Engagement and Popularity Dynamics of YouTube Videos and Sensitivity to Meta-Data

William Hoiles, *Student Member, IEEE*, Anup Aprem, and Vikram Krishnamurthy, *Fellow, IEEE*

Abstract—YouTube, with millions of content creators, has become the preferred destination for viewing videos online. Through the Partner program, YouTube allows content creators to monetize their popular videos. Of significant importance for content creators is which meta-level features (title, tag, thumbnail, and description) are most sensitive for promoting video popularity. The popularity of videos also depends on the social dynamics, i.e., the interaction of the content creators (or channels) with YouTube users. Using real-world data consisting of about 6 million videos spread over 25 thousand channels, we empirically examine the sensitivity of YouTube meta-level features and social dynamics. The key meta-level features that impact the view counts of a video include: first day view count, number of subscribers, contrast of the video thumbnail, Google hits, number of keywords, video category, title length, and number of upper-case letters in the title, respectively, and illustrate that these meta-level features can be used to estimate the popularity of a video. In addition, optimizing the meta-level features after a video is posted increases the popularity of videos. In the context of social dynamics, we discover that there is a causal relationship between views to a channel and the associated number of subscribers. Additionally, insights into the effects of scheduling and video playthrough in a channel are also provided. Our findings provide a useful understanding of user engagement in YouTube.

Index Terms—YouTube, social media, sensitivity analysis, metadata, user engagement, channel dynamics, popularity prediction, Granger causality, machine learning

1 Introduction

The YouTube social network contains over 1 billion users who collectively watch millions of hours of YouTube videos and generate billions of views every day. Additionally, users upload over 300 hours of video content every minute. YouTube generates billions in revenue through advertising and through the Partner program shares the revenue with the content creators.

The video view count is a key metric of the measure of popularity or "user engagement" of a video and the metric by which YouTube pays the content providers. A key question is: *How do meta-level features of a posted video (e.g., thumbnail, title, tags, description) drive user engagement in the YouTube social network?* However, the content alone does not influence the popularity of a video. YouTube also has a social network layer on top of it's media content. The main social component is how the content creators (also called "channels") interact with the users. So another key question is: *How does the interaction of the YouTube channel with the user affect popularity of videos?* In this paper, we study both the above questions. In particular, our aim is to examine how the individual video features (through the meta-level data) and the social dynamics contribute to the popularity of a video.

- *W. Hoiles and A. Aprem are with the Department of Electrical & Computer Engineering, University of British Columbia, Vancouver, BC V6T 1Z4, Canada. E-mail: {whoiles, aaprem}@ece.ubc.ca.*
- *V. Krishnamurthy is with the Department of Electrical & Computer Engineering and Cornell Tech, Cornell University, Ithaca, NY 14850. E-mail: vikramk@cornell.edu.*

Manuscript received 2 June 2016; revised 13 Jan. 2017; accepted 11 Mar. 2017. Date of publication 15 Mar. 2017; date of current version 1 June 2017. Recommended for acceptance by J. Tang.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TKDE.2017.2682858

1.1 Literature Review

The study of popularity of YouTube videos based on meta-level features is a challenging problem given the diversity of users and content providers. Several types of parametric models are used to characterize the popularity of YouTube videos, where the view count time series is used to estimate the model parameters. For example, ARMA time series models [1], multivariate linear regression models [2], modified Gompertz models [3], [4], have been utilized to estimate the future video view counts given past view count time series. Using only the title of the video (one of the meta-level features) [5] considers the problem of predicting whether the view count will be high or low. In a related context, [6], [7] studied the importance of tags for Flicker data. Aside from text based meta-level features (title and tags), in [8] Support Vector Regression (SVR) is proposed to predict the popularity using features of the video frames (e.g., face present, rigidity, color, clutter). It is illustrated in [8] that using the combination of visual features and temporal dynamics results in improved performance of the SVR for predicting view count compared to using only visual features or temporal dynamics alone. In the social context, the uploading behaviour of YouTube content creators was studied in [9]. Specifically, the paper finds that YouTube users within a social network are more popular compared to other users.

1.2 Main Results

In this paper, we investigate how the meta-level features and the interaction of the YouTube channel with the users

affect the popularity of videos. For convenience we summarize the main empirical conclusions of this paper:

1) The five dominant meta-level features that affect the popularity of a video are: first day view count , number of subscribers, contrast of the video thumbnail, Google hits, and number of keywords. Section 2 discusses this further.
2) Optimizing the meta-level features (e.g., thumbnail, title, tags, description) after a video has been posted increases the popularity of the video. In addition, optimizing the title increases the traffic due to YouTube search, optimizing the thumbnail increases the traffic from related videos and optimizing the keywords increases the traffic from related and promoted videos. Section 2.4 provides details on this analysis.
3) Insight into the causal relationship between the subscribers and view count for YouTube channels is explored. For popular YouTube channels, we found that the channel view count affects the subscriber count, see Section 3.1.
4) New insights into the scheduling dynamics in YouTube gaming channels are also found. For channels with a dominant periodic uploading schedule, going "off the schedule" increases the popularity of the channel, see Section 3.2.
5) The generalized Gompertz model can be used to distinguish views due to virality (views from subscribers), migration (views from non-subscribers) and exogenous events, see Section 3.3.
6) New insights into playlist dynamics. The early view count dynamics of a YouTube videos are highly correlated with the long term "migration" of viewers to the video. Also, early videos in a game playthrough typically contain higher views compared with later videos in a game playthrough playlist, see Section 3.4.
7) The number of subscribers of a channel only affects the early view count dynamics of videos in a playthrough, see Section 3.4.

All the above results are validated on a YouTube dataset consisting of over 6 million videos across 25 thousand channels. This dataset[1] was provided to us by BroadbandTV Corp. (BBTV). The dataset consists of daily samples of metadata of the YouTube videos on the BBTV platform from April, 2007 to May, 2015. BBTV is one of the largest Multichannel network (MCN) in the world.[2] The results of the paper allows YouTube partners such as BBTV to adapt their user engagement strategies to generate more views and hence increase revenue.

Caveat. It is important to note that the above empirical conclusions are based on the BBTV dataset. These videos cover the YouTube categories of gaming, entertainment, food, music, and sports as described in Table 6 of the Appendix. Whether the above conclusions hold for other types of YouTube videos is an open issue that is beyond the scope of this paper.

1. The Appendix summarizes the key features of the YouTube dataset that we have used.

2. http://variety.com/2016/digital/news/broadbandtv-mcn-disney-maker-comscore-1201696857/

The organization of the paper is as follows. In Section 2, we use several machine learning methods to characterize the sensitivity of meta-level features on the popularity of YouTube videos. In Section 3, we use time series analysis methods to investigate how the interaction of content creators with users affect the popularity of videos. Using Granger causality, we determine the causal relation between view count and subscribers for channels in Section 3.1. Section 3.2 studies the scheduling dynamics of YouTube channels. Section 3.3 addresses the problem of separating the view count dynamics due to virality (view count resulting from subscribers), migration (views from non-subscribers) and exogenous events (events other than meta-level optimization considered in Section 2), which affect the popularity of the videos using a generalized Gompertz model. In Section 3.4, we study the playlist dynamics in YouTube.

2 Sensitivity Analysis of YouTube Meta-Level Features: A Machine Learning Approach

In this section we apply machine learning methods to study how meta-level features of YouTube videos impacts the view count of the video. The machine learning methods we utilize include: the Extreme Learning Machine (ELM) [10], [11], Feed-Forward Neural Network (FFNN) [12], Stacked Auto-Encoder Deep Neural-Network [13], [14], Elasticnet [15], Lasso, Relaxed Lasso [16], Quantile Regression with Lasso [17], Conditional Inference Random Forest (CIRF) [18], Boosted Generalized Additive Model [19], [20], Bagged MARS using gCV Pruning [21], Generalized Linear Model with Stepwise Feature Selection using Akaike information criterion, and Spike and Slab Regression [22]. Additionally we use the feature selection method Hilbert-Schmidt Independence Criterion Lasso (HSIC-Lasso) [23] to study the sensitivity of meta-level features which may be highly correlated. Note that the meta-level features used for prediction in the YouTube dataset contain significant noise. For example, Fig. 1 illustrates a trace of the subscribers when the video was posted, and the associated viewcount 14 days after the video has been posted. Therefore the machine learning algorithms used must be able to address this challenging problem of mapping from these type of noisy meta-level features to the associated view count of a video. Of these methods we found that the ELM provides sufficient performance to both be used to estimate the meta-level features which significantly contribute to the view count of a video, and for predicting the view count of videos.

