Twitter-driven YouTube Views: Beyond Individual Influencers

Honglin Yu, Lexing Xie & Scott Sanner, Australian National University, NICTA

The Problem: Predicting Popularity in Social Media

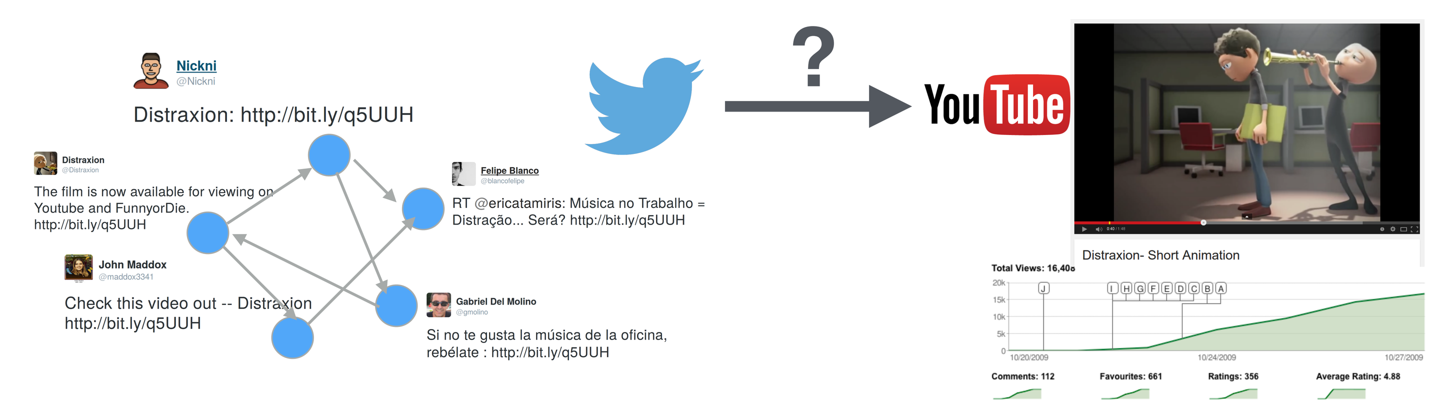


Figure 1: Problem overview: Can one use activities on Twitter to predict popularity on YouTube?

- Predicting popularity is an important open problem in social media.
- Most current methods operate under the assumption that *past popularity implies future popularity* (Figure 2 Left). However, this approach cannot account for sudden changes, or when popularity history is unknown (Figure 2 Right).
- We propose a novel system to *connect across two social networks*, and provide an affirmative answer for the question: **Can Twitter help predict YouTube popularity?**
- This work takes the first steps towards answering questions like "will an obscure video suddenly become very popular, and when", or "which videos will be the most popular in 1, 2, or 3 months"?

