Characterizing and Predicting Viral-and-Popular Video Content

David Vallet[†*], Shlomo Berkovsky[‡], Sebastien Ardon[†], Anirban Mahanti[†],
Mohamed Ali Kaafar[†]
[†]NICTA, Australia [‡]CSIRO, Australia
david.vallet@nicta.com.au, shlomo.berkovsky@csiro.au, sebastien@incoming.tv,
anirban.mahanti@nicta.com.au, dali.kafaar@nicta.com.au

ABSTRACT

The proliferation of online video content has triggered numerous works on its evolution and popularity, as well as on the effect of social sharing on content propagation. In this paper, we focus on the observable dependencies between the virality of video content on a micro-blogging social network (in this case, Twitter) and the popularity of such content on a video distribution service (YouTube). To this end, we collected and analysed a corpus of Twitter posts containing links to YouTube clips and the corresponding video metadata from YouTube. Our analysis highlights the unique properties of content that is both popular and viral, i.e., attracts high number of views on YouTube and achieves fast propagation on Twitter. With this in mind, we proceed to the predictions of popular-and-viral clips and propose a framework that can, with high degree of accuracy and low amount of training data, predict videos that are likely to be popular, viral, and both. The key contribution of our work is the focus on cross-system dynamics between YouTube and Twitter. We conjecture and validate that cross-system prediction of both popularity and virality of videos is feasible, and can be performed with a reasonably high degree of accuracy. One of our key findings is that YouTube features capturing user engagement, have strong virality prediction capabilities. This findings allows to solely rely on data extracted from a video sharing service to predict popularity and virality aspects of videos.

Keywords

Popularity; Virality; Video.

1. INTRODUCTION

Over the past decade, video sharing services like YouTube have witnessed exponential traffic growth. In these services, some videos achieve immense popularity and have over a billion views, while others fall into oblivion without attracting any interest. In this corner of the attention economy, views – sometimes called eyeballs – directly impact advertising revenues and are, therefore, of prime importance for content producers, marketing agencies, and service providers. As a result, over the recent years, the dynamics and prediction of user-generated content popularity has captured the attention of service providers, media agencies, and academics alike [26, 6, 9, 3, 22, 17].

In this paper, we study the popularity of content on a video sharing platform (YouTube), while simultaneously observing its propagation in an online social network for micro-blogging (Twitter). Our methodology allows us to evaluate the interplay between a video content *popularity*, i.e., its inherent propensity to attract views on YouTube, and content *virality*, i.e., its potential to elicit Twitter posts from its viewers. Our underlying assumption is that two main types of diffusion channel contribute in combination to a content's popularity: conventional channels and online social networking (OSN) channels. Conventional diffusion channels are essentially mass-media outlets, such as news sites, product reviews, and popular blogs, which can be considered as direct marketing efforts designed to steer traffic and attention towards the videos. OSN channels, in contrast, make use of the crowd to generate traffic toward the videos, as users leverage their social circles to share video links with friends and followers. What drives users to share videos still remains largely under-studied [20], but empirical evidence shows that only some videos (referred to as viral), generate enough interest for a user to share the video.

We use the Twitter service as a proxy for OSN channels. We collect data about tweets and re-tweets containing links to YouTube videos, while simultaneously collecting statistics pertaining to these videos (view counts, number of comments, tags, and so on) from the YouTube service. The data collected simultaneously on both systems allows us to quantify two system-specific metrics for each video: popularity and virality. The popularity reflects the exposure of the video, driven by both the conventional and OSN diffusion channels, whereas virality measures the propagation of the clip on Twitter. Note that the two metrics are measured independently on the two services, although they may be inter-related and fuel each other. Based on those metrics, we define video classes by selecting the set of most popular and most viral videos, and characterize their evolution on the YouTube and Twitter platforms. We then extract a suite of Twitter- and YouTube-specific features, and feed these into a Gradient Boosted Decision Tree classifier [13], trained to predict popularity and/or virality of videos.

---

*Now with Google.

Table 1: Overview of example video sub-groups.

popular	5,000 videos with highest view count during the labeling window
viral	5,000 videos with the highest number of mentions during the labeling window
recent	5,000 videos that are less than 2 days old at the time of data collection
random	5,000 videos randomly selected from the seed set of 200,000 videos

We make a number of observations. We demonstrate that we can produce highly accurate predictions of video popularity and virality using only a small amount of training data. For example, our classifier is able to predict with high accuracy the videos in the *viral-and-popular* class up to 7 days in advance, using only 1 day of training data. We find that only the fundamental features, such as past number of views and tweet count, are insufficient to obtain a reasonable classification accuracy for these videos. Using feature importance analysis, we show that the current tweeting rate, along with the volume of original tweets received since upload are the two most important Twitter features to consider for the classification of these videos. Further, an important result presented is on cross-system predictions: we show that YouTube features, and in particular features reflecting the level of user engagement, have a strong prediction capability for the viral-and-popular class, especially for recently uploaded videos. This result in itself motivates the application of our approach for identification of both popular and viral items solely based on data extracted from a video sharing service.

The remainder of the paper is structured as follows. Section 2 presents the datasets used in this work, describes our data collection methodology, and includes a high-level characterization of the collected data. Section 3 presents our results, analyses, and observations in detail. Then, section 4 present related work, before concluding and stating directions for future work in section 5.

2. DATASETS

In this section, we describe our data collection methodology and present a high-level characterization of our datasets. The characterization provides insights into the interplay between popularity and virality, and serves as a prelude to the modeling and analyses presented later in the paper.

