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](image)

- 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?”

Result Highlights

![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].](image)

Dataset

- Tweets = 467 million, August 1st to October 31st, 2009 [5].
- Viewcount history: the history of cumulative viewcounts, in 800 data points from the video upload date or to December 2012.
- Twitter User Graph: 41.7 million nodes and 1.47 billion edges, collected in 2009 [2].

Ealy Popularity

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 7: Box plots of mutual information grouped by feature aggregates. The most informative features (first 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.](image)

Feature Importance by Mutual Information

![Figure 6: Illustration for defining a JUMP.
- Define viewcount increment ratio $\Delta C = \frac{C - C_0}{C_0}$, where $C_0$ represents the increment in the observation period, and $C$ that of the prediction period.
- A video is considered to have gone through a JUMP if $\Delta C$ is no more than the average viewcount increment (0.16 in this work) and $\Delta C_T$ is no less than a significant fraction of all accumulated views (0.5 in this work).](image)

Method Overview

![Figure 5: A video having less than 900 views in its first 3 months, and then gaining 1.2 million views within 15 days. The inset shows a Tweet linking to this video by celebrity user Alyssa Milano.](image)

Features

- Features: Compute YouTube features. Compute Twitter user features. Feature aggregation: $(\sum_{i=1}^{n} \text{mean}(a_1, a_2, \ldots, a_n)) \times \log(a_n)$

Video Tweet Example

![Figure 3: 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.](image)

![Figure 4: A video having less than 900 views in its first 3 months, and then gaining 1.2 million views within 15 days.](image)

Results

<table>
<thead>
<tr>
<th>Features</th>
<th>Pres@1000</th>
<th>Pres@1500</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random</td>
<td>0.015</td>
<td>0.015</td>
</tr>
<tr>
<td>YT-Views</td>
<td>0.054 ± 0.002</td>
<td>0.204 ± 0.041</td>
</tr>
<tr>
<td>YT-Views+Twitter</td>
<td>0.054 ± 0.002</td>
<td>0.204 ± 0.041</td>
</tr>
<tr>
<td>YT-Views+Graph</td>
<td>0.054 ± 0.002</td>
<td>0.204 ± 0.041</td>
</tr>
<tr>
<td>YT-Views+Active</td>
<td>0.054 ± 0.002</td>
<td>0.204 ± 0.041</td>
</tr>
<tr>
<td>YT-Views+Passive</td>
<td>0.054 ± 0.002</td>
<td>0.204 ± 0.041</td>
</tr>
<tr>
<td>ALL</td>
<td>0.115 ± 0.024</td>
<td>0.444 ± 0.041</td>
</tr>
</tbody>
</table>

- User features perform significantly better than Tweet properties.
- The best predictor doubles the AP and nearly quadruples the Prec@1000 versus the baseline consisting only of viewcount history.

Summary

- User and content information from Twitter can be effectively used to predict video popularity on YouTube – popularity is 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


Demo

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