2.1 Extreme Learning Machine

The dataset of features (described in Section 2.3) and view count are denoted as $\mathcal{D} = \{(x_i, v_i)\}_{i=1}^{N}$ where $x_i \in \mathbb{R}^m$ is the feature vector, of dimension m, for video i, and v_i is the total view count for video i. Here, N is the number of videos in the training dataset (The ELM was trained for three categories of videos, for details see Section 2.3). The ELM is a single hidden-layer feed-forward neural network–that is, the ELM consists of an input layer, a single hidden layer of L neurons and an

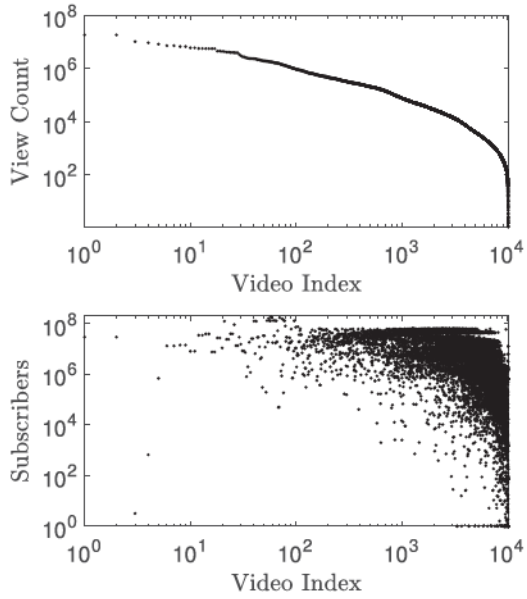


Fig. 1. The top figure shows the view count of all videos (arranged according to decreasing order of view count) after 14 days of the video being posted. The bottom figure shows the associated subscriber count when the video was posted.

output layer. Each hidden-layer neuron can have a unique transfer function. Popular transfer functions include the sigmoid, hyperbolic tangent, and Gaussian. However any non-linear piecewise continuous function can be utilized. The output layer is obtained by a weighted linear combination of the output of the L hidden neurons.

The ELM model presented in [24], [25] is given by

$$v_i = \sum_{k=1}^{L} \beta_k h_k(x_i; \theta_k), \tag{1}$$

where β_k is the weight of neuron k, and $h_k(\cdot; \theta_k)$ is the hidden-layer neuron transfer function with parameter θ_k, and L is the total number of hidden-layer neurons in the ELM. Given $\mathcal{D}$, how can the ELM model parameters β_k, θ_k, and L in (1) be selected? Given L, the ELM trains β_k and θ_k in two steps. First, the hidden layer parameters θ_k are randomly initialized. Any continuous probability distribution can be used to initialize the parameters θ_k. Second, the parameters β_k are selected to minimize the square error between the model output and the measured output from $\mathcal{D}$. Formally,

$$\beta^* \in \underset{\beta \in \mathbb{R}^L}{\operatorname{argmax}} \left\{ ||H\beta - V||_2^2 \right\}, \tag{2}$$

where H denotes the hidden-layer output matrix with entries $H_{kj} = h_k(x_j; \theta_k)$ for $k \in \{1, 2, \ldots, L\}$ and $j \in \{1, 2 \ldots, N\}$, and V the target output with entries $V = [v_1, v_2, \ldots, v_N]$. The solution to (2) is given by $\beta^* = H^{\dagger}V$ where $H^{\dagger}$ denotes the Moore-Penrose generalized inverse of H. The major benefit of using the ELM, compared to other single layer feed-forward neural network, is that the training only requires the random generation of the parameters θ_k, and the parameters β_k can be computed as the solution of a set of linear equations. The computational cost of training the ELM is $O(N^3)$ for constructing the Moore-Penrose inverse [26].

2.2 Sensitivity Analysis (Background)

There are several sensitivity analysis techniques available in the literature [27], [28] which can be classified into two groups: filter methods, and wrapper methods. The filter methods consider only the meta-level features and the viewcount without the information available from a machine learning algorithm. The wrapper methods, on the other hand, utilize the information from the machine learning algorithm. Typically, wrapper methods give a more accurate measure of the sensitivity compared to filter methods [27], [28]. However, filter methods are computationally less expensive than wrapper methods and do not require the training and evaluation of the machine learning algorithm. Given the noise present in the meta-level features (Fig. 1) and the non-linearity between the meta-level features and view count , filter methods are not suitable for the sensitivity analysis of the meta-level features. Hence, in this section we focus on two wrapper methods suitable for estimating the sensitivity of meta-level features on the view count of YouTube videos.

For the first method we focus on the ELM (1) for evaluating the sensitivity of the meta-level features, however the method can be used for any machine learning method. Given that the ELM (1) is a single feed-forward hidden layer neural network, it is possible to evaluate the sensitivity of the meta-level features by taking the partial derivative of (1) for the trained ELM. Note that this method is utilized to estimate the sensitivity of input features in neural networks [29]. The *sum of squares derivatives*, denoted by SSD_k for meta-level feature $x(k)$, is given by

$$\text{SSD}_k = \sum_{i=1}^{N} \left(\frac{\partial v_i}{\partial x(k)} \right)^2 = \sum_{i=1}^{N} \left(\sum_{k=1}^{L} \beta_k \frac{\partial h_k(x_i; \theta_k)}{\partial x(k)} \right)^2. \tag{3}$$

The variable with the largest SSD_k is most influential to the prediction of the view count using the ELM v (1). Note that since the ELM is trained using all the meta-level features, the SSD_k evaluates the average sensitivity of changes in a single meta-level feature with all other features held constant.

To account for significant interdependency relationships between meta-level features requires sophisticated methods to evaluate the meta-level feature sensitivities. A state-of-the art method which can be used for this task is the Hilbert-Schmidt Independence Criterion Lasso [23]. The main idea of this method is to use the benefits of least absolute shrinkage and selection operator (Lasso) with a feature wise kernel to capture the non-linear input-output dependency. A measure of the importance of a meta-level feature is then given by the coefficient of the centered Gram matrix used in the HSIC-Lasso.

Both of these methods will be applied to the YouTube dataset to study the sensitivity of the meta-level features of YouTube videos on the videos view count .

2.3 Sensitivity of YouTube Meta-Level Features and Predicting View Count

In this section, the ELM (1) and other state-of-the art machine learning methods are applied to the YouTube dataset to compute the sensitivity of a videos meta-level features on the view count of the video based on the feature importance measure SSD_k (3). Videos of different popularity, (i.e., highly

popular, popular, and unpopular as defined in Table 7 in the Appendix), may have different sensitivities to the meta-level features. Hence, in this paper, we independently perform the sensitivity analysis on the three popularity categories. First we define the meta-level features for each video, then evaluate the meta-level feature sensitivities on the associated view count , and finally provide methods to predict the view count of YouTube videos using various machine learning techniques. The analysis provides insight into which meta-level features are useful for optimizing the view count of a YouTube video.

2.3.1 Meta-Level Feature Construction

Each YouTube video contains four primary components: the Thumbnail of the video, the Title of the video, the Keywords (also known as tags), and the description of the video. However, in typical user searches only a subset of the description is provided to the user. Therefore, we do not consider the contents of the description to significantly affect the view count of the video. The meta-level features are constructed using the Thumbnail, Title, and Keywords. For the Thumbnail, 19 meta-level features are computed which include: the blurriness (e.g., CannyEdge, Laplace Frequency), brightness, contrast (e.g., tone), overexposure, and entropy of the thumbnail. For the Title, 23 meta-level features are computed which include: word count, punctuation count, character count, Google hits (e.g., if the title is entered into the Google search engine how many results are found), and the Sentiment/Subjectivity of the title computed using Vader [30], and TextBlob.[3] For the Keywords, seven meta-level features are computed which include: the number of keywords, and keyword length. In addition, to the above 49 meta-level features, we also include auxiliary user meta-level features including: the number of subscribers, resolution of the thumbnail used, category of the video, the length of the video, and the first day view count of the video. Note that our analysis does not consider the video or audio quality of the YouTube video. Our analysis is focused on the sensitivity of the view count based on the Thumbnail, Title, Keywords, and auxiliary channel information of the user that uploaded the video. In total 54 meta-level features are computed for each video. The complete dataset used for the sensitivity analysis is given by $\mathcal{D} = \{(x_i, v_i)\}_{i=1}^{N}$, with $x_i \in \mathbb{R}^{54}$ the computed meta-level features for video $i \in \{1, \ldots, N\}$, v_i the view count 14 days after the video is published, and $N = 10^4$, the total number of videos used for the sensitivity analysis. Note that the view count v_i is on the log scale (i.e., if a video has 10^6 views then $v_i = 6$). This is a necessary step as the range of view counts is from 10^2 to above 10^7.

