

**Stochastic dynamics and the predictability of big hits in online videos**José M. Miotto,<sup>1</sup> Holger Kantz,<sup>1</sup> and Eduardo G. Altmann<sup>1,2</sup><sup>1</sup>*Max Planck Institute for the Physics of Complex Systems, 01187 Dresden, Germany*<sup>2</sup>*School of Mathematics and Statistics, University of Sydney, NSW 2006, Australia*

(Received 8 February 2016; revised manuscript received 2 February 2017; published 9 March 2017)

The competition for the attention of users is a central element of the Internet. Crucial issues are the origin and predictability of big hits, the few items that capture a big portion of the total attention. We address these issues analyzing  $10^6$  time series of videos' views from YouTube. We find that the average gain of views is linearly proportional to the number of views a video already has, in agreement with usual rich-get-richer mechanisms and Gibrat's law, but this fails to explain the prevalence of big hits. The reason is that the fluctuations around the average views are themselves heavy tailed. Based on these empirical observations, we propose a stochastic differential equation with Lévy noise as a model of the dynamics of videos. We show how this model is substantially better in estimating the probability of an ordinary item becoming a big hit, which is considerably underestimated in the traditional proportional-growth models.

DOI: [10.1103/PhysRevE.95.032311](https://doi.org/10.1103/PhysRevE.95.032311)**I. INTRODUCTION**

YouTube is a representative example of online platforms in which items (videos in this case) compete for the attention of users [1–3]. The popularity of videos varies by orders of magnitude, resembling the fat-tailed distributions that have been reported in other online systems [4–6], in income and wealth [7], in finance [8], and in disciplines such as ecology, earth science, and physics [9]. The origin of such fat-tailed distributions is a century-old problem that lies at the heart of complex-systems science [10–15]. At the core of the different proposed models lies the idea that the current popularity (wealth) determines the future popularity gain (income) and enhances the inequality (rich-get-richer). Indeed, such (linear) proportional growth is the essential ingredient of Gibrat's law (used to describe the growth of firms [11,13] and cities [16]), the Yule-Simon model (to model species genera [10] and language [17,18]), scientific memes [19], and the preferential attachment model of network growth [20]. Proportional growth suggests that the big hits are very predictable because they originate from early advantages that are amplified over time.

The application of growth models to describe the popularity of online items brings new opportunities and challenges. On the one hand, due to the increasing availability of data sets, it becomes possible to compare models with an unprecedented accuracy. On the other hand, the expectations we have of the models are higher. For instance, a central question is how to forecast and identify the origins of the big hits [21,22], the most successful videos, which capture most of the attention and produce most of the revenue through advertisement. To address this and other questions, the characterization of the heavy-tailed distribution of aggregated activity is not enough. One has to (i) improve the description of the dynamics of individual items, and (ii) go beyond the average growth and analyze the stochastic fluctuations [23,24]. The importance of these factors is illustrated in Fig. 1, where we show trajectories (views vs. time) of videos with the same early success (the same number of views three days after publication). We see that trajectories quickly spread and that many trajectories with a weak start become popular in time. This suggests that big hits have a low predictability (i.e., they are hard to anticipate).

In this article, we investigate the predictability of big hits using stochastic models of individual items. Predictability is the possibility of anticipating the future based on present information and we confront the predictability expected from models to observations in the data. We compare traditional growth models to data ( $X_t$ , views over time) of more than  $10^6$  YouTube videos. We find that previously proposed models are unable to correctly account for the (random) fluctuations observed in the data, which we find to be described by a Lévy-stable distribution. We propose and validate a stochastic model that explains such reduced predictability by incorporating both proportional growth and Lévy noise. This shows that, even if present, proportional growth is not the only thing responsible for the origin of fat-tailed distributions. Finally, we show that our model substantially improves the prediction of the probability of big hits, but that unexpected big hits have an even higher probability in real data due to temporal correlations not accounted by this class of models.

**II. THEORETICAL FRAMEWORK**

YouTube is a web site where videos generated by third parties are shared. It is the third most visited web site of all the Internet. We collected more than  $10^6$  time series of the daily number of views of videos published between December 2011 and March 2013 [25]. The number of views a video receives depends on the interplay between its content and various factors. Videos related to ongoing events are strongly influenced by their development, their media coverage, and other factors exogenous to the online activity of users. Videos are also influenced by endogenous factors, such as the sharing and recommendation in online media, generating cascades of activity in the social network [26,27]. Additionally, a video can be viewed by following a link from a related video, i.e., hopping through the videos' network, which changes continuously according to YouTube's recommendation and promotion algorithms. The interplay and feedback between these and other factors lead to the complex dynamics we observe in the time series. Modeling specific factors [28] and differentiating between them (e.g., between exogenous and

