Mathematical Models of Fads Explain the Temporal Dynamics of Internet Memes

Christian Bauckhage
B-IT, University of Bonn
Bonn, Germany

Kristian Kersting
IGG, University of Bonn
Bonn, Germany

Fabian Hadiji
Fraunhofer IAIS
St.Augustin, Germany

Abstract

Internet memes are a pervasive phenomenon on the social Web. They typically consist of viral catch phrases, images, or videos that spread through instant messaging, (micro) blogs, forums, and social networking sites. Due to their popularity and proliferation, Internet memes attract interest in areas as diverse as marketing, sociology, or computer science and have been dubbed a new form of communication or artistic expression. In this paper, we examine the merits of such claims and analyze how collective attention into Internet memes evolves over time. We introduce and discuss statistical models of the dynamics of fads and fit them to meme related time series obtained from Google Trends. Given data as to more than 200 memes, we find that our models provide more accurate descriptions of the dynamics of growth and decline of collective attention to individual Internet memes than previous approaches from the literature. In short, our results suggest that Internet memes are nothing but fads.

Introduction

Internet memes are catch phrases or humorous or repugnant pictures or video clips that "go viral" among Internet users.

While the phenomenon of viral content can be traced back to the early days of the Web, it is because of the interactive and participatory nature of modern social media such as blogs, wikis, or social networking sites that Internet memes have become a staple of contemporary Web culture. They typically originate from platforms like *4chan*, *tumblr*, or *youtube*, gain notoriety via social news and entertainment sites such as *reddit*, *failblog*, or *memegenerator*, and then spread through the social Web at large (Bauckhage 2011).

Due to their popularity and proliferation, Internet memes are beginning to get noticed in traditional media (Pogue 2011) and websites such as such as *knowyourmeme* or *memebase* promote them as a form of artistic expression. In addition, public relation professionals are hitchhiking the trend, trying to design Internet memes for viral marketing or political campaigning (Burgess 2008). Consequently, Internet memes increasingly attract academic interest (Bernstein et al. 2011; Thom and Millen 2012). Yet, many aspects of the phenomenon are still poorly understood.

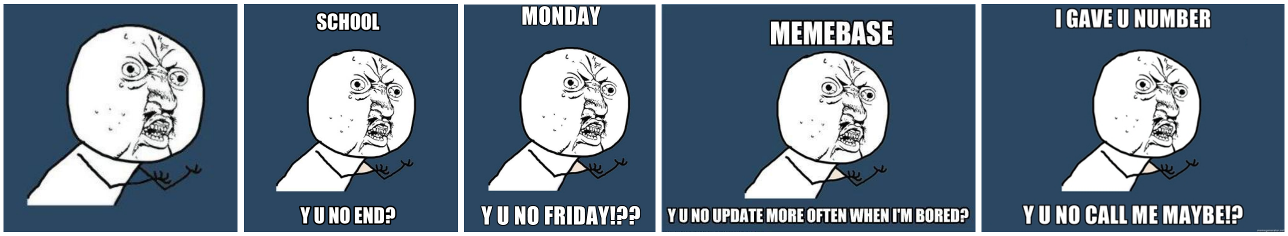
With the work reported here, we attempt to contribute to a better understanding of the nature of Internet memes. In particular, we address aspects of *meme dynamics*.

Internet memes are *dynamic media objects* that evolve through commentary or parody. Consider, for example, the "y u no" meme shown in Fig. 1. It first appeared on *tumblr* in 2010 and quickly found its way to *memegenerator* from where it spread virally. In its basic form, the meme consists of an image of a stick figure whose angry face was copied from the Japanese anime series Gantz. It typically contains a text in short messaging style that poses mundane questions as to modern life and culture (Fig. 1(b)). Mutations include self-referential variants that allude to *meme culture* (Fig. 1(c)) as well as versions that deviate from the original phenotype (Fig. 1(d)). Also, the meme occasionally occurs in media outside of the Internet but is then reported back on the Web, for instance on social networking sites (Fig. 1(e)). Internet memes therefore transgress media and cultural boundaries and can be characterized as inside jokes that many people are in on.

In addition to their content dynamics, Internet memes also show characteristic properties regarding their life cycles. While some were observed to go in and out of popularity in just a matter of weeks, others attract collective attention for extended periods of time. This is exemplified in Fig. 2 which shows meme related time series retrieved from Google Trends. The graphs indicate how worldwide interest in individual memes (measured in terms of relative search frequencies) grew and declined over time. Although details of these time series appear chaotic, there are characteristic *general trends*. After a point of onset, public interest in a meme grows explosively but once a meme has reached peak popularity, interest begins to fade more or less quickly.

Interestingly, the temporal dynamics of meme related search frequencies as in Fig. 2 resemble those of epidemic outbreaks. We noted this in (Bauckhage 2011) where we investigated meme dynamics from the perspective of epidemic modeling. While we found epidemic models to provide reasonable accounts of trends in meme related time series, the Log-Normal distribution gave more accurate descriptions. Alas, this was a purely empirical observation and we had no plausible explanation as to possible causes for this finding.

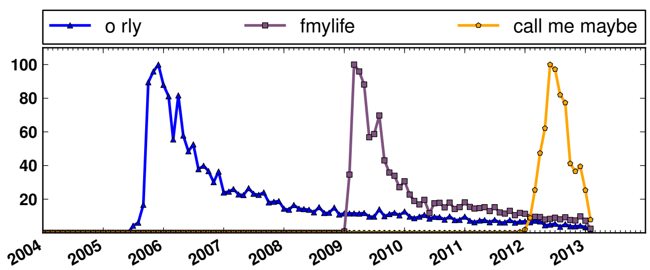
Our goal in this paper is thus to develop meaningful and interpretable models of how collective attention to Internet



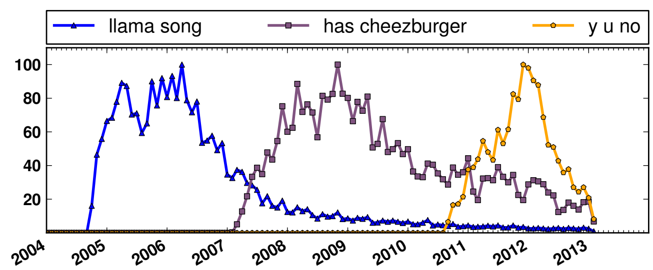
(a) the "y u no" guy (b) instances of the "y u no" meme (c) the "y u no" meme with references to the website memebase.com and the "call me maybe" meme

(d) mutations of the "y u no" meme alluding to items of popular culture such as an anime movie or a game franchise (e) the "y u no" meme appearing on a birthday cake and on the front page of a printed magazine

Figure 1: An example of a popular Internet meme. Instances of the "y u no" meme consist of a simple image macro and a grammatically carefree piece of text that calls to attention questions of everyday life and contemporary culture. The meme first appeared on tumblr.com in 2010; as of this writing, querying Google for "y u no guy" yields more than 8,000,000 results.