Prior to performing any analysis, we pre-process the meta-level features in the dataset $\mathcal{D}$. First, all the meta-level features are scaled to satisfy $x(k) \in [0, 1]$. Note that the meta-level features were not whitened (e.g., the meta-level data as not transformed to have an identity covariance matrix). The second pre-processing step involves removing redundant features in $\mathcal{D}$. Feature selection is a popular method for eliminating redundant meta-level features. In this paper we employ a correlation based feature selection based on the Pearson correlation coefficient (which was used for feature selection in [31]) to eliminate the redundant meta-level features. Of the original 54 meta-level features, $m = 29$ meta-level features remain after the removal of the correlated meta-level features. Note that removal of these features does not significantly impact the performance of the machine learning algorithms or the sensitivity analysis results.

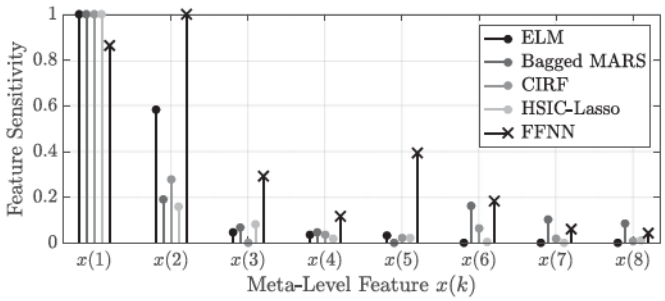


Fig. 2. Sensitivity of the meta-level features computed using the sum of squares derivatives SSD_k (3) for the ELM, Bagged MARS, CIRF, and FFN, and the associated coefficient of the centered Gram matrix for the HSIC-Lasso using the dataset $\mathcal{D}$ defined in Section 2.3.1. The meta-level features $k = 1$ to $k = 8$ are associated with: first day view count, number of subscribers, contrast of the video thumbnail, Google hits, number of keywords, video category, title length, and number of uppercase letters in the title, respectively. Similar results are obtained for highly popular, popular, and unpopular videos as defined in Table 7.

2.3.2 Meta-Level Feature Sensitivity

Given the dataset $\mathcal{D} = \{(x_i, v_i)\}_{i=1}^{N}$ constructed in Section 2.3.1, the goal is to estimate which features significantly contribute to the view count of a video. To perform this sensitivity analysis five machine learning algorithms which include: the ELM, Bagged MARS using gCV Pruning [21], Conditional Inference Random Forest [18], Feed-Forward Neural Network [12], and the feature selection method Hilbert-Schmidt Independence Criterion Lasso [23]. Each of these models is trained using a 10-fold cross validation technique, and the design parameters of each was optimized via extensive empirical evaluation. We selected the ELM (1) to contain $L = 100$ neurons which ensures that we have sufficient accuracy on the predicted view count given the features x_i, while reducing the effects of over-fitting. For the CIRF the design parameter for randomly selected predictors was set to 6, and the FFNN we have 10 neurons in the hidden-layer. The HSIC-Lasso regularization parameter was set to 100. Given the trained models, the sensitivity of the view count on the meta-level features of a video is computed by evaluating the sum of squares derivatives, SSD_k (3). Fig. 2 shows the normalized[4] SSD_k for the five highest sensitivity meta-level features of these five machine learning methods. Note that for the HSIC-Lasso we do not use the SSD_k but instead the values of the coefficient of the centered Gram matrix computed from kth feature which provides an estimate of the feature sensitivity. Recall, from Section 2.2, that larger the SSD_k value or higher the coefficient of the centered Gram matrix the more sensitive the view count is to variations in the meta-level feature. From Fig. 2, the meta-level features with the

3. http://textblob.readthedocs.io/en/dev/

4. The normalization is with respect to the highest value among the computed SSD_k.

highest sensitivities are: first day view count, number of subscribers, contrast of the video thumbnail, Google hits, number of keywords, video category, title length, and number of upper-case letters in the title respectively. Notice that all these methods have the first day view count and number of subscribers as the most sensitive meta-level features as expected. The FFNN and Bagged MARS however do not have the contrast of the video thumbnail as the third most sensitive meta-level feature compared with the other algorithms. This results as the learning method and learning rate of each of these algorithms is different which results in differences in the meta-level feature sensitivity. However as we can see from Fig. 2, the view count of a video is dependent on these eight meta-level features with the first day view count and number of subscribers being the most sensitive features.

As expected, Fig. 2 shows that if the first day view count is high then the associated view count 14 days after the video is posted will be high. Additionally, if there is a large number of subscribers to the channel that posted the video, then the associated view count after 14 days is also expected to be large. As expected, the properties of the title and keywords also contribute to the view count of the video however with less sensitivity then the thumbnail of the video. Therefore, to increase the view count of a video it is vital to increase the number of subscribers, and focus on the quality of the Thumbnail used. A surprising result is that the sensitivity of the view count resulting from changes in these meta-level features are negligible across the three popularity classes of videos (i.e., highly popular, popular, and unpopular as defined in Table 7). Therefore, regardless of the expected popularity of a video, a channel owner should focus on maximizing the number of subscribers and the quality of the thumbnail to increase the associated view count of a video.

2.3.3 Predicting the View Count of YouTube Videos

In this section we illustrate how machine learning methods can be used to the view count of a YouTube video. The machine learning methods used for prediction include: the Extreme Learning Machine (1), Feed-Forward Neural Network [12], Stacked Auto-Encoder Deep Neural-Network [13], [14], Elasticnet [15], Lasso, Relaxed Lasso [16], Quantile Regression with Lasso [17], Conditional Inference Random Forest [18], Boosted Generalized Additive Model [19], [20], Bagged MARS using gCV Pruning [21], Generalized Linear Model with Stepwise Feature Selection using Akaike information criterion, and Spike and Slab Regression [22]. For each method their predictive performance and the top-five highest sensitivity meta-level features are provided.

To perform the analysis we train each model using an identical 10-fold cross validation method with the dataset $\mathcal{D} = \{(x_i, v_i)\}_{i=1}^{N}$ with all the meta-level features included. The predictive performance of the machine learning methods are evaluated using the root-mean-square error (RMSE) and the R^2 (e.g., coefficient of determination). Note that for both training and evaluation the view count is pre-processed to be on the log scale (i.e., if the view count is 10^6, the associated label is $v_i = 6$).

TABLE 1
Performance and Feature Sensitivity

Method	RMSE	R^2	Features $x(k)$				
Extreme Learning Machine	**0.44**	0.77	1	2	3	4	5
Feed-Forward Neural Network	0.48	0.79	2	1	5	3	6
Stacked Auto-Encoder DNN	0.59	0.66	1	2	3	4	5
Elasticnet	0.57	0.64	1	2	3	4	5
Lasso	0.53	0.66	1	2	3	4	5
Relaxed Lasso	1.14	0.64	1	2	3	4	5
Quantile Regression with Lasso	0.60	0.62	1	2	3	4	5
CI Random Forest	0.47	**0.80**	1	2	6	4	5
Boosted GAM	0.50	0.77	1	2	6	4	5
Bagged MARS	0.50	0.77	1	2	6	7	8
GLM with Feature Selection	0.53	0.67	1	2	3	4	5
Spike and Slab Regression	0.53	0.67	1	2	3	4	5

The predictive performance and the top-five highest sensitivity meta-level features of the machine learning methods are provided in Table 1. In Table 1 the meta-level feature numbers are identical to those defined in Fig. 2. As seen from Table 1, the ELM has the lowest RMSE of 0.44 which is comparable to the RMSE of the Conditional Inference Random Forest and Feed-Forward Neural Network which have 0.47 and 0.48 respectively. The R^{25} of the ELM, Feed-Forward Neural Network and Conditional Inference Random Forest are also comparable with values of 0.77, 0.79, and 0.80. Therefore any of these methods could be used to estimate the view count of a YouTube video. A key question is which of the meta-level features $x(k)$ are most sensitive between these machine learning methods. As seen from the results in Table 1 the top two most important features are the first day view count and the number of subscribers, and the majority of methods suggest that the number of Google hits is also an important meta-level feature. Interestingly the Conditional Inference Random Forest, Boosted Generalized Additive Model, and the Bagged MARS using gCV Pruning do not consider the number of Google hits in the top five most sensitive features and instead use the video category. This is consistent with the result that videos in the "Music" category are the most viewed on YouTube, followed by "Entertainment" and "People and Blogs". Only the Bagged MARS using gCV Pruning considers the meta-level features of title length and number of upper-case letters in the title to be in the five most sensitive features compared with the other machine learning methods. This result suggests that the number of Google hits associated with the title significantly contributes to the video's popularity however the view count is not very sensitive to the specific length and number of upper-case letters in the title. Therefore, when performing meta-level feature optimization for a video a user should focus on the meta-level features of: first day view count, number of subscribers, contrast of the video thumbnail, Google hits, number of keywords, and video category.

