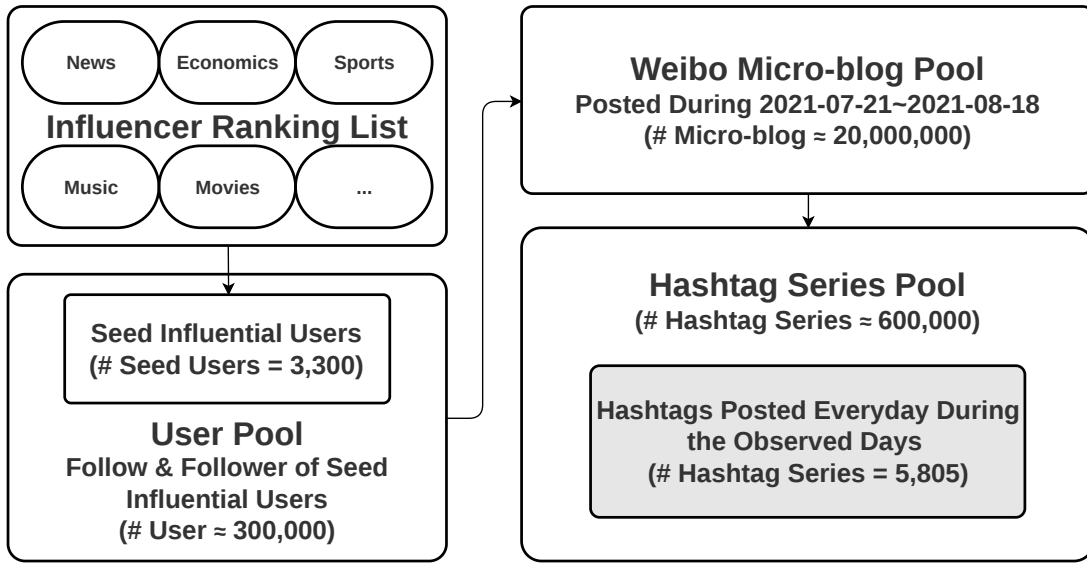


Scale-dependent Power Law Properties in Hashtag Usage Time Series of Weibo

Jiwei J. Jiang, Kenta Yamada, Hideki Takayasu, Misako Takayasu

1 Data Set

Weibo is a mainstream micro-blog social media in China, we collect Weibo data through the publicly available API. Due to the huge volume of users and the limitation of the API, it is impossible to collect all the data for analysis, so we focus on the hashtag posting behavior of sampled users with a total number of about 300,000. We collected a total of about 20 million micro-blogs from these users posted between July 21st and August 18th in 2021, and extracted about 600,000 different hashtags. We chose the 5805 hashtags that are posted every day during the observation interval (29 days) as the object for the convenience of the analysis. The detailed process of collecting data is shown in the flow chart in Figure 1.



Supplementary Figure 1: Flowchart for collecting the hashtag usage count time series with open API. Due to the huge volume of users and the limitation of the API, it is impossible to collect all the data, and we focus on the hashtags posted by sampled users. We obtain the data in three steps, first, we select 3,300 seed users from the publicly influential users' ranking lists which contain the official top 100 influences in 33 different fields. The following and followers of the seed users are sampled, getting a sample of about 300,000 users which we called a user pool; Then, from these users, we collected micro-blogs posted between July 21st and August 18th in 2021 to get our Weibo micro-blog Pool, a total of about 20 million blogs; Finally, we extract hashtag data from micro-blogs to get the hashtag pool, where about 600,000 different hashtags are extracted. The duration of hashtags ranges from 1 day to 29 days (the longest observation interval). For the convenience of the analysis, we choose 5805 hashtags that were used every day during the observation interval (29 days) as the object of analysis.

