VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

Title

Analysis of the YouTube Channel Recommendation Network

Group members

Aniket Singh – 18BCI0039
Aditya Raj – 18BCE0395
Robins Raj – 18BCE0393

Report Submitted for the Final Review of
Course Code: CSE3021- Social and Information Network

SLOT- C2+TC2

Professor: DR. GAWAS MAHADEV ANANT

Introduction

With more than 1.3 billion users, you tube is one of the largest online used video sharing websites with over 23 million YouTube channel and 30 million visitors every day. In order to upload a video on YouTube, People can create a channel. These channels serve as the home page for that account showing channel's name, description and uploaded videos on YouTube. In addition to the content available on specific channel, a channel can recommend another channel which have same or similar genre of video uploaded. This could be done in two ways: either the user can choose a feature a channel by subscribing to the channel or YouTube can recommend a channel whose genre is similar to the present channel. The 50 videos that YouTube recommended have been viewed on an average of 456 million times each.Through this paper we are interested in analysing structure of the YouTube recommendation network.

As per the data available on website (https://socialblade.com/youtube/), they have crawled different YouTube channels and created a database of 200k+ different channels, 400k+ user recommendation and 400k+ YouTube recommendation. In this paper, we have tried to present systematic analysis of the structure of YouTuberecommendation; we have created a detailed visualization,performed different centrality measures on the network and performed the motif analysis.

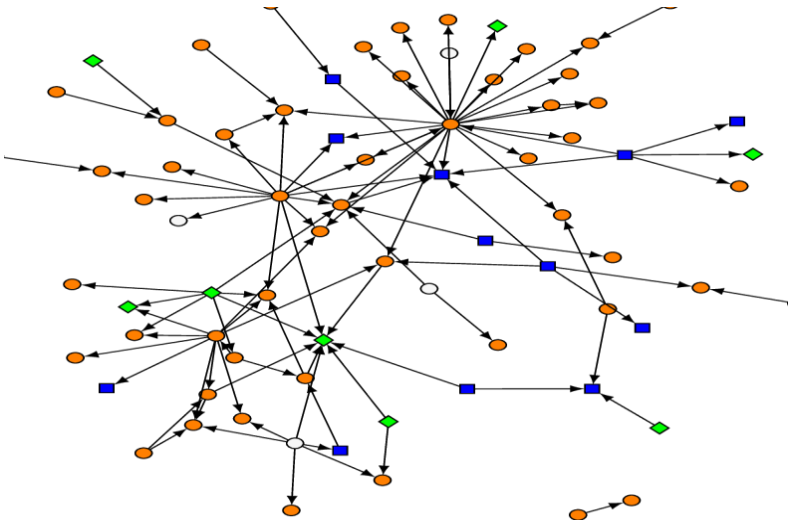
We have two primary goals for this project.
The first one is to discover the structure behind these three networks. We will find different properties of each network like centrality measures, degree and motifs.
The second goal is to compare the three models based on these properties. Using this, we hope to shed light on the network structure that YouTube recommendation used.

Literature Review

YouTube was created in 2005, since then its popularity is increasing massively. There has been many research and theory discovering the strong structure behind YouTube network. There are many research papers to analysis YouTube as a social network. Some of the research paper was published by Cheng and Paolillo. As mentioned in Cheng research paper he looked videos as nodes and recommendation to other videos as edges. While Paolillo looks channel as nodes and friends as edges. In addition to Cheng and Paolillo's papers there were other papers which examine the YouTuberecommendation network.

- Statistics of social network- YouTube (Cheng – 2008)

  Cheng present a study of various statistics of YouTube videos. According to his paper data was collected from 3 million YouTube videos and analyzing all the data he found the evidence of small world characteristics. Data collected was summarised in a directed graph where each node YouTube video was a node and an edge between two node (say a and b) exists if the latter YouTube video is in first 15 recommended videos of the former. Likewise the directed graph is completed and clustering coefficient was calculated. It was found that clustering coefficient of YouTube was much higher than any random graph. Thus the network shows the traits of small world network as it contains cliques and nodes linked to all others.