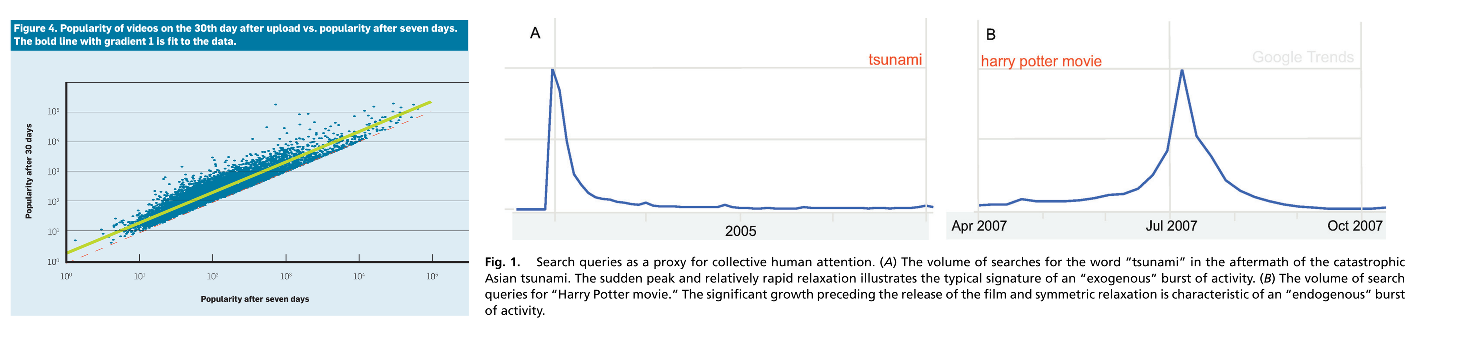


Figure 2: (Left) Direct correlation of past popularity to future popularity, courtesy of Szabo and Huberman [4]. (Right) Two dynamic classes for popularity – endogenous and exogenous, courtesy of Crane and Sornette [1].

Result Highlights

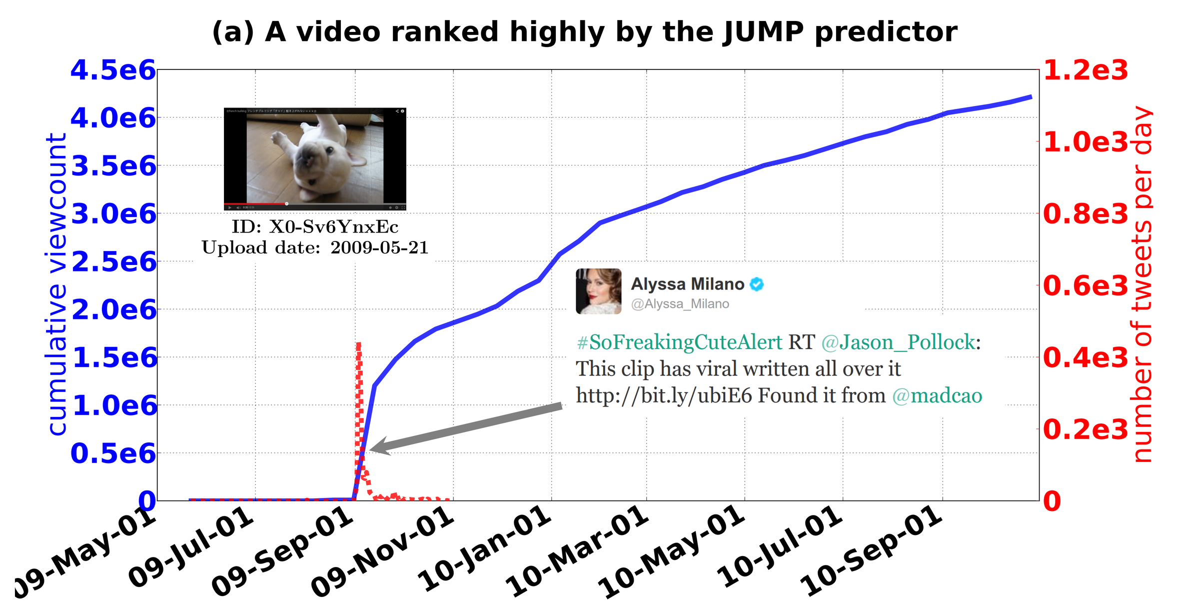


Figure 3: A video having less than 9000 views in its first 3 months and then gaining 1.2 million views within 15 days. The insert shows a Tweet linking to this video by celebrity user Alyssa Milano. X-axis date format : yy-MMM-dd.