(a) positively skewed, narrow peaks, rather long tails



(b) positively skewed, broad peaks, rather short tails

Figure 2: Examples of prototypic, meme related time series retrieved from Google Trends. While details are chaotic, there appear to be distinct patterns or *global trends* as to how collective interest in individual Internet memes grows and declines.

memes evolves over time. The empirical basis of our study consists of time series of meme related search frequencies available from Google Trends. With respect to such data, we ask: 1) What kind of social dynamics would explain the emergence of skewed time series of meme related outbreak data? 2) Which statistical distributions provide reasonable models of the corresponding mechanisms? 3) How well do theoretically plausible models of the evolution of collective attention to individual Internet memes fit empirical data?

Addressing these questions, we present simple models that describe the temporal dynamics of *fads*. We show that well established statistical distributions, namely the Weibull, the Gompertz, and the Frechet, are particular instances of these models. Given more than 200 meme related time series, we fit these distributions and compare their performance to that of the Log-Normal. We find that our models yield more accurate characterizations of the evolution of interest in Internet memes than the Log-Normal model. In short, our results suggest that Internet memes are nothing but subcultural fads. Finally, because of their rigor, simplicity, and empirical validity, we believe that the models proposed in this paper are applicable beyond the study of Internet memes. We present corresponding anecdotal evidence and analyze the dynamics of technology trends.

Our presentation proceeds as follows: First, we review existing work on analyzing Google Trends time series and argue that data from Google Trends provide a useful proxy for the study of collective attention mechanisms on the Web. Then, we discuss mathematical models of growth and decline dynamics and relate them to statistical distributions. We fit these models to meme time series, discuss our results, and review the related literature. Finally, we summarize our questions, methodology, and results and provide an outlook to promising future research.

A Proxy for Collective Attention on the Web

The work reported in this paper aims at plausible models of how collective attention to Internet memes evolves over time. The empirical basis of our analysis are time series obtained from Google Trends (formerly Google Insights for Search) which summarize meme related search behaviors of millions of Web users worldwide.

Google Trends is a service that provides statistics on queries users have entered into the Google search engine. It allows for retrieving weekly summaries of how frequently a query has been used in different regions of the world since January 1st 2004. Instead of revealing total search counts, Google Trends normalizes the data such that the peak search activity for a query corresponds to a value of 100. Data obtained from Google Trends therefore only indicates relative search frequencies rather than absolute public interest. Nevertheless, analyzing the evolution of topic specific searches is an increasingly popular approach in studies on collective preferences (Granka 2009) and in economic forecasting (Choi and Varian 2012; Da, Engelberg, and Gao 2011; Joseph, Wintoki, and Zhang 2011).

Questions as to the validity of this methodology and the significance of Internet search data have been addressed in two recent contributions: Mellon (2011) correlated Gallup surveys and Google Trends data and found that, w.r.t. political and economic issues, search frequency data provide accurate proxies of the dynamics of salient public opinions. Teevan et al. (2011) studied how people navigate the Web and found that over 25% of all queries to search engines are *navigational queries* intended to find and then access particular Web resources. A large percentage of Internet users therefore relies on Google searches rather than on bookmarks or on entering URLs in order to navigate to Web sites.

With respect to our work, both these findings suggest that data from Google Trends which aggregate information about the activity of millions of users is indeed indicative of collective interests and attention and thus forms suitable proxies for research on the dynamics of Internet memes.

Modeling the Dynamics of Fads

In this section, we present different statistical distributions that account for *general trends* or *global shapes* of skewed search frequencies such as shown in Fig. 2. In contrast to previous work on meme dynamics, our models are derived from basic principles and admit interpretations in terms of common sense or simple social mechanisms.

Our reasoning is based on properties of the dynamics of *fads*. Fads are forms of behavior related to ideas, activities, or products that are enthusiastically followed by large populations for a period of time, basically because the respective concepts are perceived as being novel. When a fad "catches on", the number of people adopting it grows rapidly. However, once the perception of novelty is gone, the behavior will fade again (Meyerson and Katz 1957).

This characterization of the dynamics of fads matches the global behavior of time series of meme related search activities as shown in Fig. 2. Moreover, dynamics like this can be modeled using growth equations (Zeide 1993).

A growth equation models how entities grow and decline over time. The corresponding function $f(t)$ is assumed to be composed of a protagonistic term $g(t)$ that represents the propensity of an entity to grow and an antagonistic term $d(t)$ that represents a propensity to decline. In their most basic form, growth equations consider two mechanisms of how $g(t)$ and $d(t)$ are coupled:

$$\text{subtraction:} \quad f(t) = g(t) - d(t) \tag{1}$$

$$\text{division:} \quad f(t) = g(t)/d(t). \tag{2}$$

In either case, $g(t)$ as well as $d(t)$ are supposed to grow monotonously. Both variants therefore describe processes where $f(t)$ will itself be growing as long as $g(t)$ grows faster than $d(t)$; yet, once the propensity to decline becomes dominant, $f(t)$ will begin to decline. Next, we discuss statistical distributions that mimic this behavior.

Statistical Models of Growth and Decline

Statistical model fitting is an exercise in trading off precision and generality. Given a collection of data, it is always possible to determine models of suitably many parameters that fit the data exactly. However, such models will likely over-fit the data and therefore not allow for general conclusions. Instead, models of only a few parameters may not exactly match the data but capture its gist. As our focus here is on *general trends* in search frequency data, we confine our following discussion to well established two-parameter life-time distributions (Lawless 2003).

The *Weibull distribution* plays an important role in life-time analysis of biological and mechanical systems (Rinne 2008). Its probability density function (pdf) is defined for $t \in [0, \infty)$ and is given by

$$f_{\mathcal{WB}}(t; \kappa, \lambda) = \frac{\kappa}{\lambda} \left(\frac{t}{\lambda}\right)^{\kappa-1} e^{-(t/\lambda)^{\kappa}} \tag{3}$$

so that its cumulative density function (cdf) amounts to

$$\begin{aligned} F_{\mathcal{WB}}(t; \kappa, \lambda) &= \int_0^t f_{\mathcal{WB}}(\tau; \kappa, \lambda)\, d\tau \\ &= 1 - e^{-(t/\lambda)^{\kappa}} \end{aligned} \tag{4}$$

where the parameters κ and λ determine shape and scale of the distribution.

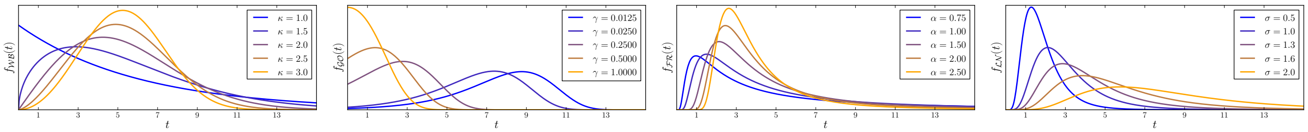
The Weibull is the type III extreme value distribution. It is typically skewed to the right, rather short tailed, and, depending on the choice of parameters, can assume various shapes and forms (see Fig. 3(a)). Note that for $\kappa = 1$, the Weibull coincides with the Exponential distribution while for $\kappa \approx 3.5$, it approaches the Normal distribution.

