Proceedings of the Tenth International AAAI Conference on
Web and Social Media (ICWSM 2016)

A Data-Driven Study of View Duration on YouTube

Minsu Park[1], **Mor Naaman**[1], **Jonah Berger**[2]
[1]Jacobs Institute, Cornell Tech
[2]The Wharton School, University of Pennsylvania
{minsu, mor}@jacobs.cornell.edu
jberger@wharton.upenn.edu

Abstract

Video watching had emerged as one of the most frequent media activities on the Internet. Yet, little is known about how users watch online video. Using two distinct YouTube datasets, a set of random YouTube videos crawled from the Web and a set of videos watched by participants tracked by a Chrome extension, we examine whether and how indicators of collective preferences and reactions are associated with view duration of videos. We show that video view duration is positively associated with the video's view count, the number of likes per view, and the negative sentiment in the comments. These metrics and reactions have a significant predictive power over the duration the video is watched by individuals. Our findings provide a more precise understandings of user engagement with video content in social media beyond view count.

Introduction

Video watching is perhaps the most popular web-based activity, through video hosting and sharing services such as YouTube, Facebook, Netflix, Vimeo, and others (Cisco 2015). As of 2015, YouTube alone has more than 1 billion viewers every day, watching hundreds of millions of hours of content (YouTube 2015). It is forecasted that video will represent 80 percent of all traffic by 2019 (Cisco 2015). Yet, little is known about how users engage with and watch online video. We use two distinct datasets from YouTube to investigate how users' engagement in watching a video (i.e., view duration) is associated with other video metrics such as the number of views, likes, comments, and the sentiment of comments.

A number of research efforts have investigated view count as a key indicator of popularity or quality of video—particularly looking at its relationships with other popularity or preference metrics (e.g., the number of likes and comments). For example, the number of comments/favorites/ratings and average rating are significant predictors of video view counts on YouTube (Chatzopoulou, Sheng, and Faloutsos 2010); the sequence of comments and its structure are strongly associated with view counts (De Choudhury et al. 2009); and view counts can be predicted through socially shared viewing behaviors around the content such as how many times a video was rewound or fast-forwarded as well as the duration of the session in a tool that allows people watch videos together in sync and real time (Shamma et al. 2011).

Although views, likes, comments, and other such measures can be considered as indicators of general popularity and preferences, there has been growing interest in using deeper post-click user engagement (e.g., how long a user watched a video) to estimate more accurate relevance and interest and to improve ranking and recommendation (Yi et al. 2014). For example, YouTube has started to use 'dwell time' (the length of time that a user spends on a video, e.g., video watching session length) instead of click events to better measure the engagement with video content (YouTube 2012). Beyond video, Facebook is using dwell time on external links to combat Clickbait—stories with arousing headlines that attract users to click and share more than usual, but are not consumed in depth (Cramer 2015; El-Arini and Tang 2014).

Understanding the relationships between watching behavior and other popularity and engagement metrics can help develop more comprehensive behavioral models of user engagement and preferences beyond view count. In this work, we explore the relationship between user engagement with video (i.e., the portion of the video watched by the user) and other activity metrics (e.g., the number of views/likes/dislikes/comments) that are often considered as indicators of popularity or collective preference (Baym 2013) and may impact or at least align with user engagement and watch duration. In fact, our basis hypothesis is that:

H1: Video view duration is positively associated with other video popularity and quality measures.

Moreover, we examine another factor which may be associated with video view duration: the sentiment of video comments. A number of recent studies in different disciplines provided potential links among emotionality of contents, comments, and user engagement. Comments are indicative of users' opinion of and attitude towards video content (Yew, Shamma, and Churchill 2011). Interest, enjoyment, and desire to find out more about online news are enticed when strong sentiment and negative connotations are present in the content (Arapakis et al. 2014). Further, emotionality of content and its intensity can be reflected on comments (Alhabash et al. 2015), or, alternatively, certain aspects of con-

tent may cause people to comment with strong sentiment. In each case, we expect that sentiment of comments are correlated with engagement in terms of view duration:

H2: Video view duration is positively associated with the sentiment (positive or negative) of comments.

To examine how those indicators of collective preferences and reactions are associated with engagement, we devise a data-driven study, using two distinct YouTube datasets: a set of randomly selected YouTube videos crawled from the Web and a set of videos watched by study participants that installed a Chrome extension tracking their behavior on YouTube. The main contribution in this work, then, is showing the robust relationships between view count, the number of likes per view, negative sentiment in comments, and video view duration. In this, we show that observed popularity metrics indeed have a significant predictive power for engagement. These findings could inform web services that require more precise understandings of engagement with social media video content beyond view count.

Method

To make our findings and analysis more robust, we used two distinct datasets.[1] For each of these datasets we computed an overlapping but not completely identical set of variables. In particular, the dependent variable representing view duration is different in both datasets, as we detail below.

