Expecting to be HIP: Hawkes Intensity Processes for Social Media Popularity

Marian-Andrei Rizoiu[†♯], Lexing Xie[†♯], Scott Sanner[‡]
Manuel Cebrian[♯], Honglin Yu[†♯], Pascal Van Henteryck[♭]
[†]Australian National University, [♯]Data61 CSIRO, [‡]University of Toronto, [♭]University of Michigan

ABSTRACT

Modeling and predicting the popularity of online content is a significant problem for the practice of information dissemination, advertising, and consumption. Recent work analyzing massive datasets advances our understanding of popularity, but one major gap remains: To precisely quantify the relationship between the popularity of an online item and the external promotions it receives. This work supplies the missing link between exogenous inputs from public social media platforms, such as Twitter, and endogenous responses within the content platform, such as YouTube. We develop a novel mathematical model, the Hawkes intensity process, which can explain the complex popularity history of each video according to its type of content, network of diffusion, and sensitivity to promotion. Our model supplies a prototypical description of videos, called an *endo-exo map*. This map explains popularity as the result of an extrinsic factor – the amount of promotions from the outside world that the video receives, acting upon two intrinsic factors – sensitivity to promotion, and inherent virality. We use this model to forecast future popularity given promotions on a large 5-months feed of the most-tweeted videos, and found it to lower the average error by 28.6% from approaches based on popularity history. Finally, we can identify videos that have a high potential to become viral, as well as those for which promotions will have hardly any effect.

1. INTRODUCTION

The popularity of an online cultural item is described by the amount of attention it receives, and the popularity dynamics refers to its evolution over time. Popularity is a critical measure of information dissemination for content producers, and a way to manage information overload for content consumers. Understanding and predicting popularity have been active topics in both research and practice, but many fundamental questions remain open, such as: What describes the most viral items? What do the popularity dynamics of news, music, films look like, and what are their differences and similarities? Can we promote an item to increase its popularity, and how much promotion is needed?

Building upon recent research progress in understanding popularity, we identify three important questions that are still open. The first one concerns modeling popularity. One set of approaches describe popularity dynamics as stylistic prototypes, such as being power-law shapes from either an exogenous shock or endogenous relaxation [13], a combination of power-law and exponential decay [24], multiple power-law decays with periodicity [27] or a collection of recurrence peaks [10]. However, one question remains: **How would popularity evolve under continuous external influence?** Especially, how one can explain complex rise and fall patterns that do not follow the prescribed prototypes. The second questions concerns virality. Content and initial diffusion have both been identified as key factors that influence popularity. Here content factors include positive sentiment [2], emotional arousal [5], publishing venue [3], visibility [6]; and factors of diffusion history include [9] network structure, information about the original poster and re-sharers, the timing of the early posts. However, describing viral content in the light of external promotions is still an open problem, and in particular: **Can something go viral if promoted?** The third questions involves predicting future popularity. It is known that the approaches that use the popularity history [30, 34] produce competitive estimates about future popularity over time. Also, timing features have been shown to be more predictable than content, structure, and user features [9], and prediction without initial history is generally shown as a hard problem [26]. However, these recent insights do not answer: **How to forecast future popularity given planned promotions?**

In this work, we answer all three questions above, using a large dataset that connects popularity in one social media platform – 81.9 million YouTube videos – to discussions about each of these digital items in an external platform – in 1.06 billion tweets over a six-month period.

To describe complex popularity dynamics under continuous external influence, we propose a new mathematical model that reveals an analytical relationship between endogenous and exogenous demand factors, called the Hawkes Intensity Process (HIP). HIP extends the well-known Hawkes point-process [19], by taking the expectation over stochastic event histories so as to describe expected event volumes, rather than a set of event times. Figure 1 illustrates the HIP model. On the top left is the volume of exogenous promotions over time, which drives the endogenous response determined by the HIP (middle); the output on the right is the

WWW 2017, April 3–7, 2017, Perth, Australia.
ACM 978-1-4503-4913-0/17/04.
http://dx.doi.org/10.1145/3038912.3052650

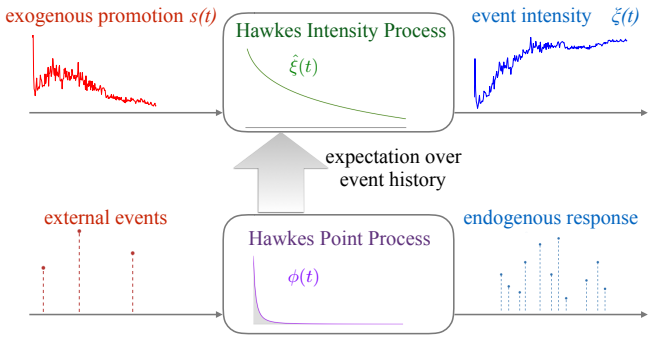


Figure 1: Linking endogenous and exogenous factors of popularity using the Hawkes Intensity Process. *Top row*: The input are volumes of exogenous promotion or discussions $s(t)$, that engender endogenous reactions from the online social networks described by the impulse response function $\hat{\xi}(t)$ (middle box, defined in Sec 2.5), to generate the total popularity series $\xi(t)$. *Bottom row*: The endogenous reactions are self-exciting point processes, widely used in recent literature [4, 23, 28, 31, 33, 39]. Here each event triggers subsequent events with memory kernels $\phi(t)$. Such point process models can incorporate individual external stimulus (show on the left) which in turn lead to a larger number of events in response (shown on the right). *Middle arrow*: The proposed HIP model is a result of taking the expectation over all stochastic event history of the Hawkes process in the bottom.

popularity series. The popularity series modeled through the Hawkes intensity process matches closely with the observed view count series, even for videos with complex popularity lifecycles (Section 2).

To answer the second question, on whether or not an item will go viral if promoted, we derive two new metrics based on HIP – the endogenous response and exogenous sensitivity. These two metrics naturally lend to a novel two-dimensional visualization tool, dubbed the *endo-exo map* (Section 4). On this map, one can identify online videos that have high potential but are not yet popular. In other words, video with high sensitivity to external promotions and high endogenous response are expected to go viral if promoted. On the other hand, one can also identify videos for which promotion is unlikely to have an effect, such as those scoring very low in either the endo- or exo- dimension.

Finally, the HIP model can be used to help forecast future popularity given (known or planned) promotions. HIP model parameters are estimated on the first 90 days of each video's history, and forecasts are made for the next 30 days. We evaluate forecasting on a collection of 13K+ most actively discussed YouTube videos over a six-month period, and found that estimates made with the HIP lower the average percentile error by 28.6% from state-of-the-art methods based on popularity history (Section 5).

The main contributions of this work include:

- The HIP model, a volume based version of the Hawkes point process. Its essential novelty is to regard popularity as externally-driven, with exogenous events activating endogenous responses inside the social environment which may, or may not, amplify the exogenous signal.
- The exogenous sensitivity and the endogenous response, two new metrics to quantify two distinct aspects of a video's inherent tendency to be popular. They are combined in the endo-exo map, a tool used to comparatively explain popularity and identify potentially viral videos.
- A method to forecast popularity gain after promotion. Evaluated on a large set of YouTube videos, it significantly outperforms approaches using popularity history.
- A new dataset of tweeted videos that links online videos to their external discussions, available at https://github.com/andrei-rizoiu/hip-popularity.