The *Gompertz distribution* is frequently used as a model of mortality in demographic or actuary studies (Kleiber and Kotz 2003). Its pdf is defined for $t \in [0, \infty)$ and amounts to

$$f_{\mathcal{GO}}(t; \eta, \gamma) = \gamma \eta e^{\gamma t} e^{-\eta(e^{\gamma t}-1)} \tag{5}$$

where η and γ are shape and scale parameters, respectively. The corresponding cdf is given by

$$F_{\mathcal{GO}}(t; \eta, \gamma) = 1 - e^{-\eta(e^{\gamma t}-1)}. \tag{6}$$



(a) Weibull pdfs ($\lambda = 6$) (b) Gompertz pdfs ($\eta = 0.5$) (c) Frechet pdfs ($\beta = 3$) (d) Log-Normal pdfs ($\mu = 0.5$)

Figure 3: Examples of possible shapes of the statistical distributions considered in this paper. Note that, depending on the choice of parameters, the Frechet and the Log-Normal distribution may look rather similar.

The Gompertz results from truncating the type I extreme value distribution at zero and is very versatile. Depending on the choice of its parameters it may be skewed to the left or to the right but it, too, is rather short tailed (see Fig.3(b)).

There is an interesting, indirect connection between the Weibull and the Gompertz distribution. Recall that if a random variable X is distributed according to $f_X(x)$, then the transformed random variable $Y = h(X)$ is distributed as

$$f_Y(y) = f_X\big(h^{-1}(y)\big) \left| \frac{d}{dy} h^{-1}(y) \right|. \tag{7}$$

Now, if X is Weibull distributed and $Y = \frac{\lambda^2}{X}$, then $f_Y(y)$ is inverse Weibull. At the same time, if X is Gompertz distributed and $Y = e^{-X}$, then $f_Y(y)$ is logarithmic Gompertz. The inverse Weibull and the logarithmic Gompertz are indeed identical and also known as the Frechet distribution.

The *Frechet distribution* is the type II extreme value distribution. Its pdf and cdf correspond to

$$f_{\mathcal{FR}}(t; \alpha, \beta) = \frac{\alpha}{\beta} \left(\frac{t}{\beta}\right)^{-\alpha-1} e^{-(t/\beta)^{-\alpha}} \tag{8}$$

and

$$F_{\mathcal{FR}}(t; \alpha, \beta) = e^{-(t/\beta)^{-\alpha}} \tag{9}$$

where α and β are shape and scale parameters.

Given what was said before, the Frechet can be seen as the Gompertz on a logarithmic scale. It can assume a large variety of shapes and has a long right tail (see Fig. 3(c)).

Finally, for baseline comparison to the previous literature, we also consider the *Log-Normal distribution* which is frequently observed in nature since it describes the outcome of multiplicative growth processes (Koch 1966). It models link distributions on the Web (Mitzenmacher 2004), amounts of popularity on social platforms (Wu and Huberman 2007), and was found to represent the shape of meme related time series (Bauckhage 2011). For $t > 0$, its pdf is given by

$$f_{\mathcal{LN}}(t; \mu, \sigma) = \frac{1}{t\sigma\sqrt{2\pi}}\, e^{-\frac{(\log t - \mu)^2}{2\sigma^2}} \tag{10}$$

where μ and σ are mean and standard deviation of $\log t$.

Just as the Frechet is a Log-Gompertz, the Log-Normal corresponds to a log-transformed Normal distribution and Fig. 3(d) indicates that it, too, may assume a variety of shapes. We note that, depending on parametrization, the Log-Normal may be easily confused with the Frechet. This observation is critical as it is likely the main reason why the Log-Normal was found to be a good model of time series that indicate the temporal dynamics of Internet memes.

Fad Representations of Growth Distributions

So far, we presented the Weibull-, the Gompertz-, and the Frechet distribution in their usual forms. Next, we show that these forms implicitly correspond to the simple variants of growth equations in (1) and (2). That is, we recast the usual forms of these three statistical distributions in terms of explicit growth and decline terms. For want of a better name, we refer to the results as *fad representations*.

In order to see how the Weibull distribution in (3) can be written as a growth function, we reduce notational clutter by substituting $b = (1/\lambda)^\kappa$ and $c = \kappa$ and prove the following

Lemma 1. *The probability density function of the Weibull distribution $f(t) = bct^{c-1}e^{-bt^c}$ can be written as*

$$f(t) = bct^{c-1} - bct^{c-1}F(t)$$

where $F(t)$ denotes the cumulative density up to time t.

Proof. The cumulative density of the Weibull is given by $F(t) = 1 - e^{-bt^c}$ so that $e^{-bt^c} = 1 - F(t)$. Thus

$$\begin{aligned} f(t) &= bct^{c-1}e^{-bt^c} \\ &= bct^{c-1}\big(1 - F(t)\big) \\ &= bct^{c-1} - bct^{c-1}F(t). \end{aligned}$$

□

We therefore find that the pdf of the Weibull distribution implicitly consists of a growth and a decline term. Growth and decline, are *polynomial* in t. Moreover, decline depends on $F(t)$ which grows monotonously. Therefore, the longer a process is running whose dynamics are characterized by a Weibull distribution, the quicker the decline.

In order to see that the Gompertz distribution in (5) can be written as a growth function, we re-parameterize $b = \eta$ and $c = \gamma$ for comparability to the previous result and prove

Lemma 2. *The probability density function of the Gompertz distribution $f(t) = bce^{ct}e^{-b(e^{ct}-1)}$ can be written as*

$$f(t) = bce^{ct} - bce^{ct}F(t)$$

where $F(t)$ denotes the cumulative density up to time t.

Proof. The cumulative density of the Gompertz is given by $F(t) = 1 - e^{-b(e^{ct}-1)}$ and hence $e^{-b(e^{ct}-1)} = 1 - F(t)$. This immediately leads to

$$\begin{aligned} f(t) &= bce^{ct}e^{-b(e^{ct}-1)} \\ &= bce^{ct}\big(1 - F(t)\big) \\ &= bce^{ct} - bce^{ct}F(t). \end{aligned}$$

□

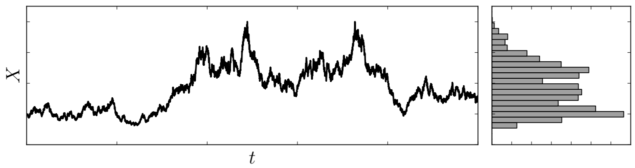


Figure 4: A random multiplicative process $X_t = \epsilon_t X_{t-1}$. While X is log-normally distributed, the series X_t is not.

Therefore, the pdf of the Gompertz distribution, too, implicitly contains a growth and a decline term and decline depends on $F(t)$. In contrast to the Weibull, the growth and decline terms of the Gompertz are *exponential* in t.