2 Functional Form of the Growth Rate Distributions

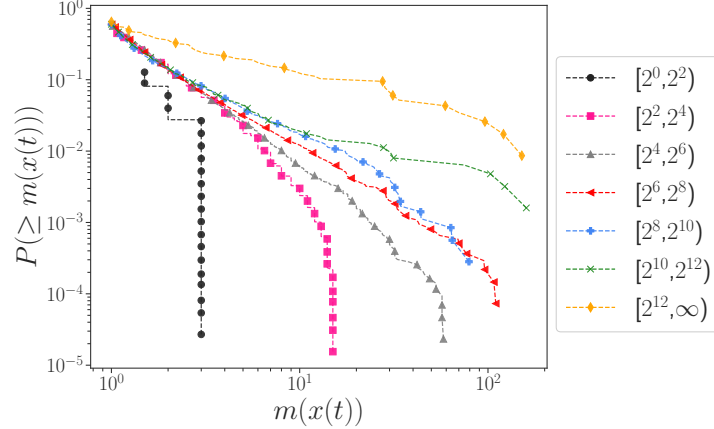
In the chapter on Dynamic Properties of main text, we observed that the logarithmic growth rate $\log b(t) = \log \frac{x(t+1)}{x(t)}$ follows the Laplace distribution as shown in Figure 4, the Laplace distribution also known as double exponential distribution, which means that $\log b$ follows exponential distribution on both sides of greater than μ and less than μ , respectively. Thus the cumulative probability density function for b is approximately power law function as follows.

$$\begin{cases} P(\leq b) \propto b^{\sqrt{2}/\sigma}, b \leq e^\mu \approx 1 \\ P(> b) \propto b^{-\sqrt{2}/\sigma}, b > e^\mu \approx 1 \end{cases} \quad (1)$$

Here $\mu \approx 0$, $\sigma \approx 0.31$ refer to the mean and the standard deviation value of $\log b(t)$. In Figure 5 (a) we demonstrated that the distributions of $b(t|x(t))$ for different intervals of $x(t)$ are asymmetric and deviate clearly from the Laplace distribution. For the cumulative distributions of $b(t|x(t)) < 1$ and $b(t|x(t)) > 1$ which are shown in Figure 5 (b)(c), the distributions at log-log scale are approximately straight lines, implies that these distributions can still be approximated by power law functions however the power law exponents are deviated from that in Equation 1. Here we fit these cumulative distributions.

2.1 Cumulative distribution functions of $b(t|x(t)) < 1$

For this case, we observe in Figure 5 (b) that for different sizes of $x(t)$, $b(t)$ is a power law like distribution with cutoff. Since it is not often to fit a power law distribution with a cumulative probability calculated by $P(\leq b) = \int_0^b p(b')db'$, we first observe and fit the cumulative probability function of $m = \frac{1}{b(t|x(t))}$, which exhibits a cumulative probability function $P(\geq m)$ symmetric to $P(\leq b)$, as shown in Supplementary Figure 2.



Supplementary Figure 2: Cumulative distribution of $m(x(t))$. Log-Log plot of cumulative distribution for m , m is calculated by inverse of growth rate $\frac{1}{b(t|x(t))}$ where $b(t|x(t)) < 1$. The cumulative probability is calculated by $P(\geq m) = \int_m^\infty p(m')dm'$.

Since for different intervals of $x(t)$, $m = \frac{x(t)}{x(t+1)}$, and the maximum value of m takes the maximum value of $x(t)$ in that interval when $x(t+1) = 1$, so we consider fitting the distribution of m as a power law function with specific cutoff, as follows.

$$P(\geq m) \propto (m + m_0)^{-\beta} - (m_c + m_0)^{-\beta}, \quad 1 < m \leq m_c \quad (2)$$

where m_c is the parameter that controls the position of cutoff, such that $P(\geq m_c) = 0$. For different interval of $x(t)$, m_c is set to maximum value of $x(t)$ minus 1 as shown in Supplementary Table 1. The