- Network and Structure of YouTube core(Paolillo -2008)

  The paper by Paolillo illustrates how YouTube has social core among its users. He tried to related uploaded videos on YouTube to the social interaction on the network. Data was collected from over 1000 videos and he identified 80 thousand users and 270 thousand edges directed as friend edges. His paper shows the interaction between groups of user watching similar kind of videos. These social interactions are similar to social media networks but with more coherence on content.

- Finding and Evaluating Community Structure in Networks (Newman and Girvan - 2003)

  They give three algorithms to determine community structure in networks using different edge betweenness. The main feature of his algorithm is that they break up a network into communities by removing edges one by one, and then betweenness is calculated again after removing edges. The general community structure algorithm using consists of calculating betweenness score for all edges, removing the edge with highest score and recalculating the betweenness. Paper recommends using short path measure. The algorithm used has a runtime of O (m^2n).

- Unfolding of Communities in Large Networks (Blondel - 2008)

  This paper provides us with an algorithm to extract communities from large networks. In first phase of the algorithm different communities are assigned to node of the network. Hen for every node we consider a neighbour and then calculate modularity by removing a node and replacing it with its neighbour. In second phase new network is build according to the data in first phase. We can again apply first phase to find any other improvement. The accuracy of this algorithm is very good as compared to other community detection algorithms.

- FANMOD: Motif Detection (Wernicke – 2006)

  Wernicke and Rasche together developed a tool for fast motif detection in large networks known as “FANMOD”. FANMOD works on an algorithm devised by Wernicke. The algorithm developed by Wernicke is fast and unbiased. The main feature of FANMOD is that it compares the frequency of a motif in originalnetwork to the random network. FANMOD runs orders of magnitude faster than similar tools on large networks.

Data Collection

With the help of Google Forms, we have asked different question regarding YouTube and user recommendations. We got 104 responses.

First question: What people would prefer either YouTube recommendation or user recommendation?
Output: 61 People chooseYouTube recommendation while other 43prefers user recommendation.

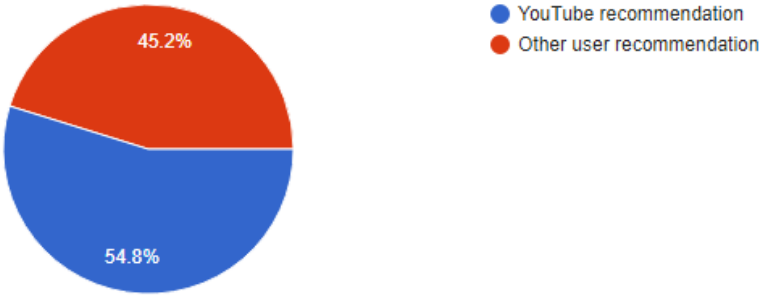
Second question: Does YouTube recommend similar kind of videos watched before or not?
Output:39.4% People believe that mostly YouTube recommend similar kind of videos while 40.4% people believe that only sometimes YouTube recommend similar videos.

Third question:How frequently people used YouTube recommendation sidebar to watch videos?
Output: 74.15 People use YouTube recommendation while others don’t use YouTube recommendation.

Fourth question: It is about people satisfaction towards YouTube recommendations.
Output: People were very satisfied withYouTuberecommendation while only 35.6% people thinks YouTube still needs to improve their recommendation algorithm.

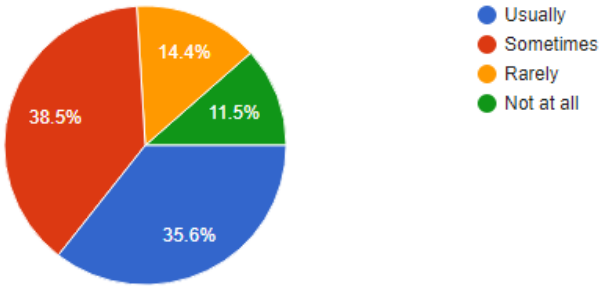
For watching a YouTube video what would you prefer?