In both cases considered so far, the coupling between growth and decline corresponds to the model in (1). The Frechet distribution, on the other hand, implies the growth model in (2). Setting $b = (1/\beta)^{-\alpha}$ and $c = \alpha$, we prove

Lemma 3. *The probability density function of the Frechet distribution* $f(t) = bct^{-(c+1)}e^{-bt^{-c}}$ *can be written as*

$$f(t) = \frac{F(t)}{bct^{c+1}}$$

where $F(t)$ *denotes the cumulative density up to time* t.

Proof. The cumulative density of the Frechet is given by $F(t) = e^{-bt^{-c}}$. This immediately leads to

$$f(t) = bct^{-(c+1)}F(t) = \frac{F(t)}{bct^{c+1}}.$$

□

For the Frechet distribution, the tendencies for growth and decline are thus coupled through division. The growth term simply corresponds to $F(t)$, a monotonously growing, time-dependent function. The decline term is *polynomial* in t.

Implications

While each of the above distributions is skewed and may therefore fit general trends of time series as shown in Fig. 2, we just saw that the Weibull, Gompertz, and Frechet implicitly model growth and decline dynamics. The Log-Normal, on the other hand, cannot be expressed a growth equation.

This can be confusing, because Log-Normal distributions arise from random multiplicative processes where variables increase or decrease proportional to their previous values. These processes are governed by a time-dependent random variable ϵ_t such that $X_t = \epsilon_t X_{t-1}$. However, what is log-normally distributed in such a process is the value of X not the shape of the time series X_t (see Fig. 4).

In light of this, it appears implausible to consider the Log-Normal as a reasonable, i.e. interpretable, model of the global shape of meme related time series. However, if the Weibull, Gompertz, or Frechet would fit meme time series, this could be explained as the interplay of processes of growing and declining attention. Moreover, when written as a growth equation, each of these distributions expresses the amount of attention $f(t)$ a meme attracts at time t in terms of the overall interest $F(t)$ attracted so far. If these models were to fit meme related time series, this would mean, that Internet memes were fads whose attraction depends on their perceived novelty or, vice versa, on the amount of interest they have received so far.

Empirical Analysis

In this section, we analyze time series for a collection of 214 Internet memes listed in Tables 1 and 2.

Our data was gathered from Google Trends. We retrieved weekly summaries of worldwide, meme related English queries for the period from January 1st 2004 to February 8th 2013. Since Google only reveals search frequencies, the data neither allows for estimating absolute collective interest nor for inferring who was interested. Our data thus provides averaged, compartmentalized indicators as to how a meme's popularity develops over time and results obtained therefrom have to be understood as statistical expectations.

We compute monthly averages and obtain discrete time series $\boldsymbol{z} = [z_1, z_2, \ldots, z_{109}]$ where z_1 represents meme related activities in January 2004 and z_{109} represents frequencies for January 2013. Since not every meme in our sample was active during this whole period, we use CUSUM statistics to determine onset times t_o and obtain truncated time series $\boldsymbol{h} = [z_{t_o}, \ldots, z_{109}]$. Finally, we apply multinomial maximum likelihood (Jennrich and Moore 1975) to fit continuous Weibull ($f_{\mathcal{WB}}$), Gompertz ($f_{\mathcal{GO}}$), Frechet ($f_{\mathcal{FR}}$), and Log-Normal ($f_{\mathcal{LN}}$) distributions.

Since statistical tests such as the χ^2 test underestimate the quality of fits to time series (Gleser and Moore 1983), we use the Kullback-Leibler (KL) divergence

$$D_{KL}(\boldsymbol{h}|\boldsymbol{f}) = \sum_t h_t \log \frac{h_t}{f_t} \tag{11}$$

between empirical data $\boldsymbol{h}$ and model $\boldsymbol{f}$ sampled at times t to test goodness of fit. Since the KL divergence measures loss of information if $\boldsymbol{h}$ is represented by $\boldsymbol{f}$, a low divergence indicates a well fitting model.

To present our results, we split our data set of 214 meme time series into two disjoint subsets. The larger set (**set 1**) contains 204 time series (see Tab. 1) with a "complete" view of the process of growth and decline of collective attention. Search frequencies in this set are observed to rise, to peak, and to (begin to) decline during the observation period from January 2004 to February 2013. For statistical parameter estimation, this provides enough information to unambiguously determine which of our models gives the best fit.

The smaller set (**set 2**) contains 10 time series (see Tab. 2) which provide only "incomplete" views on evolving interest in a meme. This is because these memes appeared prior to 2004 so that their Google Trends time series are truncated from below. In other words, any evolution of attention to such a meme prior to 2004 is not visible in our data. From the point of view of statistical model fitting, conclusions as to goodness of fit to truncated time series are less reliable.

Note that in Tables 1 and 2, we have garbled a few memes (marked "XXX") because they are of controversial nature. They either are *gross out memes* which often center around

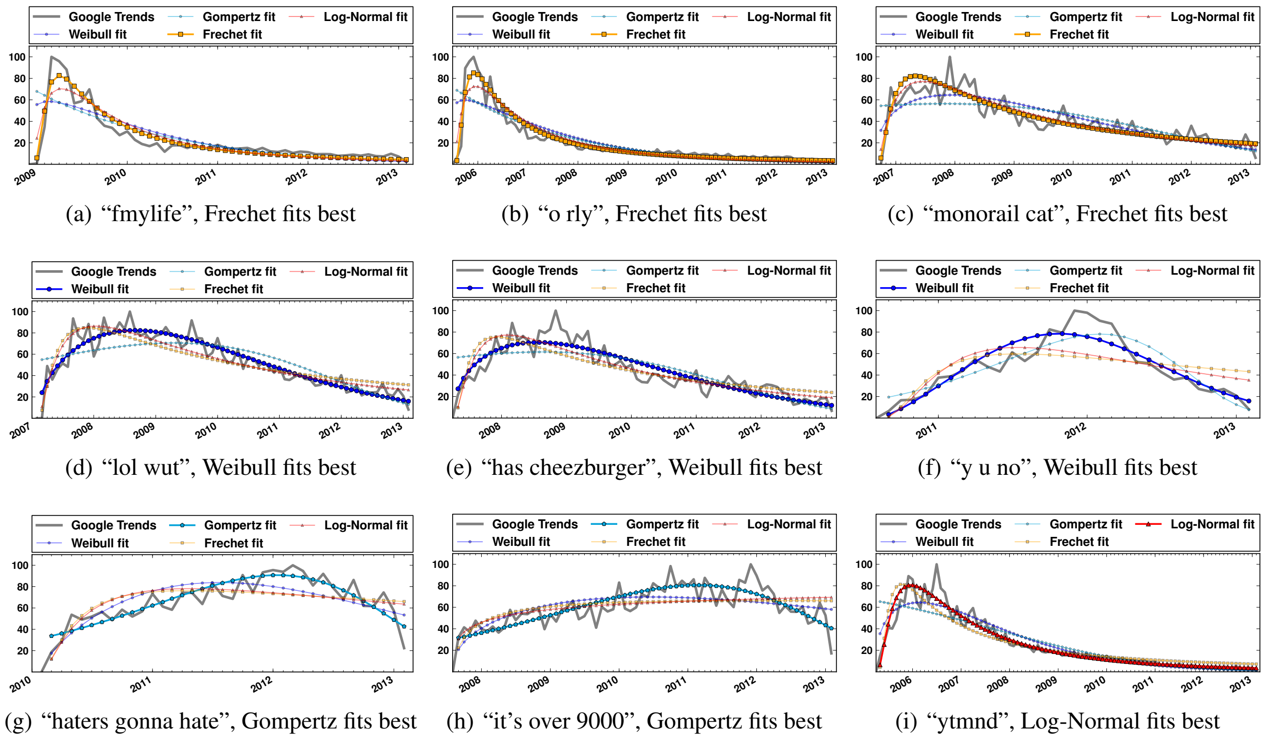
Table 1: 204 Internet memes (set 1) and best fitting models.