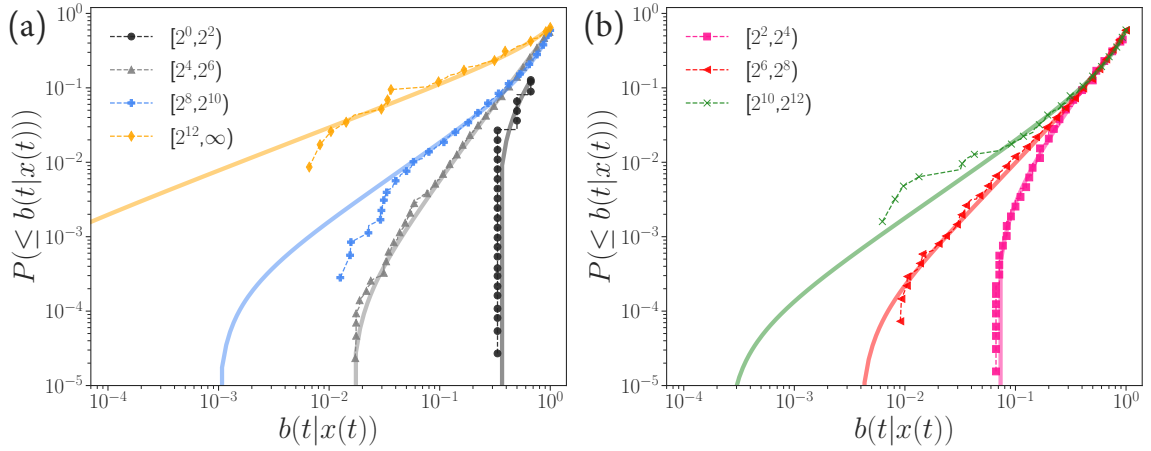
other two parameters m_0 and β are estimated by the maximum likelihood method. The estimated parameters for each interval of $x(t)$ are presented in Supplementary Table 1. After approximating the functional form of the cumulative distribution of m , we can obtain the functional form of the cumulative distribution of $b(t|x(t))$ as follows.

$$P(\leq b) = P(\geq \frac{1}{b}) = P(\geq m) \propto (\frac{1}{b} + m_0)^{-\beta} - (m_c + m_0)^{-\beta}, \quad \frac{1}{m_c} \leq b < 1 \quad (3)$$

The three parameters are the same as the cumulative distribution of m , as shown in Supplementary Table 1. Supplementary Figure 3 shows the fitting result of $b(t|x(t))$.

Supplementary Table 1: Estimated result of the parameters of $b(t|x(t))$ for the case of $b(t|x(t)) < 1$.

$\backslash$	$[2^0, 2^2)$	$[2^2, 2^4)$	$[2^4, 2^6)$	$[2^6, 2^8)$	$[2^8, 2^{10})$	$[2^{10}, 2^{12})$	$[2^{12}, +\infty)$
m_0	0.60	0.53	0.11	-0.36	-0.72	-0.68	-0.56
β	2.74	2.43	2.09	1.51	0.99	1.01	0.57
m_c	$2^2 - 1$	$2^4 - 1$	$2^6 - 1$	$2^8 - 1$	$2^{10} - 1$	$2^{12} - 1$	∞



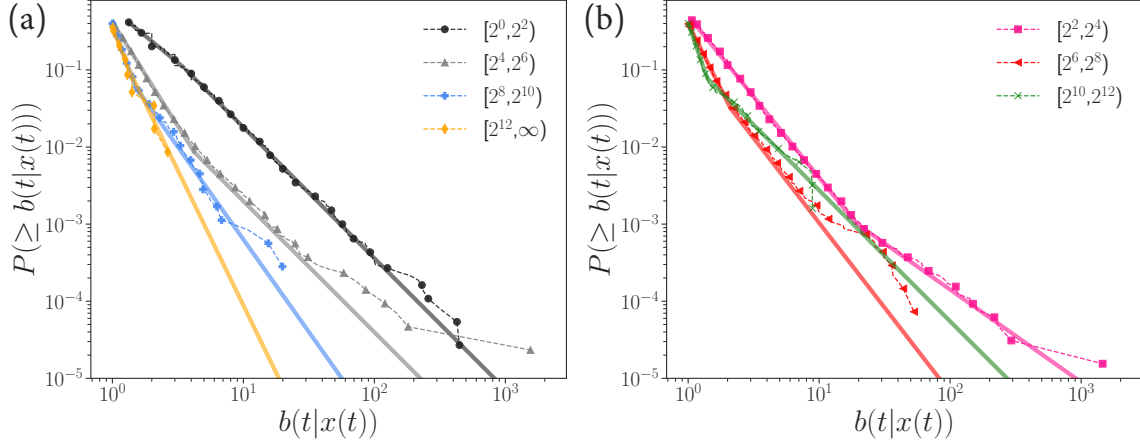
Supplementary Figure 3: The fitting results of the cumulative distribution of $b(t|x(t))$ for the case of $b(t|x(t)) < 1$. The lines with graphs are of real data, the solid lines are of fitted function forms from Equation 3. (a) Distributions for the case of $x(t)$ in intervals of $[2^0, 2^2)$, $[2^4, 2^6)$, $[2^8, 2^{10})$ and $[2^{12}, +\infty)$; (b) Distributions for the case of $x(t)$ in intervals of $[2^2, 2^4)$, $[2^6, 2^8)$ and $[2^{10}, 2^{12})$.