2.1 Data collection

We initially selected a set of videos to study by sampling the Twitter stream for 24 hours searching for links to YouTube videos. Specifically, we used the Twitter streaming API with the `expanded_url` keyword matching with the domain names used by YouTube, like youtube.com and youtu.be. Note that the Twitter streaming API is able to resolve shortened URL links, so tweets containing a shortened, e.g., through bit.ly, URL to a YouTube video were also captured by our search. Using this search facility over a 24-hour period, we collected links to a large number of videos from which we randomly selected 200,000 videos to form the seed set for our dataset. Then, for a period of two
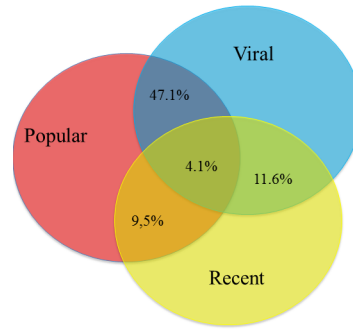

Figure 1: Venn diagram showing the overlaps among the popular, viral, and recent sub-groups.
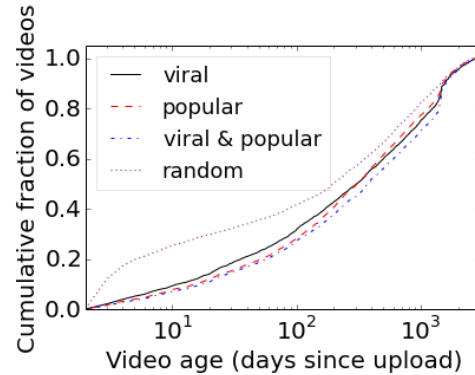

Figure 2: Distribution of video age for different classes.

weeks we collected YouTube and Twitter data for each video included in the seed set. Specifically, we collected statistics from YouTube once per day and concurrently tracked mentions of these videos using the Twitter streaming API. A detailed description of the collected data and extracted features will be presented in Section 3.1.

Out of the 200,000 seed videos, we focus on four sub-groups, described in Table 1, which are then used in the remainder of this paper. The *popular* sub-group contains the 5,000 videos that attracted the largest number of views on YouTube. Specifically, among the 200,000 videos in the seed set, the popular videos account for the most viewed 2.5% of videos. We also define the *viral* sub-group, which consists of the 5,000 videos that were mentioned the most on Twitter[1]. The *recent* sub-group consists of 5,000 randomly selected videos, which are two days old or younger, at the time of data collection. Finally, the *random* sub-group consists of a sample of 5,000 videos selected at random from our initial seed set of 200,000 videos. The relative size of the intersections of these sub-groups is shown in Figure 1.

Given these sub-groups, we define another sub-group, denoted by *viral-and-popular*. As the name suggests, this sub-group includes videos that are both viral and popular, i.e., videos in the intersection of the viral and popular video sets. Note that the cardinality of the viral-and-popular sub-group is close to a half of either sub-group individually.

---

[1] We define popularity and virality based on YouTube views and Twitter mentions captured during the observation period, referred to as the *labeling* window. More details on the labeling will be given in section 3.1.
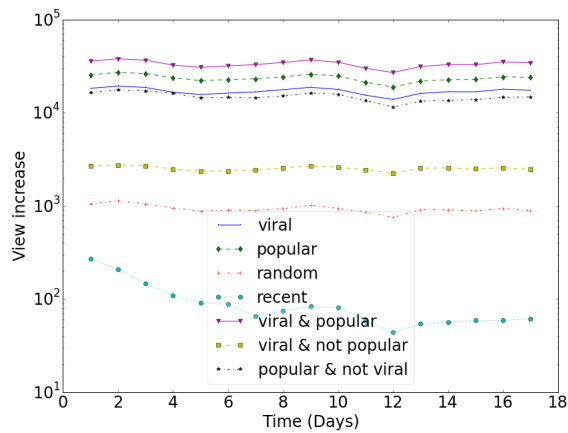
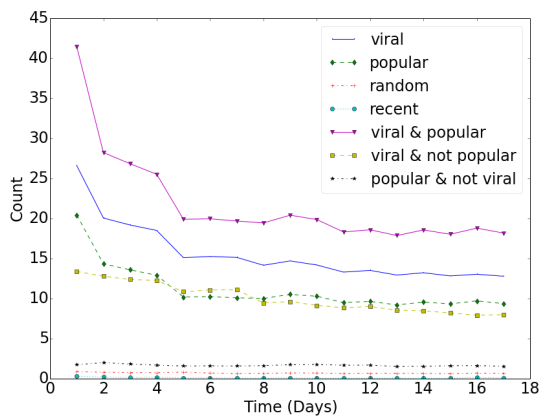Figure 3: Average daily increase in view count for the various video sub-groups.

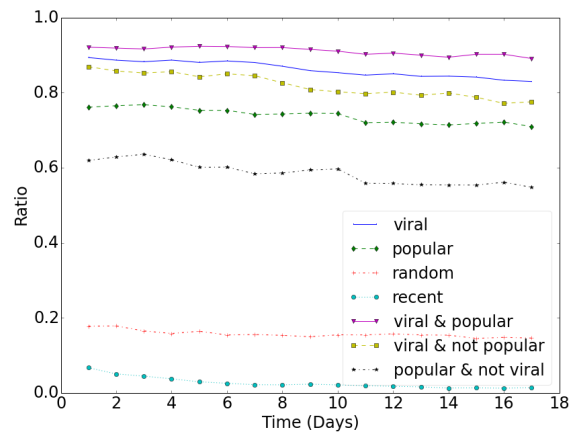Figure 4: Average number of tweet mentions for the various video sub-groups.

Figure 5: Ratio between the original and overall number of tweets averaged per day and video class.
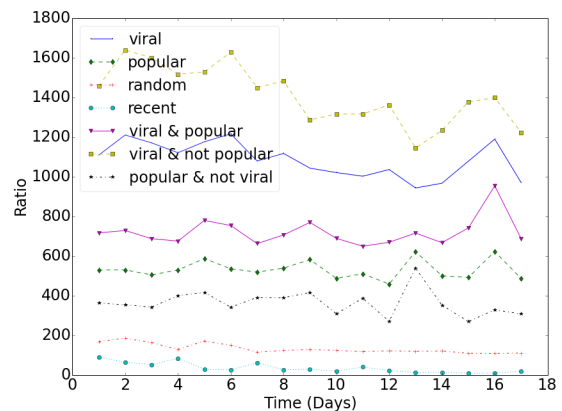
Figure 6: Average number of users reached per day by tweets for videos in the various sub-groups.

2.2 High-Level Characterization

This section presents a high-level characterization of the datasets used in this study. Figure 2 shows the video age distribution, for the random, viral, popular, and viral-and-popular sub-groups. We observe that the video age has a wide range of values, and varies across the three classes, with videos in the random class skewed towards younger videos more than in the other three classes. This shows that, as expected, a random sample of YouTube videos has a higher portion of young videos than the viral and popular sub-groups. Somewhat surprisingly, we find that some very old videos, which are over 7 years old, are amongst the viral-and-popular videos. On investigation, we find that these older videos are music video clips, with a significant number being Michael Jackson's videos that were released in 2009.