meme	onset	D_{KL} based ranking				meme	onset	D_{KL} based ranking				meme	onset	D_{KL} based ranking			
		1st	2nd	3rd	4th			1st	2nd	3rd	4th			1st	2nd	3rd	4th
XXX	2007	FR	LN	WB	GO	XXX	2010	WB	GO	LN	FR	noob tube	2009	FR	LN	WB	GO
3 wolf moon	2009	FR	LN	WB	GO	full of win	2009	WB	GO	LN	FR	not sure fry	2011	GO	WB	LN	FR
56 stars	2009	FR	LN	WB	GO	funtwo	2006	FR	LN	WB	GO	numa numa	2004	FR	LN	GO	WB
angry german kid	2006	FR	LN	WB	GO	fus ro dah	2011	FR	LN	WB	GO	nyan cat	2011	FR	LN	WB	GO
arrow in the knee	2011	FR	LN	WB	GO	gangnam style	2012	WB	LN	FR	GO	o rly	2005	FR	LN	WB	GO
art student owl	2011	FR	LN	WB	GO	giant enemy crab	2006	FR	WB	LN	GO	oh hai	2008	GO	WB	FR	LN
asian backstreet boys	2005	FR	LN	WB	GO	giga pudding	2010	FR	LN	WB	GO	ok go treadmill	2006	FR	LN	WB	GO
ask a ninja	2006	LN	FR	WB	GO	gimme pizza	2010	FR	LN	GO	WB	om nom nom	2007	WB	LN	GO	FR
awww yeah	2010	WB	GO	LN	FR	goatse.cx	2004	WB	LN	FR	GO	one red paper clip	2005	FR	LN	WB	GO
balloon boy	2009	FR	LN	WB	GO	good guy greg	2011	WB	LN	FR	GO	oolong rabbit	2004	FR	LN	WB	GO
bananaphone	2004	GO	FR	LN	WB	gordo granudo	2010	WB	GO	LN	FR	owling	2011	FR	LN	GO	WB
bed intruder	2010	FR	LN	WB	GO	gunther ding dong	2004	FR	LN	WB	GO	pants on the ground	2009	FR	LN	GO	WB
benny lava	2007	FR	LN	WB	GO	has cheezburger	2007	WB	LN	FR	GO	peanut butter jelly song	2005	FR	LN	GO	WB
bert ernie rap	2008	FR	LN	WB	GO	haters gonna hate	2010	GO	WB	FR	LN	XXX	2008	GO	WB	LN	FR
XXX	2010	FR	LN	WB	GO	herp derp	2010	WB	LN	GO	FR	XXX	2009	FR	LN	GO	WB
boom goes the dyn.	2005	FR	WB	LN	GO	hipster kitty	2010	WB	GO	LN	FR	picard facepalm	2008	LN	FR	WB	GO
boxxy	2008	FR	LN	WB	GO	hipster mermaid	2011	LN	WB	GO	FR	planking	2011	LN	FR	WB	GO
breading cats	2011	FR	LN	WB	GO	XXX	2009	FR	LN	WB	GO	pork and beans	2008	FR	LN	WB	GO
brian peppers	2005	FR	LN	WB	GO	hopkin green frog	2004	FR	LN	WB	GO	powerthirst	2007	LN	FR	WB	GO
brony	2011	GO	WB	LN	FR	hover hand	2010	FR	LN	WB	GO	pownage	2008	GO	WB	LN	FR
buscemeyes	2011	FR	LN	GO	WB	hurr durr	2009	GO	WB	LN	FR	pure pwnage	2005	WB	LN	GO	FR
business cat	2011	WB	LN	FR	GO	hurra torpedo	2005	FR	LN	WB	GO	push but. receive bacon	2009	FR	LN	WB	GO
call me maybe	2012	WB	LN	FR	GO	i am on a boat	2009	FR	LN	WB	GO	qwop	2010	FR	LN	WB	GO
candlejack	2007	WB	GO	LN	FR	i dunno lol	2009	GO	WB	LN	FR	rage comics	2011	WB	GO	LN	FR
caramelldansen	2008	FR	LN	GO	WB	i eated it	2005	LN	FR	WB	GO	rage guy	2009	GO	WB	LN	FR
caturday	2006	FR	LN	WB	GO	i just met you	2012	LN	WB	FR	GO	ran ran ru	2007	FR	LN	WB	GO
ceiling cat	2006	WB	GO	LN	FR	i love bees	2004	FR	LN	WB	GO	raymond crowe	2007	FR	LN	WB	GO
chad vader	2006	FR	LN	WB	GO	i own a horse	2010	FR	LN	WB	GO	red solo cup	2011	FR	LN	WB	GO
challenge accepted	2010	GO	WB	LN	FR	impossibru	2011	WB	LN	FR	GO	rick roll	2007	FR	LN	WB	GO
charlie bit me	2007	FR	LN	GO	WB	in soviet russia	2009	WB	GO	LN	FR	rofl copter	2004	GO	WB	LN	FR
chocolate rain	2007	FR	LN	WB	GO	int. crocodile allig.	2008	FR	LN	WB	GO	rules of the internet	2008	GO	WB	LN	FR
chris crocker	2007	FR	LN	WB	GO	XXX	2007	FR	LN	WB	GO	salad fingers	2004	LN	FR	WB	GO
christian bale rant	2009	FR	LN	WB	GO	it is over 9000	2007	GO	WB	LN	FR	scarlet takes tumble	2008	FR	LN	WB	GO
chuck norris facts	2005	FR	LN	WB	GO	jejemon	2010	FR	LN	GO	WB	serious cat	2006	WB	LN	GO	FR
cinnamon challenge	2011	FR	LN	WB	GO	XXX	2008	FR	LN	WB	GO	shoop da woop	2006	WB	GO	LN	FR
XXX	2007	FR	LN	WB	GO	jk wedding dance	2009	FR	LN	GO	WB	simons cat	2009	GO	WB	LN	FR
come at me bro	2010	WB	GO	LN	FR	karate kyle	2011	FR	LN	WB	GO	soc. awkward penguin	2009	GO	WB	LN	FR
coneing	2011	FR	LN	WB	GO	ken lee	2008	FR	LN	WB	GO	standing cat	2010	FR	LN	WB	GO
courage wolf	2008	WB	GO	LN	FR	keyboard cat	2009	FR	LN	WB	GO	stoner comics	2010	WB	LN	FR	GO
crank that	2007	FR	LN	WB	GO	kitler cats	2006	FR	LN	WB	GO	subservient chicken	2004	FR	WB	LN	GO
crasher squirrel	2009	FR	LN	WB	GO	la caida de edgar	2006	FR	LN	WB	GO	surprised kitty	2009	FR	LN	GO	WB
crazy frog	2004	FR	LN	WB	GO	lazy sunday	2005	FR	LN	WB	GO	talking twin babies	2011	FR	LN	GO	WB
cupcake dog	2008	FR	LN	WB	GO	leave britney alone	2007	FR	WB	LN	GO	techno viking	2007	WB	LN	FR	GO
daft bodies	2007	FR	LN	GO	WB	leek spin	2006	FR	LN	WB	GO	tech. impaired duck	2010	LN	FR	WB	GO
daft hands	2007	FR	LN	GO	WB	leeroy jenkins	2005	FR	LN	WB	GO	the internet is for porn	2005	FR	LN	GO	WB
demotivational	2009	GO	WB	LN	FR	like a sir	2011	GO	WB	LN	FR	the last lecture	2008	FR	LN	WB	GO
depression dog	2008	GO	WB	LN	FR	line rider	2006	FR	LN	GO	WB	this is sparta	2007	FR	LN	GO	WB
derpina	2011	WB	GO	LN	FR	literal music video	2009	FR	LN	GO	WB	this land is your land	2004	FR	WB	LN	GO
XXX	2006	FR	LN	WB	GO	llama song	2004	LN	WB	FR	GO	tinaecmusic	2007	FR	LN	WB	GO
diet coke mentos	2006	FR	LN	GO	WB	loituma	2006	FR	LN	GO	WB	XXX	2008	GO	WB	LN	FR
dis gon b gud	2011	WB	LN	FR	GO	lol wut	2007	WB	LN	FR	GO	trolldad	2010	LN	WB	FR	GO
do a barrel roll	2011	FR	LN	WB	GO	lolcats	2007	WB	GO	LN	FR	trollface	2010	GO	WB	LN	FR
don't tease me bro	2007	FR	LN	WB	GO	loneleygirl15	2006	FR	LN	GO	WB	true story bro	2011	LN	FR	WB	GO
double rainbow	2010	FR	LN	GO	WB	lurk moar	2006	LN	FR	WB	GO	u mad bro	2011	WB	GO	LN	FR
dramatic chipmunk	2007	FR	LN	WB	GO	magibon	2007	FR	LN	WB	GO	united breaks guitars	2009	FR	LN	WB	GO
eat da poo poo	2010	FR	LN	WB	GO	maru the cat	2009	GO	LN	WB	FR	XXX	2007	FR	LN	GO	WB
engrish funny	2008	FR	LN	WB	GO	meanwhile in	2010	GO	LN	WB	FR	vernon koekemoer	2008	FR	LN	WB	GO
epic beard man	2010	FR	LN	WB	GO	million $ homepage	2005	FR	LN	GO	WB	we are not afraid	2005	FR	LN	WB	GO
epic fail	2009	GO	WB	LN	FR	monorail cat	2006	FR	LN	WB	GO	XXX	2011	LN	WB	FR	GO
epic win	2009	LN	WB	FR	GO	montauk monster	2008	FR	LN	GO	WB	wii hula girl	2008	FR	GO	LN	WB
ermahgerd	2012	LN	WB	FR	GO	mudkips	2007	WB	GO	LN	FR	winnebago man	2010	FR	LN	WB	GO
evolution of dance	2006	FR	LN	WB	GO	music is my hot sex	2007	LN	WB	FR	GO	x all the y	2011	LN	FR	WB	GO
failboat	2007	WB	LN	GO	FR	my new haircut	2007	FR	LN	WB	GO	y u no	2010	WB	GO	LN	FR
fayul	2010	FR	LN	WB	GO	nek minnit	2011	FR	LN	WB	GO	ya rly	2005	FR	LN	WB	GO
flying spag. monster	2005	FR	LN	WB	GO	nevada tan	2007	FR	LN	WB	GO	yao ming face	2011	WB	GO	LN	FR
fmylife	2009	FR	LN	GO	WB	ninja cat	2008	FR	LN	GO	WB	yes this is dog	2011	FR	LN	WB	GO
forever alone	2010	GO	WB	LN	FR	no wai	2005	FR	LN	WB	GO	yo dawg	2008	FR	LN	WB	GO
fsjal	2009	FR	LN	WB	GO	noah takes a photo	2007	FR	LN	WB	GO	ytmnd	2005	LN	WB	FR	GO

