Social Networking and YouTube

Chapter-I

**Introduction:**

YouTube is the most popular video sharing website on the internet today. It’s the biggest user driven video content provider in the world (Wattenhofer, M. 2012). For this research I chose YouTube as it is one of the biggest and most popular Social Networks. “500 Years of YouTube video are watched on Facebook every day, and over 700 YouTube videos are shared on Twitter each minute” (YouTube Statistics\*). The main aim of this project is to determine what role other social networks like Facebook and Twitter have in the popularity of the videos on YouTube which become viral. Facebook allow users to share the video links on their Facebook home page called wall or post a link on other user’s wall. Those videos can be viewed within the Facebook page where the video plays in a player embedded within the page and can be shared using the “Share” option under the video and the video becomes available on the user`s home page wall where his friends can view and share it.

We also study from where these videos are getting the maximum views, and the effect on its popularity. YouTube provides the statistical information where we can view the video link has been visited from which includes referral from other videos, post on other social networks, blogs and sites, view from subscriber module and view from mobile devices. A crawler was specifically designed to gather the required information from YouTube videos which had more than a hundred thousand views. The tasks after getting the data were:

1. Get the information about the videos that were embedded and viewed from Facebook. What relationship can we establish with Facebook and those videos?
2. Where are those videos getting their maximum views from? Are they more popular means of sharing the videos?
3. What other things can we determine from that data?

This will help us analyse what is the significance of Facebook in the popularity of the viral videos on YouTube and which “Category” do they belong to. This project is based on studying the aspects of YouTube which have not been examined in other YouTube researches but involves a small dataset compared to them.

**About other Social Networks and Facebook**

Social Networking has always been an important topic for research and has gained much popularity in the last decade. Today a large amount of internet population is based on file sharing and social networking. The purpose of online social networks is to allow users to access each other`s information as long as they have a friend relationship with each other. Almost every person who accesses the internet uses social network to connect with others. These networks are as old as the internet itself, for example the graph formed by a simple exchange of email message or a chat message among users forms an online social network (Mislove et al 2007).

The most popular social networks today are MySpace, Facebook, Twitter, LinkedIn, Google+, and Orkut etc. YouTube is a video sharing website where users can upload and view videos. Unlike other conventional social networks YouTube is primarily based on sharing videos, but it also provides features like messaging, subscribing, adding friends, sharing and adding comments. We will discuss about the details later in this paper.

Many social networking websites like Facebook allow users to hide information from unknown users who are not in their friends list. Users can assign friends to specific groups and allow each group some attributes giving privilege to users of groups to view content and share information accordingly, for example one group user might be able to access complete information while others may get partial or less information to see. Users also have the option to display complete available information to other users who are not in the friend list, but such action is avoided as it is critical to safety and security of the user. On the other hand YouTube also has a user profile page available known as Channel of the user where other users can visit and access the information. YouTube videos are publically available to users and guest users of YouTube. Some videos might not be viewable if the users are not logged in so as to validate age authentication. So basically social networks are composed of user accounts where users are connected through links with each other. Information will be available depending upon the users profile preferences. Users can search for other users and form a link by sending a friend request. To establish a friendship link between users, a mutual consent is required from both the target user and the sender (Mislove et al 2007).

**About YouTube and other Video Sharing sites:**

YouTube is a platform which does not only provide a video sharing service but a service which facilitates all the qualities of a social network. There are two ways where users can connect with each other, either a user subscribes to other users channel or add the other user as a fiend in their own friend list (based on mutual agreement). There are some major differences between the two connections:

Subscriber will get updates from users when they upload a new video on their channel. It does not matter if they are friends with them or not, whereas the friends of the users can also be subscribers and will get the update on the new videos uploaded by the user if they are subscribers. Friends can accept, share and invite other friends but users who are only subscribers do not have this option. Friends can send, receive messages and receive comments on the channel while Subscribes do not have this option.

