On the Prediction of Popularity of Trends and Hits for User Generated Videos

Flavio Figueiredo
flaviov@dcc.ufmg.br
Computer Science Department
Universidade Federal de Minas Gerais

ABSTRACT

User generated content (UGC) has emerged as the predominant form of media publishing on the Web 2.0. Motivated by the large adoption of video sharing on the Web 2.0, the objective of our work is to understand and predict popularity trends (e.g, will a video be viral?) and hits (e.g, how may views will a video receive?) of user generated videos. Such knowledge is paramount to the effective design of various services including content distribution and advertising. Thus, in this paper we formalize the problem of predicting trends and hits in user generated videos. Also, we describe our research methodology on approaching this problem. To the best of knowledge, our work is novel in focusing on the problem of predicting popularity trends complementary to hits. Moreover, we intend on evaluating efficacy of our results not only based on common statistical error metrics, but also on the possible online advertising revenues our predictions can generate. After describing our proposal, we here summarize our latest findings regarding (1) uncovering common popularity trends; (2) measuring associations between UGC features and popularity trends; and (3) assessing the effectiveness of models for predicting popularity trends.

Categories and Subject Descriptors

C.4 [**Computer Systems Organization**]: Performance of Systems—*Measurement techniques*; H.3.5 [**Information Storage and Retrieval**]: Online Information Services—*Web-based services*

Keywords

UGC; video; popularity; trends

General Terms

Human Factors; Design; Measurement

1. INTRODUCTION

On the Web 2.0, user generated content (UGC) has become the de-facto form of media publishing on some of the most popular Internet applications nowadays [6]. Focusing on video content, websites such as YouTube [1] receive over 800 million unique users monthly, attracting over 1 million different advertisers [23]. Even niche applications, such as Vimeo [2], which focuses on independent filmmakers, manage to attract over 70 million unique users monthly [19].

Given the success of such applications and the current large volume of videos consumed daily, understanding how users find such content and how content popularity evolves provide valuable insights for content generators, online advertisers and Internet service providers (ISPs), amongst others. For instance, from a systems perspective, understanding these properties may drive the design of better analytic tools, a major market segment nowadays. Online advertisers may also benefit from this information to better place contextual advertisements, while ISPs could exploit it to develop more cost-effective content delivery platforms and caching systems. From a social perspective, understanding the properties of video popularity trends could be used to better comprehend the human dynamics of consumption processes [7]. Also, content producers could use insights on how user collaboration and collaborative social activities on Web 2.0 applications may impact content popularity, providing information on aspects related to their own fame on video sharing applications.

Most previous efforts, which are focused on predicting the popularity of a piece of content measured at a specific future date [16–18, 22], are still preliminary, as they provide limited knowledge on which features and system mechanisms (e.g., search, related videos, etc) contribute the most to popularity growth. Analyzing the importance of such features to popularity growth is key to provide scalable alternatives to service design, as solutions based on content analysis are less scalable in (user generated) videos. Moreover, there is little effort towards predicting popularity evolution (or trends), which may also provide valuable knowledge. For instance, online advertisers and content delivery systems could benefit more from predicting not only a final popularity measure for UGC, but also whether its popularity trend is increasing and how stable it is likely to be over time.

In sum, our proposed research aims at understanding the importance and utility of various features, particularly referrers (i.e, incoming links to videos), on the popularity evolution of individual user generated videos and exploiting them to develop methods to predict future popularity measures and trends of those videos.

The rest of this paper is organized as follows. Section 2 describes our problem statement and research goals. We describe our current methodology on addressing our goals on Section 3. The current state of our research is described in Section 4 while our related work is addressed in Section 5. Section 6 concludes this paper.

[1] `http://youtube.com`
[2] `http://vimeo.com`

WSDM'13, February 4–8, 2013, Rome, Italy.

