methods: We crawled the data from Sina Weibo Hot Search List (HSL), with a frequency of $\Delta t = 5$ minutes from 22 May 2020 to 29 September 2020. We use the rank diversity [1] measure to capture the ranking characteristics of the hashtags at different ranks on the HSL. A popular hashtag i enters the HSL at time t_i at enter-rank $r_i(t_i)$ with $1 \leq r_i(t_i) \leq L$ and disappears from it at time T_i at leave-rank $r_i(T_i)$. During the period $t_i \leq t \leq T_i$ the rank of this hashtag changes with time producing a trajectory $r_i(t)$ on the HSL. The rank trajectory $r_i(t)$ is uniquely defined for $\forall$ hashtag i. Some hashtags have short lifetime on the HSL, others can attract popularity for a longer period of time; some go rapidly to high ranks, others never reach that level. To study whether there are similarities between different shapes of the trajectories, we use machine learning techniques to find characteristic patterns in these rank trajectories. In order to deal with rank time series of different lengths, we use Dynamic Time Warping (DTW) [2] as a similarity measure between two time series. DTW computes the best possible alignment between two time series. Then we use k-means clustering to find clusters of characteristic shapes. The computation was done using python tslearn package [3].

References

[1] Morales JA, Sánchez S, Flores J, Pineda C, Gershenson C, Cocho G, Zizumbo J, Rodríguez RF, Iñiguez G. Generic temporal features of performance rankings in sports and games. EPJ Data Science. 2016 Dec;5:1-6.
[2] Müller M. Dynamic time warping. Information retrieval for music and motion. 2007:69-84.
[3] Tavenard R, Faouzi J, Vandewiele G, Divo F, Androz G, Holtz C, Payne M, Yurchak R, Rußwurm M, Kolar K, Woods E. Tslearn, a machine learning toolkit for time series data. The Journal of Machine Learning Research. 2020 Jan 1;21(1):4686-91.

Figures

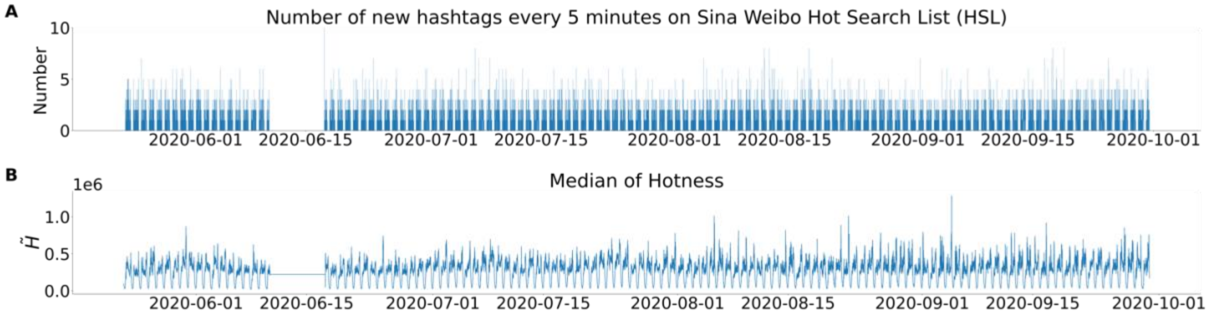

Figure 1. Circadian patterns of the Sina Weibo Hot Search List (HSL). (A) Increment of number of new hashtags per 5 minutes on the HSL during the observation period from 22 May 2020 to 29 September 2020. (B) Time series of the median of search volume index of all hashtags on the HSL at a timestamp, advertisement rank positions excluded. $\widetilde{H}$represents the median value hotness H of hashtags on Sina Weibo HSL at a timestamp. In both (A) and (B) the one-week gap due to the suspension of HSL by the cyberspace authority of China is visible.