Figure 3 shows the average *increase* in the number of YouTube views per day for the video sub-groups introduced above. We notice that with the exception of the videos in the recent sub-group, videos in all other sub-groups add views at a steady rate, showing that the popularity of these sub-groups is not affected during the observation window. Focusing on the standalone viral and popular sub-groups, we notice that, on average, popular videos receive a larger number of additional views per day than those in the viral category.

However, it is interesting to note that videos categorized as viral-and-popular add more views per days than any other sub-group, which presumably stems from their definition of being viral and popular at the same time.

Figure 4 shows the average number of Twitter mentions observed, for the various video sub-groups, during our measurement window. As expected, videos in the standalone viral sub-group are the most active ones in Twitter, and are mentioned more often than those that are categorized as popular only. Note that videos in the random and recent sub-groups, on average, are hardly mentioned in Twitter. The important takeaway here, though, is that videos that are both viral-and-popular are mentioned, on average, six times as much as videos that are only popular but not viral, and, on average, twice as much as videos that are only viral but not popular. This observation highlights again our proposition regarding the special attention that needs to be paid to the viral-and-popular sub-group.

Figure 5 further investigates the observed tweeting activity across the video sub-groups, focusing on the level of user engagement. For a given video, we consider the ratio of original tweets to the overall number Twitter mentions, i.e., the ratio between the number of tweets and the sum of the number of tweets and re-tweets. We observe that this ratio is relatively steady across all the video sub-groups. However,

Feature	Description
yt_views_n	Number of views added on day n
yt_likes_n	Number of likes added on day n
$yt\_ratings_n$	Number of ratings added on day n
$yt\_comments_n$	Number of comments added on day n
$yt\_uploaded$	Video upload timestamp
$yt\_category$	Video category
yt_rating	Average rating (0-5)
$yt\_uploader.uploads$	Number of videos uploaded by the uploader
$yt\_uploader.favorites$	Number of times the uploader was favorited
$yt\_uploader.subscribers$	Number of uploader's subscribers

Table 2: YouTube features for a video ($n \geq 1$).

Individual features	Description
tw_tweets_n	Number of tweets on day n
$tw\_orig\_tweets_n$	Number of original tweets on day n
$tw\_retweets_n$	Number of retweets posted on day n
$tw\_replies_n$	Number of tweet-reply posted on day n
$tw\_en\_language$	Number of tweets in English
$tw\_user.en\_language$	Number of users who set English as language
$tw\_user.statuses$	Number of user's status updates
$tw\_user.friends$	Number of user's friends (i.e., 'followees')
$tw\_user.followers$	Total number of followers
$tw\_user.favorites_n$	Number of tweets that were favorited on day n

Table 3: Twitter features for tweets mentioning a video ($n \geq 1$)

we observe also here that a video simply being popular is not sufficient for it to be highly mentioned on Twitter. The viral-and-popular videos not only receive a larger fraction of original tweets than viral and popular sub-groups individually, but also substantially larger than those videos that are only popular but not viral.

Figure 6 depicts how the Twitter OSN facilitates the spread of content. For every day of the observations, we compute the fan-out of the tweets, i.e., the average number of users reached by tweets mentioning videos in various sub-groups. We observe that this number is relatively stable over time in all the sub-groups. Note that the average number of users reached by tweets mentioning viral videos (both popular and not popular) is higher than that of the non-viral videos. Interestingly, the number of users reached by tweets mentioning viral-and-popular videos is inferior to the one reached by tweets mentioning only viral but not popular videos. We posit that this is explained by the large number of short and low-quality videos that are often tweeted by users with many followers. As we show later in the analyses, the number of followers, which is a proxy for a number of users reached by a tweets, is indeed an important feature for detecting content virality.

3. CLASSIFICATION AND PREDICTION RESULTS

In this section, we present performance evaluation of the classifiers trained to predict virality and/or popularity of videos, and analyze cross-system interactions between YouTube and Twitter. We also discuss the results of the feature importance analysis that provide valuable insights on the performance of the classifiers and underpin the observed cross-system interactions. Finally, we present a sensitivity analysis assessing the effect of the training and labeling windows sizes on the classification performance.

3.1 Methodology and Experimental Setting

Our analysis consists of a classification task, where we aim to classify each video, based on a set of input features, as viral and/or popular. For this, we train two binary classifiers capable of separately predicting the virality and popularity label for any given video. Given the outputs of these classifiers, we define the viral-and-popular class as the intersection of the viral and popular video sets. We implement the classifiers using a Gradient Boosted Decision Tree, widely used for general classification problems [13].

Table 2 shows a sample of YouTube features extracted per video. Some features refer to the video characteristics and some to the uploader of the video. It is important to note that the YouTube API returns the cumulative number of

video views, likes, ratings, and comments from the upload to the time of query. The value of the desired counters for any given day or period can be derived from the cumulative counts. Table 3 shows a sample of Twitter features extracted for an individual tweet mentioning a YouTube video. Again, some of them refer to the tweet itself and some to the user who posted the tweet.

In addition to the above features, we derive additional features, which we denote using the following modifiers:

- The superscript *ratio* is used to relate a feature value to its magnitude. For example, $yt\_views_n^{ratio}$ is the ratio between the number of views on day n, yt_views_n, and the total number of views since upload.

- The superscript *acc* denotes the acceleration of the feature, i.e., $yt\_views_n^{acc}$ is the ratio between yt_views_n and yt_views_{n-1}.

- The subscript *dif* denotes the difference between the *accumulated* value of a feature at the first and last day of a period, e.g., yt_views_{dif}. For instance, if collecting data over a three-day period, $yt\_views_{dif} = yt\_views_3$ - yt_views_1.

- The superscript *age_ratio* is used to relate a feature value to its age. For example, $yt\_views_n^{age\_ratio}$ is the ratio between the number of views on day n, yt_views_n, and the time since $yt\_uploaded$.

The classifiers use these features to predict binary labels of virality and popularity. We define the *training window* as the period used to train the classifier. When the training window is greater than one day, we populate also the above ratio, acceleration, daily, difference, and age ratio features. The ground truth[2] regarding the viral and popular (and the derived viral-and-popular) labels of a video is determined by its uptake during the *labeling window* that comes after the training window. By the uptake we refer to the number of YouTube views as the indicator of popularity and number of tweets mentioning the video as the indicator of virality. The period between the training and labeling windows is referred to as the *offset*. The offset can be set to 0, when predicting the virality and popularity classes of a video immediately after the training period, or to, e.g., 7 days, when predicting these one week after the training period.