2. THE MODEL

We introduce a model for the evolution of online attention under external influence. We start by discussing the problem setting of aggregated attention under external promotion (in Sec. 2.1), the key concepts of the Hawkes process and its use to link the ongoing effect of external stimuli to the word-of-mouth spread of attention (Sec. 2.2). Next, we propose HIP, a model to explain the observed popularity history from daily volumes when the underlying viewing events are unobserved (Sec. 2.3). Lastly, we introduce two key metrics derived from the HIP model, the endogenous response and exogenous sensitivity, to quantify the viral potential of a video (Sec. 2.5).

2.1 Problem setting: views under promotion

This paper aims to model the popularity of videos under external promotion. Here popularity is measured in the number of total views after the video being online for a certain number of days (e.g. up to 120 days). External promotion is harder to measure, since by definition, it needs to capture data from other platforms. In this paper, we have two different views of promotion, due to the data collection setting described in Sec 3. The first is *shares*, tracked by YouTube via the *share* button under each video that allows a user to share a link of the video on a selection of popular social network sites – 13 at the time of this writing. The second view is *tweets*, tracked with twitter streaming API with keyword filters that retrieve tweets that link to a video. Neither source is complete – with the distributed nature of the Internet, one can see that a complete capture of all discussions is practically impossible. The *shares* captures external promotions from a diverse set of sources, but is far from complete in any one source. The *tweets* captures an almost-complete feed of video promotions in one platform. In the rest of this paper, both of these sources are collectively referred to as external *promotions* about a video. In our evaluations, the results obtained using each source are presented separately.

2.2 Hawkes process for social events

We model online attention as an exogenously-driven self-exciting process – each viewing *event* is triggered either by a previous event or as a result of external influence. We assume that viewing events of a YouTube video follow a Hawkes point process [19], a type of non-homogeneous point process in which the arrival of an event increases the likelihood of future events. Although variants of point processes have recently been used to model events in social media, all existing work focus on learning point process model from one information source, such retweeting [39, 23], arrival of citations [33], or endogenous response after an initial external shock [13]. To the best of our knowledge, this is the first work that models the continuous interaction of two sources – exogenous stimuli and endogenous response.

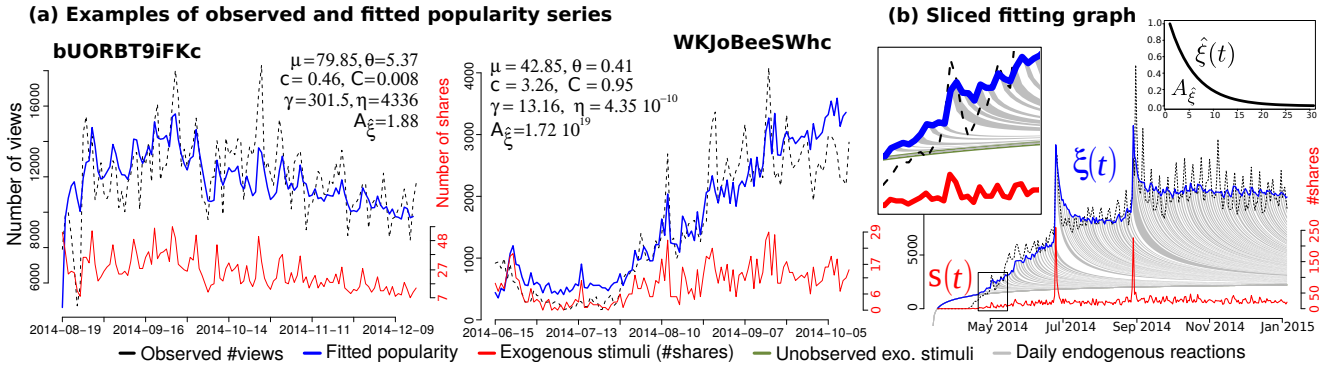


Figure 2: Explaining popularity dynamics using the Hawkes intensity model. **(a)** Number of shares (red), observed popularity history (black dashed) and popularity as explained by the HIP (blue) on two examples videos: a `music` video *bUORBT9iFKc* and a `News & Politics` video *WKJoBeeSWhc*. The multi-phased popularity history cannot be explained by current models such as [13], while the HIP tracks the complex dynamics well. **(b)** A *sliced* fitting graph of a music video (Youtube ID *0bR4L0Y94AQ*) – using the impulse response $\hat{\xi}(t)$ and exogenous stimuli $s(t)$ to explain observed popularity. Each alternating gray and white area under the fitted (blue) curve is a *slice* of endogenous reaction generated by the external influence in a given day. The left inset zooms-in one of the early months in the video's evolution, in May 2014. The total event intensity (blue solid line) is a sum of temporally shifted and scaled versions of $\hat{\xi}(t)$, which tracks the long-term trends in observed popularity well (dashed line). The period around the first larger exogenous peak is shown magnified so that its corresponding endogenous response is clearly visible. (right inset) Example of the impulse response $\hat{\xi}(t)$ to one unit of external excitation. The area under this function, $A_{\hat{\xi}}$, quantifies the endogenous reaction of a video – it is the total number of views after each unit of exogenous excitation.

In particular, the arrival rate of viewing events $\lambda(t)$, a measure of how likely a viewing event will occur in a infinitesimal interval around time t, is determined by two additive components in Eq (1). The first component is proportional to a measure of external influence $s(t)$ scaled by a constant μ. Here $s(t)$ represents the volume of external discussion (or promotion) over time. The second component represents the rate of views triggered by a previous event i, which occurred at time t_i with magnitude $m_i > 0$, according to a time-decaying triggering kernel $\phi_{m_i}(t - t_i)$. Furthermore, each event $t_i < t$ adds to $\lambda(t)$ independently. The following equations describe the event rate of such a *marked* Hawkes process:

$$\lambda(t) = \mu s(t) + \sum_{t_i < t} \phi_{m_i}(t - t_i) \tag{1}$$

$$\phi_m(\tau) = \kappa\, m^{\beta}\, (\tau + c)^{-(1+\theta)},\ \tau \in \mathcal{R}^+ \tag{2}$$

Eq. 2 describes the triggering kernel $\phi(\tau)$. In this work it is designed to capture several key quantities influencing popularity. Parameter κ is a scaling factor for video *quality*. m describes the relative *influence* of the user who generated the event, i.e., m_i in Eq. 1 when multiple events are concerned. The user influence exponent β, newly introduced in this work, accounts for the nonlinearity between observed metrics of influence (such as the number of followers) and popularity. This particular form allows both flexibility in modeling how much effect some observed metric of *influence* (e.g. number of followers) has on views (e.g. $\beta = 0$ would be no effect), and at the same time computing expectations over stochastic event history analytically, as will be shown in the next subsection. Time interval $\tau = t - t_i$ is the elapsed time since the parent event at t_i; $c > 0$ is a cutoff term to keep $\phi_m(\tau)$ bounded when τ is small; $1 + \theta$ (for $\theta > 0$) is the power-law exponent for social memory – the larger θ is, the sooner the reaction to an event will stop. We use a power-law kernel for $\phi_m(\tau)$, as recent work [28] observed it to have better performance on social media data than popular variants like the exponential kernel.

This model is an instance of a marked Hawkes process [19]. An illustration of the Hawkes process with external excitation is in the bottom row of Figure 1. A set of input events of different magnitudes trigger new events through the kernel $\phi(t)$, which then trigger offspring events themselves, resulting in the observed event sequence.

2.3 From Hawkes to HIP

The Hawkes point-process faces a few modeling challenges in large-scale applications. In terms of data source, what we often observe is the volume of total attention in a given interval (e.g. daily views on YouTube), rather than the times and properties of individual actions, due to constraints in user privacy and data volume. In terms of computation, full estimation of the Hawkes process is quadratic in the number of events. Therefore, the full estimation quickly becomes expensive when the number of events is in the hundred thousands or millions – this is where the most popular videos are (see Sec 3.2). It is very desirable if one could estimate video popularity with daily data, which is typically a few dozens to a few hundred data points.