bizarre sexual practices or *screamer memes* that are intended to shock their audience or mock people or beliefs.

Table 1 summarizes the results of our trend analysis for the 204 memes in set 1; models are ranked w.r.t. their D_{KL} values. We observe the Frechet to give the best fit in the majority of cases (62%). The Weibull is the second best fitting function followed by the Gompertz and the Log-Normal. In order to visualize these results, Fig. 5 displays prototypic examples of time series and graphs of the corresponding best fitting version of each of the four models.

Table 2 summarizes the results of our trend analysis for the 10 memes in set 2. Again, we rank our models w.r.t. their KL-divergence between data and fit. For the "incomplete" data in this set, the Frechet distribution provides the best fitting model in 50% of all cases. The Weibull distribution is the second best fitting function; the Gompertz and Log-Normal perform equally. Visualizations of how the models fit truncated meme time series are shown in Fig. 6.



(a) "fmylife", Frechet fits best (b) "o rly", Frechet fits best (c) "monorail cat", Frechet fits best

(d) "lol wut", Weibull fits best (e) "has cheezburger", Weibull fits best (f) "y u no", Weibull fits best

(g) "haters gonna hate", Gompertz fits best (h) "it's over 9000", Gompertz fits best (i) "ytmnd", Log-Normal fits best

Figure 5: Example of time series of relative search frequencies related to different Internet memes and results of global trend analysis using the life-time distributions considered in this paper; the graph of the respective best fitting model is emphasized.