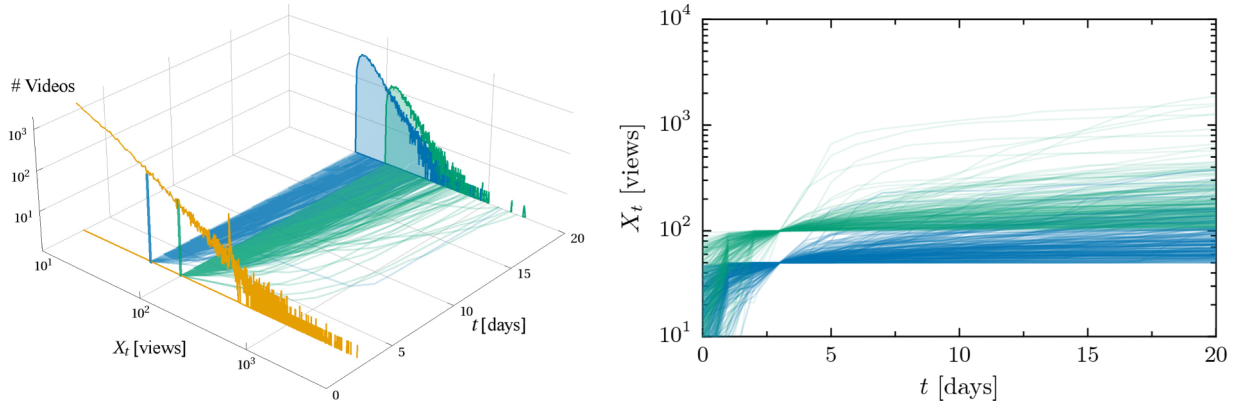


FIG. 1. Evolution of videos' views  $X_t$  as a function of the time  $t$  after publication. After  $t = 3$  days the distribution of views is already heavy tailed (orange histogram). Videos having initially the same amount of views show very distinct evolutions. This is illustrated here by highlighting two groups of videos with the same number of views at  $t = 3$  (blue  $X_{t=3} = 50$  and green  $X_{t=3} = 100$ ). Each line (at the bottom and in the right plot) corresponds to the trajectory of one video. On the back, the histograms of the two groups of videos at  $t = 20$  days are shown.

endogenous factors [26,29,30]) are topics of recent research. This approach is difficult to be pursued because it requires detailed information of user activities and the possibility of isolating the factors. Instead, here we aim at a coarse-grained description of the dynamics of attention in which the combination of the different factors described above are effectively accounted for by deterministic and stochastic terms.

Let  $X_t$  be the cumulative number of views that a video received in the first  $t$  days after its release. A very general stochastic model for the growth of  $X_t$  in  $t$  is the diffusion process [23,31,32]

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dW_t, \quad (1)$$

where  $W_t$  is a Wiener process ( $\langle W_t \rangle = 0$  and  $\langle W_t^2 \rangle = dt$ ),  $\mu(t, X_t)$  is the average growth, and  $\sigma(t, X_t)$  scales the fluctuations; an additional cutoff in  $dW_t$  is added to ensure that  $dX_t > 0$ . We consider all videos to be indistinguishable so that variations in the behavior of videos with the same  $X_t$  should be accounted for by the stochastic term  $\sigma(t, X_t)dW_t$ . Extensions

of our model could consider  $\mu(t, X_t)$  and  $\sigma(t, X_t)$  to depend on properties of the video and on  $X_{t'}$  for  $t' < t$ .

### III. DATA ANALYSIS

We now analyze the data in order to identify the functions  $\mu(t, X_t)$  and  $\sigma(t, X_t)$ . Since the minimum resolution of our data is  $\Delta t = 1$  day, the models we propose aim to fit the quantity  $\Delta X_t = X_{t+1} - X_t$ , the number of views obtained exclusively in the day  $t + 1$ . We first focus on the deterministic term of Eq. (1),  $\mu(t, X_t)$ . In linear proportional growth models the average growth is proportional to the views,  $\mu(t, X_t) = \mu_t X_t$ , where the temporal dependence on  $\mu_t$  accounts for the decay in the attention gain [2]; this decay is very strong in the first weeks, so we will focus on the days up to  $t = 30$ .

This condition is consistent with our data: in Fig. 2(a) we see for a fixed  $t$  that the dependence of the conditional average  $\langle \Delta X_t | X_t \rangle$  (computed in windows of  $N$  videos) with  $X_t$  is roughly a line with slope 1, a standard method to check for proportional growth [15,33]. We now repeat the analysis for

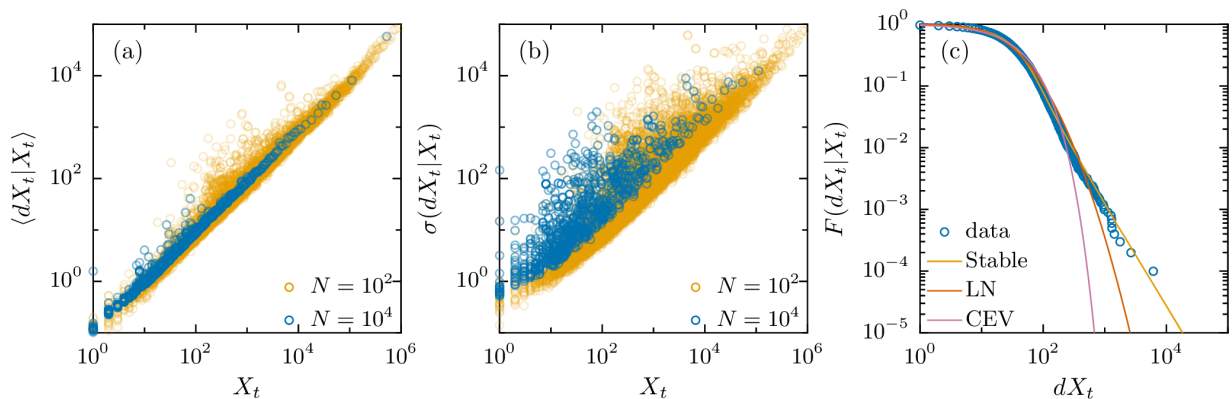


FIG. 2. Average and fluctuations in the growth of YouTube videos. (a) Mean  $\langle \Delta X_t \rangle$  and (b) standard deviation  $\sigma(\Delta X_t)$  for videos with  $X_t$  views  $t = 3$  days after publication. Both  $\langle \dots \rangle$  and  $\sigma$  are computed in windows centered at  $X_t$  and containing  $N$  items (see legend). (c) Complementary cumulative density function  $F(\Delta X_t | X_t)$  for  $X_t \in [499, 513]$  and  $t = 3$ . Data (blue circles) is compared with fits of three distributions (S: Lévy-stable, LN: log-normal, CEV: constant elasticity of variance) and confirms the existence of heavy tails, which are best described by the Lévy distribution.