There are many other websites which provide video sharing; some of the most popular ones are MetaCafe, DailyMotion and Flickr. Compared to its competitors YouTube is more popular as it allows users to upload unlimited number of videos as long as the content is not copyrighted. The video length provided for all users is up to a maximum of 15 minutes. Users can upload longer videos but the need to go through a verification process where the user account should be in good standing which is determined by YouTube community guidelines (having no copyright or community guideline violations), verifying the account with a mobile phone SMS verification and the account has no worldwide content ID block on any of the videos.

Chapter – II

**Related Work**

Chapter – III

**Details about the YouTube Crawler**

Chapter – IV

**Statistics and Analysis**

**Statistics of the Crawled Data**

The YouTube crawler ran for six days crawling over 62,799 videos out of which 31,302 videos were added in the database. Out of the total videos crawled, 12,280 videos did not had any statistics available. 3,850 videos had Facebook view count available and 18,892 videos had been viewed from a mobile device. Twitter view count was only available for only 56 videos out of the total crawled videos which constitute 0.18% of the total data. The little exceptions with the crawled data is that those videos which did not had any statistics available might be having views from Facebook, Twitter and mobile devices, but these videos were excluded from the comparison for the analysis of the dataset. YouTube allow users to hide the statistics of their uploaded videos from the public view. The other exception with the crawled data is that the videos which did not had any Facebook view-count available might not necessarily be lacking views from Facebook but those numbers must be very insignificant that they were not available in the statistics section of the video and can only be seen by the user of the channel. The statistics of every video are updated once in a week on YouTube servers. This information comes in detailed statistics known as YouTube ‘insight data’. With the help of YouTube API users can only retrieve the insight data of only those videos which are owned by them, thus restricting them to view the insight data from other user’s channel using the API of course. But the insight data of other user`s videos are available to view on the video page, if and only the user of the particular video has allowed them to be publicly available. They can be viewed by clicking a small button on the video page, which can be seen under the video player next to the total view-count as shown in the figure 5(a).