To this end, we introduce the Hawkes intensity $\xi(t)$, the expectation of the event rate $\lambda(t)$ over the event history $\mathcal{H}_t$, consisting of the set of (random) event times and magnitudes up to time t.

THEOREM 2.1. HAWKES INTENSITY PROCESS (HIP) *Given a marked Hawkes process described in Equations (1) and (2). Its event history*

$$\mathcal{H}_t = \{(t_1, m_1), \ldots, (t_n, m_n)\}_{t_n < t}$$

contains all event times and marks before time t, where each mark m is drawn iid from a power-law distribution $p(m) = (\alpha - 1)m^{-\alpha}$. We define event intensity as the expectation of the event rate over the event history $\xi(t) = \mathbb{E}_{\mathcal{H}_t}[\lambda(t)]$, then

$\xi(t)$ follows the following self-consistent integral equation:

$$\xi(t) = \mu s(t) + C \int_0^t \xi(t-\tau)(\tau + c)^{-(1+\theta)} d\tau \ . \qquad (3)$$

Here constant $C = \frac{\kappa(\alpha-1)}{\alpha-\beta-1}$, and κ and β are as in Eq (2).

Intuitively, this expression of event intensity $\xi(t)$ at time t is determined by the external stimulus $s(t)$, and a convolution of its own history with a power-law memory kernel $(\tau+c)^{-(1+\theta)}$. Theorem 2.1 can be intuitively understood by breaking down the expectation into several parts. Note $\mu s(t)$ is non-random and does not change after expectation. We compute analytically the expectation over stochastic history, with a random number of events at random times, by decomposing $E_{\mathcal{H}_t}$ into expectations over binary variables dNt, which indicates whether or not there is an event in a small interval around time t. This trick discretizes time, and converts the sum over past events in Eq (1) into an integration seen in Eq (3). Note that the expectation of the user influence warping term m^β over the power-law distribution of the mark m has an analytical form, leading to the constant C. Due to space limitations, we include the full proof in the online appendix [1].

Here (μ, θ, C, c) are video-dependent parameters estimated from the popularity history of each video. Note that $\alpha > 0$ is the power-law exponent of user influence distribution, estimated as $\alpha = 2.016$ from a large Twitter sample using standard fitting procedures [11]. The two power law exponents α and θ in HIP are distinct in meaning and function, θ defines memory decay over time, while α is determined by the user distribution at large.

Compared to existing models of data volume, HIP captures the ongoing interactions of exogenous and endogenous effects. Hence it is able to explain complex popularity series with multiple rises and falls (as shown in Figure 2). Helmstetter and Sornette [20] fit the observed event rate after an initial shock, and Crane and Sornette [13] produce a curve fit on the long-term approximation of the endogenous decay with no exogenous input. SpikeM [27] models volumes of events both prior and after a single considered shock, without accounting for external influences. The work most related to ours on computing expectations over stochastic event histories is th work of Farajtabar *et al.* [16], who modeled co-excitation on Twitter and computed the equivalent of $\xi(t)$ on multivariate Hawkes process with exponential kernels, which admits a closed-form solution. In contrast, our work uses a univariate Hawkes process focused on modeling the impact of Twitter on individual Youtube videos and a power law kernel. De *et al.* [14] further develop the work in [16] by combining a Markov process with a multivariate Hawkes process for modeling opinion dynamics.

2.4 Estimating HIP from data

We discuss key steps for estimating HIP from observed series of views and external promotions over time.

Discretizing over time. We observe that behavioral statistics are aggregated over fixed and discrete intervals – for YouTube, the public API provides the daily history of the number of views $\bar{\xi}[t]$ and number of shares $\bar{s}[t]$ for $t = 1, \ldots, T$. Expressing HIP (Eq (3)) over discrete time gives:

$$\xi[t] = \mu s[t] + C \sum_{\tau=1}^{t} \xi[t-\tau](\tau + c)^{-(1+\theta)} \ . \qquad (4)$$

Here we use square brackets to denote discrete time, e.g. $\xi[t]$, and round brackets to denote continuous time, e.g. $\xi(t)$.

Accounting for unobserved external influence. In addition to the observed external promotions $\bar{s}[t]$ in tweets or shares, we model the unobserved external excitation as an initial shock (at $t = 0$) and a constant background excitation (for $t > 0$).

$$s[t] = \frac{\gamma}{\mu} \mathbb{1}[t = 0] + \frac{\eta}{\mu} \mathbb{1}[t > 0] + \bar{s}[t] \ , \qquad (5)$$

where $\mathbb{1}(arg)$ is the standard impulse function – taking the value 1 when arg is true and 0 otherwise. In the absence of a parametric model of generic external influence, the initial impulse and the constant component require the least amount of assumptions about how unobserved influence evolves. Here γ and η are additional parameters estimated from data. In our experiments, adding estimates for such unobserved influence components improves the fitting for a large number of videos.

The loss function For each video, we find an optimal set of model parameters (μ, θ, C, c) and of unobserved external influence (γ and η). This is done by minimizing the square error between the observed viewcount series $\bar{\xi}[t]$ and the model $\xi[t]$, $t = 1 : T$. The corresponding optimization problem is as follows:

$$\min_{\mu,\theta,C,c,\gamma,\eta} J = \frac{1}{2} \sum_{t=0}^{T} \left(\xi[t] - \bar{\xi}[t] \right)^2 \qquad (6)$$

We use `L-BFGS` [25] with analytical gradients and random restarts to minimize this non-linear loss function. Gradient computation is detailed in the appendix [1].

Three example fits are shown in Figure 2. Visibly, the event intensity model in Equation 3 links the exogenous and the endogenous effects of the social system, resulting in a tight fit between the model and the observed popularity history. For the Brazilian music video *bUORBT9iFKc* the memory kernel decays fast ($\theta = 5.37$), and the resulting intensity series tracks the temporal dynamics of the stimuli closely. For news video *WKJoBeeSWhc*, the memory kernel decays slowly ($\theta = 0.41$), hence the delayed accumulation of exogenous promotion via the memory kernel results in an overall rising trend. We can see that only by capturing the non-obvious joint effects from within and outside a social network can a model produce both fine-grained short-term dynamics and accurate long-term trends.

2.5 Properties of the HIP

In this section, we examine the key property of HIP of being a linear time-invariant system, which leads to two important metrics for measuring two distinct aspects of a video's viral potential – the exogenous sensitivity and the endogenous response.

Exogenous sensitivity μ. As shown in Eq 3, the total attention that a video receives consists of two parts: the input from the exogenous stimuli, and the endogenous response corresponding to non-linear effects accumulated through the integral equation. The scaling parameter μ quantifies a video's sensitivity to external stimuli $s(t)$. When $\mu \to 0$, external promotion would have no effect; when μ is large, each unit of external promotion leads to a large number of new views.

HIP as an LTI system. We observe an important property of the HIP model.

COROLLARY 2.2. *The HIP model,* as defined in Eq (4) and (3)*, is a linear time-invariant (LTI) system for* $t > 0$.

Being an LTI system [29] is to say that if $\xi[t]$ is the event intensity function for input $s[t]$, then (for the same video) the event intensity function for a shifted and scaled version of the input $as[t - t_0]$ is $a\xi[t - t_0]$ for $a > 0, t_0 \geq 0$, i.e., scaled and shifted by the same amount.

It is easy to see linearity holds by multiplying both sides of Eq 3 by the same constant. For time invariance, change of variable and then using the fact that $\xi[t] = 0$ when $t < 0$. A full proof is in the appendix [1].

