Identifying and Characterizing Sleeping Beauties on YouTube

Sandipan Sikdar [a]
Indian Institute of Technology
Kharagpur, India

Abhijnan Chakraborty [a]
Indian Institute of Technology
Kharagpur, India

Anshit Chaudhary
Indian Institute of Technology
Kharagpur, India

Gaurav Kumar
Indian Institute of Technology
Kharagpur, India

Shraman Kumar
Indian Institute of Technology
Kharagpur, India

Abhijeet Patil
Indian Institute of Technology
Kharagpur, India

Niloy Ganguly
Indian Institute of Technology
Kharagpur, India

Animesh Mukherjee
Indian Institute of Technology
Kharagpur, India

[a] These two authors contributed equally

Abstract
The generally accepted notion about popularity dynamics of user generated contents (e.g., tweets, videos) is that such contents attain their peak popularity within first few days and then gradually fade into oblivion. However, analyzing more than $350K$ videos on YouTube, we find that more than 10% of them obtain their peak popularity after at least one year from being uploaded. We term such videos as *Sleeping Beauties* and observe that these videos engage users more compared to other videos on YouTube. We further observe that sleeping beauties can retain their popularity to a greater extent following their peak popularity compared to other videos. We believe that identifying such videos will not only benefit the advertisers, but also the designers of recommendation systems who seek to maximize user satisfaction. Through this interactive poster, we bring the presence and characteristics of sleeping beauties in front of the research community and simultaneously identify few factors which can trigger the awakening of such videos.

Introduction
With the exponential increase in both number of videos and the amount of traffic to YouTube, different videos face wide variations in popularity. While some videos gain immense popularity, others fail to generate any interest. As the popularity of individual videos directly impact the content producers as well as marketing agencies, the dynamics and

Figure 1: Daily viewing patterns of different videos (solid green lines). Screenshots are grabbed from YouTube pages for specific videos.

prediction of video popularity have captured much attention of research community in the recent years [1, 4]. The general observation in most of the prior works is that a particular video gains popularity within a few days from being uploaded and the typical lifetime is only a few months.

However, analyzing $350K+$ videos with average age of more than 5 years, we find that more than 10% videos attain the peak in popularity after at least one year from being uploaded. Taking a cue from bibliometric literature [3], we term such videos as *'Sleeping Beauties'*. Figure 1(A) and Figure 1(B) show the popularity variations of two sleeping beauties. It is indeed intriguing to find that after receiving negligible views for several years, such videos suddenly start drawing user attention and become very popular. On the other hand, Figure 1(C) shows the more typical popularity pattern, where the peak in popularity occurs within a week and then there is a quick drop in the number of views.

Presence of sleeping beauties on YouTube not only questions the existing models of popularity dynamics, but also calls for identifying the basic factors which contribute to the popularity of YouTube videos. In this work, we take the first step towards that direction by identifying $38K+$ sleeping beauties on YouTube and comparing them with videos which attain peak popularity at the early stage in their lifetimes. We find that sleeping beauties drive more active user engagements in terms of liking and commenting and they remain popular for longer durations after the peak. In terms of categorical compositions of the videos, we notice substantial difference in the composition of sleeping beauties compared to other videos on YouTube. Finally, we manually identify the possible reasons for awakening of few sleeping beauties. We believe that this work has implications not only for the advertisers on YouTube videos, but also for designers of different recommendation frameworks which try to maximize user satisfactions.

Dataset and Methodology

In this section, we first describe the dataset and then outline how we identify sleeping beauties from the daily view (i.e., the number of views accrued on a day) statistics.

Dataset

In the individual webpages for YouTube videos, YouTube shows the daily distribution of views starting from the upload date. We first gather $500K+$ video ids reported in past

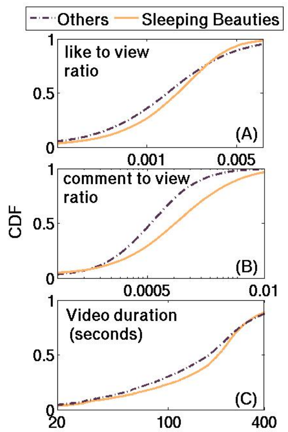

Figure 2: CDF of (A) like-to-view ratio, (B) comment-to-view ratio and (C) video duration for both sleeping beauties and other videos. All x-axes are in logarithmic scale.

literatures [1, 4] and then scrape the daily view counts for each video. Removing the videos, where the statistics or the videos themselves are no longer available, we finally get more than $350K$ videos with average age more than 5 years. The videos included in the dataset are a mixture of videos featured in "Recently Featured", "Most Viewed", "Top Rated" and "Most Discussed" sections during the years 2007-2008, along with the videos most shared on Twitter during June – July, 2009. In addition to the daily view counts, we also collect different metadata (e.g., the title, description, video duration, comments, #likes etc.) for these $350K+$ videos using YouTube Data API[1].

Identifying sleeping beauties on YouTube

We define sleeping beauties as the videos which attain peak in popularity at least after 1 year from the upload date. To obtain the peaks in popularity, we extend the peak detection algorithm presented in [2] and identify sleeping beauties based on the position of the peaks.