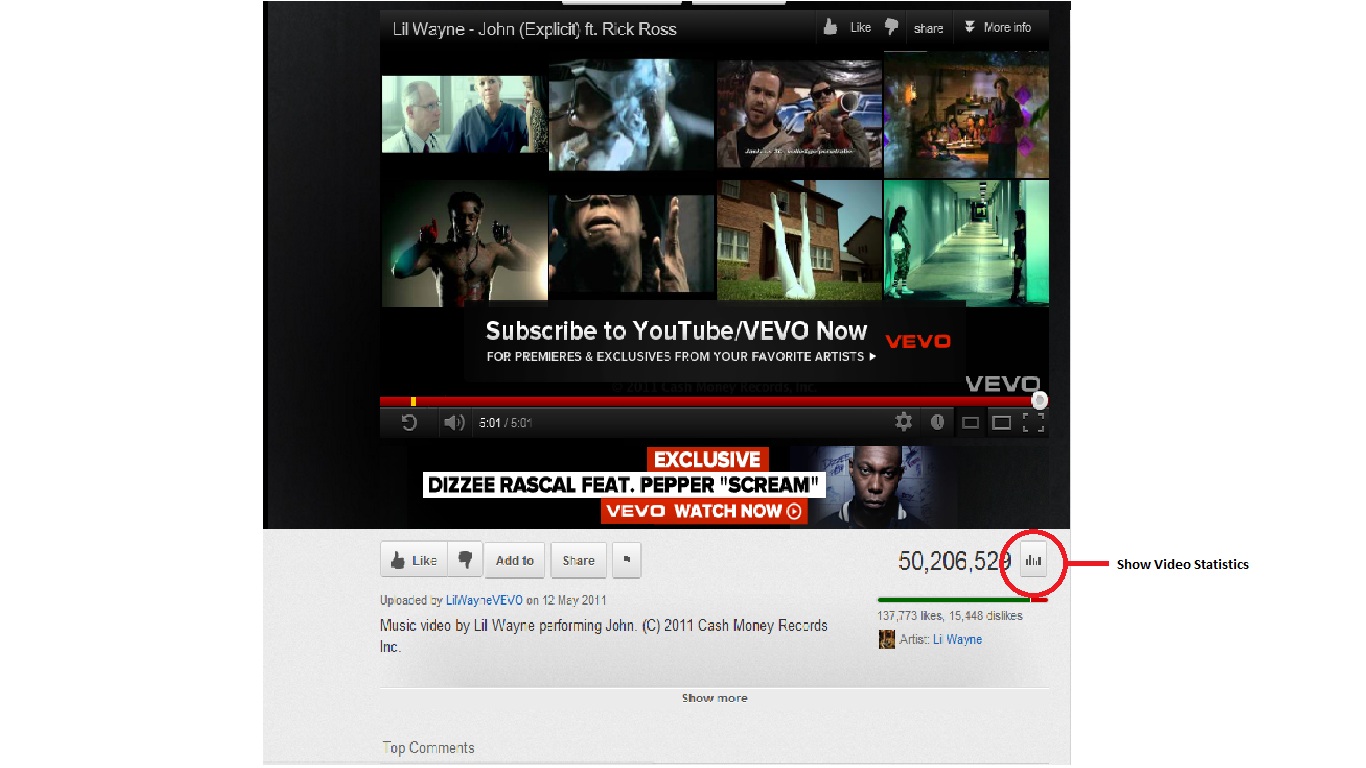


Fig. 5(a)

More detailed information can be seen after clicking the button.