Impulse response function $\hat{\xi}[t]$. One important descriptor of an LTI system is the impulse response function, the response to the unit impulse function $\mathbb{1}[t]$, which takes the value 1 when $t = 0$, and 0 otherwise. We define $\hat{\xi}[t]$ as the impulse response of the HIP model. It follows from Eq. (4) that $\hat{\xi}[t]$ is the solution to the following self-consistent equation:

$$\hat{\xi}[t] = \mathbb{1}[t] + C \sum_{t=0}^{T} \hat{\xi}[t - \tau](\tau + c)^{-(1+\theta)} d\tau \ , \qquad (7)$$

For each video, $\hat{\xi}[t]$ completely characterizes the endogenous response of the HIP model:

LEMMA 2.3. SLICED RESPONSES *The intensity function* $\xi[t]$ *of HIP can be written as the sum of impulse responses, scaled and shifted by the corresponding external input.*

$$\xi[t] = \sum_{\tau=0}^{T} s[\tau]\hat{\xi}[t - \tau] \qquad (8)$$

To see that this is true, first notice that external input $s[t]$ can be expressed as a sum of shifted and scaled impulses.

$$s[t] = \sum_{\tau=0}^{T} s[\tau]\mathbb{1}[t - \tau] \qquad (9)$$

Combining Eq (7) and (9) will lead to Eq (8). In other words, the total popularity at time T can be obtained as the sum of the unfolding through the endogenous reaction, of the external stimuli having occurred at times $1, 2, \ldots, T-1$. Fig 2(b) illustrates this property using a *sliced* and *stacked* popularity graph. The alternating white and gray *slices* are scaled (and shifted) versions of the impulse response represented in the right inset. For each discrete time point t' corresponds a slice, scaled by the external stimuli $s(t')$, which adds to the slices constructed at previous times $t < t'$. Adding all these slices together recovers the overall intensity $\xi(t)$ as in Eq 3 (blue line), which tracks closely the long-term dynamics of the observed popularity (dashed line). The LTI property and its related quantities provides the mathematical ground to define our second important measure.

Endogenous response $A_{\hat{\xi}}$. We define the total *endogenous response* generated from a single unit of exogenous excitation, computed as $A_{\hat{\xi}} = \sum_{t=0}^{\infty} \hat{\xi}[t]$. In this work, we compute $A_{\hat{\xi}}$ by taking the sum over 10,000 time steps. $A_{\hat{\xi}}$ is finite when the underlying HIP is so-called *sub-critical*. Other HIP-derived quantities, such as scaling parameter C or memory exponent θ could potentially serve to describe video virality. We find, however, that despite being related, the non-linear interactions among HIP parameters render them inaccurate in explaining popularity compared to $A_{\hat{\xi}}$. Detailed discussions on the convergence criteria for $A_{\hat{\xi}}$, and visualizations of other parameters are in the appendix [1]. Together with exogenous sensitivity μ, this is the second key quantity for measuring video virality. They will be used to compare individual and collections of videos in Sec. 4.

3. THE TWEETED VIDEOS DATASET

A key component in linking the exogenous influence and the endogenous response is to obtain data for the exogenous component, preferably both inside and outside the studied social network. We describe a new dataset across Twitter and Youtube networks, linked via the unique video ids, in which the volumes of tweets and Youtube shares serve as exogenous signals. We then introduce the popularity scale, a mapping between the number of views (or shares, or tweets) and the percentile ranking of a video, which will be used for visualizing popularity and for evaluating popularity forecast.

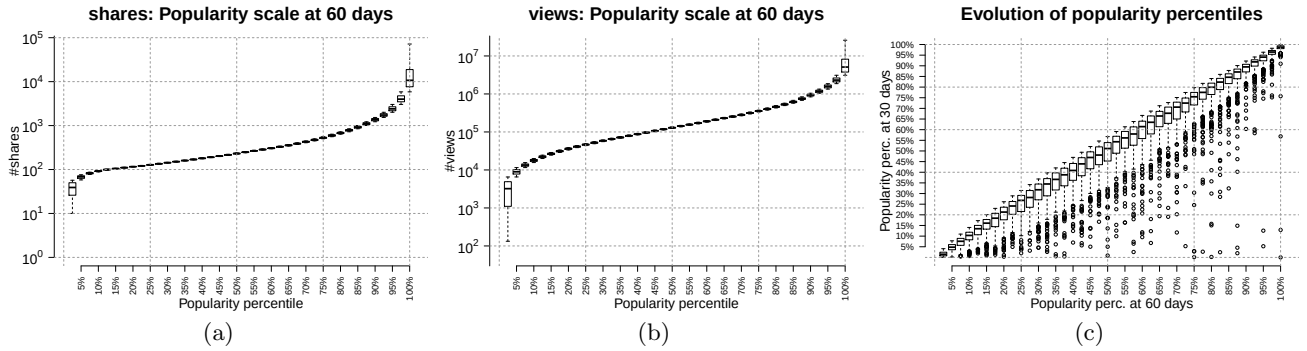
3.1 Dataset construction

We collect a dataset of *tweeted videos* by streaming tweets (via Twitter API) published between 2014-05-29 and 2014-12-26 which mentions YouTube videos. This yields a large and diverse set of over 81.9 million videos mentioned in 1.06 billion tweets. We obtain from YouTube their video metadata, including upload date, author and video category, as well as the time series consisting of the daily number of views and shares. The video categories are a one-level YouTube classification of videos, example of such categories being `Music`, `Gaming` or `Film & Animation`. Along with the daily number of tweets, we have three attention-related time series for each video: ($views[t]$, $shares[t]$ and $tweets[t]$), where t indexes time with the unit of a day.

In order to study videos with non-trivial popularity and promotion activities, we construct a subset, denoted as the ACTIVE dataset, by restricting to videos that are still online and that have their popularity and sharing series at least 120 days long, since the upload and until the crawling date. Furthermore, we restrict the set of videos to those that received at least 100 tweets and 100 shares by the 120th day, in order to obtain videos twitted and shared enough to estimate the external influence on popularity. We also remove 6 rare categories containing less than 1% videos (and their corresponding videos). The ACTIVE dataset contains 13,738 videos across 14 categories and it is used in both explaining and forecasting popularity in Sec. 5. Reasons for the drastic dataset reduction from 81M to ACTIVE include: videos uploaded earlier than 2014-05-29 (and hence without a complete tweet history), videos that are no longer online, those do not make viewcount history public, and the long-tailed distribution of tweets and shares – more than half of the videos are tweeted only once. Note that when they exist, the popularity and the sharing series do not contain missing data. A profile of the tweeted videos dataset and more details about its construction are given in the appendix [1]. We use the first 90 days of each videos' viewing and sharing/tweeting history to estimate the HIP parameters.