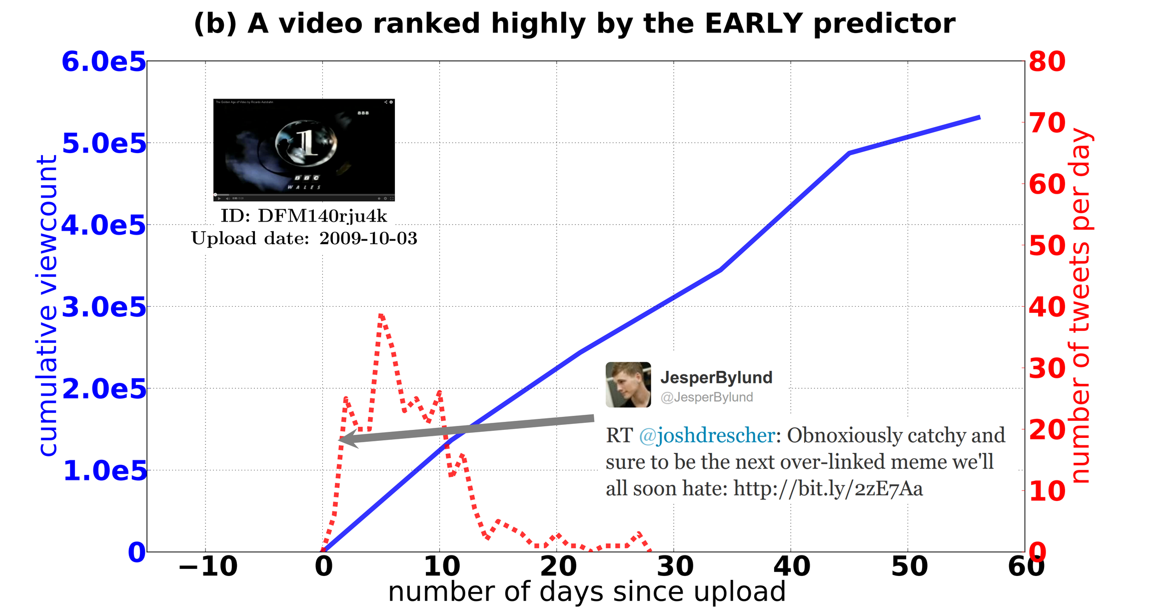


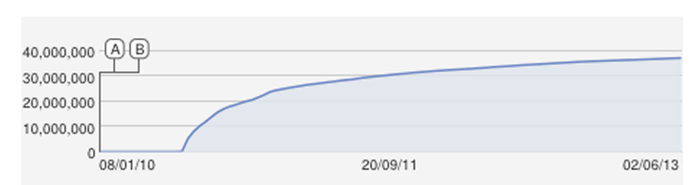
Figure 4: A video with a few dozen Twitter mentions and nearly 200,000 views in its first 15 days. Note that the video popularity continues to rise even after the tweet volume has tapered off, illustrating the prediction power of Tweets that happened early in a video's lifecycle.

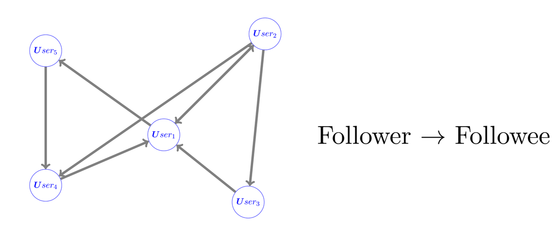
Demo

http://cantabile.cecs.anu.edu.au/yt/demo/

Dataset

- **Tweets** : 467 million, August 1st to October 31st, 2009 [5].
- **Viewcount history** : the history of cumulative viewcounts, in 100 data points from the video upload date to December 2012.
- **Twitter User Graph** : 41.7 million nodes and 1.47 billion edges, collected in 2009 [2].





Viewcount JUMP

Observing Tweets about a video for 15 days, we predict whether or not the video will go through a viewcount JUMP in the next 15 days.