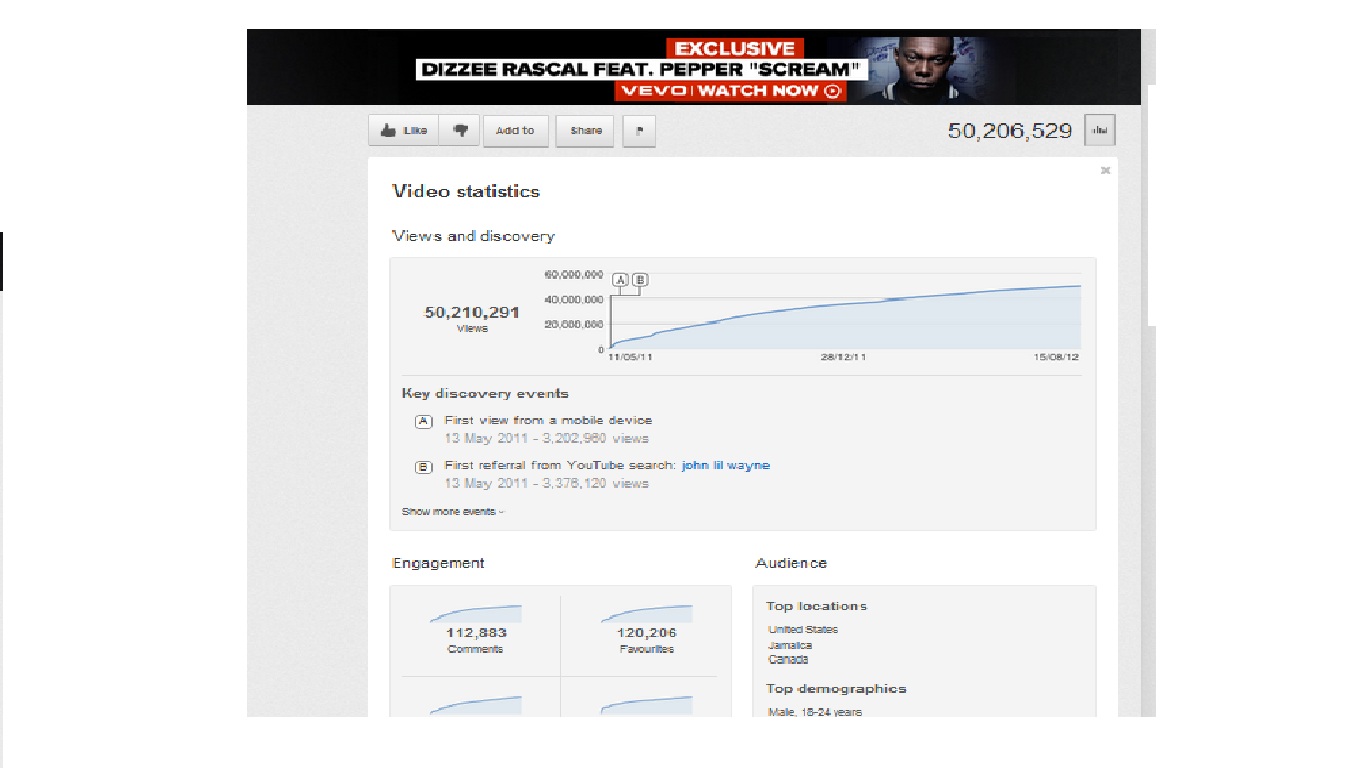


Fig. 5(b)

This information is not presented in the HTML when viewed in the source code viewer of the browser as YouTube restricts users to scrap this data. So it is impossible to retrieve the insight data of other users from the API and scrapping it from the web page.

But as we know that the insight report can still be seen, there is another way around to get this data and scrap it. After researching a little more, I used the Google Chrome Browser`s Tools->Developer Tools to see what action does take place when we click the show statistics button. I was able to retrieve the insight link which was getting the insight information: “[**http://www.youtube.com/insight\_ajax?action\_get\_statistics\_and\_data=1&v=(VIDEO\_ID)**](http://www.youtube.com/insight_ajax?action_get_statistics_and_data=1&v=(VIDEO_ID))**”** from the resources section of the Developer tool tab. This link was hidden inside the following information in the resources tab (shown in highlighted text):

("yt.www.watch.actions.stats",function(a){qn(a)&&(ln(),X**("/insight\_ajax",{**format:"XML",method:"GET",o**:{action\_get\_statistics\_and\_data**:1,v:L("VIDEO\_ID")},j:function(a,c){nn(c.html\_content);vard=G("stats-opt-out-chbox");d&&S(d,"change",function(){Ym(!d.checked)})},A:on}))});t("yt.www.watch.actions.unlike",function(){kn();var a=1==en();gn(a?2:1)});

The above code can was taken from the Google Resources tab shown below.

**

Fig. 5 (c)

When using the insight link inside the browser, it automatically downloads an XML file which contains the HTML tags with data, opened and closed by the XML tags as shown in the following code example.

**<root>**

**<html\_content>**

**<![CDATA[**

<div class="watch-actions-stats"> <div class="stats-header"> <h1> Video statistics </h1> </div> <p>

…………………………………………………………Some HTML Tags and Data……………………………...............

<span>First embedded on:</span> <span class="extra"> <a rel="nofollow" href="http://facebook.com" dir="ltr">**facebook.com**</a> </span> </p> <p class="sub-data">**12 Jul 2012 - 10,009 views** </p></dd> </dl>

…………………………………………………………Some HTML Tags and Data……………………………...............

**]]>**

**</html\_content>**

**<return\_code>**

**<![CDATA[ 0 ]]>**

**</return\_code>**

**</root>**

Fig. 5(d)

Within this content all the insight information was available which could be seen on the video page. Removing the XML tags from the above content and saving it as a HTML file displayed the following information which is shown on the video page.