2.2 Cumulative Distribution Functions of $b(t|x(t)) > 1$

For this case, we find that for different intervals of $x(t)$, $b(t)$ is a power law like distribution, and we find that the power law exponent changes for large value $b(t)$ as shown in Supplementary Figure 4, we consider the following functional form to fit the distribution,

$$P(\geq b) \propto \begin{cases} b^{-\gamma_1} & , 1 < b < b_{mid} \\ b^{-\gamma_2} & , b \geq b_{mid} \end{cases} \quad (4)$$

where b_{mid} , γ_1 and γ_2 are parameters, we obtain the estimation of the three parameters by minimizing the sum of squared errors in the log-log scale. The results are presented in Supplementary Table 2. The distributions of these function form is shown in Supplementary Figure 4.



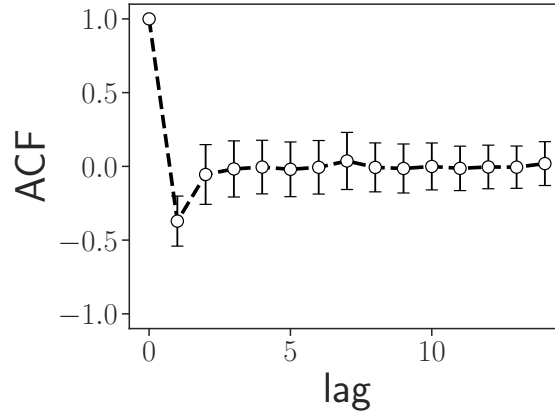
Supplementary Figure 4: The fitting results of the cumulative distribution of $b(t|x(t))$ for the case of $b(t|x(t)) < 1$. The lines with graphs are of real data, the solid lines are of fitted function forms from Equation 4. **(a)** Distributions for the case of $x(t)$ in intervals of $[2^0, 2^2)$, $[2^4, 2^6)$, $[2^8, 2^{10})$ and $[2^{12}, +\infty)$; **(b)** Distributions for the case of $x(t)$ in intervals of $[2^2, 2^4)$, $[2^6, 2^8)$ and $[2^{10}, 2^{12})$.

Supplementary Table 2: Estimated result of the parameters of $b(t|x(t))$ for the case of $b(t|x(t)) > 1$.

$\backslash$	$[2^0, 2^2)$	$[2^2, 2^4)$	$[2^4, 2^6)$	$[2^6, 2^8)$	$[2^8, 2^{10})$	$[2^{10}, 2^{12})$	$[2^{12}, +\infty)$
b_{mid}	3.5	19.1	4.2	2.1	1.5	1.4	1.3
γ_1	1.4	2.1	2.7	3.4	4.7	4.9	4.5
γ_2	1.7	1.2	1.7	2.2	2.4	1.7	3.5