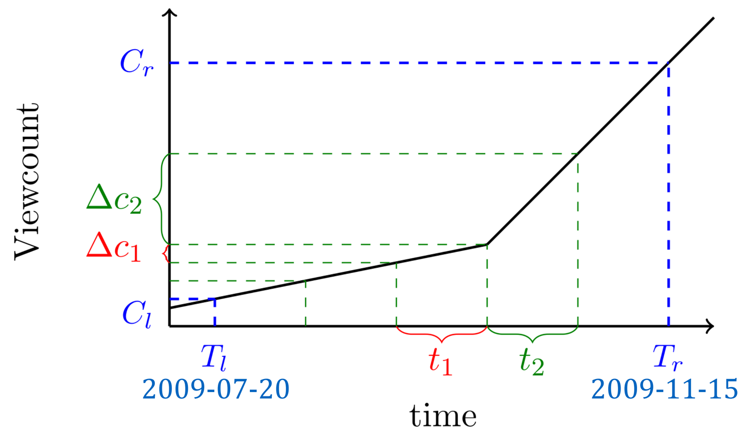


Figure 5: Illustration for defining a JUMP.

- Define viewcount increment ratio $\Delta r_i = \frac{\Delta c_i}{C_r - C_l}$, here Δr_1 represents the increment in the observation period, and Δr_2 that of the prediction period.
- A video is considered to have gone through a JUMP if Δr_1 is no more than the average viewcount increment (0.16 in this work) and Δr_2 is no less than a significant fraction of all accumulated views (0.5 in this work).

EARLY Popularity

Observing Tweets about a video during its first 15 days, we predict the top 5% videos after being online for D days, where $D = 30, 60, 90$.

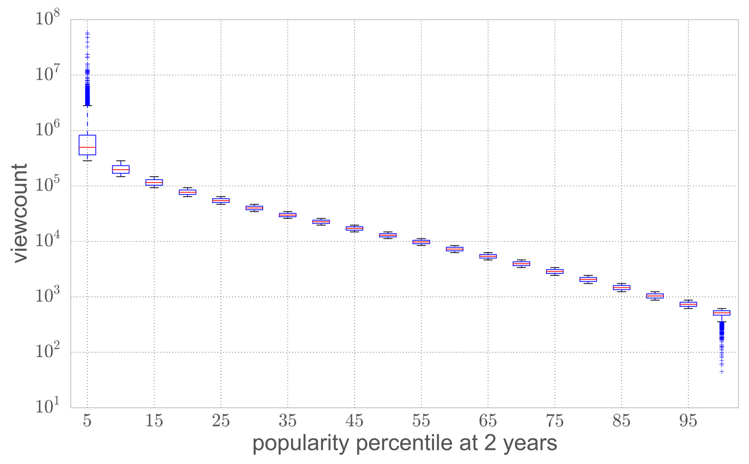
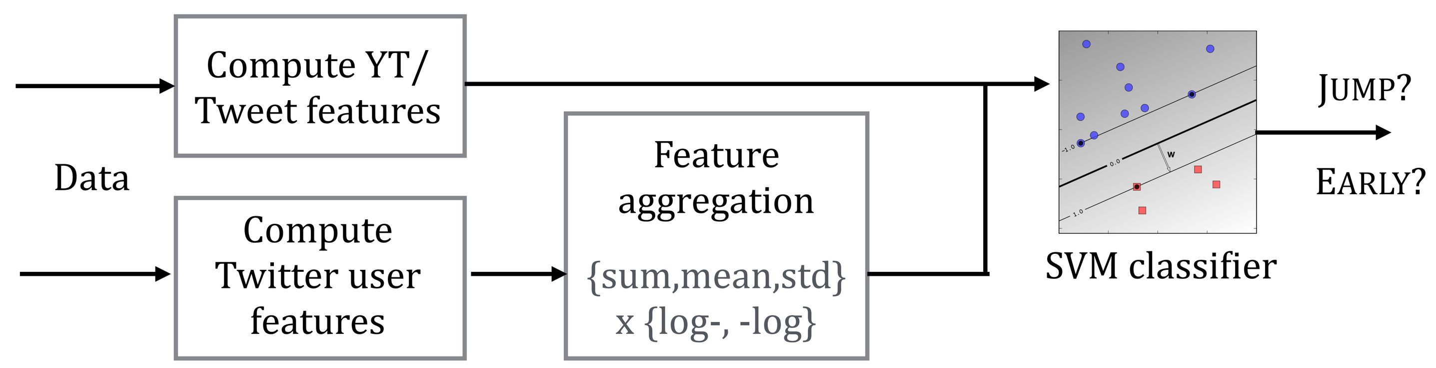