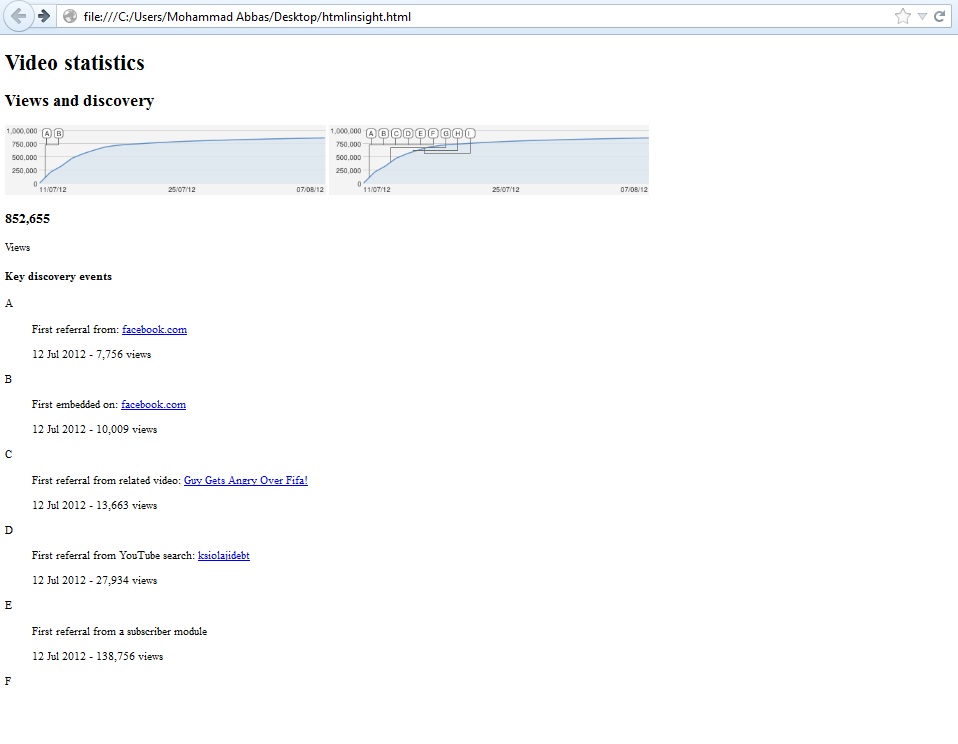


Fig. 5(e)

After getting the data from the insight link it was possible to parse the XML page in Java and scrap the data within the HTML content. These details are already discussed in the “Crawler” chapter.

**Analysis of the Crawled Data**

Crawling all videos and retrieving complete information from YouTube would certainly take several months as YouTube is a swarm of millions of videos. Due to the small nature of this project we take a small dataset of videos to examine the patterns of viral and most popular videos which had total views over a hundred thousand and the effect of Facebook and Twitter on their popularity.

As the database had 31,300 videos in the database, the distribution of videos belonging to each category is shown in the graph 5(f).

Fig. 5(f)

*Formula Used:*

*Percentage of Videos by category = (Total Videos in the Category/Total Number of Videos in Database) X 100*

The maximum numbers of videos crawled are from “Music” and “Entertainment” genres while the least number of videos in the database are from “Trailers”, “Shows” and “Unspecified Category”.

1. *Video Categories and Ratings*

YouTube allows its users to tag their videos under 27 different genres in general and 19 different genres within the movie-genre. The reach of my crawler ranged up to 17 different genres and one unspecified genre where it was unable to get the information about the genre of the video. There are only 40 videos which were stored in the unspecified category. The Average Rating for a video ranges from 1 to 5 and is calculated by this formula:

*Average Rating = ((Number of Likes \* 5) + Number of Dislikes)/Total Number of Ratings*

*Total Number of Ratings (Raters) = Number of Likes + Number of Dislikes*

*Average rating for each category = (Total Ratings of videos in the category/Total number of videos) X 100*

Fig. 5(g)

The graph above describes the Average rating of the videos in different categories of the entire dataset. “Trailers” and “Music” categories were the ones that were most rated with an average rating of approximately 4.8 whereas News had the least average rating of 3.9. The third most highly rated category is “Shows” with an average rating of approximately 4.7. Categories in between “Gaming” and “Travel” had an average rating ranging between 4 and 4.5. According to the above scale “Trailers”, ”Movies” and “Shows” seems to be the most popular category, but to support this assumption we also need to take a look at the number of raters for each category as a video can have 10 likes and 1 dislike to have more average rating while the other which has more raters will become insignificant although having more views.