2. PROBLEM STATEMENT AND RESEARCH GOALS

In general terms, the problem we intend to address is defined as follows. Given a collection V of user generated videos, we define for each video $\mathbf{v} \in V$ a set of referrer, popularity, content and social features related to $\mathbf{v}$. We are interested in understanding the importance of each of these features when applied to the task of predicting how the popularity of individual videos will evolve from a reference time t to a target time $t + \delta$, considering that the information related to each of these features is available only up to the reference time t and δ is a time interval. For example, referrer features of a video only account for the views a video received via some incoming links up to time t. Moreover, we intend to develop methods to predict future popularity measures and popularity trends of individual videos. Thus, for each video we also define $\mathbf{s}_v$ as a vector of views received by the video at different dates (e.g, each week). As stated, referrer features capture information on incoming links to videos. We are particularly interested in these features since they provide valuable information on how users reach UGC content on the Web 2.0. As a basis for comparison, we also plan to study: (1) popularity features, which are related to temporal data about the evolution of popularity of individual videos; (2) content features, such as the category and tags of video; and (3) social features, which capture information about the user who created and uploaded the video (e.g., number of followers or user-rank) [20].

More specifically, our research goals (RGs) are the following:

RG1 - Understanding Referrer Importance: Our first research goal focuses on characterizing and modeling: (1) how video popularity evolves over time; and, (2) how video popularity evolution correlates with the referrers that lead users to videos (as well as with other content, social and popularity features). We note that, unlike previous work, which correlated different features of YouTube with final popularity [2, 20], here we are concerned with measuring the impact of referrers on how the popularity of each video evolves over time [9, 20].

RG2 - Predicting Video Popularity: After data characterization and modeling, we intend to exploit the available data to answer the following question: Is it possible to predict how the popularity of individual videos evolves over time? In other words, we want to know if it is possible to model the popularity curve (or trend) of each video. Although previous studies point out that modeling the popularity trend of individual videos is intractable [1], some recent results we obtained suggest that this might be possible. We shall detail more of this in Section 3. We also intend to investigate whether more effective methods to predict the popularity measure of a video at a target date can be devised by exploiting the developed popularity trend prediction models (e.g., by building specialized models to pre-defined popularity trends).

RG3 - Applications for Popularity Prediction: As an applied step of our research, we intend to investigate whether our popularity prediction methods could be exploited for generating revenue in advertising campaigns. Online advertisement is a multi-billion dollar industry [15] and in UGC applications, traditional keyword-targeted or content-targeted advertising may not be suitable since viral content attracts millions of views at small time windows [4]. Thus, popularity prediction might be useful to supplement traditional tools and provide advertisers with more information to determine how much they are willing to pay per click per content. As far as we know, this has not been investigated before since research on UGC popularity prediction is still at its infancy. Thus, it is a good research opportunity and a means to validate the effectiveness of our prediction methods.

3. METHODOLOGY

The initial step in studying the challenges described above is data collection. In order to understand popularity growth of videos we can collect data from applications like YouTube or Vimeo. Both provide on their websites time series of popularity of different videos. They also provide data on the top 10 referrers that led users to videos. Moreover, some notion of popularity can be inferred from the timestamps of user comments, a piece of temporal information largely available in many Web 2.0 applications. Although availability of only the top 10 referrers per video poses as a limit, we note that these referrers account for a non-negligible amount of video views (in the order of billions) [9], which can be made useful for popularity time series prediction tasks. Most of the work we have done so far is based on data collected in this fashion. That is, we already have a dataset containing the popularity curves and top 10 referrers related to more than a billion views for YouTube videos. We note that we also intend to investigate the possibility of monitoring and collecting traffic from a large local network, specifically our university campus network (as in [14]).

After data collection, we will characterize the importance of referrer features, as well as other popularity, content and social features, for the evolution of popularity of videos. We plan to proceed as follows. Initially we will determine classes of popularity trends based on a recent time series- clustering algorithm [21]. We then plan to correlate these classes and popularity measures at specific target times with the selected features (RG1). For prediction (RG2), traditional machine learning techniques (e.g., classification and regression models) will be used to infer not only the final popularity measure at a target time but also the popularity trend (i.e., class) of individual videos, using, as input, feature values collected up to a reference time. For example, we plan to use these classes to train supervised classifiers to predict the trends of individual videos. Class specific regression tasks [17, 18] can then be used, after classification, to build specialized models that predict popularity measures for videos with specific popularity trends. Indeed, we have some preliminary results that show that some accuracy can be achieve in both the classification and regression tasks. One particular aspect to consider in our evaluation of the prediction methods is whether accurate predictions can be made using only feature values collected prior to the video's popularity peak. This is particularly important as many videos exhibit a popularity distribution with a clear single peak [7, 9]. Thus, it is desirable to produce accurate predictions before the interest in the content starts diminishing.