We would like to highlight the importance of parameters pertaining to the size of these windows. Clearly, the most

---

[2]As presented earlier, we label top 2.5% most-viewed videos as popular and top 2.5% most-tweeted videos as viral.

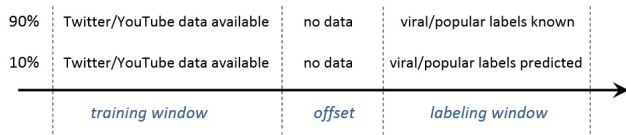| 90% | Twitter/YouTube data available | no data | viral/popular labels known |
| 10% | Twitter/YouTube data available | no data | viral/popular labels predicted |
| | *training window* | *offset* | *labeling window* |

Figure 7: Time line of the experimental methodology and data availability.

challenging predictions are those aiming to predict the virality and/or popularity of new videos as long as possibly in advance, i.e., small to none training window and large offset. This is the use case for predictive analytics in applications like advertisement and content placement. The size of the labeling window is not as important, and generally ranges from days to several weeks. Due to practical limitations, the data collection period was limited in our study to one month. Hence, in the sensitivity analysis we vary the size of the training window, offset, and labeling window from 1 to 7 days. To assess the accuracy of predictions for recently uploaded videos, we break the analysis into two groups: videos that are younger than 2 weeks and older videos.

For the evaluation of the predictive accuracy, we use the 10-fold validation methodology. That is, for 90% of videos we have the data referring both to the training window and the labeling window, such that we know what videos qualified to the viral and popular classes. For the remaining 10% of videos we have the training window data only and predict the the viral and/or popular class labels at the labeling window. The predicted and real labels of these 10% are used to compute the precision and recall scores. Also, the performance of the classifier is quantified using the area under the precision-recall curve (AUC) and the mean F1 score. We then repeat the experiment 10 times for differently chosen 10% chunks and average the obtained accuracy scores across these experiments. Figure 7 visualizes the time line of the windows and the evaluation methodology.

3.2 Prediction Accuracy

We start with examining the accuracy of predicting virality or popularity, or both. Here, we set the training window size to 2 days and the labeling window size to the immediately following 7 days, i.e., no offset. We use all the above mentioned YouTube and Twitter features to train the classifiers. As a comparison baseline, we chose a simple classifier that only uses two features: the number of original tweets and the number of views, i.e., $tw\_orig\_tweets^{cnt}$ as the Twitter feature and yt_views as the YouTube feature. This reflects a simple baseline classifier that uses only the raw statistics on the uptake of the videos in both systems. The analysis splits the videos into two groups, according to their age: those uploaded less than 14 days prior to data collection (referred to as *recent*) and the rest (*others*).

In Figure 8 we show the obtained precision-recall graphs. The rows show the source of features used by the classifiers (Twitter, YouTube, both) and the columns show the video class being predicted (viral, popular, viral-and-popular). In every graph, we depict 4 curves: classifier considering all the features vs. only the baseline counter features $tw\_orig\_tweets^{cnt}$ and yt_views, and accuracy of recent videos' classification vs. older videos.

3.2.1 Virality or popularity predictions

First, we focus on predicting the standalone viral and popular class labels. Prior work has shown that high prediction accuracy can be achieved by using simple features [25, 29, 10]. Our results reaffirm this finding, as shown by the high AUC values obtained when predicting virality with Twitter features (top-left graph in Figure 8) and popularity with YouTube features (central graph). The bottom-left and bottom-center graphs in Figure 8 show the performance when features from both systems are used to predict virality and popularity, respectively. In comparison to the previous Twitter- and YouTube-only graphs, we observe that there is little difference in performance. This suggests that, as expected, Twitter features are predominant when predicting virality on Twitter, whereas YouTube features are the most important for YouTube popularity predictions.

Note that very close AUC values are achieved when using only the baseline features from both systems, which suggests that applying regression to the training data is sufficient when predicting viral or popular videos. However, it should be highlighted that the AUC values obtained for predicting recent videos are substantially lower than those obtained when predicting older videos. This is explained by the abundance and reliability of data available for older videos, which allows achieving more accurate predictions. This result indicates that early detection of popularity and/or virality of recent videos, which may be the ultimate goal of content providers, is a rather challenging task that requires mobilizing many predictive features.

3.2.2 Cross-system predictions

Next, we proceed to cross-system predictions and assess the feasibility of using YouTube features to predict virality on Twitter, and vice versa, of using Twitter features to predict popularity on YouTube. This scenario investigates the interplay between Twitter and YouTube in terms of similarities between the evolution of content virality and popularity. The results are shown in the top-central and middle-left graphs in Figure 8.

We observe that both cross-system predictions are less accurate than the direct using data from the same system, i.e., virality predictions using Twitter data and popularity predictions using YouTube data. However while the obtained AUC values are lower than previously, we note that a reasonable accuracy is still achieved for all features and not-recent videos. At recall of 0.5, the precision is about 0.75 for both the virality and popularity predictions. For recent videos though, the accuracy deteriorates and the precision drops below 0.5 at a recall of 0.5. Focusing on predictions using all the available features, we observe that virality predictions using YouTube data are more accurate than popularity predictions using Twitter data. We posit that this is due to the fact that some YouTube views can be attributed to Twitter propagation, which is a strong predictive feature of virality. We also note that the baseline features perform considerably worse than all the features, which indicates that features mined from the original data improve the accuracy of cross-system predictions.

