CS 6474 - Social Computing Project Proposal

Characteristics of Viral Videos on YouTube

1. Main Idea

Youtube is a video sharing platform, which is used by a wide range of content creators. There is increasing interest amongst these video creators to find ways to optimize the reach of their video-based on attributes such as the usage of tags. In this project, we aim to identify the common characteristics (eg: tags and category) exhibited by the daily trending videos of Youtube, factoring in the country of origin.

2. Significance

The study of virality, in general, is of great importance since it allows us to glimpse at the ways in which the online and offline world affect one another. For example, it is widely believed that the 2008 presidential elections in the USA were significantly influenced by the virality of the "Yes we can" video of Obama on Youtube. The content that goes viral on any social computing platform allows us to have a better understanding of prevailing issues concerning people offline. Additionally, what goes viral on social computing systems could shape interactions within society and affect policymaking as well.

There are several reasons why certain posts/videos go viral on social computing systems. In this study, we aim to see what type of content (and the characteristics of those videos) usually goes viral on Youtube. This would help us understand for what purposes Youtube is used for (Example: If the trending videos are mainly regarding politics, then it could be logically inferred that Youtube is mainly used by its users for consuming political content). It could also help us have a better understanding of the reasons why some videos go viral, while others do not.

There have been comparable studies on virality that have been performed on other platforms such as Reddit and Twitter. A study on Youtube may provide different results from these other studies since users in many instances compartmentalize their use of social computing. By that, we mean it is possible for an individual to primarily use Facebook to connect with friends, use Twitter to consume political news and

Youtube for viewing Music videos. Therefore the characteristics of virality of videos in Youtube may differ from these other social computing sites.

3. Related Work

User-Generated Content (UGC) has been the driving force behind popular online social media platforms, especially video content systems, like YouTube, that almost entirely rely on it. In "I Tube, You Tube, Everybody Tubes: Analyzing the World's Largest User Generated Content Video System," [1] Cha et al. (2007) study the impact of YouTube and the different properties associated with the content of its videos. Through a data-driven analysis, they find that a peer-to-peer (P2P) communication system is imperative for massive content distribution and describe how the YouTube structural framework could benefit from it. Since the publishing of this work, YouTube has made improvements regarding P2P communication which caused content to be shared more rapidly. This study was extensive and focused on several aspects of YouTube and other UGC video systems such as user behavior, interface design, and content aliasing. Because of the massive scale of YouTube, their dataset was limited to just include two categories: "Entertainment" and "Science & Technology." This could have led to important information not being found in their study, which could serve as a major limitation due to the diversity of genres that actually exist on Youtube. Our study considers all categories of content. One main idea from this work that relates to our study is the popularity distribution and evolution of YouTube videos. Our study mostly focuses on the characteristics of trending YouTube videos and considers other external factors that could affect the popularity of videos such as significant current events.

De Choudhury et al. (2009) are more specific in their work, "What Makes Conversations Interesting? Themes, Participants and Consequences of Conversations in Online Social Media,"[2] which studies what causes a conversation to be interesting as it pertained to user participation in discussions under YouTube videos. They believe that user participation in discussions depends on the conversation theme and the familiarity between participants. In this work, a computational model is developed to detect the theme,

determine the “interestingness”, and establish the consequential impact of “interestingness” of the conversations being held. They establish that “interestingness” of conversations is essential to the success of online social media platforms because it promotes diversity among conversation participants and themes. However, similarly to the previous work discussed, the dataset was limited to the “Politics” category of YouTube videos, so it didn’t capture the conversations being discussed from other categories and the conversation themes would have hence been limited. To fill this gap and to provide another aspect of this work, our study intends to identify similarities and differences in themes of viral videos among different countries while taking into account the external factors mentioned in this work.

In ‘What makes a video go viral? An analysis of emotional contagion and Internet memes’ [3], the authors, Rosanna E.Guadagno et al. explain the various factors that cause a video to go viral. They especially touch on the various emotional and psychological factors that cause people to forward certain types of videos over others. The authors’ analysis methodology is through a survey. The authors use their self-assigned categories, such as ‘cute’, ‘humorous’ along with other factors such as emotion associated with the videos to measure the likelihood of forwarding the video for participants. They also go into theorizing the impact of factors such as social validation, the source of the video to determine why participants chose to forward certain video content over others.