Most previous studies of popularity prediction [16,18,22] are focused only on prediction accuracy metrics such as coefficient of determination or squared errors. We here intend to not only use these traditional metrics but also assess if advertising markets [8, 13, 15] can benefit from our popularity prediction models (RG3). For example, we aim at understanding if, even with errors, our predictions generate revenue. In traditional keyword-targeted (see Chapter 15 of [8]) or content-targeted [15] advertising, the popularity evolution of query terms and web pages can be ignored since such content is expected to remain popular for a large period of time. In the context of user generated videos, this is not the case [4, 7, 9]. Viral videos can garner millions of views at small time windows [4] and even those videos that are not viral may exhibit single-peaked based popularity distributions. The question we raise here is: How can advertisement-campaigns make use of content popularity pre-

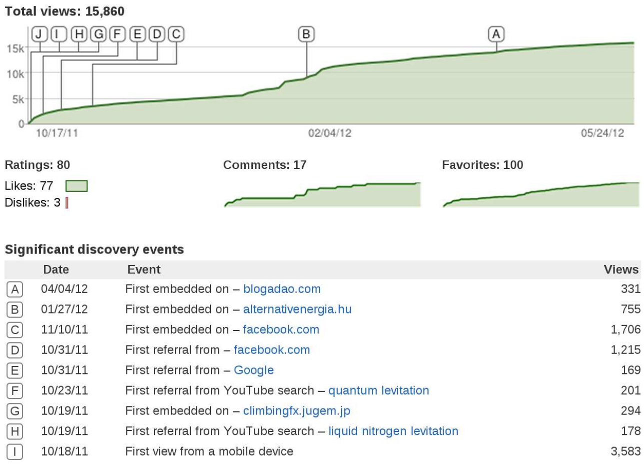

Figure 1: Example of Video Features

dictions in user generated video applications? We note that this is different from the task of assigning relevant-ads to content [15]. In user generated video applications, textual information such as tags or comments [10, 11] can be explored for this task.

A common setting in online advertising is the pay-per-click approach. In this setting, advertisers pay content providers per click which generates traffic to their website [15]. Here, we plan to study if the prediction of popularity measures (i.e., hits) and trends (evolution) can be used in conjunction with simple models of revenue estimation [13]. In simple terms, we want to verify if, in the case where click prices is determined by popularity predictions, these predictions are good enough in order to generate revenue. To this end, we plan to compare if our predictions generate revenues as good as the ones expected from the actual popularity evolution of each video. We intend to use revenue estimation models that assume fixed revenue per click for all videos, fixed revenue per click for each video (i.e., the revenue per click varies with the video's popularity) as well as models that assume revenue per click evolving based on temporal and popularity information [13]. Our evaluation will be performed in several synthetic (but relevant) scenarios, built from the model parameters [13].

4. CURRENT RESULTS

We now discuss some current results focused on research goals RG1 and RG2. We refer the reader to our published work [9] for a in-depth characterization of our datasets.

Our work analyzes features that are public available at YouTube. An example of such data is provided in Figure 1: for each video, YouTube provides the cumulative popularity time series as well as what they call *Significant Discovery Events*, which represent the the top 10 referrers (i.e., incoming links) that attracted more views to the video. Along with the number of views each referrer was responsible for, YouTube also provides the date of the first event triggered by that referrer. Moreover, the application also provides the cumulative time series related to other video features, such as the number of comments and the number of times the video was marked as favorite. Using HTML scrapping or the YouTube's API[3], it is also possible to collect other pieces of information, such as the category to which the video belongs (e.g., music, comedy, pets, autos etc.).