3.2 The popularity scale

It is well-known that network measurements such as the number of views and shares follow a long-tailed distribution. We quantify video popularity on an explicit popularity percentile scale, with 0.0% being the least popular, and 100% being the most popular. Fig. 3(a) and (b) show the popular-



(a) (b) (c)

Figure 3: The popularity scale of YouTube videos, computed on the ACTIVE dataset. The total numbers of `shares` (a) and `views` (b) obtained by each video in the first 60 days after upload are divided into 40 equally spaced bins (i.e. each with 2.5% of the videos). Boxplots of shares/views in each bin are shown. The 2.5% most popular videos span more than one order of magnitude for both views and shares. Note that outliers in this bin are not represented, as the most popular videos in the collection have $\sim 10^8$ views and $\sim 10^6$ shares. (c) Evolution of the views popularity between 30 (y-axis) and 60 (x-axis) days. Boxplots show where each 2.5% of videos at 60 days came from (in terms of percentile position at 30 days). The outliers are videos that have improved their popularity significantly.

ity scale as boxplots (in log-scale) over the ACTIVE dataset, after 60 days of video life for shares and views, respectively. The shape of the scale is similar for both shares and views, and it reflects their long tail distribution. The only notable difference is the scale of the y-axis, as videos tend to accumulate less shares than views. The popularity scale for tweets is very similar to the one for shares, and shown in the appendix [1]. Based on the shares and views popularity scales, we define two mapping functions $S_t(x), P_t(x) : \mathbb{R}_+ \longrightarrow [0, 1]$. Each function takes an argument – the number of shares for $S_t(x)$ or the number of views for $P_t(x)$ – and outputs the percentile value on the corresponding popularity scale constructed at time t.

In Fig. 3(c) we explore the change of views popularity of each video from 30 days (y-axis) to 60 days (x-axis). Formally, we plot the relation between $P_{30}\left(\sum_1^{30} \bar{\xi}[t]\right)$ and $P_{60}\left(\sum_1^{60} \bar{\xi}[t]\right)$, where $\bar{\xi}[t]$ is the number of views at time t (here the t-th day). Note that most videos retain a similar rank (in the boxes along the 45 degree diagonal line), or have a slight rank decrease as they are overtaken by other videos (slightly above the diagonal in the plot). No outliers exist in the upper-left part of the graph, since a video cannot lose viewcount that it already gained. Most notably, we can see that videos from any bin can *jump* to the top popularity bins between 30 and 60 days of age, such as the outliers for the few boxes on the far right. This phenomenon elicits two important questions: how did these videos go viral, and is this phenomenon related to external promotions?

4. THE ENDO-EXO MAP

Using two quantities defined in Sec 2.5, we construct a 2-dimensional map with *endogenous response* $A_{\hat{\xi}}$ as the x-axis and *exogenous sensitivity* μ as the y-axis. We call this plot the *endo-exo map*. This section presents example uses of this map for explaining video popularity, and identifying videos that are not promotable.

Explaining popularity. Intuitively, a video with a large endogenous response $A_{\hat{\xi}}$ and a high exogenous sensitivity μ has high potential to become viral. Specifically, each unit of exogenous excitation will generate $\mu A_{\hat{\xi}}$ events through the Hawkes intensity process. On the endo-exo map, videos in close proximity have similar potentials to become popular and the differences in their popularity would be due solely to the difference in exogenous attention. Fig 4(a) illustrates this phenomena using four videos. Videos v_1 and v_2 are very similar in both $A_{\hat{\xi}}$ and μ; the fact that v_1 has 4.61x more views is explained by it receiving 3.22x more exogenous promotions. On the same map, v_4 received a similar amount of promotion as v_1 and their differences in popularity are explained by v_4 being less endogenously responsive (smaller $A_{\hat{\xi}}$) than v_1. Moreover, v_3 has a similar endogenous response and sees similar amounts of promotion as v_1; the difference between their popularities is explained by v_3 being less exogenously sensitive, with a lower μ. The endo-exo map provides two distinct aspects from which a video's popularity can be analyzed, which are detailed next.

What describes the most popular videos? One may wonder whether higher popularity can be attributed to higher exogenous sensitivity, higher endogenous response or a combination of both. We examine a collection containing diverse video categories and find that the explanation varies. We draw on the endo-exo map all the videos that belong to the same category in the ACTIVE dataset and we visualize them as two-dimensional density plots. Fig. 4 (c) and (d) compares the plots for the videos in `Gaming` and `Film & Animation`, to that of the top 5% most popular videos in these two categories, respectively. Visibly, while most popular videos in `Film & Animation` are described by higher exogenous sensitivity (shifting upwards), the most popular `Gaming` videos have higher endogenous response – their density mass is shifted to the right of the endo-exo map. Other categories such as `Comedy` or `News & Politics` (shown in the appendix [1]) present two dense regions, one for higher $A_{\hat{\xi}}$ and one for higher μ. These observations show that the most popular videos in different categories differ in terms of the two main factors that drive popularity.

Identifying unpromotable videos. The endo-exo map can be used to readily identify an interesting class of videos: the ones which are very difficult to promote. Given that the quantity $\mu A_{\hat{\xi}}$ describes the number of views that one unit of external promotion (via sharing or tweeting) will generate under the joint influence of endo- and exo- factors – a very small $\mu A_{\hat{\xi}}$ (e.g., $\mu A_{\hat{\xi}} < 1e-3$) is a hallmark of a video