Figure 6: Boxplots of video viewcounts, across the *popularity scale* with bins of 5%, or 8000+ videos each. Viewcounts of the 5% most popular and least popular videos span more than three orders of magnitude, while videos in the same bins (from the 10^{th} to 95^{th} percentile) are within 30% views of each other. This shows that **the top 5% of videos deviates from the general power-law behavior of rank vs popularity, and can be worthy prediction targets.**

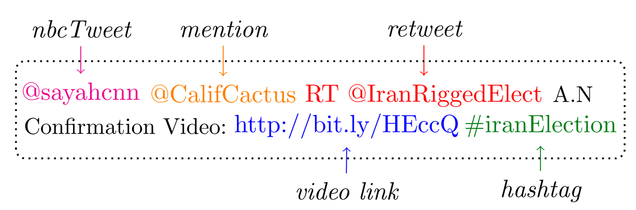
Method Overview



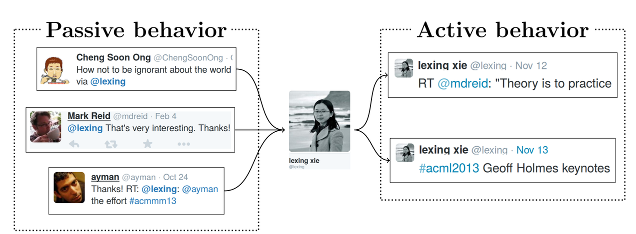
Features

Feature group	Feature name	Meaning
YT-VIEWS	viewcount	Vector of previous viewcounts
Tweet Feature: TWEET	T.tweet T.hashtag T.mention T.nbcTweet T.RT	Five counting metrics that describe the properties of video tweets about video v in the observation interval.
Twitter User Feature: GRAPH	G.outdegree G.pagerank G.hubauthority	Features on the Twitter user graph, describing all users tweeting about video v.
Twitter User Feature: ACTIVE	A.tweet A.hashtag A.mention A.nbcTweet A.RT	Five features that describe the behaviors of users U who tweet about video v.
Twitter User Feature: PASSIVE	P.mention P.nbcTweet P.RT	Three features that describe the interactions users U receive from other users.

Video Tweet Example



User Behavior



Results

Features	Avg Prec	Prec@100
Random	0.012	0.012
YT-VIEWS	0.056 ± 0.006	0.125 ± 0.028
YT-VIEWS+TWEET	0.058 ± 0.002	0.204 ± 0.041
YT-VIEWS+GRAPH	0.097 ± 0.007	0.406 ± 0.023
YT-VIEWS+ACTIVE	0.105 ± 0.003	0.432 ± 0.057
YT-VIEWS+PASSIVE	0.104 ± 0.005	0.444 ± 0.044
ALL	**0.113** ± 0.008	**0.460** ± 0.053

Table 1: Performance for JUMP prediction.

- User features perform significantly better than Tweet properties.
- The best predictor doubles the AP and nearly quadruples the Precision@100 versus the baseline consisting only of viewcount history.

$\hat{\tau}$	Feature	Avg Prec	Prec@100
all	Random	0.053	0.053
15-d	TWEET	0.248 ± 0.142	0.450 ± 0.229
15-d	GRAPH	0.382 ± 0.030	0.646 ± 0.044
15-d	ACTIVE	**0.441** ± 0.027	**0.702** ± 0.058
15-d	PASSIVE	0.375 ± 0.055	0.656 ± 0.088
15-d	ALL	0.463 ± 0.029	0.750 ± 0.045
30-d	ACTIVE	0.421 ± 0.023	0.686 ± 0.060
60-d	ACTIVE	0.435 ± 0.024	**0.722** ± 0.018
90-d	ACTIVE	0.424 ± 0.026	**0.720** ± 0.043

Table 2: Performance for EARLY prediction.