Due to space constraints, we shall discuss our results in only one

[3] **`https://developers.google.com/youtube/`**

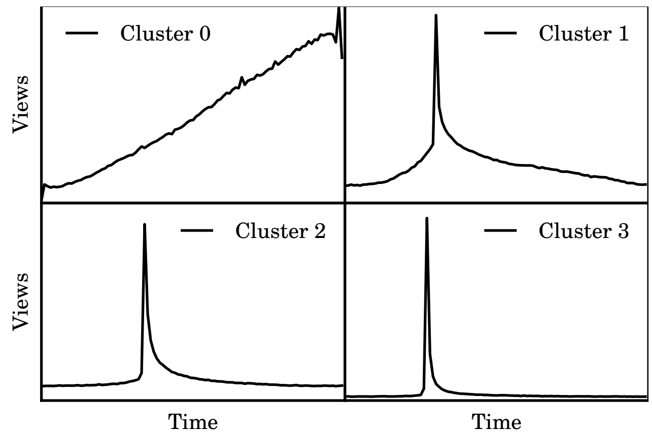

Figure 2: Discovered Trends by the KSC Algorithm.

Table 1: Summary of Popularity Trends

	$C0$	$C1$	$C2$	$C3$
# of Videos	4,023	6,718	5,031	3,790
Avg. # of Views	305,130	108,844	64,274	127,768
Avg. Change Rate #Views	47	7	4	4
Avg. Peak Fraction	0.03	0.03	0.08	0.28

of our three distinct video datasets [9]. This dataset is composed of 24,482 videos collected as results of random queries submitted to YouTube's search API. For each video, the datasets contain the video category and the statistics depicted in Figure 1, namely: the time series of the number of views, comments and favorites as well as the top-10 referrers, along with the total number of views each referrer was responsible for and the date of the first access from it.

4.1 Trend Discovery

The initial step of our study is the extraction of temporal patterns of popularity evolution. For this task we implemented a recently proposed time series clustering algorithm, called K-Spectral Clustering (KSC)[4], which was shown to be effective in detecting popularity patterns of other types of UGC content [21]. We selected this time series clustering algorithm, as opposed to alternative methods (e.g., Dynamic Time Warping or analytical modeling [7]) because the clustering is performed independently of shifts (i.e., dates) and scale (i.e., volume), focusing rather on the overall shape of the time series. The algorithm make's use of video's popularity time series $\mathbf{s_v}$ as input. The number of clusters was defined based on common cluster quality measures [21].

Figure 2 shows the discovered trends which govern popularity evolution on YouTube videos. Each graph shows the number of views as function of time. Moreover, Table 1 presents, for each cluster, the number of videos that belong to it as well as the average number of views, the average change rate in the number of views, and the fraction of views at the peak time window of these videos. The average change rate in the number of views of a given video v is computed as the average difference between two (non-cumulative) measures taken in successive time windows. Thus, it captures the trend in the number of views of the video: a positive (negative) change rate indicates that the number of views measured during each time window is increasing (decreasing) with time, whereas a change rate equal to 0 indicates stability. Table 1

[4] We have implemented a parallel version of the KSC algorithm which is available at **`http://goo.gl/0dmHc`**

Table 2: Summary of Features

Feature Class	Feature Name	Type
Content	Video category	Categorical
	Upload date	Numerical
	Video age	Numerical
	Time window size (w)	Numerical
Link	Referrer first date	Numerical
	Referrer number of views	Numerical
Popularity	# of views	Numerical
	# of comments	Numerical
	# of favorites	Numerical
	Avg. change rate in # of views	Numerical
	Avg. change rate in # of comments	Numerical
	Avg. change rate in # of favorites	Numerical
	Peak fraction	Numerical

shows the average change rate computed over the total duration of the video's lifespan. The peak fraction, also shown in Table 1, is the ratio of the maximum in $\mathbf{s}_v$ divided by the total of views of the video (i.e., the sum of all elements in $\mathbf{s}_v$).