5. The R^2 is a popular measure of the goodness of fit. It is given by the ratio of the variation (measured using sum of squares) explained using a model to the total variation in the data. The important property of R^2 is that it is bounded between 0 and 1. A high value of R^2 implies that the variation in the data can be explained using the model in question.

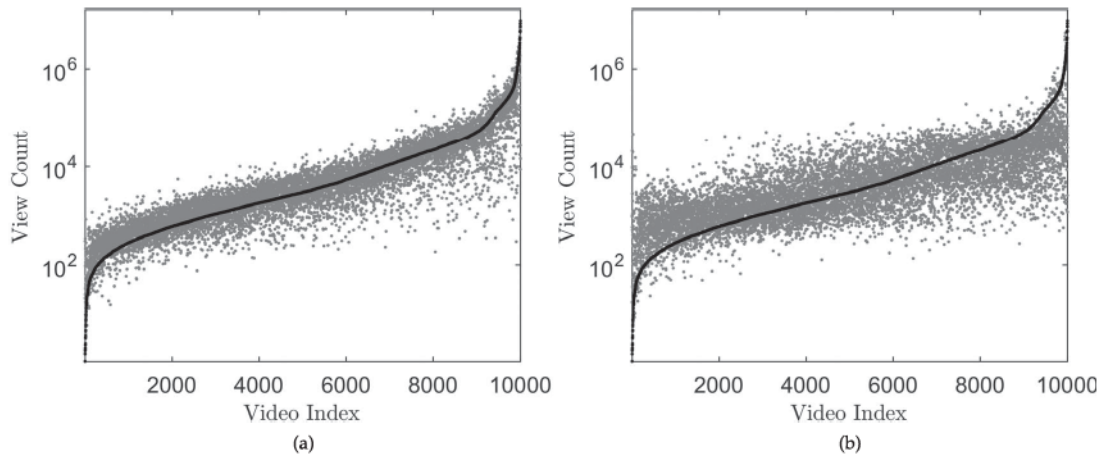


Fig. 3. Predictive view count (gray dots) using an ELM versus the actual view count (black dots). Fig. 3a illustrates the results for a trained ELM (1) using all 29 meta-level features defined in Section 2.3. Fig. 3b illustrates the results for a trained ELM (1) using the 28 meta-level features (first day view count removed from the 29 meta-level features defined in Section 2.3).

To estimate the view count of an unpublished video (a video that is about to be posted for the first time) we can not utilize the most sensitive meta-level feature of the machine learning algorithms which is the first day view count. Is it still possible to estimate the view count with the remaining meta-level features? To answer this question we compare the performance of the ELM using the 28 meta-level features with the view count on the first day removed. Fig. 3a shows the predicted view count of the ELM trained using 29 meta-level features, and Fig. 3b shows the predicted view count using the 28 meta-level features. As expected, Fig. 3 illustrates that the predictive accuracy of the ELM decreases if the view count on the first day is removed. Though there is a drop in the predictive accuracy of the ELM trained using the 28 meta-level features, it still contains sufficient predictive accuracy to aid in the selection of the meta-level features to increase the view count of a video. Note that similar performance results are obtained for the Feed-Forward Neural Network and Conditional Inference Random Forest when performing the prediction with the first day view count removed. Therefore, these prediction methods can be used to optimize the meta-level features of unpublished videos where the optimization can focus on the meta-level features of: number of subscribers, contrast of the video thumbnail, Google hits, number of keywords, and video category.

2.4 Sensitivity to Meta-Level Optimization

Section 2.3 described how meta-level features (e.g., number of subscribers) can be used to estimate the popularity of a video. In this section, we analyze how changing meta-level features, after a video is posted, impacts the user engagement of the video. Meta-level data plays a significant role in the discovery of content, through YouTube search, and in video recommendation, through the YouTube related videos. Hence, "optimizing" the meta-level data to enhance the discoverability and user engagement of videos is of significant importance to content providers. Therefore, in this section, we study how optimizing the title, thumbnail or keywords affect the view count of YouTube videos.

To perform the analysis, we utilize the dataset (see Table 8 in the Appendix), and remove any time-sensitive videos. Time-sensitive videos are those videos that are relevant for a short period of time and the popularity of such videos cannot be improved by optimization. We removed the following two time-sensitive categories of videos: "politics" and "movies and trailers". In addition, we removed videos (from other categories) which contained the following keywords in their video meta-data: "holiday", "movie", or "trailers". For example, holiday videos are not watched frequently during off-holiday times.

Let $\hat{\tau}_i$ be the time at which the meta-level optimization was performed on video i and let s_i, denote the corresponding sensitivity. We characterize the sensitivity to meta-level optimization as follows:

$$s_i = \frac{\left(\sum_{t=\hat{\tau}_i}^{\hat{\tau}_i+6} v_i(t)\right)/7}{\left(\sum_{t=\hat{\tau}_i-6}^{\hat{\tau}_i} v_i(t)\right)/7}. \tag{4}$$

The numerator of (4) is the mean value of the view count 7 days after optimization. Similarly, the denominator of (4) is the mean value of the view count 7 days before optimization. The results are provided in Table 2 for optimization of the title, thumbnail, and keywords.

As shown in Table 2, at least half of the optimizations resulted in an increase in the popularity of the video. In addition, compared to videos with no optimization, the meta-level optimization improves the probability of increased popularity by 45 percent. This is consistent with YouTube and BBTV recommendation to optimize meta-level features to increase user engagement. However, some class of videos benefit from optimizing meta-data much more than others. The effect may be due to small user channels, which have limited number of videos and subscribers, gain by optimizing the meta-level data of the video compared to hugely popular channels such as Sony or CNN. The highly popular channel (e.g., Sony or CNN) upload videos frequently (even multiple times daily), so video content becomes irrelevant quickly. The question of which class of users gain by optimizing the meta level features of the video is part of our ongoing research.

TABLE 2
Sensitivity to Meta-Level Optimization

Optimization	Fraction of Videos with increased popularity
Title change	0.52
Thumbnail change	0.533
Keyword change	0.50
No change[6]	0.35

The table shows that for more than 50 percent of the videos, meta-level optimization resulted in an increase in the popularity of the video.

TABLE 3
Sensitivity of Various Traffic Sources to Meta-Level Optimization, for Videos with Increased Popularity

Optimization	Related	Promoted	Search
Title change	1.13	NA[a]	**1.24**
Thumbnail change	**1.20**	NA[a]	1.125
Keyword change	**1.10**	**1.16**	1

[a] *Not enough data available: A binomial test to check for the true hypothesis with 95 percent confidence interval requires that the sample size, n, should be at least $(\frac{1.96}{0.04})^2 p(1-p)$. With $p = 0.5$, $n > 600$. The title optimization resulted in significant improvement (approximately 25 percent) from the YouTube search. Similarly, thumbnail optimization improved traffic from the related videos and keyword optimization resulted in increased traffic from related and promoted videos.*

Table 3 summarizes the impact of various meta-level changes on the three major sources of YouTube traffic, i.e., YouTube search,[7] YouTube promoted[8] and traffic from related videos.[9] For those videos where meta-level optimization increased the popularity (the ratio of the mean value of the views after and before optimization is higher than one), we computed the sensitivity for various traffic sources as in (4). Table 3 summarizes the median statistics of the ratio of the traffic sources before and after optimization.

The title optimization resulted in significant improvement (approximately 25 percent) from the YouTube search. Similarly, thumbnail optimization improved traffic from the related videos and keyword optimization resulted in increased traffic from related and promoted videos.