- ACTIVE is the best-performing feature group, and it is inexpensive to obtain.
- $P@100 \geq 0.7$. This accuracy is maintained for predicting future popularity for 30 to 90 days.

Feature Importance by Mutual Information

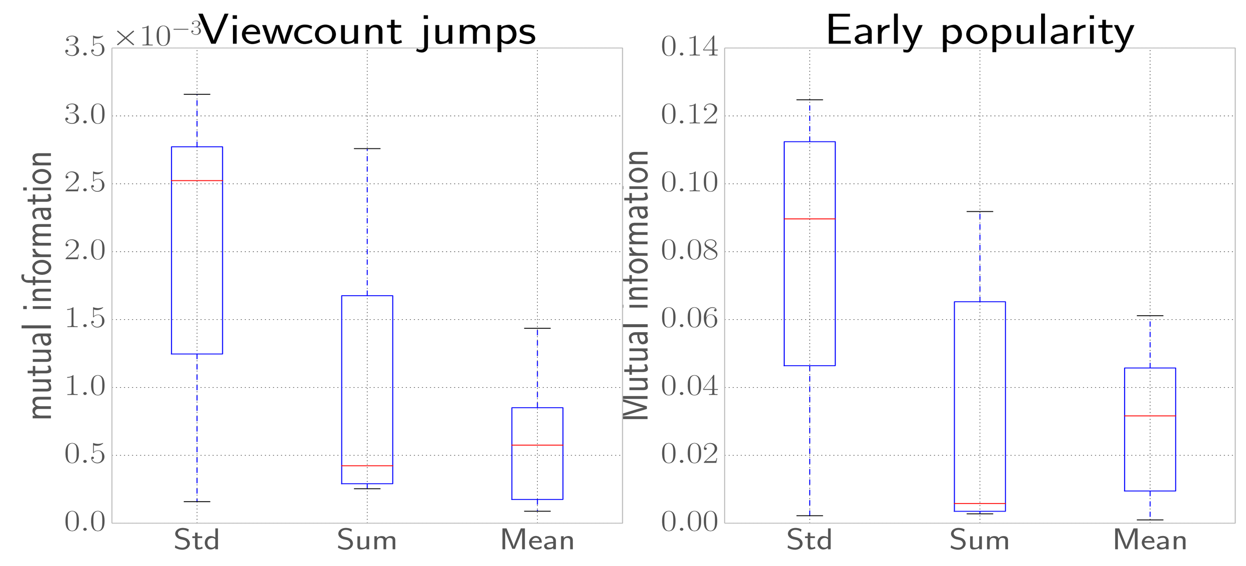


Figure 7: Box plots of mutual information grouped by feature aggregates. The most informative features (best 1/6) are generated by *std* aggregation for both JUMP and EARLY . This implies that having a diverse set of users (as reflected by a large *std*) mentioning an item is helpful for improving its popularity.

Summary

- User and content information from Twitter can be effectively used to predict video popularity on YouTube – popularity are predictable from one external source alone.
- Twitter user features associated with tweeting activities are more informative than graph features. Having a diverse range of users and associated tweeting activities is more informative than the total or average volume of activities of these users.
- Future work include leveraging diffusion patterns to further improve popularity prediction, or selecting a good set of users for prediction.

References

[1] R. Crane and D. Sornette. Robust dynamic classes revealed by measuring the response function of a social system. Proc. Natl. Aca. Sci., vol 105 no 41, pages 15649-53, 2008

[2] H. Kwak, C. Lee, H. Park, and S. Moon. What is twitter, a social network or a news media? In *WWW*, pages 591–600. 2010.

[3] H. Pinto, J. M. Almeida, and M. A. Gonçalves. Using early view patterns to predict the popularity of youtube videos. WSDM '13.

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

[5] J. Yang and J. Leskovec. Patterns of temporal variation in online media. WSDM 2011.