3.2.3 Viral-and-popular predictions

Videos in this class are usually mentioned and viewed more frequently than those in the standalone viral and popular classes. The accuracy of the classifier, shown in the right column in Figure 8, suggests that the highest accu-
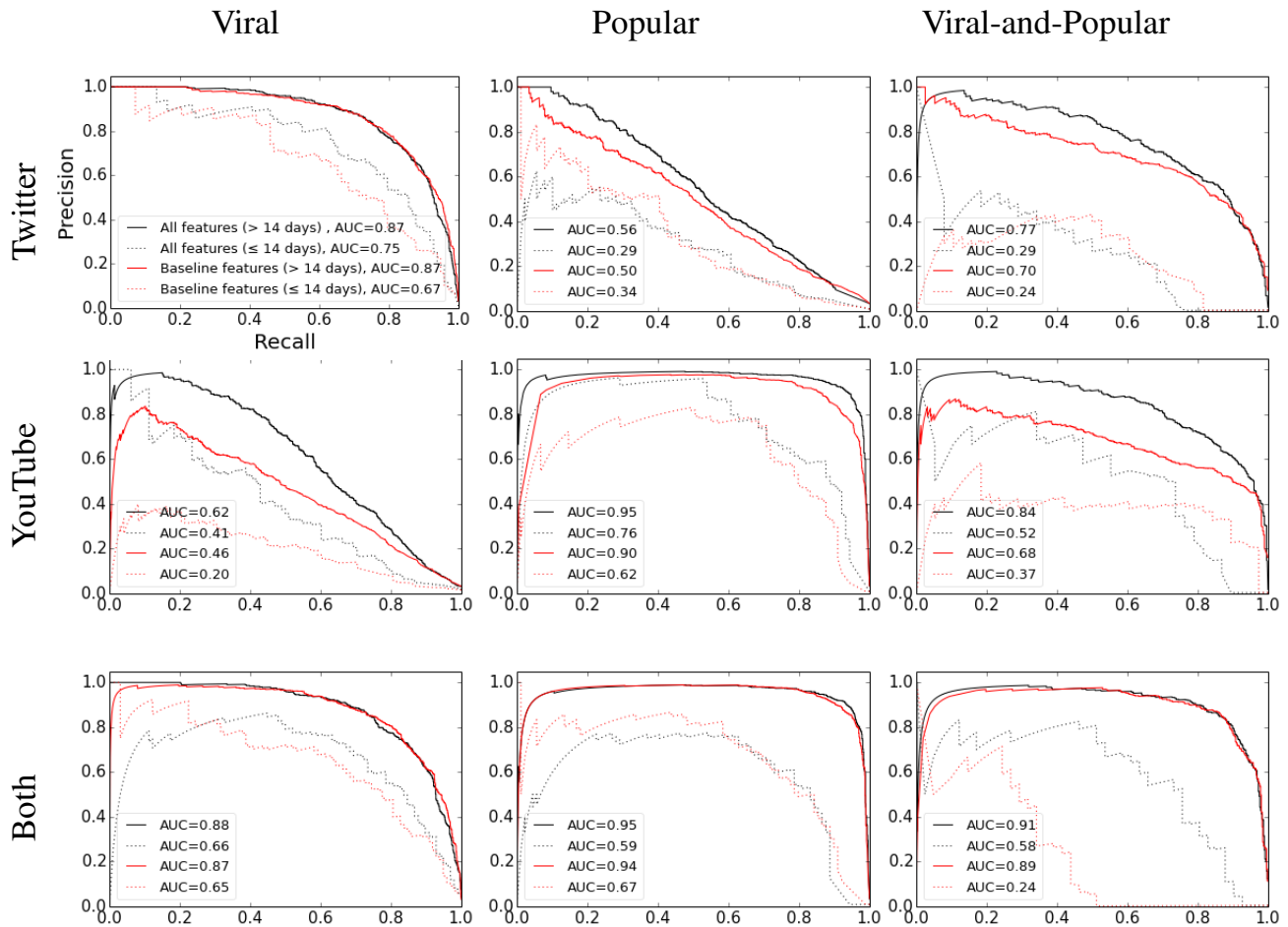
Figure 8: Prediction of top 5,000 (2.5%) most viral and popular videos. Precision-recall graphs and AUC values for all features vs baseline features and recent vs other videos.

racy is achieved when using both Twitter and YouTube features. Comparing the all features and the baseline features classifiers, we observe a substantial difference between the two when only Twitter features (top-right graph) or only YouTube features (middle-right) are used. This indicates that in this scenario, just like in the cross-system prediction task, the mined features substantially improve the accuracy of the classification.

On the contrary, the baseline feature are sufficient when using both YouTube and Twitter features, and the difference observed between the two classifiers is minor. This is also consistent with previous analysis that showed that the baseline features led to reasonably accurate classification. This observation, however, does not hold when focusing on the recent videos, where the gap between the two classifiers is wider. Our analysis reaffirms that early prediction of recently uploaded videos is challenging and, in this case, the baseline features need to be augmented with additional features mined from the original data. Moreover, virality-and-popularity predictions for recent videos achieve lower accuracy than their standalone prediction, showing that this task is hard even with the additional features.

3.3 Feature Importance Analysis

In this section, we focus on the importance of features for the classification task. We infer the feature importance scores[3] from the outputs of the Gradient Boost Decision Tree. They communicate the contribution of features to classifying each instance correctly (readers may refer to [19] for more details). Like in the experiments in section 3.2, the feature importance scores are computed for the setting of 2 training days, 7 labeling days, and no offset between the two. Hence, only simple *dif* and *ratio* features are considered. Since the direct predictions of virality using Twitter data (Figure 8, top-left) and of popularity using YouTube data (Figure 8, central graph) achieved high AUC values, we focus here on the cross-system and viral-and-popular predictions.

3.3.1 Cross-system predictions

Feature importance values when predicting video popularity on YouTube using Twitter features are shown in Table 4. For recently uploaded videos, we find that the current tweet-

---

[3]Note that the importance scores of the features do not sum up to 100%. Instead, the most important feature is marked as 100% and the importance of other features is computed relatively to the importance of the top feature.

Other (> 14 days)		Recent (≤ 14 days)	
feature	import.	feature	import.
$tw\_orig\_tweets_2$	100%	$tw\_orig\_tweets_2$	100%
$tw\_orig\_tweets_1^{ratio}$	69.6%	$tw\_user.favorites_{dif}$	66.6%
tw_tweets_{dif}	69.6%	$tw\_user.friends_{dif}$	60.9%
$tw\_orig\_tweets_1$	60.5%	$tw\_user.favorites_2$	57.1%
$tw\_orig\_tweets_2^{ratio}$	51.7%	$tw\_user.statuses_{dif}$	47.9%
$tw\_user.statuses_{acc}$	47.2%	$tw\_orig\_tweets_1$	47.3%

Table 4: Popularity predictions using Twitter features.

Other (> 14 days)		Recent (≤ 14 days)	
feature	import.	feature	import.
yt_likes_2	100%	$yt\_ratings_{dif}$	100%
yt_likes_1	61.9%	yt_likes_1	86.1%
yt_views_2	58.6%	$yt\_views_2^{age\_ratio}$	81.0%
yt_views_1	49.5%	yt_likes_2	69.6%
$yt\_user.uploads$	31.5%	yt_views_2	67.8%

Table 5: Virality predictions using YouTube features.