Summary. This section studied the sensitivity of view count with respect to meta-level optimization. The main finding is that meta-level optimization increased the popularity of videos in the majority of cases. In addition, we found that optimizing the title improved traffic from YouTube search. Similarly, thumbnail optimization improved traffic from the related videos and keyword optimization resulted in increased traffic from related and promoted videos.

3 Social Interaction of the Channel with YouTube Users

In this section, we use time series analysis methods to determine how the social interaction of a YouTube channel with its viewers affects the view count dynamics. This section is organized as follows. Section 3.1, characterizes the causal relationship between the subscribers and view count of a channel using Granger causality test. In Section 3.2, we investigate how the popularity of the channel is affected by the scheduling dynamics of the channel. When channels deviate from a regular upload schedule, the view count and the comment count of the channel increase. In Section 3.3, we address the problem of separating the view count dynamics due to virality (viewcount resulting from subscribers) and migration (views from non-subscribers) and exogenous events using a generalized Gompertz model. Finally, Section 3.4, we studies the effect of video playlists on the view count. The main conclusion outlined in Section 3.4 is that the dynamics of the view count in a playlist is highly correlated and the effects of "migration" causes the view count of videos to decrease even with an increase in the subscriber count.

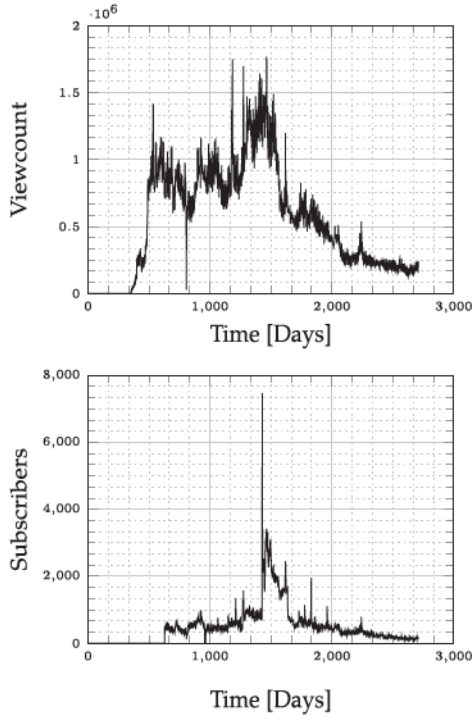


Fig. 4. Viewcount and subscribers for the popular movie trailer channel: VISOTrailers. The Granger causality test for view counts "Granger causes" subscriber count is true with a p-value of 5×10^{-8}.

3.1 Causality Between Subscribers and View Counts in YouTube

In this section the goal is to detect the causal relationship between subscriber and viewer counts and how it can be used to estimate the next day subscriber count of a channel. The results are of interest for measuring the popularity of a YouTube channel. Fig. 4 displays the subscriber and view count dynamics of a popular movie trailer channel in YouTube. It is clear from Fig. 4 that the subscribers "spike" with a corresponding "spike" in the view count. In this section we model this causal relationship of the subscribers and view count using the Granger causality test from the econometric literature [32].

The main idea of Granger causality is that if the value(s) of a lagged time-series can be used to predict another time-series, then the lagged time-series is said to "Granger cause" the predicted time-series. To formalize the

6. "No change" was obtained by randomly selecting 10^4 videos which performed no optimization and evaluating s_i 3 months from the date of posting the video.
7. Video views from YouTube search results.
8. Video views from an unpaid YouTube promotion.
9. Video views from a related video listing on another video watch page.

Granger causality model, let $s^j(t)$ denote the number of subscribers to a channel j on day t, and $v_i^j(t)$ the corresponding view count for a video i on channel j on day t. The total number of videos in a channel on day t is denoted by $\mathcal{I}(t)$. Define,

$$\hat{v}^j(t) = \sum_{i=1}^{\mathcal{I}(t)} v_i^j(t), \tag{5}$$

as the total view count of channel j at time t. The Granger causality test involves testing if the coefficients b_i are nonzero in the following equation which models the relationship between subscribers and view counts

$$s^j(t) = \sum_{k=1}^{n_s} a_k^j s^j(t-k) + \sum_{i=k}^{n_v} b_k^j \hat{v}^j(t-k) + \varepsilon^j(t), \tag{6}$$

where $\varepsilon^j(t)$ represents normal white noise for channel j at time t. The parameters $\{a_i^j\}_{\{i=1,\ldots,n_s\}}$ and $\{b_i^j\}_{\{i=1,\ldots,n_v\}}$ are the coefficients of the AR model in (6) for channel j, with n_s and n_v denoting the lags for the subscriber and view counts time series respectively. If the time-series $\mathcal{D}^j = \{s^j(t), \hat{v}^j(t)\}_{t\in}$ $\{1,\ldots,T\}$ of a channel j fits the model (6), then we can test for a causal relationship between subscribers and view count. In equation (6), it is assumed that $|a_i| < 1$, $|b_i| < 1$ for stationarity. The causal relationship can be formulated as a hypothesis testing problem as follows:

$$H_0 : b_1 = \cdots = b_{n_v} = 0 \text{ versus } H_1 : \text{Atleast one } b_i \neq 0. \tag{7}$$

The rejection of the null hypothesis, H_0, implies that there is a causal relationship between subscriber and view counts.

First, we use Box-Ljung test [33] is to evaluate the quality of the model (6) for the given dataset $\mathcal{D}^j$. If satisfied, then the Granger causality hypothesis (7) is evaluated using the Wald test [34]. If both hypothesis tests pass then we can conclude that the time series $\mathcal{D}^j$ satisfies Granger causality–that is, the previous day subscriber and view count have a causal relationship with the current subscriber count.

A key question prior to performing the Granger causality test is what percentage of videos in the YouTube dataset (Appendix) satisfy the AR model in (6). To perform this analysis we apply the Box-Ljung test with a confidence of 0.95 (p-value = 0.05). First, we need to select n_s and n_v, the number of lags for the subscribers and view count time series. For $n_s = n_v = 1$, we found that only 20 percent of the channels satisfy the model (6). When n_s and n_v are increased to 2, the number of channels satisfying the model increases to 63 percent. For $n_s = n_v = 3$, we found that 91 percent of the channels satisfy the model (6), with a confidence of 0.95 (p-value = 0.05). Hence, in the below analysis we select $n_s = n_v = 3$. It is interesting to note that the mean value of coefficients b_i decrease as i increases indicating that older view counts have less influence on the subscriber count. Similar results also hold for the coefficients a_i. Hence, as expected, the previous day subscriber count and the previous day view count most influence the current subscriber count.

The next key question is does their exist a causal relationship between the subscriber dynamics and the view count dynamics. This is modeled using the hypothesis in (7). To test (7) we use the Wald test with a confidence of 0.95 (p-value = 0.05) and found that approximately 55 percent of the channels satisfy the hypothesis. For approximately 55 percent of the channels that satisfy the AR model (6), the view count "Granger causes" the current subscriber count. Interestingly, if different channel categories are accounted for then the percentage of channels that satisfy Granger causality vary widely as illustrated in Table 4. For example, 80 percent of the Entertainment channels satisfy Granger causality while only 40 percent of the Food channels satisfy Granger causality. These results illustrate the importance of channel owners to not only maximize their subscriber count, but to also upload new videos or increase the views of old videos to increase their channels popularity (i.e., via increasing their subscriber count). Additionally, from our analysis in Section 2 which illustrates that the view count of a posted video is sensitive to the number of subscribers of the channel, increasing the number of subscribers will also increase the view count of videos that are uploaded by the channel owners.

TABLE 4
Fraction of Channels Satisfying the Hypothesis: View Count "Granger Causes" Subscriber Count, Split According to Category

Category [a]	Fraction
Gaming	0.60
Entertainment	0.80
Food	0.40
Sports	0.67

[a] *YouTube assigns a category to videos, rather than channels. The category of the channel was obtained as the majority of the category of all the videos uploaded by the channel.*

3.2 Scheduling Dynamics in YouTube

In this section, we investigate the scheduling dynamics of YouTube channels. We find the interesting property that for popular gaming YouTube channels with a dominant upload schedule, deviating from the schedule increases the views and the comment counts of the channel.

Creator Academy[10] in their best practice section recommends to upload videos on a regular schedule to get repeat views. The reason for a regular upload schedule is to increase the user engagement and to rank higher in the YouTube recommendation list. However, we show in this section that going "off the schedule" can be beneficial for a gaming YouTube channel, with a regular upload schedule, in terms of the number of views and the number of comments.