the stochastic term  $\sigma(X_t, t)$  of Eq. (1). A natural proposal for  $\sigma$  is  $\sigma(X_t, t) = \sigma_t X_t^\beta$  [32], where the  $\beta$  parameter allows us to model a possible fluctuations' scaling in the form of the Taylor's Law [34]. In particular, the  $\beta = 1$  case used in Ref. [23], is equivalent to  $Y_t = \ln X_t$  exhibiting constant fluctuations, and corresponds to a geometric Brownian motion. The simplest way to evaluate the stochastic term in this context is to repeat what was done for the mean and measure the standard deviation  $\sigma$  in a window of  $N$  items centered around  $X_t$  [23]. This is equivalent to the standard estimation of the drift and diffusion coefficients in a Fokker-Planck equation [35]. Results in Fig. 2(b) confirm the roughly linear scaling in the double logarithmic scale, in agreement with  $\sigma(X_t, t) \propto X_t^\beta$  with  $\beta \approx 1$ . However, in contrast to the case of  $\mu(X_t, t)$  shown in Fig. 2(a), the data show strong fluctuations across  $X_t$  and depend on the sample size  $N$  (the larger the  $N$  the larger the measured  $\sigma$ ).

The observations above motivate us to look at the full probability distribution  $\mathbb{P}(\Delta X_t | X_t)$  [36]. In Fig. 2(c) we see in the particular histogram  $\mathbb{P}(\Delta X_3 | X_3 \approx 500)$ , that the distribution has a heavy tail; this explains the observation that  $\sigma$  grows with  $N$ , i.e.,  $\mathbb{P}(\Delta X_t | X_t)$  has a diverging second moment [37]. Heavy-tailed fluctuations of  $\Delta X_t$  may still be compatible with Eq. (1) if one considers that the temporal interval used in our analysis is not infinitesimal  $\Delta t = 1$  day  $\gg dt$ ; indeed, Gaussian fluctuations are expected only when  $\Delta t \rightarrow 0$ . In this case, the stochastic differential equation has to be integrated up to  $\Delta t$ , so the fluctuations predicted from Eq. (1) can be log-normal (for  $\beta = 1$ ) or a distribution arising from the constant elasticity of variance model (CEV, for  $\beta \neq 1$ ) [38], as shown in Appendix A. Besides Eq. (1), classical models associated with Gibrat's law (Champernowne-Gabaix or Yule-Simon) predict  $\mathbb{P}(\Delta X_t | X_t)$  to have either short tails or log-normal distributions (see Appendix B). Beyond the log-normal and CEV distributions, which follow from Eq. (1), we consider also the Lévy-stable distribution (S) because it originates from the generalized central limit theorem for variables without finite variance [39]. In Fig. 2(c) we show the fits of discretized versions of these three distributions to the particular histogram discussed above. The best fit is obtained by the (completely asymmetric) Lévy-stable distribution (with a difference in the Bayes information criterion [40,41], BIC, of 178 and 175, with respect to the log-normal and CEV models). This result, which is confirmed below for different  $X_t$  and  $t$ , indicates that the fluctuations observed in the data are not compatible with the Wiener process  $W_t$  in Eq. (1), and that the analysis of the mean and standard deviation done for Figs. 2(a) and 2(b) may be not enough to define the functions of Eq. (1).

#### IV. ALTERNATIVE MODEL

Motivated by the better fit of the Lévy distribution and by the linear scaling of  $\mu$  and  $\sigma$  with  $X_t$  (as shown in Fig. 2), we propose as an improvement of Eq. (1) [42]

$$dX_t = \mu_t X_t dt + (a_t X_t + b_t) dL_t, \quad (2)$$

where  $L_t$  is an  $\alpha$ -stable Lévy process, analogous to the Wiener process, except that the distribution of  $dL_t$  follows a Lévy-stable distribution with index  $\alpha$ , asymmetry 1, location parameter 0 and scale 1 (using parametrization 1 of Ref. [43]).

TABLE I. Summary of models, see Appendix A for details.

Name	$\mathbb{P}(\Delta X_t   X_t)$ functional form	Parameters
LN	log-normal	$\mu_t, \sigma_t$
CEV	CEV	$\mu_t, \sigma_t, \beta_t$
S	Lévy-stable	$\alpha_t, \mu_t, a_t, b_t$

A cutoff in the noise term is added as above to ensure  $dX_t \geq 0$ , so  $\langle dL_t \rangle > 0$  and  $\langle \Delta X_t \rangle$  is not given alone by the deterministic term  $\mu_t X_t$  (even if  $dL_t$  is understood in the Ito sense, as we do here [44]). The parameters  $\alpha, \mu, a$ , and  $b$  depend on time  $t$  ( $b_t$  is important only for small  $X_t$  and  $t$ ). Table I summarizes all models.