Other (> 14 days)		Recent (≤ 14 days)	
feature	import.	feature	import.
yt_views_2	100%	yt_views_2	100%
yt_views_1	29.0%	yt_likes_2	12.7%
$tw\_orig\_tweets_2$	18.8%	$yt\_comments_2$	6.42%
$yt\_views_{ageRate}$	14.3%	$tw\_en\_language_{acc}$	4.98%
$yt\_comments_1$	13.2%	$tw\_user.en\_language_{acc}$	4.45%

Table 6: Virality-and-popularity predictions using all the features.

ing rate, captured by $tw\_orig\_tweets_2$ is amongst the most predictive features. (Recall that training window size is 2 days, and, thus, the number of tweets on the second day is the measure of the current tweeting rate.) We also note that the propagation measured by the volume of users reached by tweets over the training period, $tw\_user.friends_{dif}$, is an important feature. Features capturing the user engagement with the Twitter eco-system, such as the number and ratio of tweets that were favorited, also play an important role in the video popularity classification accuracy. When considering predictions for less recent videos, we observe that active user engagement indicated by user's propensity to write original tweets, $tw\_orig\_tweets_2^{ratio}$, is also an important feature. Summarizing, we find that there are three key factors in predicting a future video via Twitter: the current tweeting rate, the propagation effect, and the user's engagement with Twitter content. A combination of these factors communicates the magnitude of audience that can potentially be achieved by a tweet.

When predicting virality of recently uploaded videos using YouTube features (shown in Table 5), we observe that the difference in terms of ratings added to the video, $yt\_ratings_{dif}$, is the most dominant predictive feature. Similarly, the rates capturing the number of likes per day are also important features. It could be argued that rating and liking videos in YouTube is similar in flavor to tweeting activity, which explains their predictive importance. Another relevant feature in the virality classification relates to the ratio between the number of views at the time of classification and the time since the video upload, $yt\_views_2^{age\_ratio}$. In essence, the growth rate of likes, ratings, and likes are the predominant features for predicting the virality of recent videos based solely on YouTube features. Similarly, the number of likes and views seem to be the predominant features when predicting the virality of not-recent videos. In addition, we observed that the number of video uploaded by the uploader plays a role in such predictions.

3.3.2 Viral-and-popular predictions

Lastly, Table 6 shows feature importance values obtained when predicting the viral-and-popular class. As can be seen,

YouTube features generally dominate Twitter features, with the number of YouTube views in the second training day substantially outweighing other features for predictions of both recent and other videos.

Recall that the viral-and-popular videos are the sub-group that adds, on average, the largest number of views per day, and are also the videos that, on average, receive the highest number of tweets per day. However, for the purposes of prediction, these results indicate that a YouTube video's most recent viewing rate may have sufficient predictive capability for this particular class of videos (as also supported by AUC scores in the right column in Figure 8). Our analysis also indicates that the addition of Twitter features to YouTube features yields a better accuracy of the viral-and-popular classifier. Hence, Twitter features contribute additional knowledge that facilitates more accurate predictions.

All in all, we observe the dominance of the recent features in the predictive mechanism. This is not surprising, considering that recent activity signal are normally reliable indicators for the near future behavior. Hence, we point now to the investigation of the effect ofthe training window size, the labeling window size, and the offset between the two.

3.4 Sensitivity Analysis: Effect of Training and Labeling Window Size

In these section, we vary the size of the training and labeling windows, and the offset between them. We compute the mean F1 score as the performance metric. Again, we focus on the cross-system and viral-and-popular predictions.

3.4.1 Cross-system predictions

Table 7 shows the impact of training, labeling, and offset window sizes on the accuracy of virality predictions using YouTube features. As can be seen, wider training window steadily yields higher F1 scores, where using a training window of more than one day produces a noticeable improvement, since the difference between the training window of 3 and 7 days is minor. On top of the mere data availability, training window of more than one day allows for the inclusion of the difference features (subscript dif), which were found to be influential in the feature importance analysis.

Turning to the labeling window size (left part of table), we observe that windows of 3 and 7 days allow for more accurate predictions than a 1-day window. We posit that this is explained by the more stable nature of the virality labels based on 3 and 7 days of data. The observed impact of increasing the offset window from 1 day to 7 days (right part of the table) is minor. This finding reaffirms our earlier observation regarding the stability of content virality over reasonably short periods of time, in particular, over a short period closely following the training window.

	Label window size (offset = 1)			Offset (label window size= 3)		
Training window size	1 day	3 days	7 days	1 day	3 days	7 days
1 day	0.336	0.382	0.407	0.383	0.385	0.400
3 days	0.440	0.496	0.510	0.496	0.494	0.484
7 days	0.447	0.501	0.517	0.506	0.495	0.494

Table 7: Effect of training, labeling, and offset window size on F1 of virality predictions using YouTube data.

	Label window size (offset = 1)			Offset (label window size= 3)		
Training window size	1 day	3 days	7 days	1 day	3 days	7 days
1 day	0.418	0.426	0.430	0.428	0.412	0.413
3 days	0.464	0.468	0.467	0.469	0.466	0.452
7 days	0.487	0.477	0.480	0.480	0.474	0.479

Table 8: Effect of training, labeling, and offset window size on F1 of popularity predictions using Twitter data.

Table 8 shows a similar analysis conducted for popularity predictions using Twitter data. Again, popularity classification accuracy increases with the training window size, although not to the extent observed for virality predictions. This is in line with our previous observations regarding the greater stability of the popularity scores over time. The main difference with respect to virality predictions refers to the weak impact of the labeling window size (left part of the table), which reaffirms that YouTube popularity is more stable and less susceptible to fluctuations than Twitter virality. Like in the previous case, changes in the offset window size have little impact on the F1 scores.

3.4.2 Viral-and-popular predictions

Finally, Table 9 shows the training and labeling window size analysis when predicting the viral-and-popular class, using features from both YouTube and Twitter. We observe that the impact of the training window size here is weaker than in the previous case of cross-system predictions. We posit that this is explained by the higher ratio of data, which complements and hardly augments previously available data. As such, little new information is introduced over over time and increasing the training window size does not substantially improve the accuracy.

On the contrary, we observe here a stronger impact of the labeling window size and of the offset. The accuracy of the predictions is found to decrease with the offset between the training and test data, indicating that the viral-and-popular class label is more volatile over time than the standalone viral and popular classes, and, therefore, it is more difficult to predict at greater offsets. Since the viral-and-popular class is more volatile, larger labeling window produces a more stable labeling data and leads to a higher predictive accuracy, as we observe in this experiment.

4. RELATED WORK

This section provides an overview of research related to various aspects of our work.