3 Auto-correlation function (ACF) of Growth Rate

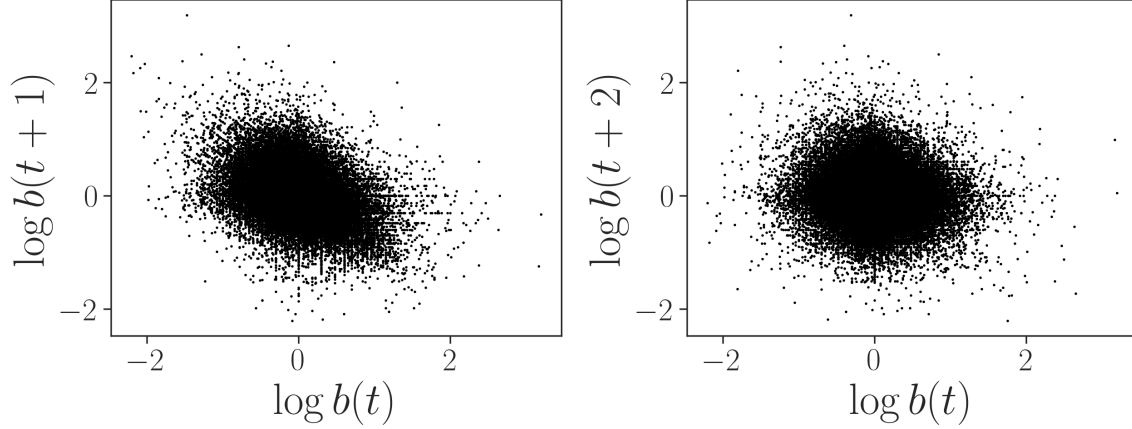
In the Numerical Simulation Results section of the main text, we describe that the logarithmic growth rate, $\log b(t)$, exhibits a negative auto-correlation at a time lag of 1 day, and for the time lag greater than one day, the auto-correlation coefficient is almost 0, as shown in Supplementary Figure 5 below.



Supplementary Figure 5: Auto-correlation function of $\log b(t)$. For all time series $x_i(t)$, the mean value of auto-correlation functions of $\log b_i(t) = \log \frac{x_i(t+1)}{x_i(t)}$ is plotted, and the error bar indicates the standard deviation.

Although we neglected the auto-correlation in the numerical simulations, we add here more details and examples. As shown in Supplementary Figure 6 below, we plot the scatter plot of the size of $\log b$ at time t and time $t+1$. Compare to the plot between time t and time $t+2$, we can find that there is

a negative correlation of $\log b(t)$ at the time lag of 1. The scatter plots with a lag greater than 2 are similar to that with a lag of 2 and are therefore omitted here.



Supplementary Figure 6: Scatter plot of $\log b(t)$ with $\log b(t + \text{lag})$. The graphs with lags of 1 and 2 are plotted. The distribution with a lag greater than 2 is similar to the distribution with lag a of 2 and is therefore omitted.

This negative auto-correlation at the lag of 1 day indicates a tendency for the variable $\log b(t)$ to decrease immediately after an increase, and vice versa. We selected hashtags for which the auto-correlation coefficient of lag 1 day is less than 0.5, and observe the time series of the hashtag usage count, $x(t)$, and growth rate, $\log(b(t))$. The time series of the number of users of the corresponding hashtags and their growth rate of them are also observed. Typical three examples are given in Supplementary Figure 7.

- (a) (b) (c) shows the time series of the hashtag '#Photography'