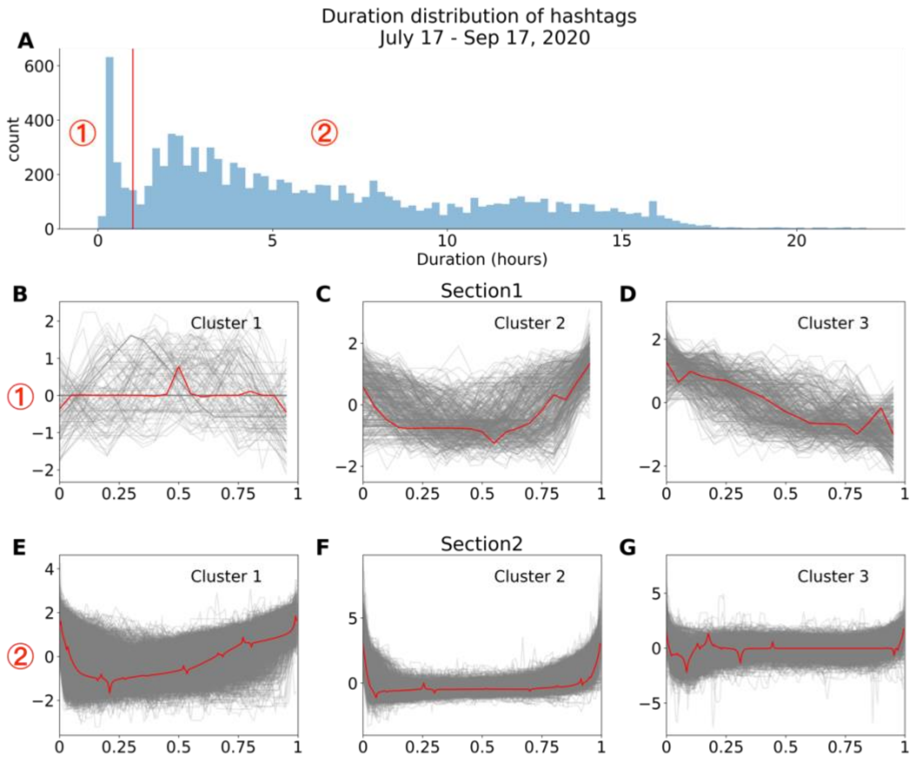

Figure 2. Clustering patterns of hashtag rank trajectories on the Sina Weibo HSL. (A) Distribution of hashtag duration on the HSL, divided into two sections based on local minima at 1 hour. Results of k-means clustering with 3 clusters in each section for time series data are shown, metric is Dynamic Time Warping (DTW) distance, y-axis is normalized to the mean and the standard deviation and the x-axis by d_i. (B), (C), (D) correspond to duration interval from 0 to 1 hour (Section 1). (E), (F), (G) correspond to duration interval larger than 1 hour (Section 2). Red curves depict clustering centers (centroid), computed as the barycenters with respect to DTW.

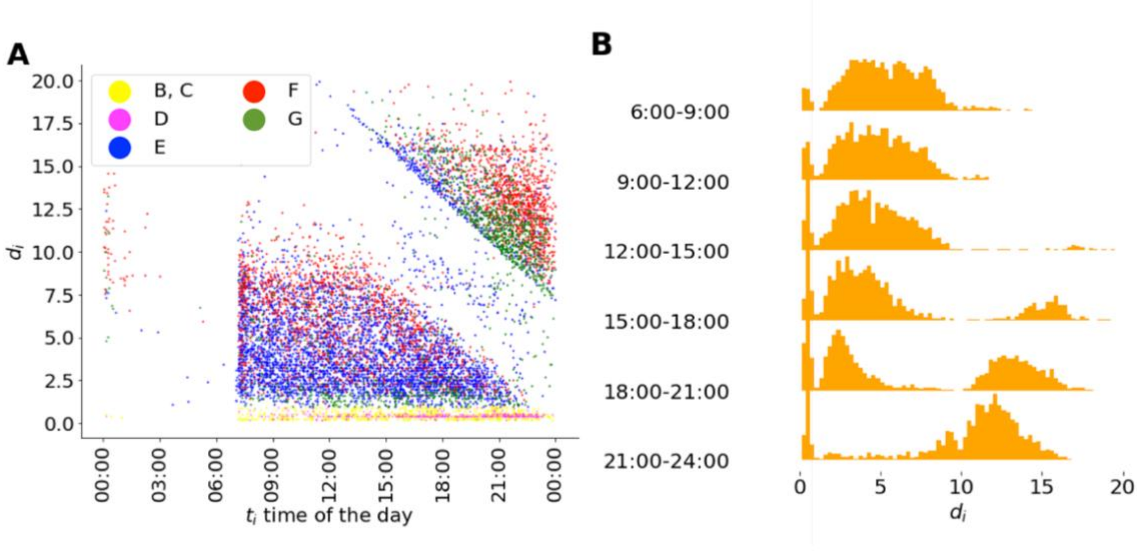

Figure 3. Relationship between hashtags' duration on the HSL and the time t_i. (A) Scatter plot of hashtags' duration on the HSL and the time of the day they first appear on the HSL. Each point is a hashtag, colored by the category it is clustered in Fig. 2. (B) Distribution of hashtags' duration on the HSL according to different time intervals during the day of first appearance on HSL.

