Social Networking and YouTube

Chapter-I

**About Social Networks**

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 users who are not in the friend list, but such action is avoided as it is critical to safety and security of the user. One 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).

**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). 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

**About YouTube**

Chapter – III

**Related Work**

Chapter – IV

**YouTube Crawler**

Chapter – V

**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).

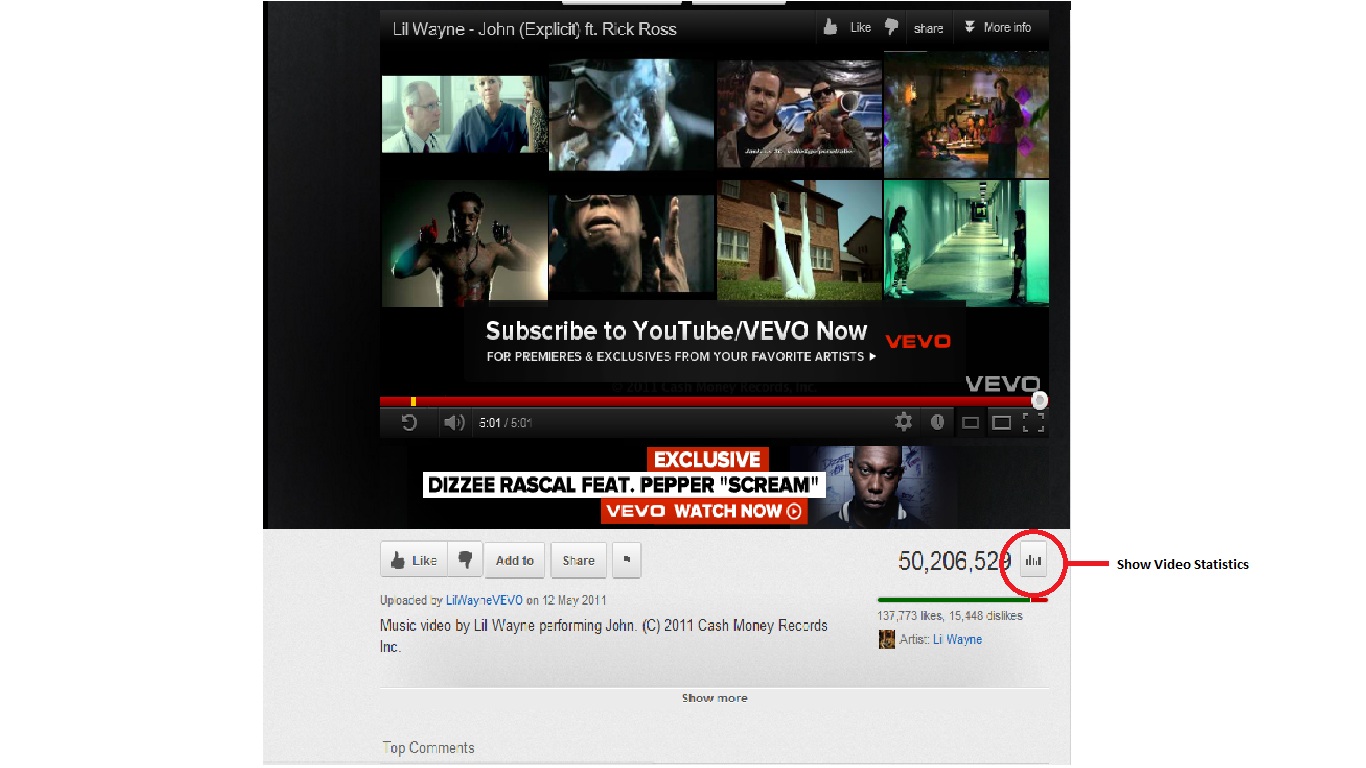


Fig. 5(a)

More detailed information can be seen after clicking the button.