#### V. IMPROVED DATA ANALYSIS

We now discuss how to determine the parameters of the two models derived from Eq. (1) and of the alternative model in Eq. (2) and to test which model best describes the data. The likelihood  $\mathcal{L}_t$  of the models, for a fixed day  $t$ , is the product of the likelihoods of each distribution of  $\Delta X_t$  conditioned on  $X_t$  with respect to the parameters of the model  $\theta$  as

$$\ln \mathcal{L}_t = \sum_{X_t} \sum_{\Delta X_t} N(\Delta X_t, X_t) \ln f(\Delta X_t | \theta, X_t), \quad (3)$$

where  $N(\Delta X_t, X_t)$  is the observed number of videos with a given  $\Delta X_t, X_t$ , and  $f$  is the probability density function proposed by the models. The best parameters  $\theta$  are obtained maximizing  $\ln \mathcal{L}_t$  [45] and the models are compared based on their (maximum) likelihood, penalizing the addition of parameters (using the BIC [41]). The distributions  $f$  we test are the same as above: log-normal (LN) and constant elasticity of variance (CEV), obtained from Eq. (1), and Lévy-stable (stable), obtained from Eq. (2). The latter is the  $f$  resulting from Eq. (2) because, since it seems to be a better fit to the data, we consider in this model that the time step  $\Delta t$  is small such that  $\Delta X_t \approx dX_t$ , making the distribution to be fitted exactly the Lévy. Each of the distributions  $f$  has different parameters that depend on  $\theta$  and  $X_t$ , as summarized in Table I and detailed in Appendix A. Our approach based on Eq. (3) considers all conditional distributions  $\mathbb{P}(\Delta X_t | X_t)$ , avoiding the difficulties and arbitrary choices involved in the grouping of data in windows [46] as done in previous estimations and in Fig. 2.

The application of the analysis described above to the YouTube data leads to significant evidence in favor of the Lévy-stable model, Eq. (2). Figure 3 shows how this model allows for the collapse of the many  $\mathbb{P}(\Delta X_t | X_t)$  in a single curve, well described by a Lévy-stable curve. More formally, the BIC [41] difference of the stable model with respect to the other models is above  $10^5$  for all  $0 \leq t \leq 30$  (inset of Fig. 3), indicating very strong statistical support for our model.

##### A. Dependence of parameters with respect to $t$

The parameters of the S model [Eq. (2)] are explicitly time dependent, so we repeat the previous procedure for each of the days considered. In Fig. 4 the values of these are shown for the first 30 days after the publication of the videos. The parameters

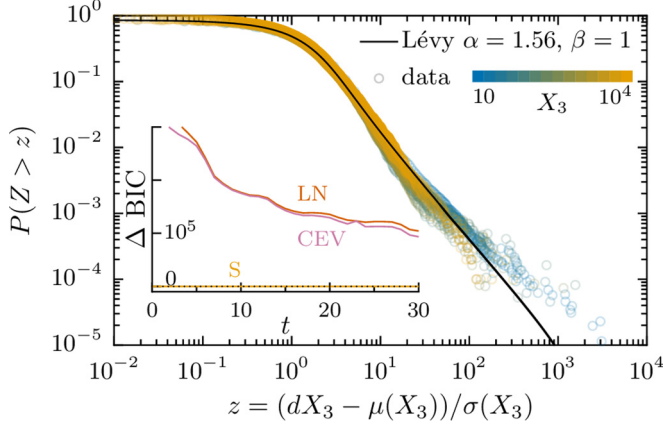


FIG. 3. Agreement of the model with respect to data. Main panel: complementary cumulative distribution of the views rescaled by the fitted parameters for  $t = 3$ . The rescaled histograms  $\mathbb{P}(\Delta X_3 | X_3)$  are plotted as points, where each color corresponds to a different value of  $X_3$ ; the black line is the Lévy-stable distribution with location 0 and scale 1. Inset: BIC difference with respect to the S model [41].

show a strong dependence in  $t$  in the first week. In particular,  $\mu_t$  decays in this period (reflecting a decay in the gain of views) and  $\alpha_t \approx 1.75$  for  $t > 5$ . It is worth to be noted is that the value of  $\mu_t$  becomes negative; while apparently in contradiction with the positive slope of  $\langle \Delta X_t | X_t \rangle$  [see Fig. 2(a)], it has to be recalled that the distribution is truncated at  $\Delta X_t = 0$ . The values of the averages from the data can be recovered through an exact, numerical computation.

If wanted, a model of the temporal dependence of  $\alpha_t, \mu_t, a_t, b_t$  can be introduced; in that case, it is possible to sum the likelihoods in Eq. (3) over  $t$  and therefore to reduce the number of parameter of the models by avoiding independent fittings for each  $t$ .

Altogether, these analysis support our proposal of stochastic differential equation with Lévy noise, Eq. (2), to describe the dynamics of popularity in YouTube.

## VI. PREDICTION OF BIG HITS

We now focus on the estimation of the probability of an item becoming a big hit after a given time. We define as a big hit at

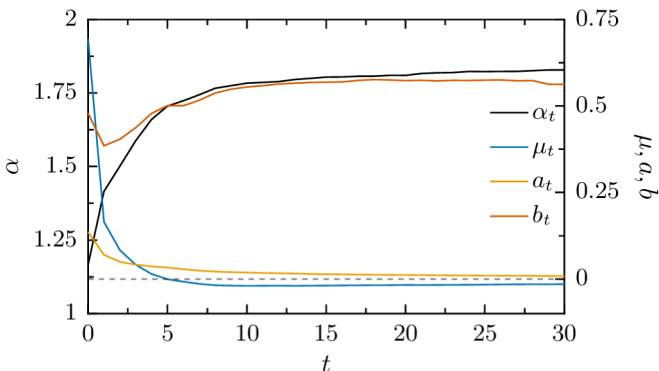


FIG. 4. Evolution of the parameters for the S model, Eq. (2), in the first 30 days after the release.