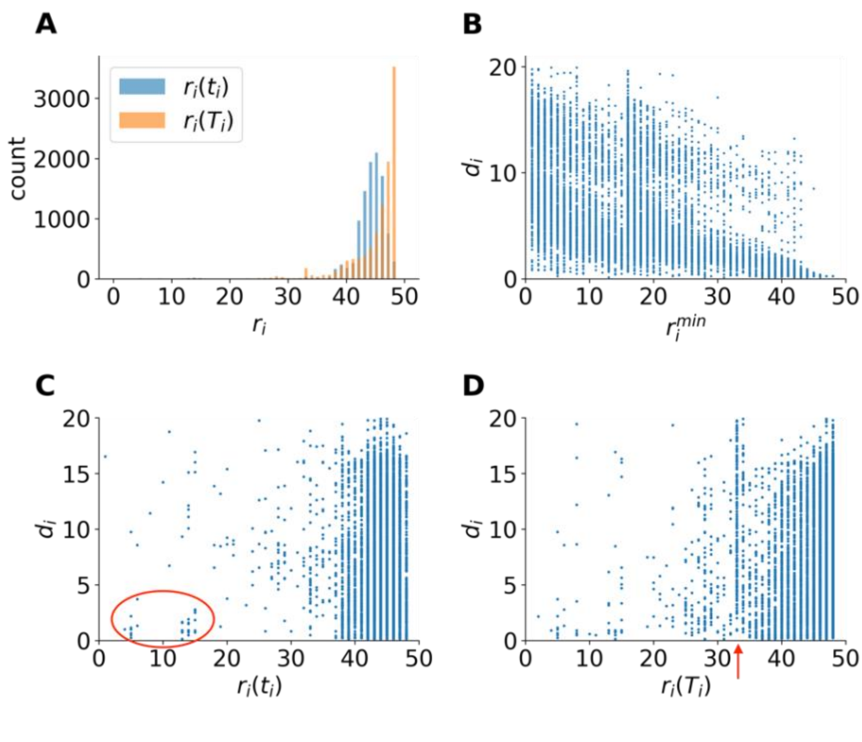

Figure 4. Ranking dynamics characterization of hashtags on the Sina Weibo HSL from 17 July 2020 to 17 Sep 2020. (A) Distribution of $r_i(t_i)$ and $r_i(T_i)$. (B) Scatter plot of r_i^{min} and d_i. (C) Scatter plot of $r_i(t_i)$ and d_i, hashtags with high enter-rank and short duration are circled in red. (D) Scatter plot of $r_i(T_i)$ and duration d_i, rank 33 marked by red arrow.

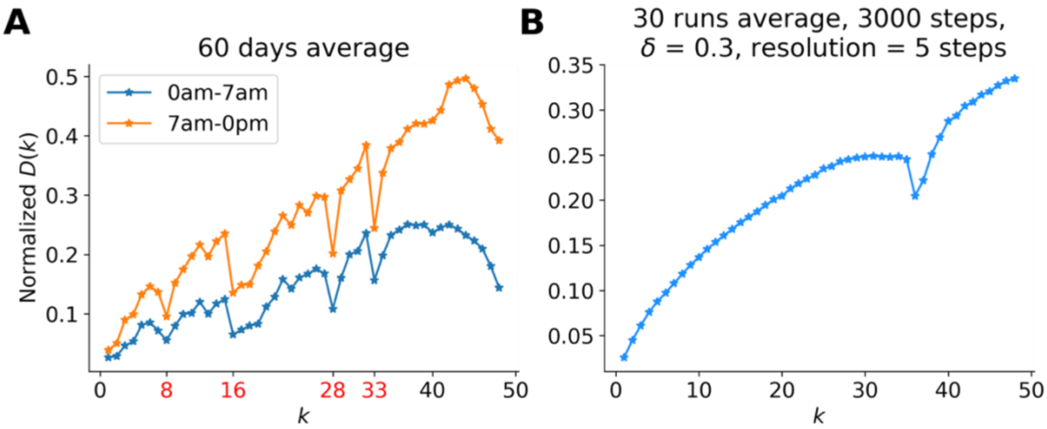

Figure 5. Rank dynamics between empirical data and a ranking model with anchoring. (A) Empirical rank diversity separated for day (upper line) and night (lower line). The sudden drops are at ranks 8, 16, 28, and 33. (B) Simulated rank diversity with the anchor effect.

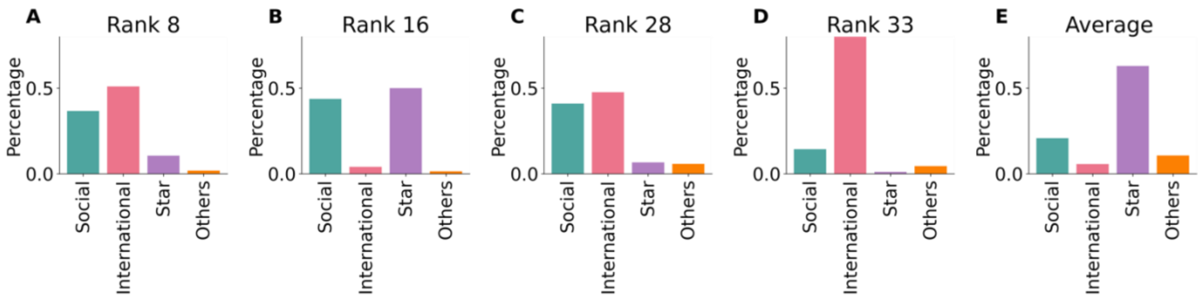

Figure 6. Categorized proportion of hashtags that have stayed at certain ranks on HSL for longer than 2 hours. (A)(B)(C)(D) show the content distribution of hashtags at ranks 8, 16, 28, and 33 respectively, corresponding to the sudden drops in Fig. 5A. (E) Averaged proportion of hashtags by content category at ranks 5, 12, 21, 25, 30, 37.