From the daily distribution of views on a video, we consider a day to have a local peak if the video accrues more views on that day than in the preceding and the succeeding day. In this way, we obtain a set of local peaks for each video. If μ and σ are the mean and standard-deviation of these set of peak values, we consider a peak to be significant if the corresponding number of views is greater than $\mu + \sigma$. Out of all such significant peaks, if for a video the first significant peak occurs after one year, we classify it as sleeping beauty. Using the above mentioned algorithm, we identify around $38K$ sleeping beauties in our dataset, which is more than 10% of all videos.

[1] https://developers.google.com/youtube/v3/

Analysis

In this section, we compare sleeping beauties with other videos in our dataset. Our analysis centers on the following questions:

Q1. Do sleeping beauties engage users more?

Similar to other social media platforms, YouTube allows its users to like, dislike and post comments on individual videos. Typically liking or commenting indicate more active engagements of the users compared to only viewing passively. In order to compare user engagements between different types of videos, we normalize the number of likes or comments on a video by the number of views on that video. Higher like to view ratio as well as comment to view ratio would indicate more participation on the part of the users. As we can see in Figure 2(A) and Figure 2(B), sleeping beauties attract more likes and comments compared to other videos in our dataset. As commenting is the most active form of user participation, we further looked at the average length of the comments (i.e. the number of words in the comments) posted on different videos. We find that the average comment length is 21 words in case of sleeping beauties, whereas for others it is 17 words. Hence, we can see that even though sleeping beauties take a while to attract users but once such videos manage to do so, they engage users more than other videos.

Q2. Can sleeping beauties retain popularity after peak?

Every video tend to lose their popularity after attaining the peak popularity. Figure 4 shows the average retention of views on the days following the peak. We observe that sleeping beauties are able to hold their popularity for a longer period following the peak while popularity drops more rapidly for others. Interestingly, we also find a periodic pattern with comparatively higher views every seventh day. We plan to investigate this phenomenon in more detail in future.

Q3. Which type of videos become sleeping beauties?
Every video on YouTube is assigned a category by the uploader. We were curious to know about the categorical composition of different types of videos. Figure 3 shows the distribution of categories in both sleeping beauties and other videos. We can see that around 40% of sleeping beauties are Music videos, with other significant contribution by Entertainment and Sports videos. However, if we look at the overall compositions, there are significant differences between sleeping beauties and other videos.

Q4. Do longer videos have more chance of becoming sleeping beauties?
To answer the question, we plot the distribution of video durations for different videos in Figure 2(C). We observe that for sleeping beauties, the average duration is marginally higher compared to other videos.

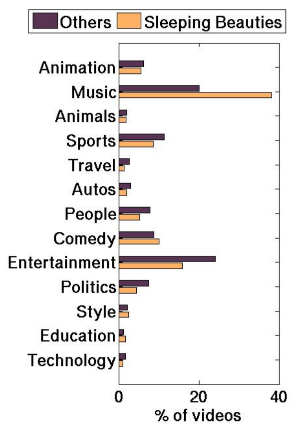

Figure 3: Category wise distribution of sleeping beauties and other videos.

Discussion

In this work, analyzing an extensive dataset on YouTube videos, we identified a significant fraction of videos which become popular after more than a year from upload date. We termed these videos as sleeping beauties. On analyzing such videos, we observed that sleeping beauties drive more user engagements in forms of getting more likes and comments than others. We further observed that the sleeping beauties can retain their popularity for a longer period following its peak while others fail to do so. All these observations lead us to believe that identifying such videos can be of immense interest to both the advertising agencies for better targeting their advertisements as well as to the designers of video recommendation systems as these videos drive more user engagement.

It would certainly be more helpful if one can identify such videos even before the peak in popularity occurs. Towards that direction, we manually looked into some of the sleeping beauties to investigate the reasons for their sudden popularity. We observe that in some cases popularity can be triggered by an external event like the death of a celebrity (e.g., the videos related to Michael Jackson attracted many visitors after his sudden death); while in some other cases, videos became popular after some celebrities picked these videos and tweeted about them drawing immediate attention of their followers. In future, we plan to develop a generic framework to automatically identify such reasons and subsequently we will try to identify sleeping beauties ahead of their peak popularity.

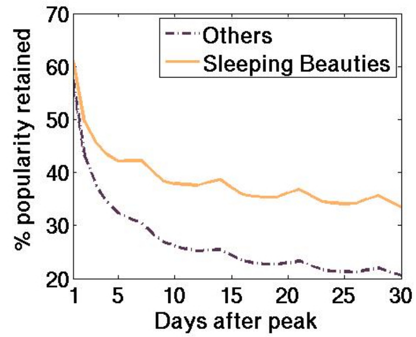

Figure 4: Percentage of popularity retained following the peak.

References

[1] X. Cheng, C. Dale, and J. Liu. Statistics and social network of youtube videos. IEEE Transactions, 2008.
[2] R. Schneider. Survey of peaks/valleys identification in time series, 2011.
[3] A. F. J. van Raan. Sleeping beauties in science. 2004.
[4] H. Yu, L. Xie, and S. Sanner. The lifecyle of a youtube video: Phases, content and popularity. In *ICWSM*, 2015.