Viral Video Style:
A Closer Look at Viral Videos on YouTube

Lu Jiang, Yajie Miao, Yi Yang, Zhenzhong Lan, Alexander G. Hauptmann

School of Computer Science, Carnegie Mellon University
Outline

- Introduction
- CMU Viral Video Dataset
- Statistical Characteristics
- Peak Day Prediction
- Conclusions
Outline

- Introduction
- CMU Viral Video Dataset
- Statistical Characteristics
- Peak Day Prediction
- Conclusions
What is a viral video?

- A viral video is a video that becomes popular through the process of (most often) Internet sharing through social media.

Gangnam Style
What is a viral video?

- A viral video is a video that becomes popular through the process of (most often) Internet sharing through social media.

Charlie bit my finger  
Gangnam Style  
Missing pilot MH370
Social Validity

• Viral videos have been having a **profound impact** on many aspects of society.
• Politics:
  – Pro-Obama video “Yes we can” went viral (10 million views) in 2008 US presidential election [Broxton 2013].
Social Validity

• Viral videos have been having a profound impact on many aspects of society.

• Politics:
  – Pro-Obama video “Yes we can” went viral (10 million views) in 2008 US presidential election [Broxton 2013].
  – We found that Obama Style and Mitt Romney Style went viral (30 million views in the month of Election Day), and peaked on Election Day.
Social Validity

Number of Views vs Days after the upload

- Mitt Romney Style
- Obama Style

2012 Election Day
Social Validity cont.

- Viral videos have been having a profound impact on many aspects of society.

- **Financial marketing:**
  - Old Spice’s campaign went viral and improved the brand’s popularity among young customers[West et al 2011].
Social Validity cont.

- Viral videos have been having a profound impact on many aspects of society.
- Financial marketing:
  - Old Spice’s campaign went viral and improved the brand’s popularity among young customers [West et al 2011].
  - Psy’s commercial deals has amounted to 4.6 million dollars from Gangnam Style.
Motivation

• Existing studies are conducted on:
  – Small set (tens of videos) \( \rightarrow \) biased observations?
  – Large-scale Google set \( \rightarrow \) confidential!
• A relatively large and public dataset on viral videos would be conducive.
• Solution: CMU Viral Video Dataset.

Beware the content are hilarious!
Outline

- Introduction
- CMU Viral Video Dataset
- Statistical Characteristics
- Peak Day Prediction
- Conclusions
CMU Viral Video Dataset

• By far the **largest** public viral video dataset. Time’s list was the largest dataset (50 videos).

• Videos are **manually selected by experts**
  – Editors from Time Magazine, YouTube and the viral video review episodes.

• Statistics: 446 viral videos, 294 quality 19,260 background videos.

10+ million subscribers!
CMU Viral Video Dataset Cont.

- For a video, it includes:
  - Thumbnail
  - Video and user metadata
CMU Viral Video Dataset Cont.

• For a video, it includes:
  – Thumbnail
  – Video and user metadata
  – Insight data: historical views, likes, dislikes, etc.
CMU Viral Video Dataset Cont.

• For a video, it includes:
  – Thumbnail
  – Video and user metadata
  – Insight data: historical views, likes, dislikes, etc.
  – Social data: #Inlinks, daily tweeter mentions (pending)

<table>
<thead>
<tr>
<th>Date</th>
<th>#Tweets with Video ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>12/18/2012</td>
<td>0</td>
</tr>
<tr>
<td>12/18/2012</td>
<td>79</td>
</tr>
<tr>
<td>12/19/2012</td>
<td>14</td>
</tr>
<tr>
<td>12/20/2012</td>
<td>3</td>
</tr>
<tr>
<td>12/21/2012</td>
<td>7</td>
</tr>
</tbody>
</table>

[Number of In-links] 624000
CMU Viral Video Dataset Cont.

- Near duplicate videos (automatic detection + manual inspection).
Outline

- Introduction
- CMU Viral Video Dataset
- Statistical Characteristics
- Peak Day Prediction
- Conclusions
Statistical Characteristics

- Observations agree with existing studies including the study on Google’s dataset
  - Short title, Short duration.
- Less biased.