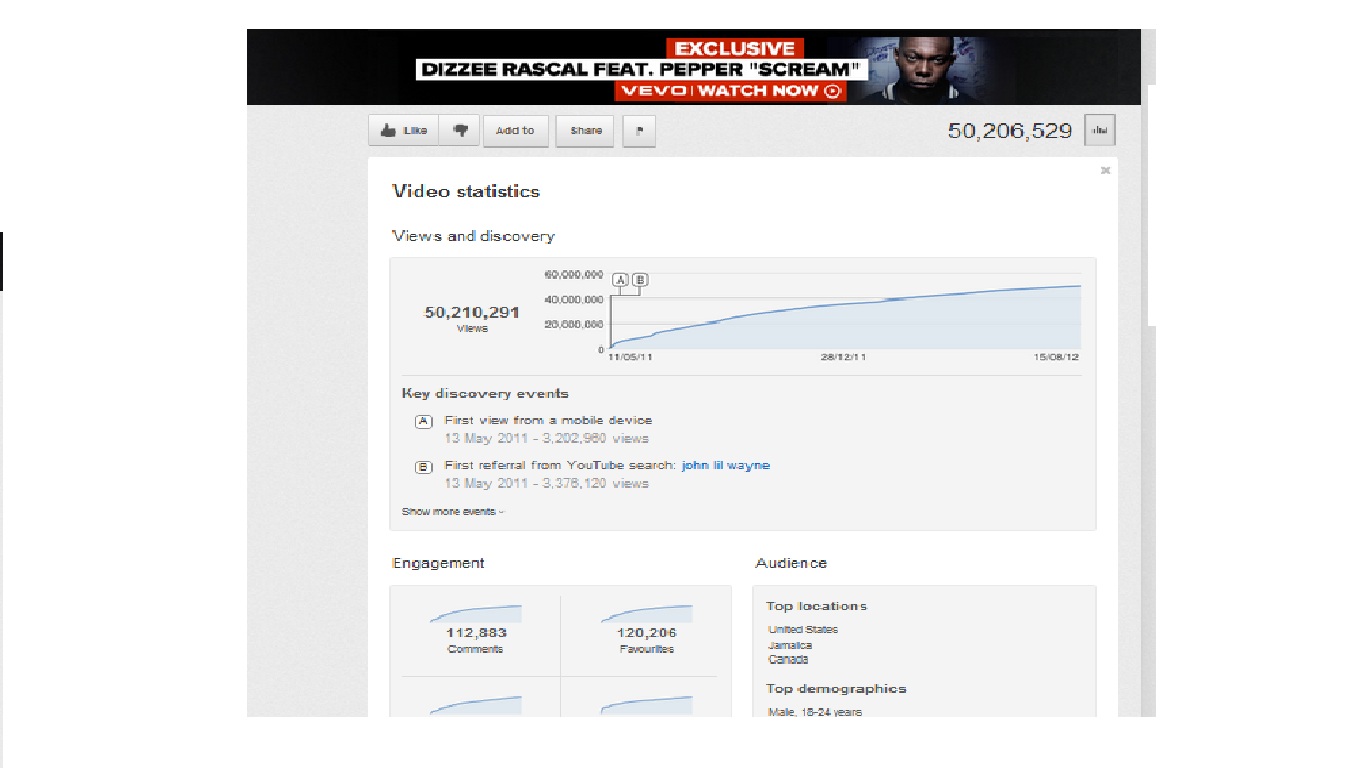


Fig. 5(b)

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. Researching a little more, I used the Google Chrome`s Tools->Developer Tools (Shortcut: Ctrl+Shift+I) 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:

("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);var d=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)});

**

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 next figure.