104 responses



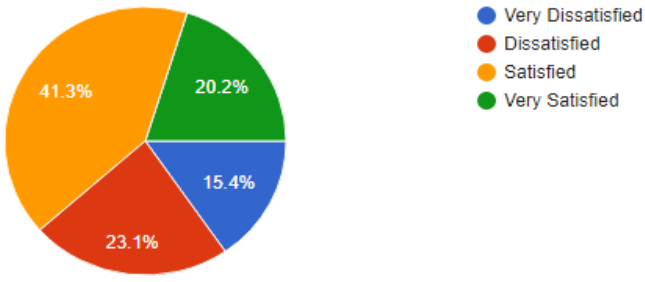
How often do you click on videos which is shown in sidebar in amid viewing video?

104 responses



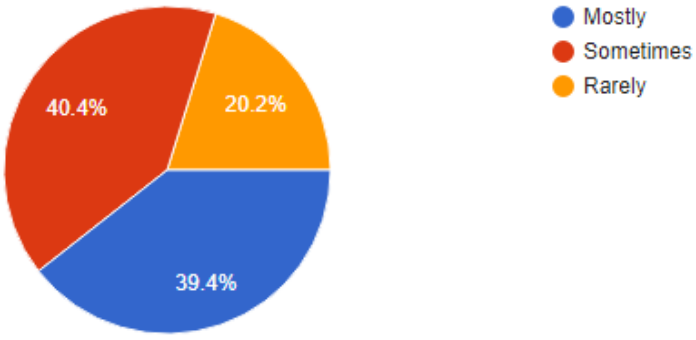
Are you satisfied with YouTube recommendations?

104 responses



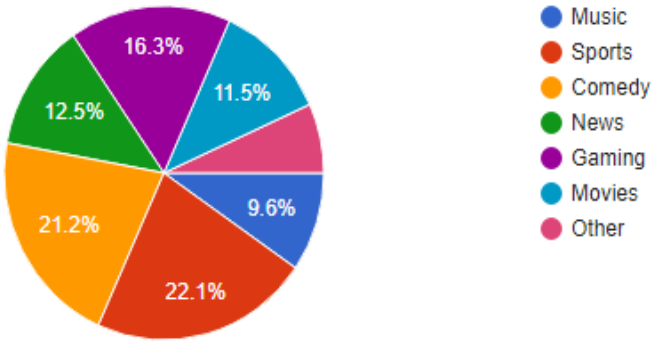
Does YouTube recommends similar videos to the videos watched before?

104 responses



What type of video you generally watch on YouTube?

104 responses



Functional module

We have divided this project to many modules to keep track of development of project and also it helps in time management. We have basically divided this project in three component or module. In first module we try to know our project scope, decide project deadline of each module in our own mind so that we can give time properly. In this module we tried to look at different way in which we can collect data and do analysis on that. After lot of discussion we chose one of trusted and public engaging way that is Google form. In Google form we made question which can be helpful in fulfilling project idea. So basically, we have formed multiple choice question and in that we chose option wisely so that every dimension can be covered. In this we have read ten research paper to get idea and know all thing to understand network. In report also we have mentioned research paper which was really helpful in doing project.

In second module our aim is to analyse this data that we collected. We have successfully made 120 people to fill this form. Here I will discuss first question because we got interesting result. Question was “for watching a YouTube video what would you prefer” option was YouTube recommendation and other user recommendation. We got 54 percent in favour of other user recommendation. So they are taking recommendation more from friend, dear people around them. After this we tried to analyse all this data on Gephi app. Using data from Google form we tried to analyse data on network analysis parameter like degree report, Eigen centrality, average in degree, average out degree, modularity analysis etc. Data that we collected on that 104 node was there. Average degree is measure of how many edges are in set compared to number of vertices in set. Average degree that we got is 3.452. we analysed average in degree which is average incoming edge in directed graph. We also analysed average out degree which is average of edge that is coming out of node. We also modularity report and modularity report is 0.779 in our analysis. Modularity report is nothing but analysis of structure of network. For this data we got average clustering coefficient equal to 0.238 .in this module we also explored all option available in Gephi.