The total number of likes and dislikes for all the categories were 230,938,984. The graph 5(h) describes the percentage of ratings for each category (total number of raters for each category).

*Formula Used:*

*Total Raters = Number of Likes + Number of Dislikes*

*Raters Percentage= (Total Raters in the category/ Total Raters of all categories) X 100*

Fig. 5(h)

It is quite clear that “Music” has the maximum number of raters while the other categories are nowhere near to it. The numbers of raters for “Trailers” are very less compared to other categories because of the less number of videos crawled in this category, so it will not be possible to measure its popularity in this case. Keeping in mind that only those videos were crawled which had more than a hundred thousand views, “Music” seems to be most popular when we measure it using ratings and view counts.

1. Facebook Views and Categories

In this section we will take a look at the number of views for each video category directed from Facebook. Out of 31,300 videos in the database, 3,850 videos had Facebook view-count available and the total view-count from those videos is approximately 496.9 million. The graph 5(i) represents the statistics of the Facebook view-count compared to the categories.

*Formula Used:*

*Facebook View-Count percentage by category = (Facebook Views from the category/Facebook Views from all the categories) X 100*

Fig. 5(i)

It is very intriguing to know that with only around three thousand videos; approximately 497 million views are directed from Facebook for the available categories and approximately 325 million views are from music videos, which is quite large and keeping the fact in mind that these are only a small set of videos which had over a hundred thousand views. In this aspect we infer that “Music” has more Facebook views compared to other categories like “Comedy”, “Entertainment” and “Gaming”. To prove this we need to take a look at more detailed information about the number of videos we have in database. Within this total segment of videos in the database with Facebook view-count, 163 are from “Car” category, 657 from “Comedy”, 526 from “Entertainment”, 151 from “Film”, 134 from “Gaming”, 967 from “Music”, 346 from “People”, 152 from “Pets”, 138 from “Science” and 292 from “Science”. The remaining categories had less than a 100 videos in the database with Facebook view-count. More detailed information can be seen in Fig 5(j).

*Formula Used:*

*Videos with Facebook view-count in percentage (Category based) = Videos with Facebook view-count in the category/Total Videos with Facebook view-count) X 100*

Fig. 5(j)

If we compare the values in graphs 5(i) and 5(j) we can get the average Facebook views of a video in each category. For this we use the formula:

***Average Facebook Views of a video in a Category = Total Facebook Views in the Category / Total number of videos with Facebook view-count in the Category***

From the above formula we calculated the values and obtained graph 5(k).

Fig. 5(k)

We can see that an average Facebook view for a Music video has near about 340 thousand views whereas rest of the categories do not have more than 100 thousand views. Therefore we deduce that “Music” category is not only the most popular, but they are shared more on Facebook and are likely to be viewed more when shared on it. “Travel” is the next most shared and viewed video category on Facebook with an average view for each video is more than a 100 thousand. The remaining categories like “Entertainment”, “Pets” and “News” etc. have average views in between 50 and 100 thousand.

I search for popular music videos myself and found out that **Vevo** which has different multiple YouTube channels with name of different artist and is currently the most popular uploader. Vevo does have its own website where users can view videos of popular artists. Vevo also has a Facebook Page where they share the links of the latest videos they upload on YouTube and it has millions of subscribers on Facebook as well. More detailed information about it can be viewed on Wikipedia.

1. Facebook View-Count Versus Total View-Count

In this part we will examine the contribution of Facebook in the total view-count of the videos. We will discover to what extent Facebook helps the videos to become viral or popular. For this we take the total Facebook view-count and the total view-count of the videos which have Facebook view-count available in our database. Then we take out the percentage of Facebook view-count by the formula:

***Facebook view-count percentage of the category = (Total Facebook Views of the Category/Total Views of the category) X 100***

Graph 5(j) represents the Facebook View percentage of the total view-count of individual categories.