From the dataset, we 'filtered out' video channels with a *dominant* upload schedule, as follows: The dominant upload schedule was identified by taking the periodogram of the upload times of the channel and then comparing the highest value to the next highest value. If the ratio defined above is greater than 2, we say that the channel has a dominant upload schedule. From the dataset containing 25 thousand channels, only 6,500 channels contain a dominant upload schedule. Some channels, particularly those that contain high amounts of copied videos such as trailers, movie/TV snippets upload videos on a daily basis. These have been removed from the above analysis. The expectation is that by doing so we concentrate on those channels that contain only user generated content.

10. YouTube website for helping with channels.

We found that channels with gaming content account for 75 percent of the 6,500 channels with a dominant upload schedule[11] and the main tags associated with the videos were: "game", "gameplay" and "videogame".[12] We computed the average views when the channel goes off the schedule and found that on an average when the channel goes off schedule the channel gains views 97 percent of the time and the channel gains comments 68 percent of the time. This suggests that channels with "gameplay" content have periodic upload schedule and benefit from going off the schedule.

3.3 Modeling the View Count Dynamics of Videos with Exogenous Events

Several time-series analysis methods have been employed in the literature to model the view count dynamics of YouTube videos. These include ARMA models [1], multivariate linear regression models [2], hidden Markov models [35], normal distribution fitting [36], and parametric model fitting [3], [4]. Though these models provide an estimate of the view count dynamics of videos, we are interested in segmenting view count dynamics of a video resulting from subscribers, non-subscribers and exogenous events. Exogenous events are due to video promotion on other social networking platform such as Facebook or the video being referenced by a popular news organization or celebrity on Twitter. This is motivated by two reasons. First, removing view count dynamics due to exogenous events provides an accurate estimate of sensitivity of meta-level features in Section 2. Second, extracting the view count resulting from exogenous events gives an estimate of the efficiency of video promotion.

The view count dynamics of popular videos in YouTube typically show an initial viral behaviour, due to subscribers watching the content, and then a linear growth resulting from non-subscribers. The linear growth is due to new users migrating from other channels or due to interested users discovering the content either through search or recommendations (we call this phenomenon *migration* similar to [3]). Hence, without exogenous events, the view count dynamics of a video due to subscribers and non-subscribers can be estimated using piecewise linear and non-linear segments. In [3], it is shown that a Gompertz time series model can be modeled the view count dynamics from subscribers and non-subscribers, if no exogenous events are present. In this paper, we generalize the model in [3] to account for views from exogenous events. It should be noted that classical change-point detection methods [37] cannot be used here as the underlying distribution generating the view count is unknown.

To account for the view count dynamics introduced by exogenous events we use the generalized Gompertz model given by

$$\begin{aligned} \bar{v}_i(t) &= \sum_{k=0}^{K_{\max}} w_i^k(t) u(t-t_k), \\ w_i^k(t) &= M_k\Big(1-e^{-\eta_k\left(e^{b_k(t-t_k)}-1\right)}\Big)+c_k(t-t_k), \end{aligned} \tag{8}$$

where $\bar{v}_i(t)$ is the total view count for video i at time t, $u(\cdot)$ is the unit step function, t_0 is the time the video was uploaded, t_k with $k \in \{1,\ldots,K_{\max}\}$ are the times associated with the $K_{\max}$ exogenous events, and $w_i^k(t)$ are Gompertz models which account for the view count dynamics from uploading the video and from the exogenous events. In total there are $K_{\max}+1$ Gompertz models with each having parameters t_k, M_k, η_k, b_k. M_k is the maximum number of requests not including migration for an exogenous event at t_k, η_k and b_k model the initial growth dynamics from event t_k, and c_k accounts for the migration of other users to the video. In (8) the parameters $\{M_k,\eta_k,b_k\}_{k=0}$ are associated with the subscriber views when the video is initially posted, the parameters $\{t_k,M_k,\eta_k,b_k\}_{k=1}^{K_{\max}}$ are associated with views introduced from exogenous events, and the views introduced from migration are given by $\{c_k\}_{k=0}^{K_{\max}}$. Each Gompertz model (8) captures the initial viral growth when the video is initially available to users, followed by a linearly increasing growth resulting from user migration to the video.

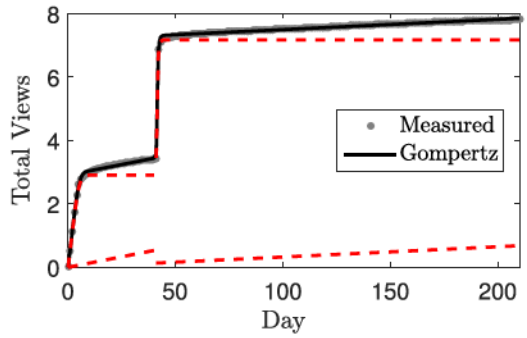


Fig. 5. Due to an exogenous event on day 41, there is a sudden increase in the number of views. The total view count fitted by the Gompertz model $\bar{v}_i(t)$ in (8) is shown in black with the virality (exponential) and migration (linear) illustrated by the dotted red.

The parameters $\theta_i = \{a_k, t_k, M_k, \eta_k, b_k, c_k\}_{k=0}^{K_{\max}}$ in (8) can be estimated by solving the following mixed-integer non-linear program

$$\begin{aligned} \theta_i &\in \arg\min\left\{\sum_{t=0}^{T_i}\left(\bar{v}_i(t)-v_i(t)\right)^2+\lambda K\right\} \\ K &= \sum_{k=0}^{K_{\max}} a_k, \qquad a_k \in \{0,1\} \qquad k \in \{0,\ldots,K_{\max}\}, \end{aligned} \tag{9}$$

with T_i the time index of the last recorded views of video v_i, and a_k a binary variable equal to 1 if an exogenous event is present at t_k. Note that (9) is a difficult optimization problem due to the presence of the binary variables a_k [38]. In the YouTube social network when an exogenous event occurs this causes a large and sudden increase in the number of views, however as seen in Fig. 5, a few days after the exogenous event occurs the views only result from migration (i.e., linear increase in total views). Assuming that each exogenous event is followed by a linear increase in views we can estimate the total number of exogenous events $K_{\max}$ present in a given time-series by first using a segmented linear regression method, and then counting the number of segments of connected linear segments with a slope less then $c_{\max}$. The parameter $c_{\max}$ is the maximum slope for the views to be considered to result from viewer migration. Plugging $K_{\max}$ into (9) results in the optimization of a non-linear program

11. This could also be due to the fact gaming videos account for 70 percent of the videos in the dataset.

12. We used a topic model to obtain the main tags.

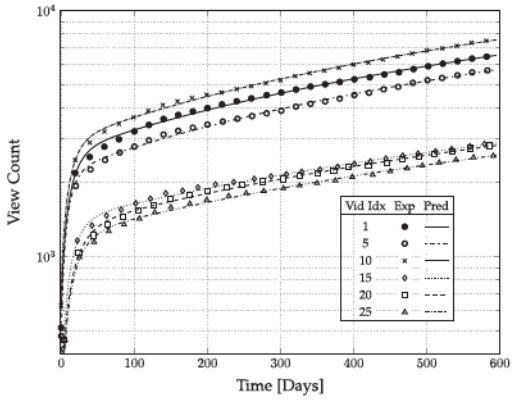
for the unknowns $\{t_k, M_k, \eta_k, b_k, c_k\}_{k=0}^{K_{\max}}$. This optimization problem can be solved using sequential quadratic programming techniques [39].

To illustrate how the Gompertz model (8) can be used to detect for exogenous events, we apply (8) to the view count dynamics of a video that only contains a single exogenous event. Fig. 5 displays the total view count of a video where an exogenous event occurs at time $t = 41$ (i.e., $t_1 = 41$ in (8)) days after the video is posted.[13] The initial increase in views for the video for $t \leq 7$ days results from the 2,910 subscribers of the channel viewing the video. For $7 \leq t \leq 41$, other users that are not subscribed to the channel migrate to view the video at an approximately constant rate of 13 views/day. At $t = 41$, an exogenous event occurs causing an increase in the views per day. The difference in viewers, resulting from the exogenous event, is 7,174. For $t \geq 43$, the views result primarily from the migration of users to approximately 2 views/day. Hence, using the generalized Gompertz model (8) we can differentiate between subscriber views, views caused by exogenous events, and views caused by migration.