Table 2: 10 Internet memes (set 2) and best fitting models.

meme	onset	D_{KL} based ranking			
		1st	2nd	3rd	4th
all your base	2000	$\mathcal{FR}$	$\mathcal{LN}$	$\mathcal{WB}$	$\mathcal{GO}$
XXX	1996	$\mathcal{FR}$	$\mathcal{LN}$	$\mathcal{GO}$	$\mathcal{WB}$
badger badger	2003	$\mathcal{WB}$	$\mathcal{GO}$	$\mathcal{LN}$	$\mathcal{FR}$
bert is evil	1998	$\mathcal{WB}$	$\mathcal{LN}$	$\mathcal{FR}$	$\mathcal{GO}$
bubb rubb	2003	$\mathcal{GO}$	$\mathcal{LN}$	$\mathcal{WB}$	$\mathcal{FR}$
dancing baby	1996	$\mathcal{FR}$	$\mathcal{LN}$	$\mathcal{WB}$	$\mathcal{GO}$
hamster dance	1998	$\mathcal{FR}$	$\mathcal{LN}$	$\mathcal{WB}$	$\mathcal{GO}$
n00b	1988	$\mathcal{FR}$	$\mathcal{LN}$	$\mathcal{WB}$	$\mathcal{GO}$
schfifty five	2003	$\mathcal{LN}$	$\mathcal{WB}$	$\mathcal{FR}$	$\mathcal{GO}$
weebl and bob	2002	$\mathcal{WB}$	$\mathcal{LN}$	$\mathcal{FR}$	$\mathcal{GO}$

Table 3: Percentages of best fit for the four different models.

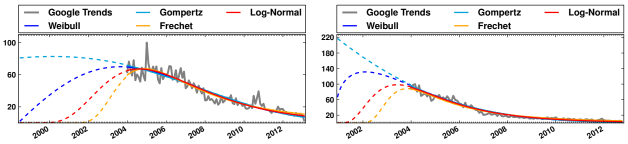
	$f_{\mathcal{WB}}$	$f_{\mathcal{GO}}$	$f_{\mathcal{FR}}$	$f_{\mathcal{LN}}$
204 memes (set 1)	16.6%	12.2%	62.0%	9.2%
10 memes (set 2)	30.0%	10.0%	50.0%	10.0%
all memes (set 1 ∪ set 2)	17.2%	12.1%	61.4%	9.3%

Discussion

Table 3 summarizes the performance of the four models in terms of percentages of best fits. Looking at these results suggests the following conclusions:

Overall, the Weibull, Gompertz, and Frechet provide better models of general trends in time series of meme related search activities than the Log-Normal distribution. While the latter cannot be explained in terms of a growth equation, we have shown that the former implicitly describe processes of growth and decline. Given these empirics, it appears that there exist general mechanisms that govern how collective attention to particular Internet memes evolves over time. In other words, the finding that the three distributions which represent growth dynamics provide particularly good descriptions of the trends in most of our data cannot be attributed to chance. The expressive power of the four tested models is equivalent; neither has more degrees of freedom than any of the others and neither completely fails to explain our data. Yet, as the Weibull, the Gompertz, and the Frechet provide significantly better fits, they appear to encode a latent mechanism that accounts for how collective attention to memes develops over time. This mechanism seems not to be encoded in the Log-Normal.

Although the growth and decline dynamics represented by the Weibull, Gompertz and Frechet are of different types (polynomial in t, exponential in t, and inverse polynomial in t, respectively), they all depend on $F(t)$, the amount of attention attracted so far. Not only do these models provide good characterizations of general trends in time series related to collective interest in Internet memes, they also describe these trends in terms of basically the same mechanism: the sum total of attention attracted so far influences the current and future development of the popularity of a meme. In terms of basic everyday experience this translates to the statement, that Internet memes undergo a hype cycle. The cumulative densities in the growth equations of these three



(a) “hamster dance” (b) “all your base”

Figure 6: “Predictions” of the past evolution of Internet memes that appeared prior to 2004. Solid curves show fits to the tails of truncated time series; dashed curves represent corresponding extrapolations into the past.

distributions act as a momentum term whose drag increases over time. The more a population gets used to a meme or the less novel it appears, the quicker it looses its appeal. Internet memes thus appear to be fads.

Among the three fad distributions we considered, the Frechet provides the best model of the temporal evolution of attention for most of the 214 popular memes whose dynamics we analyzed. Examples of time series that are best explained by the Frechet are shown in Fig. 5(a)–(c).

As the Frechet is characterized by a rather steep initial increase, a narrow mode, and a considerably long tail, it appears that the majority of memes in our data set quickly reached peak popularity and began to slowly decline shortly after. Yet, since the Weibull and the Gompertz, too, account well for a substantial percentage of meme related time series, it seems that there are at least two different kinds of Internet memes. On the one hand, there are short-lived memes such as the “o rly” meme (Fig. 5(b)) that go in and out of popularity rather quickly. On the other hand, memes like the “has cheezburger” meme (Fig. 5(e)) persist for extended periods of time. This points to interesting questions for future research regarding the content of memes: what is it, that causes memes to be short- or long-lived?

Finally, addressing the issue of the previously reported good performance of the Log-Normal in modeling meme dynamics, we evaluate the potential for the Frechet and the Log-Normal to be confused. We determine the probability for the Log-Normal to yield the second best fitting model whenever the Frechet yields the best fit as well as the probability of the inverse case and find

$$p(\text{2nd best fit} = \mathcal{LN} \mid \text{best fit} = \mathcal{FR}) = 0.95 \quad \text{and}$$
$$p(\text{2nd best fit} = \mathcal{FR} \mid \text{best fit} = \mathcal{LN}) = 0.50.$$

Thus, both distributions may indeed be confused for each other. Since similarly behaving functions may lead to ambiguous statistical conclusions (cf. the discussion by Rinne (2008)), it is pivotal to focus on appropriate models. Given our discussion, the Frechet clearly seems more appropriate.

Related Work

The study of fads has a venerable history across several disciplines (Meyerson and Katz 1957). Models of the dynamics of fads are important in economics and finance where they are used to analyze market trends. This involves differential equations that couple price and popularity effects (Tassier 2003) or probabilistic reasoning about information cascades or social signals in networks (Bikhchandani, Hirshleifer, and Welch 1998; Scharfstein and Stein 1990). Although simpler growth equations feature prominently in actuarial sciences or biology (Zeide 1993), we are not aware of their use as models of fads or Internet phenomena.

The dynamics of cultural fads have previously been modeled using differential equations (Acerbi, Ghirlanda, and Enquist 2012) that are akin to epidemic models (Britton 2010). With respect to Internet memes, such approaches have been discussed in (Bauckhage 2011) but were found to perform worse than simpler statistical models. Distributions such as the Weibull are well known in sociology and the political sciences (Zorn 2000) where they are used to model duration data. The Weibull was also found to account well for dwell times on Web sites (Liu, White, and Dumais 2010) or for the times people spend playing online games (Bauckhage et al. 2012). Yet, we are not aware of reports where the Weibull or related distributions have been used as fad models.

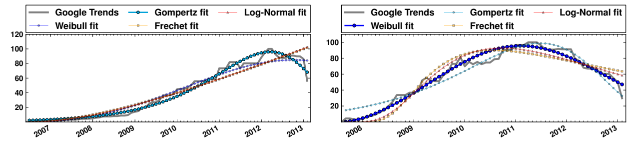
Work most closely related to what is reported here is due to (Bauckhage 2011; Leskovec, Adamic, and Huberman 2007; Wu and Huberman 2007) who study social media dynamics. While Leskovec et al. (2007) apply epidemic models, Wu and Huberman (2007) and Bauckhage (2011) make use of the Log-Normal. For the former, this is reasonable since they study distributions of amounts of *diggs* for news items; for the latter, we have shown here that better models exists for characterizing the dynamics of Internet memes.