Data Collection

Random Videos. This dataset is a sample of 1,125 random videos from YouTube. We first obtained 100,000 randomly sampled YouTube videos through Random YouTube, a site that collects IDs of the 10 most recently uploaded videos every fifteen minutes.[2] We then retrieved the video properties for each of the videos through the YouTube API. To find the average watch duration for each video, we scraped the YouTube video page to get the aggregate watch duration available on the video page statistics tab using tools from (Yu, Xie, and Sanner 2015). This step resulted in 44,766 videos—other videos were either private videos, were deleted, or did not provide aggregate statistics. Finally, to ensure proper measurement of sentiment, we required videos in our sample to have at least five English comments, resulting in a final set of 1,125 videos.

For the *Random Videos* dataset, our view duration dependent variable was computed using the video's *average view duration*: the aggregate view duration of the video (as reported by YouTube) divided by the view count, both reported over the video's lifetime.

Individual Logs. The *Individual Logs* dataset is a set of 1,814 videos watched over several weeks by a sample of 158 participants recruited from Task Rabbit.[3] To obtain this individual-level data, we developed and deployed a Chrome extension that collected how long each user spent each YouTube video page they visited, as well as other YouTube video information available at the time of the viewing. The extension recorded a view timestamp, video properties, and dwell time in each viewing session. The extension was installed and used by 189 participants who had a previous experience with YouTube, over a span of at least two weeks (our sample included 77 male and 112 female participants; there were no significant gender differences in usage of the extension). Using this extension, we collected a total of 17,599 video view sessions. We filtered videos which had missing information, were private, deleted, or did not have at least five English comments. We only kept videos where the participant's dwell time was not greater than the video duration to avoid unusually long dwell time due to potential errors (e.g., unexpected disconnections).[4] The final dataset, as mentioned above, included 1,814 video view sessions.

For the *Individual Logs* dataset, our view duration dependent variable was computed differently. In this case, we have used an individual, but approximate, view duration measurement. In particular, we used the user's dwell time for each video on the video's page, as was measured by the extension, as an approximation for the actual view time for the video by that user.

To summarize, we used two distinct datasets and collected or computed variables in slightly different fashion for each. The *Random Videos* dataset provides an aggregate understanding of average view duration that may capture life cycle dynamics of videos (e.g., older videos are potentially viewed differently over time) whereas the *Individual Logs* dataset provides an understanding of individual's viewing behavior. Note that videos viewed by our participants tend to be more popular than the videos in the *Random Videos* dataset. As we show below, while these two datasets resulted in a different distribution of videos as reflected in the video metadata variables, our results, associating these variables with the dependent variable of view duration, are fairly robust across datasets, adding confidence in the outcome.

Sentiment Analysis and Mixed-effects Model

In order to gauge the intensity and direction of emotionality of comments, we performed sentiment analysis on all of the comments for a given video by using VADER, a lexicon and rule-based sentiment analyser (Hutto and Gilbert 2014). This tool is specifically attuned to sentiment expressed in social media and sensitive both the polarity and the intensity. Thus, it is ideal for use with comments on YouTube which often use less formal language. We considered only positive and negative sentiment for our analysis.[5]

[1] Both datasets are available on http://github.com/sTechLab/YouTubeDurationData

[2] http://randomyoutube.net

[3] http://taskrabbit.com

[4] We examined different cutoffs on dwell time from 1.5 times to 3 times longer than the video duration with truncation at 1 and found no impact on our key findings.

[5] We also computed emotionality of comments by Linguistic Inquiry and Word Count (LIWC) (Tausczik and Pennebaker 2010). While LIWC is widely used, it does not include consideration for sentiment-bearing lexical items such as acronyms, initialisms, emoticons, or slang, which are known to be important for sentiment analysis of social media text. The correlations between sentiments computed by LIWC and Vader were 0.84 (positive) and 0.83 (negative) for the *Individual Logs* dataset, and 0.82 (positive)

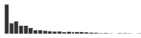
	Random Videos	Max	Individual Logs	Max
View Duration Prop.		1		1
Video Properties				
Duration (Sec)		15,229		7,267
Elapsed Time (Days)		3,059		3,140
Metrics & Sentiments				
View Count		56M		678M
Comments per View		1.33		0.075
Discussion Density		1.11		0.71
Positive Emotion		0.84		0.57
Negative Emotion		0.53		0.46
Likes per View		0.98		0.86
Dislikes per View		0.26		0.18
Like-Dislike Ratio		1		1
Shares per View		0.18		0.016

Table 1: Distribution of user engagement and independent variables used to estimate the engagement in each dataset. The distributions accompanying each variable begin at zero and end at the adjacent maximum.