### Table 3: Basic statistics about viral videos.

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Viral</th>
<th>Quality</th>
<th>Background</th>
</tr>
</thead>
<tbody>
<tr>
<td>View Count Median</td>
<td>3,079,011</td>
<td>55,455,364</td>
<td>7,528</td>
</tr>
<tr>
<td>Title length</td>
<td>5.0±0.1</td>
<td>5.4±0.1</td>
<td>7.0±0.1</td>
</tr>
<tr>
<td>Duration(s)</td>
<td>138.6±16.0</td>
<td>248±3.9</td>
<td>252±24.6</td>
</tr>
<tr>
<td>Average Rate</td>
<td>4.69±0.03</td>
<td>4.73±0.03</td>
<td>4.04±0.08</td>
</tr>
<tr>
<td>Rater/View</td>
<td>0.54±0.03%</td>
<td>0.38±0.01%</td>
<td>0.87±0.07%</td>
</tr>
<tr>
<td>Cor(inlinks, view)</td>
<td>0.54</td>
<td>0.25</td>
<td>0.28</td>
</tr>
<tr>
<td>Days-to-peak Median</td>
<td>24</td>
<td>63</td>
<td>30</td>
</tr>
<tr>
<td>Lifespan Median</td>
<td>7</td>
<td>166</td>
<td>10</td>
</tr>
</tbody>
</table>
Observation I

- The days-to-peak and the lifespan of viral videos decrease over time.

**Table 4: Evolution of viral videos.**

<table>
<thead>
<tr>
<th>Statistics</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>Days-to-peak Median</td>
<td>24</td>
<td>14</td>
<td>9</td>
</tr>
<tr>
<td>Lifespan Median</td>
<td>8</td>
<td>6</td>
<td>3</td>
</tr>
</tbody>
</table>
Observation II

- The popularity of the uploader is a more substantial factor that affects the popularity than the upload time.
- Upload time is believed to be the most important factor in background videos [Borghol et al. 2012], coined as First-mover advantage.

An example: official music videos
Outline

- Introduction
- CMU Viral Video Dataset
- Statistical Characteristics
- Peak Day Prediction
- Conclusions
Peak Day Prediction

- Forecast when a video can get its peak view based on its historical view pattern.
Peak Day Prediction

• Forecast when a video can get its peak view based on its historical daily view pattern.
Peak Day Prediction

- Forecast the date a video can get its peak view based on its daily view pattern.
Peak Day Prediction

• Forecast when a video can get its peak view based on its historical daily view pattern.
Peak Day Prediction

• Forecast the date a video can get its peak view based on its daily view pattern.

• This application is significant in supporting and driving the design of various services:
  – Advertising agencies: determine timing and estimate cost
  – YouTube: recommendation
  – Companies/Politicians: respond viral campaigns
HMM Model

• Model daily views using HMM model:
• Two types of states:
  – hibernating $\rightarrow$ less views
  – active $\rightarrow$ more views
HMM Model

• Model daily views using HMM model:
• Two types of states:
  – hibernating $\rightarrow$ less views
  – active $\rightarrow$ more views
HMM Model

- Model daily views using HMM model:
- Two types of states:
  - hibernating → less views
  - active → more views
HMM Model

• Model daily views using HMM model:
• Two types of states:
  – hibernating $\rightarrow$ less views
  – active $\rightarrow$ more views
HMM Model

• Model daily views using HMM model:
• Two types of states:
  – hibernating → less views
  – active → more views
HMM Model

• Model daily views using HMM model:

• Two types of states:
  – hibernating $\rightarrow$ less views
  – active $\rightarrow$ more views

• Novel modifications:
  • Incorporate metadata in the prediction. Other work only use the pure view count [Pinto et al. 2013].
  • Smooth transition probability by a Gaussian prior.
Experimental Results

Considering metadata in peak day prediction is instrumental.
• Early warning system for viral videos.
• Detect viral videos and forecast their peak dates.
Outline

- Introduction
- CMU Viral Video Dataset
- Statistical Characteristics
- Peak Day Prediction
- Conclusions
Conclusions

• A few messages to take away from this talk:
  – CMU Viral Video Dataset is by far the largest open dataset for viral videos study.
  – This paper discovers several interesting characteristics about viral videos.
  – This paper proposes a novel method to forecast the peak day for viral videos. The preliminary results look promising.
References


THANK YOU.

Q&A?