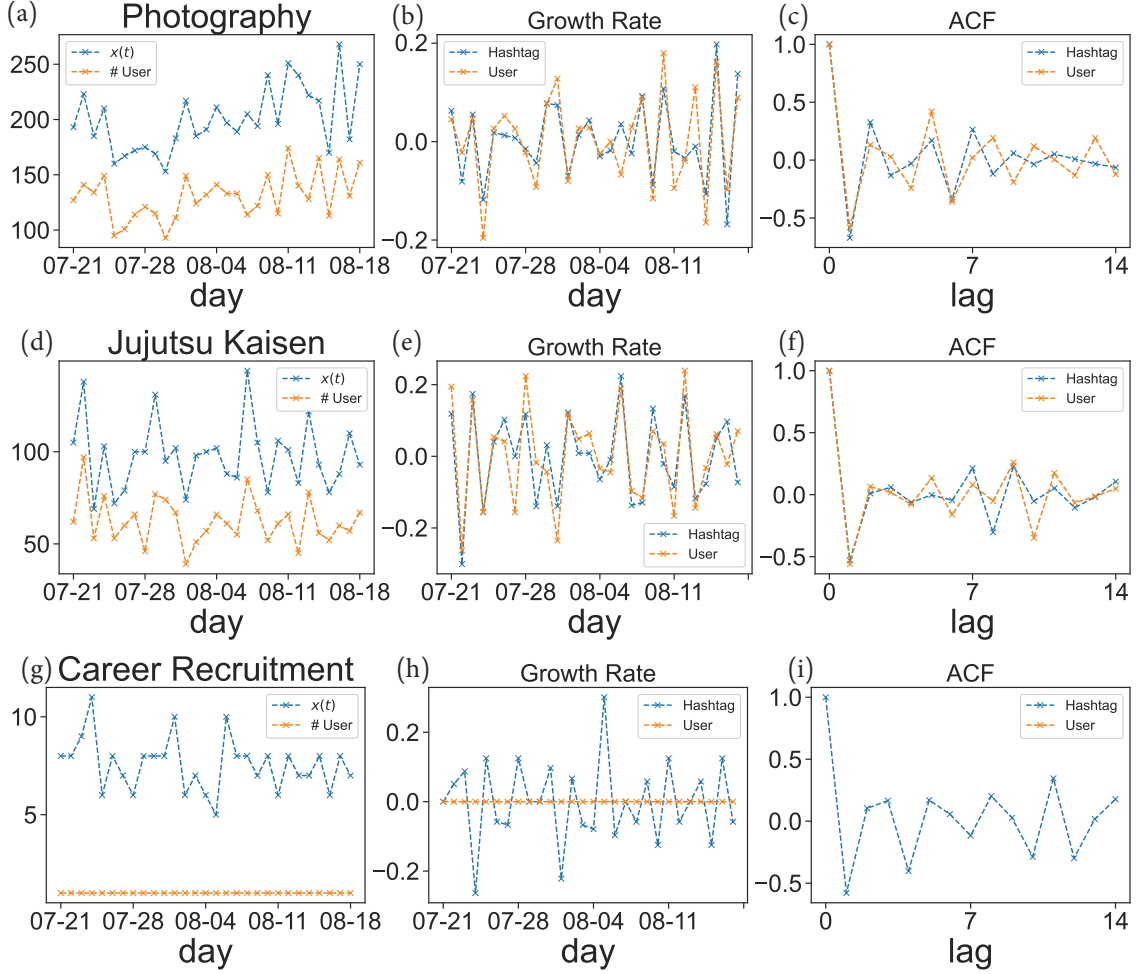
As shown in (c), the ACF of the growth rate of the number of users who use this hashtag is similar to the ACF of the growth rate of the hashtag usage count, and we infer that the negative auto-correlation of lag one day of $\log b(t)$ is due to the negative auto-correlation of the growth rate of the number of users.

- (d) (e) (f) show the time series of the hashtag '#Jujutsu Kaisen'

Hashtag '#Jujutsu Kaisen' is the name of a Japanese anime, from the time series of $x(t)$ in (d), it exhibits a clear feature of spikes, which we inferred is the reason for the negative auto-correlation of lag 1 day of $\log b(t)$.

- (g) (h) (i) show the time series of hashtag '#Career Recruitment'

As shown in the time series of the number of users of this hashtag in Figure (g), this is a hashtag used by only one user, and the number of daily usages is often stable, with a sudden increase in usage count on individual days, often accompanied by a decrease in submissions on the next day, which we inferred is the reason for the negative auto-correlation of lag 1 day of $\log b(t)$.



Supplementary Figure 7: Time series and ACF examples. Three examples of different hashtags are listed in different rows. The blue dotted lines represent hashtags, the first column is the daily usage count time series of hashtags, the second column is the time series of the corresponding growth rate, and the third column is the ACF of the growth rate. The orange dotted lines represent the number of users, the first column is the time series of the number of users of the corresponding hashtag, the second column is the corresponding growth rate, and the third column is the ACF. (a)(b)(c) Hashtag of '#Photography'; (d)(e)(f) Hashtag of '#Jujutsu Kaisen'; (g)(h)(i) Hashtag of '#Career Recruitment'.