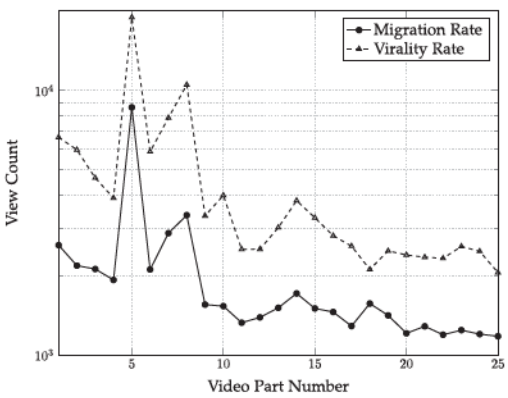
3.4 Video Playthrough Dynamics

One of the most popular sequences of YouTube videos is the video game "playthrough". A video game playthrough is a sequence of videos for which each video has a relaxed and casual focus on the game that is being played and typically contains commentary from the user presenting the playthrough. Unlike YouTube channels such as CNN, BBC, and CBC in which each new video can be considered independent from the others, in a video playthrough the future view count of videos are influenced by the previously posted videos in the playthrough. To illustrate this effect we consider a video playthrough for the game "BioShock Infinite"-a popular video game released in 2013. The channel, popular for hosting such video playthroughs, contains close to 4,500 videos and 180 video playthroughs. The channel is highly popular and has garnered a combined view count close to 100 million views with 150 thousand subscribers over a period of 3 years. Fig. 6 illustrates that the early view count dynamics are highly correlated with the view count dynamics of future videos. Both the short term view count and long term migration of future videos in the playthrough decrease after the initial video in the playthrough is posted. This results for two reasons, either the viewers purchase the game, or the viewers leave as the subsequent playthroughs become repetitive as a result of game quality or video commentary quality. A unique effect with video playthroughs is that though the number of subscribers to the channel hosting the videos in Fig. 6 increases over the 600 day period, the linear migration is still maintained after the initial 50 days after the playthrough is published. Additionally, the slope of the migration is related to the early total view count as illustrated in Fig. 6 b.

13. Due to privacy reasons, we cannot detail the specific event. Some of the reasons for the sudden increase in the popularity of the video include: Another user on YouTube mentioning the video, this will encourage viewers from that channel to view the video, resulting in a sudden increase in the number of views. Another possibility is that the channel owner or a YouTube Partner like BBTV did significant promotional initiatives on other social media sites such as Twitter, Facebook, etc. to promote the channel or video.



(a) Actual and predicted view count of playthrough. We plot the 1st, 5th, 10th, 15th, 20th and 25th video from the playlist containing 25 videos. In the legend, *Exp* and *Pred* corresponds to the actual and the predicted value using (8), respectively. The figure shows that the view counts decreases for subsequent videos in the playlist.



(b) The virality rate specifies the early views due to subscribers, and the migration rate (in units of views/1000 days) specifies the subsequent linear growth due to non-subscribers.

Fig. 6. Actual and predicted view count of a playthrough containing 25 YouTube videos for the game "BioShock Infinite". The predictions are computed by fitting a modified Gompertz model (8) to the measured view count for each video in the playthrough.

4 Conclusion

In this paper, we conducted a data-driven study of YouTube based on a large dataset (see Appendix for details). First, by using several machine learning methods, we investigated the sensitivity of the videos meta-level features on the view counts of videos. It was found that the most important meta-level features include: first day view count , number of subscribers, contrast of the video thumbnail, Google hits, number of keywords, video category, title length, and number of upper-case letters in the title respectively. Additionally, optimizing the meta-data after the video is posted improves the popularity of the video. The social dynamics (the interaction of the channel) also affects the popularity of the channel. Using the Granger causality test, we showed that the view count has a casual effect on the subscriber count of the channel. A generalized Gompertz model was also presented which can allow the classification of a videos view count

TABLE 5
Dataset Summary

Videos	6 million
Channels	26 thousand
Average number of videos (per channel)	250
Average age of videos	275 days
Average number of views (per video)	10 thousand

TABLE 6
YouTube Dataset Categories
(Out of 6 Million Videos)

Category	Fraction
Gaming	0.69
Entertainment	0.07
Food	0.07
Music	0.035
Sports	0.017

dynamics which results from subscribers, migration, and exogenous events. This is an important model as it allows the views to be categorized as resulting from the video or from exogenous events which bring viewers to the video. The final result of the paper was to study the upload scheduling dynamics of gaming channels in YouTube. It was found that going "off schedule" can actually increase the popularity of a channel. Our conclusions are based on the BBTV dataset. Extrapolating these results to other YouTube datasets is an important problem worth addressing in future work. Another extension of the current work could involve studying the effect of video characteristics on different traffic sources, for example the affect of tweets or posts of videos on Twitter or Facebook. YouTube user behavior is a valuable source of time-series data. In our ongoing work [40], we investigate the correlations among YouTube video metrics using multivariate vine copula models.

Appendix

Description of YouTube Dataset

This paper uses the dataset provided by BBTV. The dataset contains daily samples of metadata of YouTube videos on the BBTV platform from April, 2007 to May, 2015, and has a size of around 200 gigabytes. The dataset contains around 6 million videos spread over 25 thousand channels. Table 5 shows the statistics summary of the videos present in the dataset.

Table 6, shows the summary of the various category of the videos present in the dataset. The dataset contains a large percentage of gaming videos.

Fig. 7 shows the fraction of videos as a function of the age of the videos. There is a large fraction of videos uploaded within a year. Also, the dataset captures the exponential growth in the number of videos uploaded to YouTube.

Similar to [3], we define three categories of videos based on their popularity: Highly popular, popular, and unpopular. Table 7 gives a summary of the fraction of videos in the dataset belonging to each category. As can be seen from Table 7, the majority of the videos in the dataset belong to the popular category.

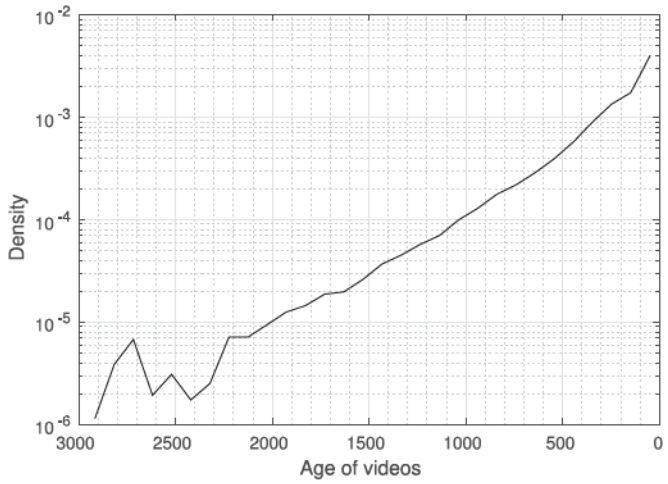


Fig. 7. Histogram of the number of videos in the dataset as a function of the age of the videos. There is a significant percentage of newer videos (videos with less age) compared to older videos. Hence, the dataset captures the exponential growth of the number of videos uploaded and is an unbiased sample of YouTube.

TABLE 7
Popularity Distribution of Videos in the Dataset

Criteria	Fraction
Highly Popular (Total Views $> 10^4$)	0.12
Popular ($150 <$ Total Views $< 10^4$)	0.67
Unpopular (Total Views < 150)	0.21

TABLE 8
Optimization Summary Statistics

Optimization	# Videos
Title change	21 thousand
Thumbnail change	13 thousand
Keyword change	21 thousand

A unique feature of the dataset is that it contains information about the "meta-level optimization" for videos. The meta-level optimization is a change in the title, tags or thumbnail, of an existing video in order to increase the popularity. BBTV markets a product that intelligently automates the meta-level optimization. Table 8 gives a summary of the statistics of the various meta-level optimization present in the dataset.

Acknowledgments

The authors are grateful to Dr. Di Xu, Dr. Lino Coria, and Dr. Mehrdad Fatourechi from BBTV for supplying the YouTube dataset described above, and also for useful discussions and reviewing a preliminary draft of this paper. They are also thankful to Prof. Mor Naaman of Cornell University for commenting on an initial draft of this paper. This research was supported by an Army Research Office grant.

References

[1] G. Gürsun, M. Crovella, and I. Matta, "Describing and forecasting video access patterns," in *Proc. IEEE INFOCOM*, 2011, pp. 16–20.