As shown in the figure, Cluster 0 consists of videos that remain popular over time, even attracting a larger number of views during each time window as time passes, as indicated by the large positive change rates shown in Table 1. The videos in cluster 0 have also no significant peaks, as the average fractions of views in the peak windows, shown in Table 1, are very small. The other three clusters are predominantly defined by a single peak in popularity followed by a steady decline. The main difference is the rate of decline, which is much slower in Cluster 1, somewhat faster in Cluster 2, and very sharp in Cluster 3. This difference is more clear if we analyze the peak fractions and the average change rates in Table 1. We note that these three clusters were previously uncovered in other YouTube datasets [7,9], being referred to as *Viral, Quality and Junk* respectively. However, unlike these previous efforts, which relied mostly on peak popularity analyzes, we here use an unsupervised learning algorithm which makes our task of discovering popularity trends more general and robust. For example, the thresholds in peak volume which define Clusters 1, 2 and 3 in these previous studies are not clearly defined. In contrast, such peaks emerge clearly in our clusters (as shown in Table 2). Moreover, no previous study that analyzed video popularity time series or other UGC time series has identified a trend similar to Cluster 0, possibly because of the models adopted, which focus on power-law like behavior [7,9].

4.2 Predicting Trends

We now focus on the task predicting popularity trends, as it complements various recent efforts towards predicting final popularity measures of UGC [16,18,22]. Our prediction model is based on extremely randomized ensemble trees [12]. We choose this technique because it has been shown to be effective on different datasets, requiring little or no preprocessing. Moreover, the learned models can be more easily interpreted, in comparison with other techniques such as Support Vector Machines

For classification, we make use of the feature vector $\mathbf{v}$. The analyzed features analyzed, which can be derived from our datasets, are grouped into three classes, namely, content features, link features, and popularity features, as shown in Table 2.

The popularity features, measured during a defined monitoring period, capture not only the total number of views, comments and times the video was marked as favorite during that period but also the trend in these measures captured by the average change rates. We also include the peak fraction as a popularity feature. Jointly, this set of features tries to capture properties of the popularity curve (during the monitoring period).

The prediction model is learned from a training set of pre-labeled videos, i.e., videos whose popularity trends are known. The effectiveness of the model is then evaluated in a different set of videos (test set). For each video $\mathbf{v}$ in the test set, the prediction is performed based on features associated with $\mathbf{v}$ collected during a given monitoring period, starting at video's upload. We then assess the effectiveness of the classifier as function of the monitoring period. We evaluate the accuracy of our prediction model using the $F1$ and Macro-$F1$ metrics using a 5-fold cross validation.

Table 3: Per-trend and Macro F1 Scores with 95% Confidence Intervals

	Monitoring Period			
Trend	1% of Lifespan	20%	50%	100%
C0	0.35±0.01	0.34±0.02	0.47±0.02	0.64±0.02
C1	0.52±0.01	0.60±0.01	0.67±0.02	0.75±0.01
C2	0.30±0.01	0.51±0.01	0.68±0.02	0.78±0.03
C3	0.43±0.02	0.75±0.01	0.86±0.01	0.91±0.03
Macro	0.40±0.01	0.55±0.01	0.67±0.01	0.77±0.02

Table 3 shows average $F1$ results for each class (i.e., popularity trend) as well as the overall Macro $F1$, along with corresponding 95% confidence intervals, for different monitoring periods, for each dataset. We start by noting that, even if the monitoring period is limited to a single window (1% of the video's lifespan), which corresponds to roughly a week, Macro $F1$ results are of 0.4. These results are much better than randomly choosing the popularity trend of a given video, which results in a Macro F1 of 0.25 for our four classes. Moreover, as the monitoring period increases, prediction accuracy also increases. For instance, if 20% of the video's lifespan is monitored, Macro F1 reaches from 0.55. Delving into the model accuracy for each class, we find that $C0$ is the hardest one to predict (lower F1 results) while $C3$ is usually the easiest one, in all three datasets.