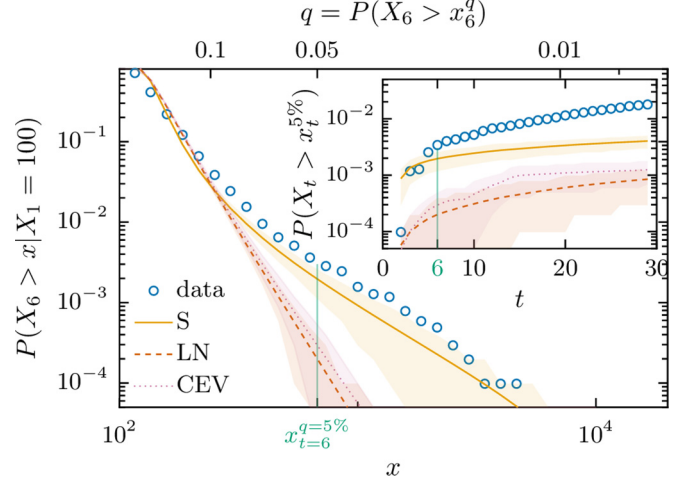


FIG. 5. Probability of videos becoming a big hit. Performance of the models evolved in time with respect to data; the selected videos had 100 views one day after their publication. Main panel: amount of videos that exceed a threshold  $x$  at  $t = 6$ . In the top axis, the quantiles  $q$  are indicated. Inset: amount of videos that enter into the 5% most viewed. Shaded areas: 95% confidence intervals, by bootstrapping.

time  $t$  the top  $q\%$  videos with highest  $X_t$  ( $X_t > x_t^q$ ). We are particularly interested in estimating the probability  $P(X_t > x_t^q | X_{t_0} = x_0)$  of videos that are not big hits at time  $t_0 < t$  (i.e.,  $x_0 < x_{t_0}^q$ ) becoming big hits at time  $t$ . This probability quantifies how unpredictable the system is. For instance, in a deterministic (proportional growth) model, the rank of the videos does not change and therefore such probability is zero. A positive probability is thus a measure of the deviation of such perfect predictability.

As an example, we select the videos that had 100 views one day after publication,  $X_1 = 100$ , which belong to a rank of  $q \approx 15\%$ . We are interested in the probability of these videos having  $X_t \gg 100$  at  $t > 1$ . To obtain the expectations of the models, we computed  $P(X_t | X_1 = 100)$  iteratively from  $P(dX_s | X_s)$  for  $s = 1, \dots, t$ , using  $X_1 = 100$  and the  $t$ -dependent parameters estimated in the previous section. The results shown in Fig. 5 (main panel) for  $t = 6$  confirm that the Lévy-stable model predicts a substantially higher probability for large  $X_t$  than alternative models. In order to investigate the temporal dependence, we focus on the probability of the videos improving their rank and being by day  $t$  in the top  $q = 5\%$ , using the previously computed probabilities from the models and the thresholds  $x_t^q$  estimated from data. The results are summarized in Fig. 5 (inset) and show that the Lévy-stable model succeeds in estimating this probability in the short term, while for the long term the data shows an even higher probability (mixing of ranks). The other models assign a video a substantially lower possibility of becoming a big hit, an effect of their highly predictable dynamics. The fact that our model provides a good account for short-time intervals but not in the long run suggests the existence of correlations in the attribution of views that span multiple days and that are not accounted for by our assumption of an independent noise.

## VII. DISCUSSION AND CONCLUSION

Our finding that the growth of views in YouTube is governed by both linear proportional growth and Lévy fluctuations has important consequences for the mathematical modeling of complex systems. First, it shows that, even if proportional growth is present, it cannot be attributed as responsible for the origin of the heavy tails because this is a feature already present in the fluctuations. Second, the use of Gaussian-based stochastic equations, such as Eq. (1) or traditional Fokker-Planck equations, overestimate the predictability of videos, by neglecting the mobility of popularity. We showed that better results are obtained in YouTube using a stochastic equation with Lévy noise, Eq. (2), an approach that has been previously used in physics [42], climate research [47], and finance [8]. Our work indicates that this formalism, and possibly also kinetic equations of the fractional type [48,49], should be considered in problems involving the dynamics of social media items and, more generally, in models of the economy of attention.

Our results bring new insights on the attention economy of the Internet. The fact that the multiple factors affecting the popularity of videos can be effectively modeled by a Lévy-stable distribution shows that the decisions of different individuals are correlated to each other and lead themselves to strong fluctuations. The Lévy-stable distribution is invariant under convolution, i.e., if  $X_1, X_2$  are stable, also  $X_1 + X_2$  is stable, and therefore it may naturally appear when multiple processes with diverging moments are combined (e.g., bursty activity patterns that characterize online social media). One challenge for future work is to identify mechanistic models of the spreading of information on the Internet (e.g., models in which viral items spread through a social network) that are compatible with these fluctuations [23,24,27]. The presented analysis of fluctuations is enabled by the large availability of data in YouTube videos and we expect similar results to hold also in more general systems in which items compete for the attention of users.

## ACKNOWLEDGMENTS

We thank I. Sokolov for helpful discussions.