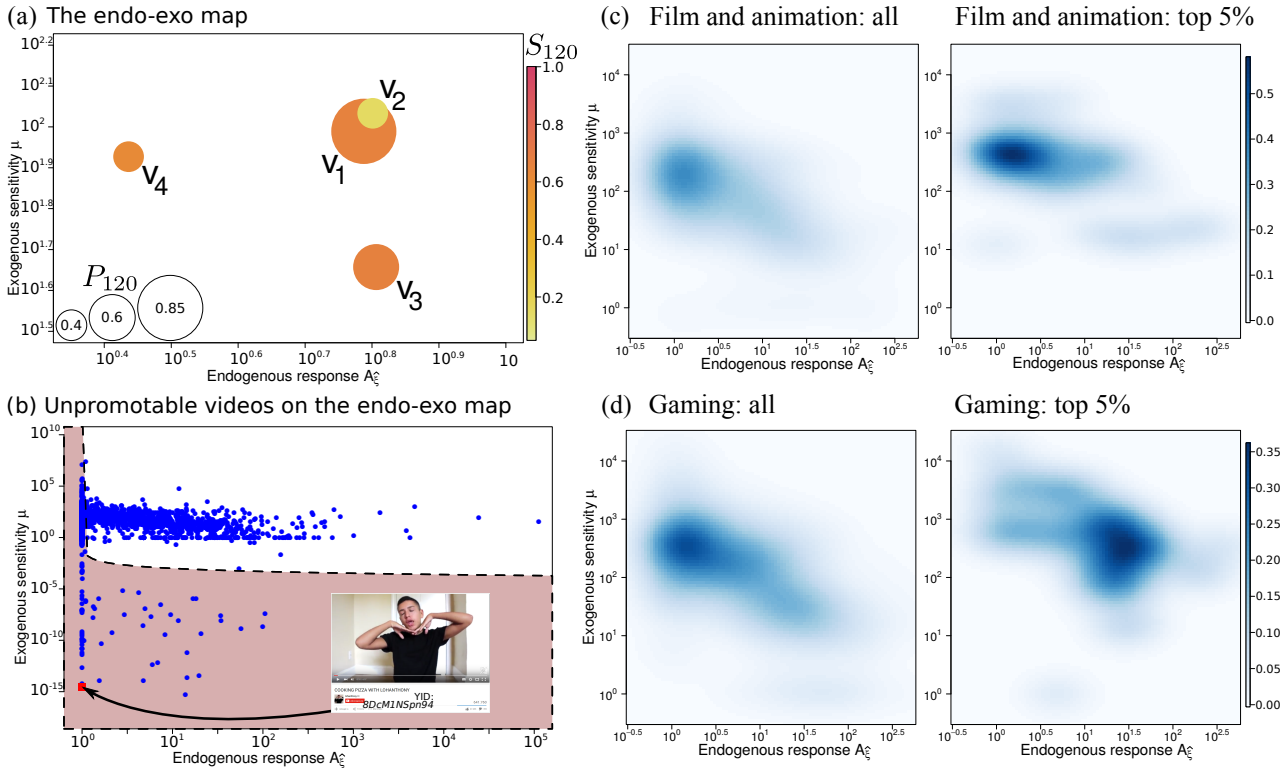


Figure 4: Visualizing video virality and video popularity using the *endo-exo map*. **(a)** Four example videos on the endo-exo map. X-axis $A_{\hat{\xi}}$: the magnitude of endogenous reaction; Y-axis μ: sensitivity to exogenous stimuli. The radius of each circle is proportional to the *popularity percentile* $P_t(\cdot)$ of each video after $t = 120$ days, with values between 0.0 (least popular) and 1.0 (most popular). The color represents the amount (percentile) of total shares received, denoted as $S_t(\cdot)$, with values between 0.0 (no promotion) and 1.0 (receiving the most promotions). v_1 and v_2 present similar endogenous reaction and exogenous sensitivity, being at the same position on the endo-exo map. The difference in their popularity (size) is explained by the fact that v_1 received 3.22 times more promotions than v_2. Both v_3 and v_4 receive similar amounts of promotion (color) as v_1, but they achieve lower popularity (smaller size) due to their less privileged position on the endo-exo map: v_3 is less sensitive to external stimuli than v_1 and v_2, while v_4 has a smaller endogenous reaction than v_1 and v_2. Information about the four example videos are as follows, with their popularity percentile P_{120} and shares percentile S_{120}: v_1 is a short `Gaming` video, YoutubeID *0lTTWeavl1c*, $P_{120}(634,370 \text{ views}) = 85\%$, $S_{120}(351 \text{ shares} = 65\%$; v_2 is a collection of "ALS ice bucket challenge" videos, YoutubeID *3hSIh-tbiKE*, $P_{120}(137,481) = 40\%$, $S_{120}(109) = 10\%$; v_3 is a funny science video, explaining types of infinity in math, YoutubeID *23I5GS4JiDg*, $P_{120}(193,052) = 60\%$, $S_{120}(356) = 65\%$; v_4 is from a Portuguese youtuber, YoutubeID *0ndmJzEIcgU*, $P_{120}(93,959) = 40\%$, $S_{120}(311) = 60\%$. **(b)** A zoomed-out scatter plot of the endo-exo map of the videos in the `People & Blogs` category. The shaded portion of this map consists of videos with low values of total response $\mu A_{\hat{\xi}} < 10^{-3}$ and hence dubbed *unpromotable* videos. Thumbnail of an example video *8DcM1NSpn94* is included, with $\mu = 2.88 \times 10^{-15}$ and $A_{\hat{\xi}} = 1$. **(c)** Density plot for all (left) vs the most popular 5% (right) Film & Animation videos. **(d)** Density plot for all (left) vs the most popular 5% (right) Gaming videos. Popular `Film and Animation` videos tend to have a higher exogenous sensitivity, while those for `Gaming` have mainly a higher endogenous response.

being *unpromotable*. Fig. 4(b) contains a zoomed-out view of the endo-exo map associated with the category `People & Blogs`. We found 63 videos ($\sim 3.2\%$) in this category to be unpromotable. Overall, 549 ($\sim 3.9\%$) videos in the ACTIVE set are deemed unpromotable. The thumbnail of one example video (a teenager video blog) is shown. It has $\mu = 2.88 \times 10^{-15}$ and $A_{\hat{\xi}} = 1$, hence each online promotion is expected to generate 0 views. In contrast, for video v_1 in Fig. 4(a), each promotion is expected to generate 598 views.

5. FORECASTING POPULARITY GROWTH

Via the endo-exo map, the Hawkes intensity process prescribes a video's expected popularity dynamics under external promotions. This section explores the predictive power of such a model. We first illustrate the setting for popularity forecasts using video examples, and then present a quantitative evaluation.

5.1 A video that will go viral

We use HIP to identify videos that are not already popular but have a high potential to become so. This is similar to the phenomenon of delayed recognition in science [21]. Note that this approach is predictive in that we aim to find such potentially viral items before they become popular, rather than a measurement-driven approach that analyzes viral items in past history. Video *1PuvXpv0yDM* in Fig. 5(a) is such an example, it received 15,687 views after being online for 90 days. The HIP model deems it to have a high endogenous response ($A_{\hat{\xi}} = 6.94 \times 10^{72}$) and a high exogenous sensitivity ($\mu = 119.02$). Between days 91 and 120, the video received

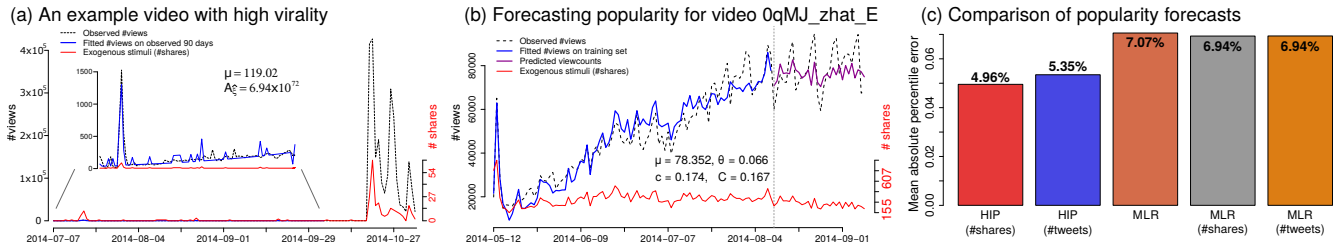


Figure 5: Popularity forecasting using the Hawkes intensity process. (a) Popularity series for video *1PuvXpv0yDM*, explaining a brain disorder. The video receives a total of 36 shares and 15,687 views in the first 90 days (see inset), it is estimated to have a high exogenous sensitivity and a high endogenous response ($\mu = 119.02$, $A_{\xi} = 6.94 \times 10^{72}$). Between day 91 and day 120, this video *jumped* from a popularity percentile of 5.85% to 94.9%, receiving 229 shares and gaining 2.42 million views. (b) Forecasting popularity for video *0qMJ_zhat_E*. Black dotted line: viewcounts series from day 1 to 120 after video upload. Red line: exogenous simuli $s(t)$, also used in parameters estimation. Left of the the gray dashed vertical line at $T \leq 90$ days: time period used for parameters estimation. Blue line: fitted viewcounts for $T \leq 90$ days, generated using Eq 3. Magenta line: viewcount forecast for day 91 to 120. (c) Comparison of average forecasting errors on the ACTIVE set. y-axis: Forecasting errors, calculated as the absolute difference between the popularity percentile at day 120 and that forecasted by each approach. x-axis, left to right: Hawkes intensity model, using either `#shares` or `#tweets` as $s(t)$; multivariate linear regression (MLR), using only popularity history, or `#shares` and `#tweets`, respectively.

an additional 229 shares, more than 6 times the number of shares during its first 90 days. Consequently, the video gained 2.42 million views, drastically improving its ranking on the popularity percentile scale from 5.85% to 94.9%.

5.2 Evaluation of forecast

The HIP model takes as input the exogenous promotion $s[t]$ to produce estimates of the viewcount $\xi[t]$. To construct $\xi[t]$ in the future, $s[t]$ needs to be either estimable or known. We call this *forecasting* popularity, as opposed to predicting popularity where no information about future exogenous stimuli is assumed. Forecasting popularity has broad applications, such as estimating the effect of intended (promotional) interventions, and making decisions about when to promote.