In third module we tried to compare result that we got for YouTube recommendation with other social network. All social platforms that we used are frontline in current day. So we compared YouTube recommendation network with Facebook network of average degree 6.602. YouTube has less interaction than Facebook because YouTube only provides video content whereas Facebook contains all type of content. We compared modularity of YouTube recommendation with Facebook and modularity of Facebook is low because the large number of active users on Facebook with more people connectivity between a user to all other people on network decreases, and for high modularity it is necessary that each node is connected every other node in the network which is not possible in Facebook network because of large number of people of Facebook network.

We chose Instagram also to compare with YouTube recommendation network. Instagram average degree is low compared to YouTube recommendation network which suggest that Instagram recommendation algorithm is weaker than YouTube .The graph of Instagram is very scattered while on the hand YouTube has well defined and precise communities which help in better recommendation to user. Instagram modularity is 0.901. We compared YouTube commendation algorithm with twitch which is similar to YouTube but twitch is used only in America whereas YouTube is worldwide used. Modularity of Twitch is 6.354 which suggest that YouTube algorithm is better than Twitch.

Individual Contribution

Aniket Singh(18BCI0039)

He has worked on Statistical Data Collection, Calculation like finding, Degree Distribution, Centrality Measure, Community Detection and Analysing the Statistical Data of the data we have collected. He has also worked on collection of Proper Research Papers, Installing of Proper Software which is necessary for the project and learned to use all that software.

Robins Raj(18BCE0393)

He has worked on Proper understanding of YouTuberecommendation method, caring about consideration of proper data, data should be taken from proper and authorised publisher. He has also worked on collection of all data, analysing of all Data and reached to any conclusion.

Aditya Raj(18BCE0395)

He has worked on taking help from Gephi Software, Google Form and analysing the data of Google Form.He has also worked on proper presentation of data, maintaining of timeline for the Project, taking Printout, installing software, and other additional help to any member and getting proper Output/Result.