[2] H. Pinto, J. Almeida, and M. Gonçalves, "Using early view patterns to predict the popularity of YouTube videos," in *Proc. 6th ACM Int. Conf. Web Search Data Mining*, 2013, pp. 365–374.
[3] C. Richier, E. Altman, R. Elazouzi, T. Jimenez, G. Linares, and Y. Portilla, "Bio-inspired models for characterizing YouTube viewcout," in *Proc. IEEE/ACM Int. Conf. Advances Social Netw. Anal. Mining*, 2014, pp. 297–305.
[4] C. Richier, R. Elazouzi, T. Jimenez, E. Altman, and G. Linares, "Forecasting online contents' popularity," arXiv:1506.00178, 2015.
[5] A. Zhang, "Judging YouTube by its covers," Dept. Comput. Sci. Eng., Univ. California, San Diego, 2015. [Online]. Available: http://cseweb.ucsd.edu/ jmcauley/cse255/reports/wi15/Angel%20Zhang.pdf
[6] T. Yamasaki, S. Sano, and K. Aizawa, "Social popularity score: Predicting numbers of views, comments, and favorites of social photos using only annotations," in *Proc. 1st Int. Workshop Internet-Scale Multimedia Manage.*, 2014, pp. 3–8.
[7] T. Yamasaki, J. Hu, K. Aizawa, and T. Mei, "Power of tags: Predicting popularity of social media in geo-spatial and temporal contexts," in *Advances in Multimedia Information Processing*. Berlin, Germany: Springer, 2015, pp. 149–158.
[8] T. Trzcinski and P. Rokita, "Predicting popularity of online videos using support vector regression," arXiv:1510.06223, 2015.
[9] Y. Ding, et al., "Broadcast yourself: Understanding YouTube uploaders," in *Proc. ACM SIGCOMM Conf. Internet Meas.*, 2011, pp. 361–370.
[10] Q. He, T. Shang, F. Zhuang, and Z. Shi, "Parallel extreme learning machine for regression based on MapReduce," *Neurocomputing*, vol. 102, pp. 52–58, 2013.
[11] A. Basu, S. Shuo, H. Zhou, M. Lim, and G. Huang, "Silicon spiking neurons for hardware implementation of extreme learning machines," *Neurocomputing*, vol. 102, pp. 125–134, 2013.
[12] C. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer-Verlag, 2006.
[13] J. Hu, J. Zhang, C. Zhang, and J. Wang, "A new deep neural network based on a stack of single-hidden-layer feedforward neural networks with randomly fixed hidden neurons," *Neurocomputing*, vol. 171, pp. 63–72, 2016.
[14] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, 2015.
[15] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. Roy. Statist. Soc.: Series B (Statist. Methodology)*, vol. 67, no. 2, pp. 301–320, 2005.
[16] N. Meinshausen, "Relaxed Lasso," *Comput. Statist. Data Anal.*, vol. 52, no. 1, pp. 374–393, 2007.
[17] B. Peng and L. Wang, "An iterative coordinate descent algorithm for high-dimensional nonconvex penalized quantile regression," *J. Comput. Graph. Statist.*, vol. 24, no. 3, pp. 676–694, 2015.
[18] T. Hothorn, K. Hornik, and A. Zeileis, "Unbiased recursive partitioning: A conditional inference framework," *J. Comput. Graph. Statist.*, vol. 15, no. 3, pp. 651–674, 2006.
[19] W. Venables and B. Ripley, *Modern Applied Statistics with S-PLUS*. Berlin, Germany: Springer, 2013.
[20] T. Hothorn, P. Bühlmann, T. Kneib, M. Schmid, and B. Hofner, "Model-based boosting 2.0," *J. Mach. Learn. Res.*, vol. 11, pp. 2109–2113, 2010.
[21] H. Drucker, "Improving regressors using boosting techniques," in *Proc. 14th Int. Conf. Mach. Learn.*, 1997, vol. 97, pp. 107–115.
[22] H. Ishwaran and S. Rao, "Spike and slab variable selection: Frequentist and Bayesian strategies," *Ann. Statist.*, vol. 33, pp. 730–773, 2005.
[23] M. Yamada, W. Jitkrittum, L. Sigal, E. Xing, and M. Sugiyama, "High-dimensional feature selection by feature-wise kernelized Lasso," *Neural Comput.*, vol. 26, no. 1, pp. 185–207, 2014.
[24] G. Huang, Q. Zhu, and C. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, no. 1, pp. 489–501, 2006.
[25] G. Huang and L. Chen, "Convex incremental extreme learning machine," *Neurocomputing*, vol. 70, no. 16, pp. 3056–3062, 2007.
[26] G. H. Golub and C. F. Van Loan, *Matrix Computations*, vol. 3. Baltimore, MD, USA: JHU Press, 2012.
[27] H. Liu and H. Motoda, *Feature Selection for Knowledge Discovery and Data Mining*, vol. 454. Berlin, Germany: Springer, 2012.
[28] U. Stańczyk and L. Jain, *Feature Selection for Data and Pattern Recognition*. Berlin, Germany: Springer, 2015.
[29] M. Gevrey, I. Dimopoulos, and S. Lek, "Review and comparison of methods to study the contribution of variables in artificial neural network models," *Ecological Model.*, vol. 160, no. 3, pp. 249–264, 2003.
[30] C. Hutto and E. Gilbert, "VADER: A parsimonious rule-based model for sentiment analysis of social media text," in *Proc. 8th Int. Conf. Weblogs Social Media*, 2014, pp. 1–10.
[31] M. A. Hall, "Correlation-based feature selection for machine learning," Ph.D. dissertation, Dept. Comput. Sci., Univ. Waikato, Hamilton, New Zealand, 1999.
[32] C. W. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica: J. Econometric Soc.*, vol. 37, pp. 424–438, 1969.
[33] G. M. Ljung and G. E. P. Box, "On a measure of lack of fit in time series models," *Biometrika*, vol. 65, no. 2, pp. 297–303, 1978.
[34] A. Wald, *Sequential Analysis*. Mineola, NY, USA: Dover, 1973.
[35] L. Jiang, Y. Miao, Y. Yang, Z. Lan, and A. Hauptmann, "Viral video style: A closer look at viral videos on YouTube," in *Proc. Int. Conf. Multimedia Retrieval*, 2014, Art. no. 193.
[36] F. Figueiredo, F. Benevenuto, and J. Almeida, "The tube over time: characterizing popularity growth of YouTube videos," in *Proc. 4th ACM Int. Conf. Web Search Data Mining*, 2011, pp. 745–754.
[37] A. Tartakovsky, I. Nikiforov, and M. Basseville, *Sequential Analysis: Hypothesis Testing and Changepoint Detection*. Boca Raton, FL, USA: CRC Press, 2014.
[38] S. Burer and A. Letchford, "Non-convex mixed-integer nonlinear programming: A survey," *Surveys Operations Res. Manage. Sci.*, vol. 17, no. 2, pp. 97–106, 2012.
[39] D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA, USA: Athena Scientific, 1999.
[40] V. Krishnamurthy and Y. Duan, "Digging into YouTube data: Dependence structure analysis using vine copula," arXiv, 2017.

William Hoiles received the MASc degree from the Department of Engineering Science, Simon Fraser University, Burnaby, Canada, in 2012 and the PhD degree from the Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, Canada, in 2015. He is currently a postdoctoral researcher in electrical and computer engineering with the University of British Columbia. His current research interests include social sensors and the bioelectronic interface. He is a student member of the IEEE.

Anup Aprem received the ME degree from the Department of Electrical and Communication Engineering, Indian Institute of Science, Bangalore, India, in 2012. He is currently working toward the PhD degree in the Statistical Signal Processing Lab., University of British Columbia. His research interests include statistical signal processing and decision making in the area of social networks and social media.

Vikram Krishnamurthy (S'90-M'91-SM'99-F'05) received the PhD degree from the Australian National University, Canberra, Australia, in 1992. He is currently a professor in the Department of Electrical and Computer Engineering and Cornell Tech, Cornell University. From 2002-2016, he was a professor and Canada research chair with the University of British Columbia, Canada. His research interests include statistical signal processing, computational game theory, and stochastic control in social networks. He served as a distinguished lecturer of the IEEE Signal Processing Society and editor-in-chief of the *IEEE Journal on Selected Topics in Signal Processing*. In 2013, he was awarded an Honorary Doctorate from KTH (Royal Institute of Technology), Sweden. He is author of the book *Partially Observed Markov Decision Processes* published by Cambridge University Press in 2016. He is a fellow of the IEEE.