5. RELATED WORK

Cha et al. [5] and Gill et al. [14] performed two of the earliest analyses of YouTube. While the first focused on data crawled from YouTube, the latter was based on traffic going to YouTube from an university campus. More recently, Wattenhofer et al. [20] analyzed the correlation between YouTube video popularity and properties of various online social networks (OSN) created among users of the system. These studies focused on single snapshots of data, thus they did not account for temporal data, thus failing to analyze the long-term popularity evolution. Broxton et al. [4] defined a mechanism to determine if a video will be viral based on the fraction of viral referrers (direct links or OSN referrers) it attracts on short periods of time. Very popular videos with a large fraction of such referrers are likely to be viral. Brodersen et al. [3] made use of the same model to determine, amongst other things, if viral videos receive most views from the same geographic region. Moreover, Borghol et al. [2] assessed the importance of referrer and content features to the final popularity of videos.

These previous studies provide some insights into which factors impact popularity growth of individual videos. However, there is still little knowledge about which video features (e.g., content, referrer and popularity features) and system mechanisms (e.g., search) contribute the most to the evolution of popularity of individual videos. That is, unlike most previous work, we here intend to correlate these features with popularity evolution trends of individual videos. In [9], we performed a preliminary characterization of popularity evolution patterns of YouTube videos and the impact of various features on such patterns. Our intention is to extend this characterization by applying a recently proposed method to identify such trends [21] as well as to exploit the knowledge obtained in more cost-effective popularity prediction methods.

Considering the task of popularity prediction, Yin et al. [22] proposed a model to predict the ranking of popular items. For this task, the authors took into account user personalities when casting votes, and developed a Bayesian model for ranking prediction. They tested their model in a popular smart phone application, JokeBox [5]. A similar effort was pursued by Lerman and Hogg [16]. The authors also made use of voting models to perform content popularity prediction, applying their model to the news sharing applications Digg [6]. The efficacy of these models on video sharing applications is debatable. Firstly, JokeBox and Digg are websites focused on user voting. Although this feature is available on video sharing applications such as YouTube, it is not one of the primary use cases of the applications. Thus, models based on this kind of data would have to deal with sparsity problems. Moreover, these methods require sensitive user data, which, due to scale or privacy issues, is not publicly available in most video sharing applications.

The studies performed by Szabo and Hubberman [18] and Pinto et al. [17] are much close to ours. Both works analyzed YouTube videos and noted that long term popularity values are related to early popularity observations at a logarithmic scale. A similar result was also reported by Borghol et al. [2]. Based on such observation, the authors proposed a simple popularity prediction model based on linear regression. Both models were based on only one kind factor, i.e., early popularity views at a specific reference date. Although very simple, these models achieve somewhat accurate predictions of popularity measures for YouTube videos. We plan to extend these models to predict not only popularity measures at specific target dates but also the popularity evolution trend (or curve) of individual videos. To our knowledge, our work is the first to address this task in user generated videos. Moreover, unlike all previous efforts which focus only on statistical error metrics (e.g. coefficient of determination-R^2 or mean squared errors-MSE), we intend to validate the effectiveness of our prediction models in real world applications, i.e., advertising [13, 15].

6. SUMMARY

Nowadays, an enormous amount of user generated content is uploaded to Web 2.0 applications on a daily basis. Given the importance and volume of this type of content, understanding how trends in UGC popularity evolves over time provides substantial knowledge about how users find and consume collaborative content on the Internet. In this work presented a methodology and some initial results on the task of predicting trends and hits in user generated videos. We believe that this research has the potential to expand the knowledge on Web 2.0 applications and user generated content, providing valuable insights and techniques for content producers, online advertisers, and ISPs, as discussed in Section 1.

As a final note we point out that although focused on YouTube videos, the knowledge we here produce could be extended to other kinds of content such as images or text. It would be interesting to see if similar referrers are responsible for the popularity of different types of content. Moreover, comparing how popularity evolves across media and identifying the factors that are responsible for this evolution can be used by content producers and viral marketeers to decide on which application they should focus their products on.

ACKNOWLEDGEMENTS

This research is partially funded by the Brazilian National Institute of Science and Technology for Web Research (MCT/CNPq/INCT Web Grant Number 573871/2008-6).