## APPENDIX A: MODELS

We compare the data collected with the distribution predicted by a series of simple models: from Eq. (1), we derive the log-normal (LN) and constant elasticity of variance (CEV) models; from Eq. (2) we derive the Lévy-stable model (S). To compare with data, we compute the distributions of  $X_{t+1}$  of the different models. Note that for the LN and CEV models, these distributions are the result of integrating Eq. (1) over a period of one day, while for the S models, this integration is not performed, i.e., we assume that in the period of one day the distribution of  $X_{t+1}$  is essentially the one of the noise.

### 1. Log-normal (LN)

The LN model is defined by considering a linear scaling of the noise term in Eq. (1)

$$dX_t = \mu_t X_t dt + \sigma_t X_t dW_t. \quad (\text{A1})$$

We integrate this equation for a time equal to one day (where we consider  $\mu_t$  and  $\sigma_t$  constant), such that  $X_{t+1}$  is distributed log-normally, with a probability density function

$$\begin{aligned} \mathbb{P}(X_{t+1} = x | X_t = x_0) \\ = \frac{\exp \left\{ - \left[ \log x - (\log x_0 + \mu_t - \sigma_t^2/2) \right]^2 / (2\sigma_t^2) \right\}}{\sqrt{2\pi} \sigma_t x}. \end{aligned} \quad (\text{A2})$$

$\Delta X_{t+1} = X_{t+1} - X_t$  is distributed also log-normally, since  $X_t$  is fixed, but a truncation at 0 is necessary. Since the data is distributed on the natural numbers, we discretize as well the distribution, normalizing by the sum of the PDF over its new domain.

### 2. Constant elasticity of variance (CEV)

If instead of Eq. (A1) the equation

$$dX_t = \mu_t X_t dt + \sigma_t X_t^\beta dW_t \quad (\text{A3})$$

is used, we have to use the distribution of the constant elasticity of variance process (CEV), described in Ref. [38]. When  $\beta < 1$ , it has the form

$$\begin{aligned} \mathbb{P}(X_{t+1} = x | X_t = x_0) \\ = 2(1 - \beta) k^{\frac{1}{2(1-\beta)}} (x z)^{\frac{1}{2(1-\beta)}} e^{-x-z} I_{\frac{1}{2(1-\beta)}}(2\sqrt{xz}) \end{aligned} \quad (\text{A4})$$

with

$$\begin{aligned} k &= \frac{\mu}{\sigma^2(1 - \beta)(e^{2\mu(1-\beta)} - 1)} \\ x &= k(x_0 e^\mu)^{2(1-\beta)} \\ z &= kx^{2(1-\beta)}, \end{aligned}$$

where  $I$  is the modified Bessel function of the first kind. The expression simplifies using the substitution  $p = 2(1 - \beta)$ :

$$\mathbb{P}(X_{t+1} = x | X_t = x_0) = p k^{\frac{1}{p}} (x z^{2p-3})^{\frac{1}{2p}} e^{-x-z} I_{\frac{1}{p}}(2\sqrt{xz}) \quad (\text{A5})$$

with

$$\begin{aligned} k &= \frac{2\mu}{\sigma^2 p (e^{\mu p} - 1)} \\ x &= k x_0^p e^{\mu p} \\ z &= k x^p. \end{aligned}$$

When  $\beta > 1$ , the distribution is the same as above but multiplied by  $-1$ . Note that the  $\beta$  parameter is the exponent of the power-law tail that the distribution has asymptotically. Here we also subtract  $X_t$  to obtain a distribution of  $\Delta X_t$ , which we also truncate, discretize, and normalize.

### 3. Lévy-stable (S)

In the S model, defined in Eq. (2),  $dX_t$  is Lévy-stable distributed with location parameter  $m = \mu_t x_0$ , scale parameter  $s = a_t x_0 + b_t$ , asymmetry  $\beta_L = 1$  and its tail decays as an  $\alpha$  power of  $dX_t$  [50]. These parameters correspond to the parametrization 1 of Ref. [43], where the characteristic function of  $dX$  (there is no explicit form of the Lévy probability

distribution function),  $\phi_{dX}(k)$  is given by

$$\log \phi_{dX}(k) = \begin{cases} imk - s^\alpha |k|^\alpha [1 + i\beta_L \tan(\frac{\pi\alpha}{2}) \text{sign}(k)] & \alpha \neq 1 \\ imk - s|k| [1 + i\beta_L \frac{2}{\pi} \text{sign}(k) \log(|k|)] & \alpha = 1. \end{cases} \quad (\text{A6})$$

We consider  $\Delta X_t \approx dX_t$ , and the parameters absorb the dependence on  $\Delta t$ . In order to get the distribution  $\mathbb{P}(X_{t+1} = x|X_t = x_0)$ , the characteristic function has to be transformed to the real space, translated on the  $X_{t+1}$  axis by an amount  $x_0$ , and then truncated at  $X_{t+1} = x_0$ , discretized, and normalized.

Numerically, the Lévy distribution is computed as:

(i) the characteristic function (its Fourier transform) is inverted numerically on a grid in  $\alpha \in (0.5, 2)$  and  $\beta \in (0, 1)$  (with a resolution of 0.05, values of  $\alpha$  below 0.5 are very unlikely and for  $\beta < 0$  the distribution can be computed from the one of  $-\beta$  using symmetry);

(ii) for general values  $\alpha, \beta$ , we compute the distribution as an interpolation of the values on the grid (using the Catmull-Rom cubic splines);

(iii) the numerical integration often becomes unstable in the tails of the distribution (large  $x$ ). In order to avoid this problem, we use the power-law approximation described in Ref. [33] to describe the distribution beyond a threshold.