We performed two mixed-effects OLS regressions, one regression for each dataset. We slightly modified the dependent variable for the regression, using the proportion of view duration for each video (rather than the view duration itself). In other words, for the *Random Videos* we used the average view duration for each video divided by the video length. For the *Individual Logs* dataset we used the user dwell time on the video page divided by video length.

In the models for both datasets, we added control variables capturing key video properties, video length and age (elapsed time since the video was published). Both these variables were added as fixed effects. The key video variables we examine in this work include view count, comments per view (the number of comments divided by the number of views), likes/dislikes/shares per view (similarly computed), and positive and negative sentiments in comments as computed by VADER and resulting in a 0–1 score. We also used discussion density (the proportion of comments that are replies to other comments) and like-dislike ratio since dense discussion and clear preference among users may help to explain view duration. We added interaction terms between the sentiment and other variables as well.

All independent variables except sentiments and like-dislike ratio were logged (base 10) to account for skewness before being standardized and centered. Multicollinearity between variables was not an issue as all variance inflation factors were less than 2. Table 1 presents the histogram distributions of variables used for our analysis.

We used a mixed-effects model to control for multiple measurements of the same individuals (*Individual Logs* dataset) and for differences between video in different categories (both datasets). Participants were added as a random effect in the individual logs model. We used the video category as a random effect as well in both models, because we noticed significant category-dependent differences in view duration. Indeed, in the *Individual Logs* dataset, for example, we found that the categories Comedy, Entertainment, People & Blogs, Film & Animation, and Gaming, were positively associated with the watch duration dependent variable (these categories were watched significantly longer compared to the baseline of the Auto & Vehicle category, even with video length and elapsed time as controls).

and 0.58 (negative) for the *Random Videos* dataset, all significant at $p < 0.001$. We therefore focus on using VADER in this work.

Results

Table 2 shows the results of the two regression models for the two different datasets. The intercepts for the regressions represent a point with all variables at zero, in which case view duration values are 49% and 66% of a video in the *Random Videos* and *Individual Logs* datasets, respectively (this disparity is indicative of the difference between the datasets). Note that in both models, the video length has a significant impact on the view duration: as expected, shorter videos were watched much longer in terms of proportion of the video viewed. Indeed, the video length accounts for much of the R^2 value (justifying our choice to use it as a control variable). However, we note that the R^2 of same models without the video length control was 0.19 for both. When excluding both control variables, the R^2 values were 0.13 (*Random Videos*) and 0.18 (*Individual Logs*). These results indicate that video popularity measures and sentiment of comments have significant explanatory power over user engagement.

We treat the results that are highly significant across both datasets as our most robust findings. These results provide partial supports for both *H1* and *H2*. Supporting *H1*, the video popularity and quality metrics of view count and likes per view were associated in both models with a higher portion of the video viewed (the ratio of people liking the video contributing more than the view count). Partially supporting *H2*, the negative sentiment (but not the positive) in comments was positively correlated with view duration in both models.

Other results were not significant in both models and we thus view them as less robust, perhaps indicative of potential association that can be exposed in a larger-scale study. The percentage of comments and positive sentiment (negative), and discussion density (positive) were correlated with view duration in the *Random Videos* dataset. Dislike percentage was negatively correlated only in the *Individual Logs* dataset. Finally, we note that there was no robust trend on Interaction terms.

Discussion and Conclusion

Drawing on two distinct YouTube datasets, this study demonstrates the relationship between popularity metrics, comment sentiment, and video watching (or the proportion of video watched by users). Our analysis shows a robust pattern: videos with higher ratio of likes per view, higher

	Random Videos		Individual Logs	
	Estimate	SE	Estimate	SE
(Intercept)	0.490	0.011	0.658	0.021
video_length	-0.146***	0.005	-0.123***	0.009
elapsed_time	-0.034***	0.005	-0.006	0.013
view_count	0.026**	0.009	0.045**	0.014
comment_percentage	-0.029**	0.011	0.010	0.014
discussion_density	0.015**	0.005	0.023	0.014
like_percentage	0.094***	0.012	0.048**	0.016
dislike_percentage	0.004	0.006	-0.042**	0.015
like_dislike_ratio	0.014*	0.006	0.004	0.012
share_percentage	0.015**	0.005	-0.010	0.009
pos	-0.014**	0.005	0.009	0.011
neg	0.019***	0.005	0.022*	0.011
like_percentage:pos	0.011	0.011	0.014	0.012
like_percentage:neg	0.020*	0.009	0.004	0.012
dislike_percentage:pos	0.005	0.004	-0.012	0.008
dislike_percentage:neg	0.005	0.006	-0.028*	0.013
share_percentage:pos	-0.004	0.005	-0.012	0.011
share_percentage:neg	0.005	0.005	-0.011	0.010
comment_percentage:pos	0.010	0.008	-0.025*	0.011
comment_percentage:neg	0.004	0.007	-0.012	0.010
discussion_density:pos	0.009	0.005	0.000	0.010
discussion_density:neg	0.000	0.005	0.002	0.011
N	1,125		1,814	
R^2	0.58		0.28	