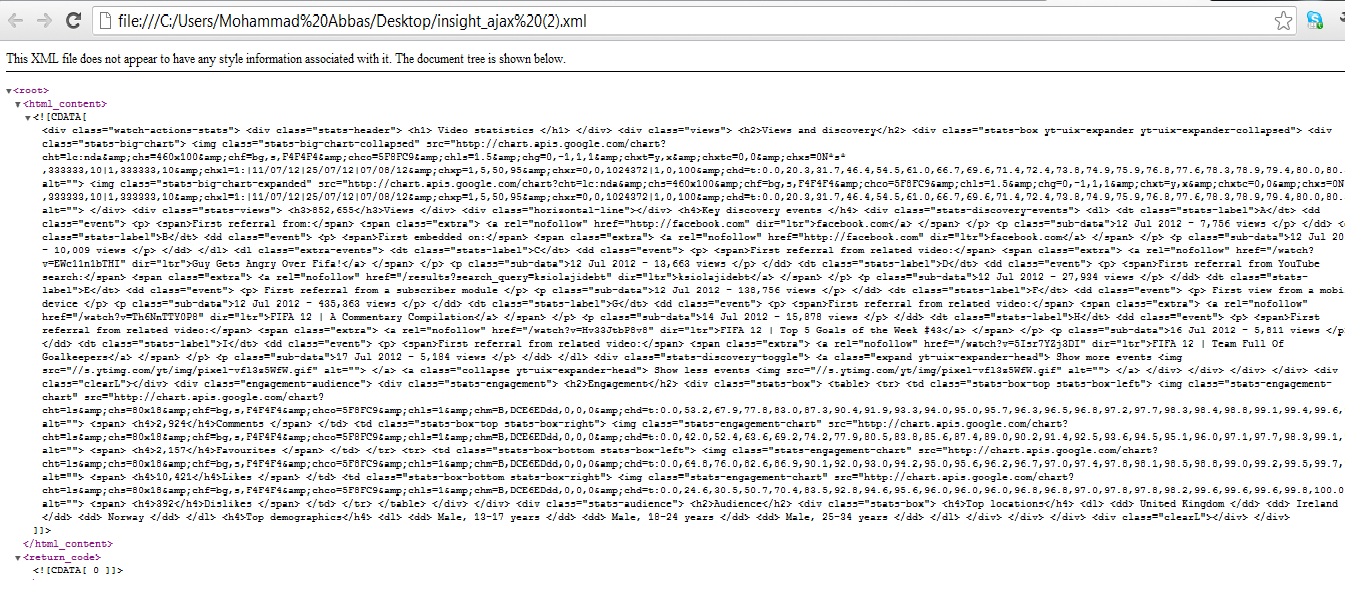


Fig. 5(d)

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.

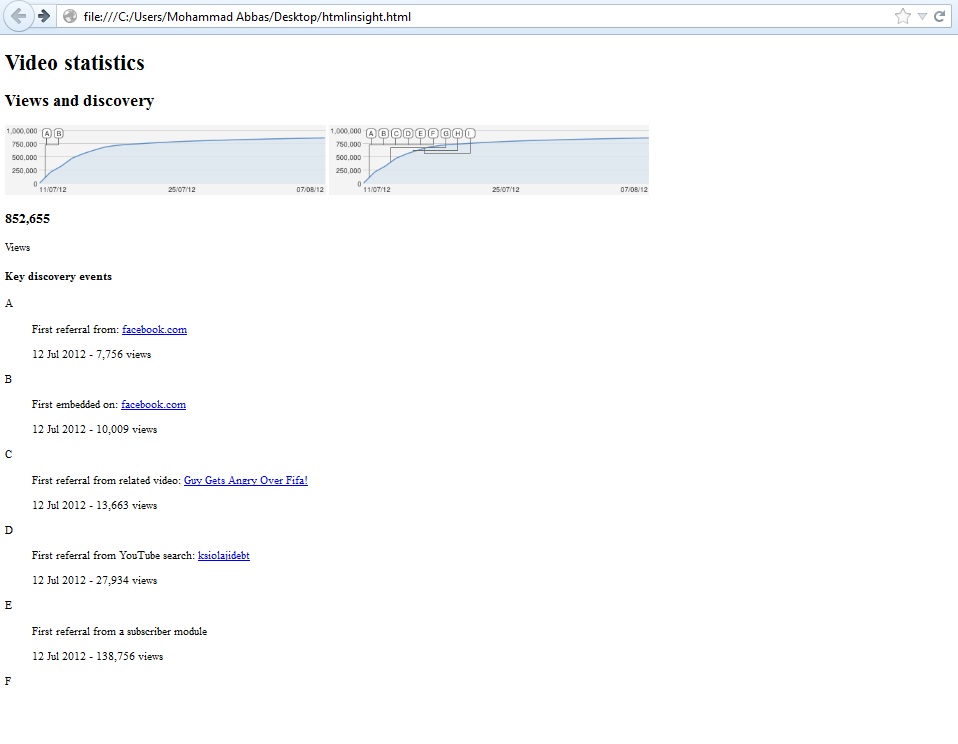


Fig. 5(e)

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 previous 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 what role do other social networks like Facebook and Twitter have 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)

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 = Number of Likes + Number of Dislikes*

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 below describes the percentage of ratings for each category which means total number of raters for each category.