One of the major limitations that the authors’ acknowledge is that the study uses a culturally homogeneous sample of participants and generalizes the results from the survey to other population. Also, the age range of the participants was also an issue: the participants in this study were undergraduates. The study is limited because of the methodology of using surveys. In contrast, we plan to use data from multiple countries and the date we’ve collected will not be constricted by age or culture.

In ‘Viral Actions: Predicting Video View Counts Using Synchronous Sharing Behaviors’ [5], David A. Shamma, et al. aim to predict the number of likes that a Youtube video might gain. They use the data from the sharing data of usage from the video-sharing environment: Zync (a plugin for yahoo messenger

app), a platform that allows users to view and interact with a video simultaneously during a chat session. The authors' argue that implicit sharing in a more accurate contributory factor to the possibility of likes that a video might get. The basic limitation of this paper is that it is basing the chat platform as the major source of video sharing that takes place. Contrary to the assumption we see that youtube videos are mostly shared on twitter and facebook. Our data sources allow us to bypass such assumption by directly accessing the view count, rather than aggregating from multiple sources of sharing.

In "Virality over YouTube: an empirical analysis"[10], Khan and Vong look at the determinants of a Youtube video's virality, by analysing the "social" (eg: number of subscribers) and "non-social capital" (eg: duration of the video) of Youtube videos and their creators. Amongst their several observations is the fact that music videos are disproportionately likely to go viral on Youtube. However, their analysis was largely restricted to English language videos and the dataset only contains 100 videos. In our study, we will be analysing viral videos from over 10 major countries and will be using a dataset that contains data for over 190000 videos.

A number of blogs exist on the internet that shares various strategies to make viral videos on YouTube. The most common strategies are - (1)Viral videos generally have an intriguing and catchy title[6]. (2)Keywords and tags go a long way in determining the category of the video and help in reaching the target audience. (3)Viral videos are generally candid, fun, entertaining and almost always convey some message[7]. This message may be direct or embedded in the video subtly. (4)Viral videos are a team effort[8]. Many individuals are involved in the background who contribute to the success of the video. (5)The platform in which the video is published makes a difference. (6)Once a video gets traction, it is likely to spread like wildfire and viewership grows exponentially - commonly known as the snowball effect. (7)Researchers have concluded that viral videos are produced by only a few YouTube channels[9]. We aim to evaluate each of the mentioned characteristics by comparing them with the results of our analysis.

4. Goals and Contributions

4.1 Goals

The goals of our project are -

- Identify what categories/genres of YouTube videos are trending in each country irrespective of related events that are occurring or have occurred within those corresponding countries.
- Determine if a correlation exists between trending YouTube videos and related events that are occurring or have occurred within those corresponding countries.
- Identify similarities or differences in the characteristics of trending YouTube videos among several countries.
- Formulate inferences regarding any other factors that could affect the virality of trending YouTube videos

4.2 Contributions

The current state of research is limited to surveys or analysis of a small number of videos. Also, the analysis done is restricted to smaller demographies. In our analysis we will be analysing common characteristics observed in more than 190 thousand viral YouTube videos from 10 different countries over a period of 200 days.

5. Outcomes

The outcomes of our study will include -

- The prevailing categories of trending videos on a per-country basis, which will give us insight into how Youtube is used in that demography.
- Temporal analysis of trending content, i.e. the factors that lead to some content staying viral for days when compared to the ones that stay viral for a very brief duration.
- The characteristics of viral videos: the ratio of likes vs dislikes for trending content and analysis for whether the poster of the video has an effect on the virality of the videos.

- Data visualizations that either support or reject our claims.

6. Outline and Evaluation Plan

6.1 Data Collection

For the purpose of this project, we will focus on Trending Youtube video statistics.YouTube maintains a list of the top trending videos on the platform which is determined by factoring in the number of likes, dislikes, comments, views and many other parameters. The dataset comprises of trending YouTube videos from 10 different countries with up to 200 listed trending videos per day. The videos in the dataset are categorised in 31 categories and each video has metrics like the number of views, likes, comments and dislikes. There are approximately 190 thousand unique trending videos on YouTube across the 10 countries in the dataset. The data we will be working on is retrieved from Kaggle [4].