Conclusion

Internet memes have become an integral part of modern Web culture. They consist of pieces of multi-medial content that spread virally on the social Web where they evolve through commentary or spoofs. Given their popularity, diversity, and proliferation, research on Internet memes is yet surprisingly scarce. In particular, plausible models that would explain their temporal dynamics have not previously been reported.

In this paper, we asked what kind of mechanisms could explain the emergence of noticeably skewed time series that characterize how collective attention to individual memes evolves over time. We considered simple growth equations that model the rise and fall dynamics of *fads* and showed that established statistical distributions, namely the Weibull, the Gompertz, and the Frechet, are particular instances of these models. In an empirical analysis of more than 200 meme related time series retrieved from Google Trends, we found that these distributions are better trend indicators than the Log-Normal that was used in the previous literature.

Our results hint at the existence of general mechanisms that govern the evolution of attention to Internet memes. Although the growth and decline dynamics encoded in the Weibull, Gompertz, and Frechet vary in detail, they all depend on attention attracted so far. That is, the sum total of attention an Internet meme has attracted so far influences its popularity. The more a population of users gets used to a meme or the less novel it appears, the quicker it looses its appeal. Internet memes therefore seem to undergo a hype cycle; they are nothing but fads.



(a) “social media” (b) “cloud computing”

Figure 7: Examples of search frequencies related to technology trends. The Gompertz (a) and the Weibull (b) fit best.

Nevertheless, as Internet memes are very popular and can be expected to stay, the phenomenon merits further research. Especially the question of what kind of features cause content to “go viral” still awaits an answer. The approaches proposed in this paper allow for distinguishing short-lived memes from long-lived ones and therefore may provide an avenue towards the automatic recognition of viral potential.

Finally, as our models are simple, interpretable, and successful, we expect them to apply to other types of fads and trends as well. Anecdotal evidence as to this belief is shown in Fig. 7 where we used our models to analyze the temporal dynamics of *technology trends*.

Acknowledgments

The work in this paper was carried out within the Fraunhofer / University of Southampton research project *SoFWIReD* which is funded by the Fraunhofer ICON initiative.

References

Acerbi, A.; Ghirlanda, S.; and Enquist, M. 2012. The Logic of Fashion Cycles. *PLoS ONE* 7(3):e32541.

Bauckhage, C.; Kersting, K.; Sifa, R.; Thurau, C.; Drachen, A.; and Canossa, A. 2012. How Players Lose Interest in Playing a Game: An Empirical Study Based on Distributions of Total Playing Times. In *Proc. CIG*.

Bauckhage, C. 2011. Insights into Internet Memes. In *Proc. ICWSM*.

Bernstein, M.; Monroy-Hernandez, A.; Harry, D.; Andr, P.; Panovich, K.; and Vargas, G. 2011. 4chan and /b/: An Analysis of Anonymity and Ephemerality in a Large Online Community. In *Proc. ICWSM*.

Bikhchandani, S.; Hirshleifer, D.; and Welch, I. 1998. Learning from the Behavior of Others: Conformity, Fads, and Informational Cascades. *J. of Economic Perspectives* 12(3):151–170.

Britton, T. 2010. Stochastic Epidemic Models: A Survey. *Mathematical Biosciences* 225(1):24–35.

Burgess, J. 2008. All Your Chocolate Rain Are Belong To Us? Viral Video, YouTube and the Dynamics of Participatory Culture. In Lovink, G., and Niederer, S., eds., *The VideoVortex Reader*. Amsterdam: Institute of Network Cultures.

Choi, H., and Varian, H. 2012. Predicting the Present with Google Trends. *Economic Record* 88(S1):2–9.

Da, Z.; Engelberg, J.; and Gao, P. 2011. In Search of Attention. *J. of Finance* 66(5):1461–1499.

Gleser, L., and Moore, D. 1983. The Effect of Dependence on Chi-Square and Empiric Distribution Tests of Fit. *The Annals of Statistics* 11(4):1100–1108.

Granka, L. 2009. Inferring the Public Agenda from Implicit Query Data. In *Proc. SIGIR*.

Jennrich, R., and Moore, R. 1975. Maximum Likelihood Estimation by Means of Nonlinear Least Squares. In *Proc. of the Statistical Computing Section*.

Joseph, K.; Wintoki, J.; and Zhang, Z. 2011. Forecasting Abnormal Stock Returns and Trading Volume Using Investor Sentiment: Evidence from Online Search. *Int. J. of Forecasting* 27(4):1116–1127.

Kleiber, C., and Kotz, S. 2003. *Statistical Size Distributions in Economics and Actuarial Sciences*. Wiley.

Koch, A. 1966. The Logarithm in Biology: 1. Mechanisms Creating Log-Normal Distributions Exactly. *J. of Theoretical Biology* 12.

Lawless, J. 2003. *Statistical Models and Methods for Lifetime Data*. Wiley.

Leskovec, J.; Adamic, L.; and Huberman, B. 2007. The Dynamics of Viral Marketing. *ACM Tans. on the Web* 1(1):5.

Liu, C.; White, R.; and Dumais, S. 2010. Understanding Web Browsing Behavior through Weibull Analysis of Dwell Times. In *Proc. SIGIR*.

Mellon, J. 2011. Search Indices and Issue Salience: the Properties of Google Trends as a Measure of Issue Salience. Sociology Working Papers 2011-01, University of Oxford.

Meyerson, R., and Katz, E. 1957. Notes on a Natural History of Fads. *American J. of Sociolog* 62(6):594–601.

Mitzenmacher, M. 2004. A Brief History of Generative Models for Power Law and Lognormal Distributions. *Internet Mathematics* 1(2):226–251.

Pogue, D. 2011. Internet Memes 101: A Guide to Online Wackiness. New York Times, September 8.

Rinne, H. 2008. *The Weibull Distribution*. Chapman & Hall.

Scharfstein, D., and Stein, J. 1990. Herd Behavior and Investment. *American Economic Review* 80(3):465–479.

Tassier, T. 2003. A Model of Fads, Fashions, and Group Formation. *Complexity* 9(5):51–61.

Teevan, J.; Liebling, D.; and Geetha, G. 2011. Understanding and Predicting Personal Navigation. In *Proc. WSDM*.

Thom, J., and Millen, D. 2012. Stuff IBMers Say: Microblogs as an Expression of Organizational Culture. In *Proc. ICWSM*.

Wu, F., and Huberman, B. 2007. Novelty and Collective Attention. *PNAS* 104(45):17599–17601.

Zeide, B. 1993. Analysis of Growth Equations. *Forrest Science* 39(3):594–616.

Zorn, C. W. 2000. Modeling Duration Dependence. *Political Analysis* 8(4):367–380.