* p < 0.05, ** p < 0.01, *** p < 0.001

Table 2: Mixed-effects linear regressions estimating the percentages of average video watching (left; random videos) or video watching of individual users (right; individual logs).

negative sentiment in comments, and higher view count are more likely to be watched longer. We note that the direction of the causal relationship between watch duration and other parameters, as well as the effect of the actual video quality, are not evident from the data. More generalizable engagement mechanisms may be exposed by a controlled experiment with a few selected videos. Theories such as excitation transfer (Alhabash et al. 2015) can inform a stronger theoretical understanding of the relationship between these various video metrics. Yet, our data-driven approach helps to understand user engagement beyond view count and shows a potential to predict engagement level by using more readily available metrics. We intend to explore those of more detailed mechanisms in future work.

An interesting question is how the fact that the user can immediately observe other popularity metrics (e.g. the view count, available from the YouTube page) can *change* the viewing behavior rather than just predict it. If users take view count or like/dislike ratio as aggregate opinion, these observable metrics may change their expectations of the video and hence their watching behaviors. Such relationships would be interesting to study in future research that explores whether and how other such factors, which we call 'social traces' (e.g., previous views, but also demographic information of previous users or other measures of their engagement) contribute to different view behaviors.

Finally, an interesting question is what precise mechanism links the negative sentiment in comments and view duration. For instance, video content that is more emotionally compelling may inspire high arousal negative emotions like anger and anxiety which are known to link to content sharing behaviors (Berger and Milkman 2012).

Acknowledgements

We thank Oded Nov and the anonymous reviewers for their helpful feedback. This work is supported by a Google Social Interactions grant with additional support from AOL and NSF CAREER 1446374.

References

Alhabash, S.; Baek, J.-h.; Cunningham, C.; and Hagerstrom, A. 2015. To Comment or Not to Comment?: How Virality, Arousal Level, and Commenting Behavior on YouTube Videos Affect Civic Behavioral Intentions. *Computers in Human Behavior*.

Arapakis, I.; Lalmas, M.; Cambazoglu, B. B.; Marcos, M.-C.; and Jose, J. M. 2014. User Engagement in Online News: Under the Scope of Sentiment, Interest, Affect, and Gaze. *Journal of the Association for Information Science and Technology*.

Baym, N. K. 2013. Data Not Seen: The Uses and Shortcomings of Social Media Metrics. *First Monday*.

Berger, J., and Milkman, K. L. 2012. What Makes Online Content Viral? *Journal of Marketing Research*.

Chatzopoulou, G.; Sheng, C.; and Faloutsos, M. 2010. A First Step Towards Understanding Popularity in YouTube. In *Proc. of INFOCOM*.

Cisco. 2015. Global IP Traffic Forecast. http://www.cisco.com/c/en/us/solutions/service-provider/visual-networking-index-vni/index.html.

Cramer, H. 2015. Effects of Ad Quality & Content-Relevance on Perceived Content Quality. In *Proc. of CHI*.

De Choudhury, M.; Sundaram, H.; John, A.; and Seligmann, D. D. 2009. What Makes Conversations Interesting?: Themes, Participants and Consequences of Conversations in Online Social Media. In *Proc. of WWW*.

El-Arini, K., and Tang, J. 2014. News Feed FYI: Clickbaiting. http://newsroom.fb.com/news/2014/08/news-feed-fyi-click-baiting.

Hutto, C., and Gilbert, E. 2014. Vader: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. In *Proc. of ICWSM*.

Shamma, D. A.; Yew, J.; Kennedy, L.; and Churchill, E. F. 2011. Viral Actions: Predicting Video View Counts Using Synchronous Sharing Behaviors. In *Proc. of ICWSM*.

Tausczik, Y. R., and Pennebaker, J. W. 2010. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of language and social psychology*.

Yew, J.; Shamma, D. A.; and Churchill, E. F. 2011. Knowing Funny: Genre Perception and Categorization in Social Video Sharing. In *Proc. of CHI*.

Yi, X.; Hong, L.; Zhong, E.; Liu, N. N.; and Rajan, S. 2014. Beyond Clicks: Dwell Time for Personalization. In *Proc. of RecSys*.

YouTube. 2012. YouTube Now: Why We Focus on Watch Time. http://youtubecreator.blogspot.com/2012/08/youtube-now-why-we-focus-on-watch-time.html.

YouTube. 2015. Statistics. https://www.youtube.com/yt/press/en-GB/statistics.html.

Yu, H.; Xie, L.; and Sanner, S. 2015. The Lifecyle of a Youtube Video: Phases, Content and Popularity. In *Proc. of ICWSM*.