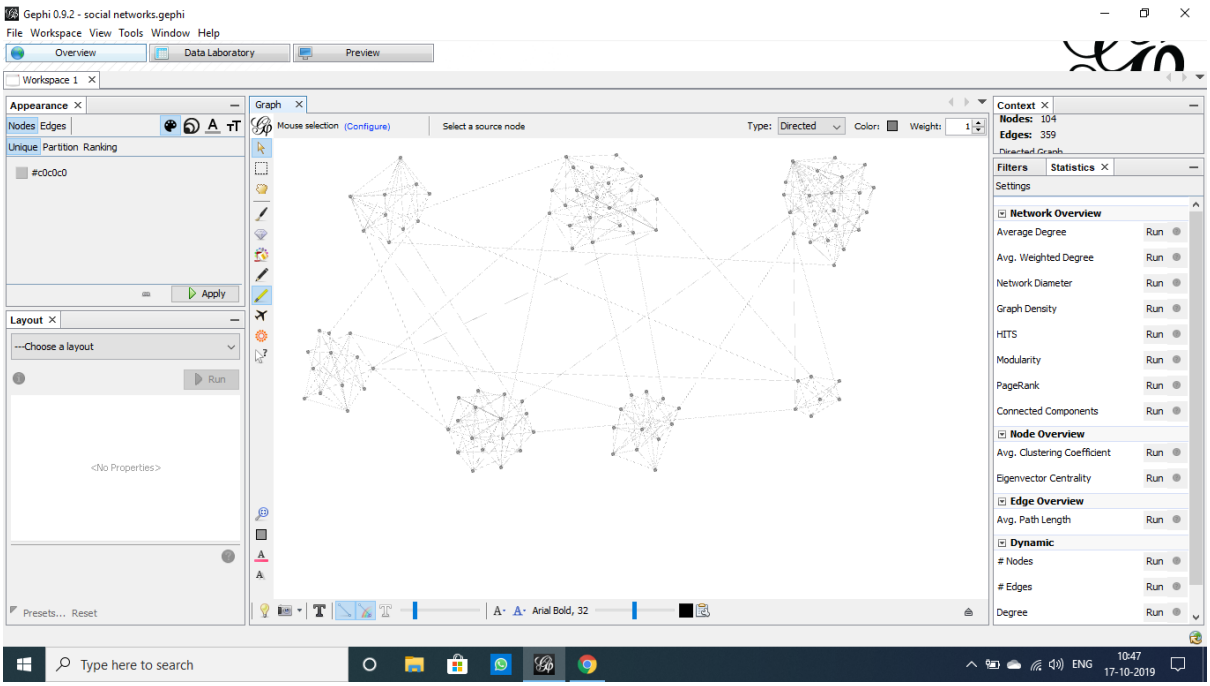
ANALYSIS USING GEPHI

We are using Gephi Software to depict YouTube channel network. The graphs consist of directed graphs where nodes are channel and there is an edge between two nodes if one channel recommends another channel. There are 7 communities which represents different genres of videos available on YouTube.

Also, there are many caveats that came with the nature of channel recommendation. There are cases when two YouTube channels collaborated with each other. In that case user recommendations changes but YouTube recommendation does not. But YouTube recommendation can be connected to each other indirectly. At some nodes in network we found that they do not connect to other node which means that that channel is deleted or does not upload more videos.