7. REFERENCES

[1] Y. Borghol, S. Mitra, S. Ardon, N. Carlsson, D. Eager, and A. Mahanti. Characterizing and modelling popularity of user-generated videos. *Performance Evaluation*, 68(11):1037–1055, Nov. 2011.

[2] Y. Borghol, N. S. Wales, S. Ardon, and D. Eager. The Untold Story of the Clones: Content-agnostic Factors that Impact YouTube Video Popularity. In *Proc. KDD*, 2012.

[3] A. Brodersen, S. Scellato, and M. Wattenhofer. YouTube around the world. In *Proc. WWW*, Apr. 2012.

[4] T. Broxton, Y. Interian, J. Vaver, and M. Wattenhofer. Catching a viral video. *Journal of Intelligent Information Systems*, pages 1–19, Dec. 2011.

[5] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, and S. Moon. Analyzing the Video Popularity Characteristics of Large-Scale User Generated Content Systems. *IEEE/ACM Transactions on Networking*, 17(5):1357–1370, Oct. 2009.

[6] G. Cormode and B. Krishnamurthy. Key Differences between Web1.0 and Web2.0. *First Monday*, 13(6), 2008.

[7] R. Crane and D. Sornette. Robust dynamic classes revealed by measuring the response function of a social system. *Proceedings of the National Academy of Sciences (PNAS)*, 105(41):15649–53, Oct. 2008.

[8] D. Easley and J. Kleinberg. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, 1 edition, July 2010.

[9] F. Figueiredo, F. Benevenuto, and J. Almeida. The tube over time: characterizing popularity growth of youtube videos. In *Proc. WSDM*, 2011.

[10] F. Figueiredo, H. Pinto, F. Belém, J. Almeida, M. Gonçalves, D. Fernandes, and E. Moura. Assessing the quality of textual features in social media. *Information Processing & Management*, Apr. 2012.

[11] K. Filippova and K. B. Hall. Improved video categorization from text metadata and user comments. In *Proc. SIGIR*, 2011.

[12] P. Geurts, D. Ernst, and L. Wehenkel. Extremely Randomized Trees. *Machine Learning*, 63(1):3–42, 2006.

[13] A. Ghose and S. Yang. An Empirical Analysis of Search Engine Advertising: Sponsored Search in Electronic Markets. *INFORMS: Management Science*, 2010.

[14] P. Gill, M. Arlitt, Z. Li, and A. Mahanti. Youtube traffic characterization: a view from the edge. In *Proc. IMC*, Oct. 2007.

[15] A. Lacerda, M. Cristo, M. A. Gonçalves, W. Fan, N. Ziviani, and B. Ribeiro-Neto. Learning to advertise. In *Proc. SIGIR*, 2006.

[16] K. Lerman and T. Hogg. Using a model of social dynamics to predict popularity of news. In *Proc. WWW*, Apr. 2010.

[17] H. Pinto, J. Almeida, and M. Gonçalves. Using early view patterns to predict the popularity of YouTube videos. In *Proc. WSDM*, 2013.

[18] G. Szabo and B. A. Huberman. Predicting the Popularity of Online Content. *Communications of the ACM*, 53(8):80–88, 2010.

[19] Vimeo. Advertise on Vimeo. http://vimeo.com/about/advertisers, 2012.

[20] M. Wattenhofer, R. Wattenhofer, and Z. Zhu. The YouTube Social Network. In *Proc. ICWSM*, 2012.

[21] J. Yang and J. Leskovec. Patterns of temporal variation in online media. In *Proc. WSDM*. ACM Press, Feb. 2011.

[22] P. Yin, P. Luo, M. Wang, and W.-C. Lee. A straw shows which way the wind blows: Ranking Potentially Popular Items from Early Votes. In *Proc. WSDM*, Feb. 2012.

[23] Youtube. YouTube Press Statistics. http://www.youtube.com/t/press_statistics, 2012.

[5] JokeBox is a smart phone application in which user share and up/down vote jokes.

[6] **http://www.digg.com**