We provide the code of this procedure in the package PYLEVY [51]. It contains routines to compute the PDF of the Lévy distribution and to fit it.

## APPENDIX B: FLUCTUATIONS EXPECTED FROM EXISTING MODELS

Here we discuss the form that  $\mathbb{P}(\Delta X_t|X_t)$  has for the classical linear proportional growth models. There are basically two schemes of implementing linear proportional growth in order to get heavy-tailed distributions.

(i) Scheme 1: Champernowne [52] introduces a lower positive limit to  $X$ . A master equation is defined to regulate the transitions to different states (amount of views), which eventually leads to a stationary distribution with a power-law tail. This argument was formalized and popularized in Refs. [16,53], using a linear stochastic differential equation (as geometric Brownian motion, GBM), which, in the limit of long time, converges to a heavy-tailed distribution; note that in this scheme, all items start with the same initial condition.

(ii) Scheme 2: Yule and Simon [10,17] design a scheme where views are added to different items while at the same time

new items are introduced, resulting in a power-law distribution. This is basically the model known as preferential attachment [20] in the context of network growth. Here the items start in different conditions, since the system is growing in the number of items, hence the first ones are privileged.

We focus on the transition probability  $\mathbb{P}(\Delta X_t|X_t)$ . Scheme 1 (GBM) is a mechanism that leads to a heavy-tailed distribution asymptotically, but also relies on the possibility of negative growth rates  $dX_t$ , which is not realistic in the context of videos' views, and more generally in the context of cumulative allocation of attention. For an infinitesimal increase of time ( $\Delta t \approx dt$ ), the distribution is normal, but if a finite time interval is considered ( $\Delta t \gg dt$ ), the integration of the model results in a log-normal distribution.

The Scheme 2, instead, is fundamentally different and can be thought of as a Pólya's urn process where at a given time  $t_0$ , a number of views  $N$  is assigned to a set of videos  $M$  that has exactly  $x_0$  views already. The probability of assigning a view to a particular video is, of course, proportional to the amount of views each video has. The distribution  $\mathbb{P}(\Delta X_t|X_t)$  of views among the videos is of the beta binomial type, with means

$$\mathbb{E}_{BB}(\Delta X_t|X_t = x_0) = \frac{N}{M} \quad (\text{B1})$$

and variance

$$\mathbb{V}_{BB}(\Delta X_t|X_t = x_0) = \frac{N(M-1)(Mx_0 + N)}{M^2(Mx_0 + 1)} \quad (\text{B2})$$

and can be roughly approximated by a normal distribution with same mean and variance. The variance scales in two regimes because of the  $Mx_0 + N$  term: when  $x_0 \lesssim N/M$ ,  $\sigma \propto x_0^{-1/2} N/M$ , and when  $x_0 \gtrsim N/M$ ,  $\sigma \propto \sqrt{N/M}$ . Notice though, that the amount of views allocated,  $N$ , is not independent. In fact, since growth is linear, we expect  $N \propto x_0 M$ , so we have in this particular case  $\sigma \propto x_0^{1/2}$ .

In conclusion, for Scheme 1, we expect normal or log-normal distributions of  $\Delta X_t$ , depending on the choice of whether integrating over a finite time or not, while for Scheme 2 we expect approximately normal distributions, with variance scaling as  $X_t$ .

[1] H. A. Simon, in *Computers, Communication, and the Public Interest* (Johns Hopkins Press, Baltimore, 1971), pp 47–72.  
 [2] F. Wu and B. A. Huberman, *Proc. Natl. Acad. Sci. USA* **104**, 17599 (2007).  
 [3] M. J. Salganik, P. S. Dodds, and D. J. Watts, *Science* **311**, 854 (2006).  
 [4] B. A. Huberman, D. M. Romero, and F. Wu, *J. Information Sci.* **35**, 758 (2009).  
 [5] J. M. Miotto and E. G. Altmann, *PLoS ONE* **9**, e111506 (2014).

[6] L. Weng, A. Flammini, A. Vespignani, and F. Menczer, *Sci. Rep.* **2**, 335 (2012).  
 [7] V. Pareto, *Cours d'Économie Politique* (Librairie Droz, Paris, 1964).  
 [8] B. Mandelbrot, *J. Bus.* **36**, 394 (1963).  
 [9] A. Clauset, C. R. Shalizi, and M. E. J. Newman, *SIAM Rev. Soc. Ind. Appl. Math.* **51**, 661 (2009).  
 [10] G. U. Yule, *Philos. Trans. R. Soc. London B* **213**, 21 (1925).