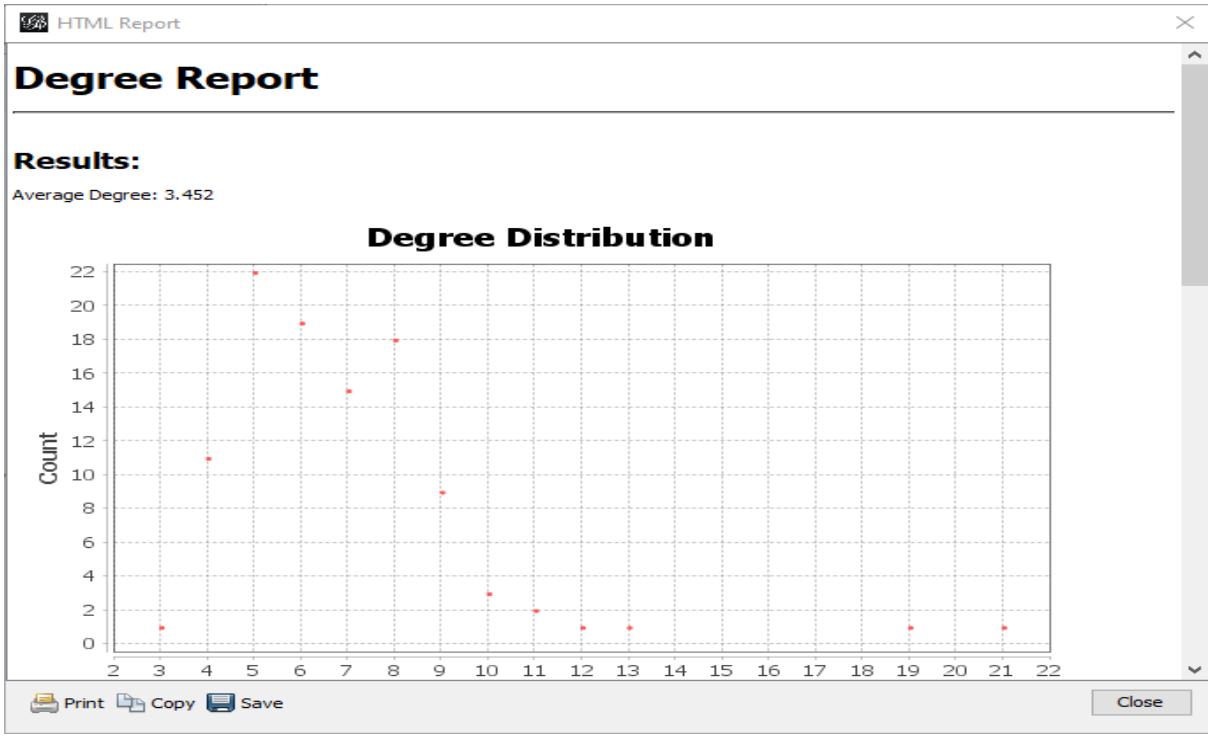
OUTPUT FROM GEPHI SOFTWARE

From Gephi we can calculate average degree, average in degree, and average out degree, modularity and clustering coefficient.

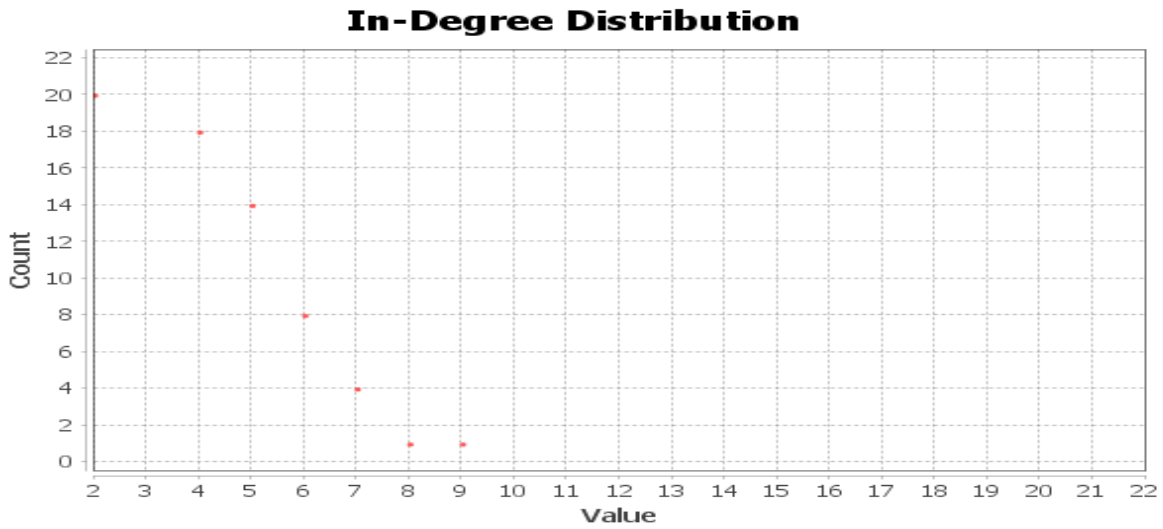


Total Node= 104
It represents individual YouTube Channel
Total Edge= 359
There is an edge between two nodes (say ‘A’ and ‘B’) only if Channel ‘A’ recommends Channel ‘B’. These recommendations can be of two type, either YouTube recommendations or User recommendations.