The extensive body of research on *video popularity prediction* spans a number of sub-disciplines. In particular, a number of prior studies analyzed the popularity of YouTube

	Label window size (offset = 1)			Offset (label window size= 3)		
Training window size	1 day	3 days	7 days	1 day	3 days	7 days
1 day	0.769	0.809	0.825	0.807	0.798	0.788
3 days	0.770	0.823	0.830	0.824	0.817	0.790
7 days	0.778	0.815	0.824	0.816	0.800	0.771

Table 9: Effect of training, labeling, and offset window size on F1 of viral-and-popular predictions.

videos, and interrelation between video popularity and OSN activity. Abisheva et al. combined user data from Twitter with video data from YouTube to discover correlations and temporal dependencies between user features, shares on Twitter, and the number of Twitter views [2]. Wang et al. designed a neural network-based learning framework for predicting the number of video views by exploring patterns of video link propagation in a microblogging system [29]. Several studies exploited the social and contextual activity related to video sharing, aiming at predicting the video view count [25, 11, 26]. Other studies measured the spread of videos within OSNs [10, 16]. Li et al. developed a propagation-based model for predicting peaks and bursts of video views basing on spread patterns [17]. Likewise, by understanding the geographic aspects of video popularity and extracting information from social activity, Scellato et al. proposed various caching strategies for improving the performance of content delivery networks [24].

Furthermore, there have been several efforts towards analyzing the popularity of online video content using early view trends and content meta-data extracted from the video provider data. Pinto et al. proposed a multivariate linear model, configured to capture information about the popularity evolution patterns of videos, which was later on used as an early indicator to predict longer-term popularity [22]. Borghol et al. characterized and modeled video popularity dynamics using a set of measures, such as the video age, churn statistics, the evolution of the viewing rate, and the distribution of view counts over time [9]. Borghol et al. also developed an approach for assessing the impacts of various content-agnostic factors on video popularity, and showed that when controlling for video content strong linear "rich-get-richer" behavior is seen with the total number of previous views as the most important predictor for future views except for very young videos [8]. Ahmed et al. developed an approach that first clustered the temporal data into discretized states, such that transition probabilities between the states were learned and then used to predict future popularity of content [3].

Many previous studies have also examined the *popularity* of social media content. For example, Lee et al. [14] and Lerman et al. [15] proposed somewhat similar models based on endogenous factors, aimed at predicting the popularity of content published in online discussion groups. Their works showed that social activity alone could be a key explanatory factor for accurate popularity predictions. Wang et al. [28] studied the interrelation between the trendiness of content in microblogging systems and its popularity in search queries, and showed that topics of interests in Twitter could serve as reliable early indicator of sudden popularity of related keywords in Google search. Romero et al. [23] focused on the connections between the social graph structure and the

topical affiliations, i.e., interests and hashtags, of Twitter users, and concluded that both the hashtags applied by users can predict their social connections and the existing social connections can predict the popularity of hashtags.

Recent studies focused on understanding *how* content spreads and becomes viral on microblogging services [5, 6, 21, 4]. For instance, Bakshy et al. defined influence in the context of Twitter and analyzed the influence of content linked by Tweets with respect to how the content spreads on the network [5]. Their findings showed that both the overall influence of the poster and the characteristics of the content impacted the content's spread range. A related study was done in the context of news sharing behavior by Berger et al. [6], and it was found that emotions evoked by news had a significant effect on their dissemination over the network. Most of the studies into content virality on Twitter assumed little or no external influence on how URLs spread over the network. Myers et al. proposed a model capable of capturing such external influence for breaking news [21]. Somewhat surprisingly, the authors concluded that roughly 30% of the URLs mentions were attributed to external sources, while 70% - exclusively to network effects.

The massive quantity of social content constitutes a gold mine for personalization services and content recommendation engines. Morales et al. [12] exploit Twitter personas to recommend relevant news to users, by studying overall topic popularity in the news and by combining the user profile and the recent interest in the social neighborhood. Twitter diffusion of information, and specifically links, was also successfully leveraged to provide personalized recommendation of trending topics [18]. Content recommendations were achieved by matching highly popular tweets to the recepient's profile, constructed by analyzing their past posts, mentions, URLs, hashtags, and other observable past behaviour [1]. An accurate model for scoring and re-ordering OSN activity feed items through modeling user preferences towards certain actions, users, and content, and matching these preferences to the available network updates, was developed by Berkovsky et al. in [7].

In contrast to the above prior works, our work examines the simultaneous evolution of the popularity and virality of video content, and highlights the inter-relations between the two platforms contributing to these characteristics. We also investigate the compound class of viral-and-popular video content items, as this class is likely to be most attractive for content producers and media agencies alike. We conjecture and verify that cross-system predictions of both popularity and virality of videos is feasible (although not straightforward in certain cases), and can be performed with a reasonably high degree of accuracy. In addition, we analyse the impact of temporal dimension on the predictions and provide insights into the features extracted from both platforms, which contribute to the accuracy of the predictions. This work represents, to the best of our knowledge, the first systematic data-driven study of the predictions of popularity and virality aspects of online video content.

5. CONCLUSIONS

Prediction of content virality on microblogging services and of video popularity on video sharing platforms have been largely investigated so far as two independent tasks. In this work, we propose a new unifying approach for predicting video content virality, popularity, as well as virality-and-popularity at the same time, using a suite of features extracted from Twitter and YouTube logs.

Our experiments indicate that these predictions are feasible, can exhibit very high levels of accuracy, and can be carried out with reasonably low amounts of training data. Specifically, YouTube popularity predictions achieve a remarkably high AUC scores (e.g., 0.95 using all the available data), whereas predictions of rather volatile Twitter virality still achieve reasonably high AUC (e.g., 0.88 using all the available data). Furthermore, we demonstrate that not only direct but also cross-system predictions are possible, although with a lower degree of accuracy. Lastly, the compound predictions of video content virality-and-popularity can also be predicted with high accuracy (e.g., AUC of 0.91, when both Twitter and YouTube features are used for the prediction), highlighting the importance of features mined from the raw data.

We proceed with an analysis of feature importance and highlight what features are predictive of future popularity and virality of content, both for content having solid prior history and for recently uploaded items, where the prediction task is more important and more challenging at the same time. We observe that recent information, and specifically current growth rate, often comes through as a highly predictive feature and that virality-and-popularity of video content is mainly driven by YouTube features. Finally, we analyze the sensitivity of the developed prediction mechanism and its dependence on the availability of reliable training data. While predictions of virality and/or popularity are feasible and accurate for non-recent content having rich prior information, we observe that having training data of only one day allows for the generation of accurate predictions for up to seven days in advance.