6.2 Evaluation

The evaluation phase will comprise of comparing the results of our analysis with various blog posts suggesting methods to create viral videos with the results obtained from our study and better understand the temporal nature of viral videos.

A number of marketing companies have shared different methods to make a video viral. Few of the methods are having the same tags and keywords as the title of the video. If a YouTube channel has published a viral video once, it is more likely to produce a viral video again in the future. Most successful videos either entertain or convey a message or both to the audience. We intend to verify the mentioned claims through the analysis of the YouTube videos dataset.

We would also like to dig deeper to understand the temporal nature of viral videos. We would like to understand how long does a video take to become viral and also observe how long does it take a video to become viral in different countries.

7. Timeline

Date	Activity
24th Sep 2019	Project Proposal
7th Oct 2019	Literature Review
21st Oct 2019	Design Finalization, Data Collection & Data Pre-processing
30th Oct 2019	Midterm Milestone Presentation & Report
4th Nov 2019	Data Analysis & Interpreting Results
18th Nov 2019	Evaluation of Results & Visualisation
1st Dec 2019	Final Project Presentation & Development
9th Dec 2019	Final Project Report

8. List of Team Members

Name	Email	GT ID
Arvind Akpuram Srinivasan	arvind_s@gatech.edu	903528961
Karthik Nama Anil	kanil3@gatech.edu	903471605
Miasia J Jones	miasia.jones@gatech.edu	903164042
Prithvi Alva Suresh	al.prithvi@gatech.edu	903541747

9. Distribution of Work

Name	Responsibilities
Arvind Akpuram Srinivasan	Literature Survey; Data Collection; Data Analysis Based on Tags; Country Wise Data Analysis; Report Documentation
Karthik Nama Anil	Literature Survey; Data Cleaning; Data Analysis Based on Categories; Country Wise Data Analysis; Data Visualisation; Report Documentation
Miasia J Jones	Literature Survey; Feature Engineering & Selection; Data Analysis Based on Tags; Country Wise Data Analysis; Report Documentation
Prithvi Alva Suresh	Literature Survey; Data Analysis Based on Tags; Evaluation of Results; Country Wise Data Analysis; Data Visualisation; Report Documentation

References

[1] Cha, M., Kwak, H., Rodriguez, P., Ahn, Y.-Y., & Moon, S. (2007). I tube, you tube, everybody tubes. *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement - IMC 07*. doi: 10.1145/1298306.1298309

[2] De Choudhury, M., Sundaram, H., John, A., & Duncan Seligmann, D. (2009). What makes conversations interesting? *Proceedings of the 18th International Conference on World Wide Web - WWW 09*. doi: 10.1145/1526709.1526754

[3] Guadagno RE, Rempala DM, Murphy S, Okdie BM: What makes a video go viral? An analysis of emotional contagion and internet memes. *Comput Hum Behav* 2013, 29(6):2312–2319. 10.1016/j.chb.2013.04.016

[4] Trending YouTube Video Statistics. (2019). Retrieved 26 September 2019, https://www.kaggle.com/datasnaek/youtube-new

[5] D. A. Shamma, J. Yew, L. Kennedy, and E. F. Churchill. Viral actions: Predicting video view counts using synchronous sharing behaviors. In ICWSM, 2011.

[6] Tyler Wes (2011). Going Viral: Factors That Lead Videos to Become Internet Phenomena.The Elon *Proceedings of Journal of Undergraduate Research in Communications Vol. 2, No. 1*

[7] Lewis Howes (2012). How To Go Viral On YouTube: The Untold Truth Behind Getting Views. https://www.forbes.com/sites/lewishowes/2012/08/09/how-to-go-viral-on-youtube-the-untold-truth-behind-getting-views

[8] Mackenie Graham (2015). 9 tips to make a video go viral on YouTube. https://thenextweb.com/insider/2015/11/20/9-things-to-help-your-youtube-video-go-viral/

[9] Adam Toren (2015). 5 Ways to Boost Your Chances of Going Viral on YouTube. https://www.entrepreneur.com/article/247661

[10] Feroz Khan, G. and Vong, S. (2014), "Virality over YouTube: an empirical analysis", Internet Research, Vol. 24 No. 5, pp. 629-647.