Evaluating popularity forecast on temporal hold-out data. We design a protocol to quantitatively evaluate the predictive power of HIP. We use historical data held-out over time, thus avoiding the practical difficulty of generating realistic promotions and responses in a large-scale social network. Using (known) exogenous promotion $s[t]$, we forecast the popularity $\sum_{91}^{120} \xi[t]$ during the evaluation period (in purple) using Eq (4). Fig. 5(b) illustrates this setting with an example music video. A vertical line divides the observation period, day 1 to 90, and the evaluation period, day 91 to 120. The viewcount and the sharing history in the observation period is used to fit model parameters and explain observed popularity (in blue). For this example, the forecast and the actual views are fairly similar.

Percentile-error metric. We obtain a predicted total viewcount at the end of evaluation period, i.e, $\sum_{1}^{90} \bar{\xi}[t] + \sum_{91}^{120} \xi[t]$, and we evaluate the performances by comparing it to the actual total viewcount $\sum_{1}^{120} \bar{\xi}$. Commonly-used performance error metrics, such as root-mean-square-error (RMSE) or the normalized RMSE, are skewed by the large number of outliers in a long-tailed viewcount distribution and we chose not to use them. Instead, we map the forecasted number views to the popularity scale described in Sec. 3.2, by applying the mapping function $P_{120}(x)$ to normalize the number of views into a metric between 0 and 1. We then compute the absolute error of the predicted percentile. When compared to the error metrics based on the difference in views (like RMSE), this metric focuses on ranking videos correctly with respect to a large collection and is as useful as the broad class of learning to rank applications.

Baseline algorithms. The state-of-the-art approach for popularity prediction uses multivariate linear regression (MLR), based on the observation that historic viewcounts are predictive for future viewcounts [30, 34]. We train linear regressors to predict daily viewcounts for each day between 91 and 120, using a 90-dimensional feature corresponding to the number of views in days 1 to 90. To give the MLR forecast the same amount of information as the HIP model, we build two enhanced baselines, denoted by MLR (#shares) and MLR (#tweets), by introducing the exogenous influence as additional variables, both in the training and in the prediction. Note that the HIP models for each video are learned and evaluated independently, all baselines are trained on ACTIVE and we obtain predictions for each video using cross-validation.

5.3 Forecasting results

Fig. 5(c) summarizes forecasting performance for HIP and the MLR baselines. The forecasts made using HIP have lower average error compared to the linear regression with or without exogenous stimuli (#shares, #tweets). The best forecast obtained an average percentile error of 4.96% (median 3%) for HIP (#shares) and 6.94% (median 3.75%) for the MLR (#shares), corresponding to a 28.6% relative reduction of error. These differences are statistically significant with paired t-test $p < 0.001$, and with a medium effect size according to Cohen's d [12]. Within the HIP variants, we found that using the number of shares generates slightly better forecast than the number of tweets, but the differences are not statistically significant at $p = 0.001$ (more details about effect sizes and statistical tests can be found in the online appendix [1]). We speculate that the difference in forecasting performance is due to the nature of the sources of exogenous excitation: shares capture the promotion behavior via a multitude of environments, whereas tweets count the volume of promotion in Twitter only.

We also observe that the performance gap doubles when forecasting popularity on more difficult videos – videos with a large exogenous shock in the forecasting period, defined as the mean number plus 100 times the standard devia-

tion of the number of shares during the observed period. Fig. 5(a) shows an example of such a video. There are 4006 such videos in the ACTIVE dataset, for which HIP (#shares) achieves a mean percentile error of 5.11% (median 3.25%), whereas MLR (#shares) achieves a mean error of 9.24% (median 6.5%). A typical situation when HIP misses the forecast is when none or very little external influence is recorded during the observed period and during which the popularity is likely to have been driven by unseen exogenous sources.

Lastly, a note about causality: HIP is linear control system with feedback loop, it is *causal* in a linear system sense [29] in that future tweets cannot change past views, but does not directly correspond to the causal inference paradigm about whether a control variable will change a response variable in the presence of other confounding factors. Nonetheless, we conducted statistical tests using the well-known Granger Causality [18] on the shares and view series (details in the appendix [1]); they do not show consistent results for either shares influencing views or vice versa.

6. RELATED WORK

Popularity modeling and prediction. Early measurement studies linked popularity with user influence in Twitter [7, 36] and with the speed and spread of information in social networks [8]. More recently, generative methods, usually based on point-processes, were introduced for popularity modeling [13, 15, 38] and prediction [4, 28]. In their seminal work, Crane and Sornette [13] showed how a Hawkes point-process can account for popularity bursts and decays. Subsequently, more sophisticated models have been proposed to model and simulate popularity in microblogs [38] and videos [15], by accounting for phenomena such as the "rich-get-richer" phenomenon and social contagion. Shen *et al.* [33] employ reinforced Poisson processes, modeling three phenomena: fitness of an item, a temporal relaxation function and a reinforcement mechanism. Zhao *et al.* [39] propose SEISMIC, which employs a double stochastic process, one accounting for infectiousness and the other one for the arrival time of events. TiDeH [23] is an extension of SEISMIC, which aims at estimating future number of views as a function of time, instead of just the final total cascade size. HIP differs from the above applications in two fundamental ways. First, most of the models [4, 23, 28, 31, 39] deal with single diffusion cascades, that is the reaction to single shocks. HIP models popularity as a continuous endogenous-exogenous intertwining, allowing it to closely fit complex evolutions. Second, typical point-process based methods require to observe each individual event during the training period, whereas HIP models volumes of attention directly.

Modeling volumes of popularity. A number of models have been proposed to describe the shape and evolution of the volume of social media activity over time. The seminal meme-tracker [24] system uses a curve with polynomial increase followed by exponential decay to describe sawtooth-shaped volume of news mentions. The SpikeM [27] system uses a fixed memory component, modulated by a periodic component, however it does not explicitly account for external influence. Most recently, Tsytsarau *et al.* [35] model the popularity volume as the convolutions two sequences, news event importance and media response, which are assumed to have predefined shapes. Yang *et al.* [37] propose a generative model to describe sequences that have multiple progression stages along with algorithms to estimate model parameters and to segment existing sequences. Being based a self-excited Hawkes process, HIP simultaneously addresses a series of shortcomings of the above approaches: it is adapted to forecast total popularity, it can recover all parameters from data, and it explains additional, non-stationary variations from linked data sources of external activities.

Influence estimation and maximization are somewhat related research problems, but distinct from the one approached in this paper. Influence estimation [17] aims to learn probabilities of influence between pairs of users, starting from a social graph and a log of actions of its users. Influence maximization [16, 22, 32] finds the subset of users who, if convinced to promote a piece of content, would maximize its diffusion. The main difference between this line of work and HIP is that we measure the volume of promotion and use it to forecast popularity, rather than taking a graph-centric view based on network structure and user interactions.

7. SUMMARY AND DISCUSSION

This research establishes a novel mathematical model to systematically link the endogenous response to the exogenous stimuli of a social system. The model developed here provides a nuanced view of the continued interactions of endogenous and exogenous effects that generate complex and multi-phased popularity dynamics over time. We validate the model on the popularity and promotion history of a large set of YouTube videos. We quantify the endogenous virality and exogenous sensitivity for each video, and we them to explain the properties of the most popular videos, as well as to identify videos that will respond well to promotions and those that will not. Such detailed analysis is possible because the aggregated attention and promotion data are available from YouTube or inferred from public sources such as Twitter. Note however that HIP does not make any platform-dependent assumption and that it can function with any popularity and promotion series generated by aggregated human behavior. We envision that the same kind of attention dynamics would hold for other content types, such as webpage views, podcasts, or blogs.