In the future, we plan to extend our model and incorporate features mirroring the social facet of the systems, e.g., user influence [6] and external influence [21]. We also intend to capitalize on more sophisticated temporal [3] and graph models [27], and further investigate the evolution of popularity and virality over time. We also intend to apply the developed predictive models to other types of content, like news and music. We believe that our work paves to way to a broad stream of future research investigating the intricacies of cross-system propagation of content, user-to-user influence, and cross-system content discovery.

6. REFERENCES

[1] F. Abel, Q. Gao, G.-J. Houben, and K. Tao. Analyzing user modeling on twitter for personalized news recommendations. In *Proceedings of the International Conference on User Modeling, Adaption, and Personalization*, pages 1–12, 2011.

[2] A. Abisheva, V. R. K. Garimella, D. Garcia, and I. Weber. Who watches (and shares) what on youtube? and when?: Using twitter to understand youtube viewership. In *Proceedings of the international conference on Web search and data mining*, pages 593–602, 2014.

[3] M. Ahmed, S. Spagna, F. Huici, and S. Niccolini. A peek into the future: Predicting the evolution of popularity in user generated content. In *Proceedings of the International Conference on Web Search and Data Mining*, pages 607–616, 2013.

[4] S. Ardon, A. Bagchi, A. Mahanti, A. Ruhela, A. Seth, R. M. Tripathy, and S. Triukose. Spatio-temporal and events based analysis of topic popularity in twitter. In *Proceedings of the International Conference on Information and Knowledge Management*, pages 219–228, 2013.

[5] E. Bakshy, J. M. Hofman, W. A. Mason, and D. J. Watts. Everyone's an influencer: Quantifying influence on twitter. In *Proceedings of the International Conference on Web Search and Data Mining*, pages 65–74, 2011.

[6] J. Berger and K. L. Milkman. What makes online content viral? *Journal of Marketing Research*, pages 1–17, 2011.

[7] S. Berkovsky, J. Freyne, and G. Smith. Personalized network updates: Increasing social interactions and contributions in social networks. In *Proceedings of the International Conference on User Modeling, Adaption, and Personalization*, pages 1–13, 2012.

[8] Y. Borghol, S. Ardon, N. Carlsson, D. Eager, and A. Mahanti. The untold story of the clones: Content-agnostic factors that impact youtube video popularity. In *Proceedings of the International Conference on Knowledge Discovery and Data Mining*, pages 1186–1194, 2012.

[9] 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, 2011.

[10] X. Cheng, H. Li, and J. Liu. Video sharing propagation in social networks: Measurement, modeling, and analysis. In *Proceedings of the International Conference on Computer Communications*, pages 45–49, 2013.

[11] 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*, 105(41):15649–15653, 2008.

[12] G. De Francisci Morales, A. Gionis, and C. Lucchese. From chatter to headlines: Harnessing the real-time web for personalized news recommendation. In *Proceedings of the International Conference on Web Search and Data Mining*, pages 153–162, 2012.

[13] J. H. Friedman. Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29:1189–1232, 2000.

[14] J. G. Lee, S. Moon, and K. Salamatian. An approach to model and predict the popularity of online contents with explanatory factors. In *Proceeedings of the International Conference on Web Intelligence and Intelligent Agent Technology*, pages 623–630, 2010.

[15] K. Lerman and T. Hogg. Using a model of social dynamics to predict popularity of news. In *Proceedings of the International Conference on World Wide Web*, pages 621–630, 2010.

[16] H. Li, J. Liu, K. Xu, and S. Wen. Understanding video propagation in online social networks. In *Proceedings of the Workshop on Quality of Service*, 2012.

[17] H. Li, X. Ma, F. Wang, J. Liu, and K. Xu. On popularity prediction of videos shared in online social networks. In *Proceedings of the International Conference on Information and Knowledge Management*, pages 169–178, 2013.

[18] H. Liang, Y. Xu, D. Tjondronegoro, and P. Christen. Time-aware topic recommendation based on micro-blogs. In *Proceedings of the International Conference on Information and Knowledge Management*, pages 1657–1661, 2012.

[19] Y. Lv, D. Lymberopoulos, and Q. Wu. An exploration of ranking heuristics in mobile local search. In *Proceedings of the International Conference on Research and Development in Information Retrieval*, pages 295–304, 2012.

[20] X. Ma, H. Wang, H. Li, J. Liu, and H. Jiang. Exploring sharing patterns for video recommendation on youtube-like social media. *Multimedia Systems*, 20(6):675–691, 2014.

[21] S. A. Myers, C. Zhu, and J. Leskovec. Information diffusion and external influence in networks. In *Proceedings of the International Conference on Knowledge Discovery and Data Mining*, pages 33–41, 2012.

[22] H. Pinto, J. M. Almeida, and M. A. Gonçalves. Using early view patterns to predict the popularity of youtube videos. In *Proceedings of the International Conference on Web Search and Data Mining*, pages 365–374, 2013.

[23] D. M. Romero, C. Tan, and J. Ugander. On the interplay between social and topical structure. In *Proceedings of the International Conference on Weblogs and Social Media*, 2013.

[24] S. Scellato, C. Mascolo, M. Musolesi, and J. Crowcroft. Track globally, deliver locally: Improving content delivery networks by tracking geographic social cascades. In *Proceedings of the International Conference on World Wide Web*, pages 457–466, 2011.

[25] D. A. Shamma, J. Yew, L. Kennedy, and E. F. Churchill. Viral actions: Predicting video view counts using synchronous sharing behaviors. In *Proceedings of the International Conference on Weblogs and Social Media*, 2011.

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

[27] A. Tiroshi, S. Berkovsky, M. A. Kâafar, D. Vallet, T. Chen, and T. Kuflik. Improving business rating predictions using graph based features. In *Proceedings of the International Conference on Intelligent User Interfaces*, pages 17–26, 2014.

[28] D. Wang, M. A. Kaafar, K. Salamatian, and G. Xie. What's trendy right now? a comparative analysis of web and microblogs trending topics. In *Proceedings of the International Conference on Internet Science*, 2013.

[29] Z. Wang, L. Sun, C. Wu, and S. Yang. Guiding internet-scale video service deployment using microblog-based prediction. In *Proceedings of the International Conference on Computer Communications*, pages 2901–2905, 2012.