- [11] R. Gibrat, *Les inégalités économiques* (Recueil Sirey, Paris, 1931).
- [12] B. Mandelbrot, *Int. Econ. Rev.* **1**, 79 (1960).
- [13] H. A. Simon and C. P. Bonini, *Am. Econ. Rev.* **48**, 607 (1958).
- [14] M. V. Simkin and V. P. Roychowdhury, *Phys. Rep.* **502**, 1 (2011).
- [15] M. Perc, *J. R. Soc. Interface* **11**, 20140378 (2014).
- [16] X. Gabaix, *Q. J. Econ.* **114**, 739 (1999).
- [17] H. A. Simon, *Biometrika* **42**, 425 (1955).
- [18] M. Gerlach and E. G. Altmann, *Phys. Rev. X* **3**, 021006 (2013).
- [19] T. Kuhn, M. Perc, and D. Helbing, *Phys. Rev. X* **4**, 041036 (2014).
- [20] A.-L. Barabási and R. Albert, *Science* **286**, 509 (1999).
- [21] D. Wang, C. Song, and A.-L. Barabási, *Science* **342**, 127 (2013).
- [22] D. J. Watts, *Everything Is Obvious: \*Once You Know the Answer* (Crown Publishing Group, New York, 2011).
- [23] T. Maillart, D. Sornette, S. Spaeth, and G. von Krogh, *Phys. Rev. Lett.* **101**, 218701 (2008).
- [24] J. Ratkiewicz, S. Fortunato, A. Flammini, F. Menczer, and A. Vespignani, *Phys. Rev. Lett.* **105**, 158701 (2010).
- [25] J. M. Miotto and E. G. Altmann, Time series of social media activity. Youtube, Usenet, Stack-Overflow, PLoS ONE., Accessed 2015 May 21. Available at <http://dx.doi.org/10.6084/m9.figshare.1160515>.
- [26] D. Sornette, F. Deschâtres, T. Gilbert, and Y. Ageon, *Phys. Rev. Lett.* **93**, 228701 (2004).
- [27] D. J. Watts, *Proc. Natl. Acad. Sci. USA* **99**, 5766 (2002).
- [28] G. Chatzopoulou, C. Sheng, and M. Faloutsos, in *INFOCOM IEEE Conference on Computer Communications Workshops, 2010* (IEEE, Piscataway, 2010), pp. 1–6.
- [29] R. Crane and D. Sornette, *Proc. Natl. Acad. Sci. USA* **105**, 15649 (2008).
- [30] F. Ghanbarnejad, M. Gerlach, J. M. Miotto, and E. G. Altmann, *J. R. Soc. Interface* **11**, 20141044 (2014).
- [31] H. Risken, *Fokker-Planck Equation* (Springer, Berlin, 1984).
- [32] A. Mollgaard and J. Mathiesen, *PLoS ONE* **10**, e0123876 (2015).
- [33] M. E. J. Newman, *Phys. Rev. E* **64**, 025102 (2001).
- [34] Z. Eisler, I. Bartos, and J. Kertesz, *Adv. Phys.* **57**, 89 (2008).
- [35] R. Friedrich and J. Peinke, *Phys. Rev. Lett.* **78**, 863 (1997).
- [36] L. A. N. Amaral, S. V. Buldyrev, S. Havlin, M. A. Salinger, and H. E. Stanley, *Phys. Rev. Lett.* **80**, 1385 (1998).
- [37] J.-P. Bouchaud and A. Georges, *Phys. Rep.* **195**, 127 (1990).
- [38] R.-R. Chen and C.-F. Lee, in *Handbook of Quantitative Finance and Risk Management* (Springer, Berlin, 2010).
- [39] V. M. Zolotarev, *One-Dimensional Stable Distributions* (American Mathematical Society, Providence, RI, 1986).
- [40] G. Schwarz, *Ann. Stat.* **6**, 461 (1978).
- [41] We compare the models with the Bayes information criterion (BIC [40]), a way of penalizing the increase of likelihood due to the addition of parameters:  $\text{BIC} = -2 \ln \mathcal{L} + k \ln N$ , where  $\mathcal{L}$  is the likelihood in Eq. (3),  $N$  is the number of data points, and  $k$  is the number of parameters.
- [42] A. Janicki and A. Weron, *Simulation and chaotic behavior of alpha-stable stochastic processes* (CRC Press, Boca Raton, 1993).
- [43] J. P. Nolan, *Stable distributions* (Birkhauser, Boston, to be published 2017).
- [44] B. Øksendal, *Stochastic differential equations* (Springer, Berlin, 2003).
- [45] This is performed numerically by minimizing  $-\log \mathcal{L}$  through the Nelder-Mead algorithm, implemented in the Python package `scipy` as `optimize.fmin`. We report the minimum value of 50 repetitions performed with different (random) initial conditions (to avoid local minima).
- [46] There are two problems with the application of maximum-likelihood estimations to  $\mathbb{P}(dX_t | X_t)$  for specific  $X_t$ : (i) The estimation is quite sensitive to fluctuations, often failing to determine the tail exponent if no threshold is set [9] (many histograms with similar tails lead to very different values of  $\alpha$ ). (ii) Once the parameters of the fits for each  $\mathbb{P}(dX_t | X_t)$  are extracted, it is not clear how to obtain the best parametrization of them, because of the presence of Lévy-distributed errors; the main limitations here are the truncation of fluctuations at  $\Delta X_t = 0$  (not to be confused with the lower positive limit on  $X_t$  of Scheme 1 described in App. A), and the discretization of the  $\Delta X_t$  variable, which prevents a fit on the rescaled variable  $z = (\Delta X_t - \mu(X_t))/\sigma(X_t)$ .
- [47] P. D. Ditlevsen, *Geophys. Res. Lett.* **26**, 1441 (1999).
- [48] R. Metzler, E. Barkai, and J. Klafter, *Phys. Rev. Lett.* **82**, 3563 (1999).
- [49] D. Brockmann and I. M. Sokolov, *Chem. Phys.* **284**, 409 (2002).
- [50] In our exploratory analysis we considered also two variants of the S-model in Eq. (2): one where  $b_t = 0$  and one where the deterministic part is  $\mu_t X_t + c_t$ . A BIC analysis shows that the quality decreases significantly in the first case, but only marginally in the second case.
- [51] J. M. Miotto, *PYLEVY*, 2016.
- [52] D. G. Champernowne, *Econ. J.* **63**, 318 (1953).
- [53] H. Kesten, *Acta Math.* **131**, 207 (1973).