There are a number of simplifying assumptions and limitations of the proposed model, which can become fruitful directions of further investigation. The Hawkes intensity process captures popularity dynamics that are reflected only in the observed external promotion series, and does not capture other factors such as (daily or weekly) seasonality. What this model also focuses on is the *expected* influence over all users rather than individual influence. Both of these observations suggest extensions that could incorporate seasonality components as well as taking into account individual influences. Lastly, media items are influenced by a variety of sources in the open world and there are many sources of online or offline promotion that are unobserved or difficult to obtain data from. A well-known example is that gaming videos are known to be discussed intensively in topic-specific forums. Tracking and estimating diverse or even unknown sources of exogenous influence is another open research question.

Acknowledgments. This material is based on research sponsored by the Air Force Research Laboratory, under agreement number FA2386-15-1-4018. We thank the National Computational Infrastructure (NCI) for providing computational resources, supported by the Australian Government.

8. REFERENCES

[1] Appendix: Expecting to be HIP: Hawkes intensity processes for social media popularity, 2017. https://arxiv.org/pdf/1602.06033.pdf#page=11.

[2] E. Bakshy, J. M. Hofman, W. A. Mason, and D. J. Watts. Everyone's an influencer. In *WSDM '11*, page 65, feb 2011.

[3] R. Bandari, S. Asur, and B. A. Huberman. The pulse of news in social media: Forecasting popularity. In *Sixth International AAAI Conference on Weblogs and Social Media*, 2012.

[4] P. Bao, H.-W. Shen, X. Jin, and X.-Q. Cheng. Modeling and Predicting Popularity Dynamics of Microblogs using Self-Excited Hawkes Processes. In *WWW*, pages 9–10, 2015.

[5] J. Berger and K. L. Milkman. What makes online content viral? *Journal of marketing research*, 49(2):192–205, 2012.

[6] J. Berger and E. M. Schwartz. What drives immediate and ongoing word of mouth? *Journal of Marketing Research*, 48(5):869–880, 2011.

[7] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi. Measuring User Influence in Twitter: The Million Follower Fallacy. In *ICWSM '10*, volume 10, pages 10–17, 2010.

[8] M. Cha, A. Mislove, and K. P. Gummadi. A measurement-driven analysis of information propagation in the flickr social network. In *WWW*, pages 721–730, 2009.

[9] J. Cheng, L. Adamic, P. A. Dow, J. M. Kleinberg, and J. Leskovec. Can cascades be predicted? In *WWW '14*, pages 925–936. ACM, 2014.

[10] J. Cheng, L. A. Adamic, J. M. Kleinberg, and J. Leskovec. Do cascades recur? In *WWW*, pages 671–681, 2016.

[11] A. Clauset, C. R. Shalizi, and M. E. J. Newman. Power-Law Distributions in Empirical Data. *SIAM Review*, 51(4):661–703, Nov. 2009.

[12] J. Cohen. *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ, 2nd edition, 1988.

[13] R. Crane and D. Sornette. Robust dynamic classes revealed by measuring the response function of a social system. *PNAS '08*, 105(41):15649–15653, oct 2008.

[14] A. De, I. Valera, N. Ganguly, S. Bhattacharya, and M. G. Rodriguez. Learning and forecasting opinion dynamics in social networks. In *NIPS'16*, pages 397–405, 2016.

[15] W. Ding, Y. Shang, L. Guo, X. Hu, R. Yan, and T. He. Video Popularity Prediction by Sentiment Propagation via Implicit Network. In *CIKM '15*, pages 1621–1630, oct 2015.

[16] M. Farajtabar, N. Du, M. G. Rodriguez, I. Valera, H. Zha, and L. Song. Shaping social activity by incentivizing users. In *NIPS'14*, pages 2474–2482, 2014.

[17] A. Goyal, F. Bonchi, and L. V. Lakshmanan. Learning influence probabilities in social networks. In *WSDM '10*, pages 241–250. ACM, 2010.

[18] C. W. Granger. Some recent development in a concept of causality. *Journal of Econom.*, 39(1):199–211, 1988.

[19] A. G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971.

[20] A. Helmstetter and D. Sornette. Subcritical and supercritical regimes in epidemic models of earthquake aftershocks. *Journal of Geophysical Research: Solid Earth*, 107(B10):ESE 10–1–ESE 10–21, 2002.

[21] Q. Ke, E. Ferrara, F. Radicchi, and A. Flammini. Defining and identifying sleeping beauties in science. *PNAS*, 112(24):7426–7431, 2015.

[22] D. Kempe, J. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In *KDD '03*, pages 137–146. ACM, 2003.

[23] R. Kobayashi and R. Lambiotte. TiDeH: Time-Dependent Hawkes Process for Predicting Retweet Dynamics. In *ICWSM 2016*, number ICWSM, 2016.

[24] J. Leskovec, L. Backstrom, and J. Kleinberg. Meme-tracking and the dynamics of the news cycle. In *KDD '09*, pages 497–506. ACM, 2009.

[25] D. C. Liu and J. Nocedal. On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, 45(1-3):503–528, aug 1989.

[26] T. Martin, J. M. Hofman, A. Sharma, A. Anderson, and D. J. Watts. Exploring limits to prediction in complex social systems. In *WWW '16*, pages 683–694, 2016.

[27] Y. Matsubara, Y. Sakurai, B. A. Prakash, L. Li, and C. Faloutsos. Rise and fall patterns of information diffusion: Model and implications. KDD '12, 2012.

[28] S. Mishra, M.-A. Rizoiu, and L. Xie. Feature Driven and Point Process Approaches for Popularity Prediction. In *CIKM '16*, page 10, 2016.

[29] A. Oppenheim, A. Willsky, and S. Nawab. *Signals and Systems*. Prentice Hall, 1997.

[30] H. Pinto, J. M. Almeida, and M. A. Gonçalves. Using early view patterns to predict the popularity of youtube videos. In *WSDM '13*, pages 365–374. ACM, 2013.

[31] J. C. L. Pinto, T. Chahed, and E. Altman. Trend detection in social networks using hawkes processes. In *ASONAM '15*, pages 1441–1448, 2015.

[32] V. Raghavan, G. Ver Steeg, A. Galstyan, and A. G. Tartakovsky. Modeling Temporal Activity Patterns in Dynamic Social Networks. *IEEE Transactions on Computational Social Systems*, 1(1):89–107, mar 2014.

[33] H.-W. Shen, D. Wang, C. Song, and A.-L. Barabási. Modeling and Predicting Popularity Dynamics via Reinforced Poisson Processes. In *AAAI*, page 291, 2014.

[34] G. Szabo and B. A. Huberman. Predicting the popularity of online content. *Com. of the ACM*, 53(8):80–88, 2010.

[35] M. Tsytsarau, T. Palpanas, and M. Castellanos. Dynamics of news events and social media reaction. KDD '14, 2014.

[36] J. Weng, E.-P. Lim, J. Jiang, and Q. He. Twitterrank: finding topic-sensitive influential twitterers. In *WSDM '10*, pages 261–270. ACM, 2010.

[37] J. Yang, J. McAuley, J. Leskovec, P. LePendu, and N. Shah. Finding progression stages in time-evolving event sequences. In *WWW '14*, pages 783–794, 2014.

[38] L. Yu, P. Cui, F. Wang, C. Song, and S. Yang. Uncovering and predicting the dynamic process of information cascades with survival model. *Know. and Inf. Syst.*, page 10, 2016.

[39] Q. Zhao, M. A. Erdogdu, H. Y. He, A. Rajaraman, and J. Leskovec. SEISMIC: A Self-Exciting Point Process Model for Predicting Tweet Popularity. In *KDD '15*, 2015.