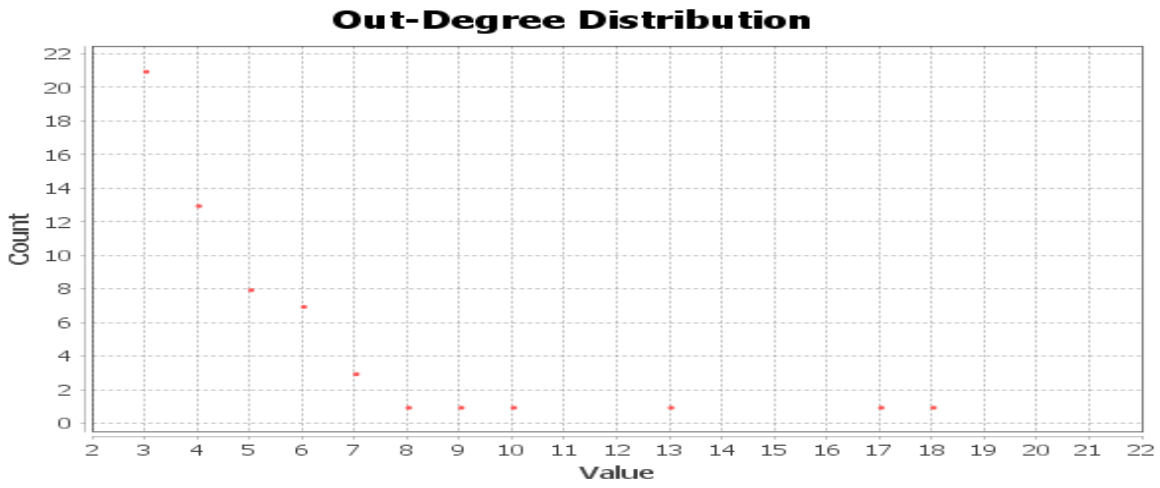
Average degree: 3.452



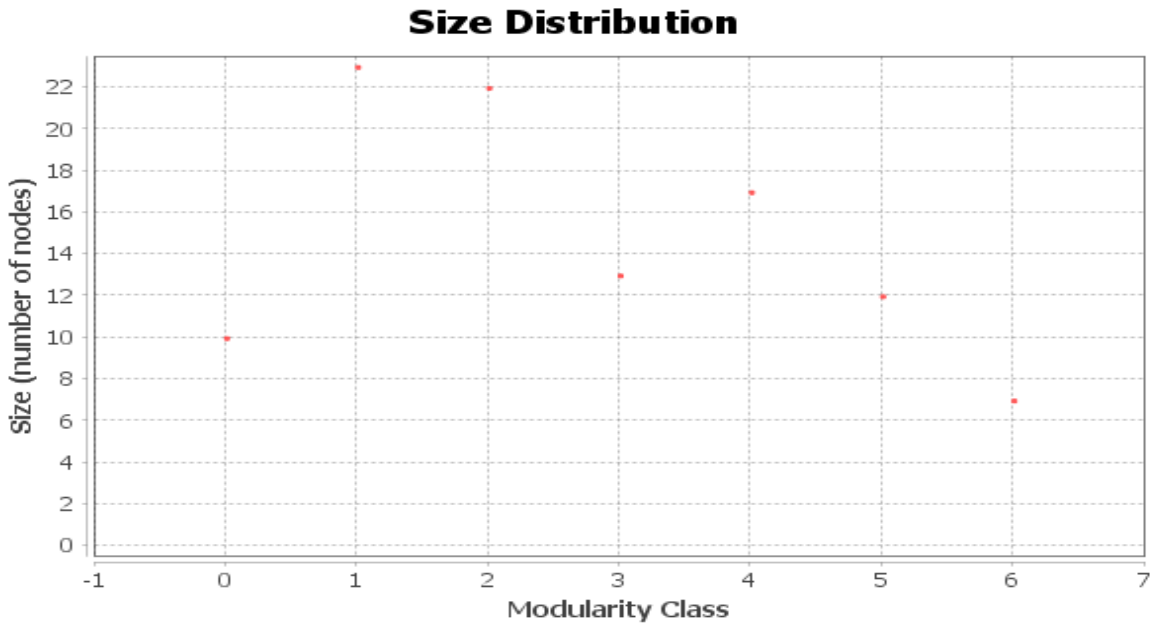
Average in degree



Average out degree



Modularity: 0.779



Parameters:

Randomize: On
Use edge weights: On
Resolution: 1.0

Results:

Modularity: 0.779
Modularity with resolution: 0.779
Number of Communities: 7

Clustering Coefficient:

Parameters:

Network Interpretation: directed

Results:

Average Clustering Coefficient: 0.238

Comparison with other Social Networks

Facebook

The YouTube recommendation network created has an average degree of 3.342 while the Facebook network has an average degree of 6.602 which is very high as compared to other social network. The reason for high average degree of Facebook is the more number of users than other social networks. With more people high connectivity is formed between people. According to a survey conducted by 'The Verge' a user on Facebook has an average of 300 friends which is the main reason for the high average degree. A YouTube network has less interaction with other users as YouTube only allows video content while Facebook allows video, photos and text content which is the main reason for the more interaction and hence high connectivity. Despite of the fact that Facebook has almost twice the average degree of YouTube, both the network Facebook and YouTube recommendation network has same nearly same modularity. Facebook has a modularity of 0.701 while YouTube recommendation network has modularity of 0.779. The reason for low modularity of Facebook network is the large number of active users on Facebook with more people connectivity between a user to all other people on

network decreases, and for high modularity it is necessary that each node is connected every other node in the network which is not possible in Facebook network because of large number of people of Facebook network.

Instagram

Instagram has an average degree of 2.234 which is very low as compared to YouTube recommendation network, this is because number of Instagram user are very less as compared to Facebook and YouTube and Instagram is newly released social network so the algorithm used is not that strong and need improvement overtime. Instagram too has a recommendation algorithm and network which is not very efficient. On comparing the network graph of YouTube and Instagram we found that Instagram does not have vey defined communities. The graph of Instagram is very scattered while on the hand YouTube has well defined and precise communities which help in better recommendation to user and well connectivity of the users to each other. Instagram has modularity of 0.901 which is very high for any social network. The main reason for the high modularity is high connectivity between all users because of less number of users. Due to this every user has a chance to connect to all other user directly or indirectly.

Twitch

Twitch is another platform for videos sharing similar to YouTube but YouTube is worldwide while twitch is mostly used in North America. YouTube has all genres of video while twitch focus mainly on gaming. Twitch has an average degree of 3.21 which is same as that of YouTube recommendation network. This suggests that both YouTube and twitch have similar type of recommendation algorithm and network. Twitch has a modularity of 6.354 which is moderate modularity for any social network. Reason behind this moderate modularity is that twitch is only popular in North American countries. Twitch never expanded outside North America which is the solid reason for the moderate modularity. On the other hand YouTube is worldwide popular has all type of videos.

Results

We have used Gephi to make the graph of the YouTube recommendation network. There are seven communities which represent different genre of YouTubevideos (like music, sports, comedy, news, gaming, movies etc). Each node in these communities are connected to each other or connected to other communities by edges which represent YouTube recommendation or user recommendation or both. From Gephi we depict a YouTube network which has average degree 3.452 which is much higher than any random graph or network. We calculated the modularity which is 0.779 which is good to have a well-connected graph. High modularity helps in linking and combining nodes to form a completed graph.

Conclusion

YouTube provides us an amazing recommendation algorithm which follows some network structure. As we have usually seen YouTube recommends channel which got similar type of content. One way in which we can enhance our work by using more data which will not keep us in confused state but more specific. This can be done by exploring more channels and crawling more website. We could explore whole network rather subset of network. Another thing we can do is to analyse development of network over time as network is dynamic which increases every time someone clicks on it. So we can capture snapshot of the network at different time and analysing its difference at different moment.

References

Research paper 1: Published date: 8 Apr 2011 Publisher: Anjana susarla, Jeong-Ha oh, Yong Tan Title: Social Networks and the Diffusion of User-Generated Content: Evidence from YouTube

Research paper 2: Publisher name: Marcelo Maia, Jussara Almeida, Virgílio Almeida Topic: Identifying User Behaviour in Online Social Networks

Research paper 3: Published date: April 2013 Publisher name: Sharon S. Phate

Research paper 4: Publisher name: James Davidson, Benjamin Liebald Topic name: The YouTube Recommendation System

Research paper 5: John C. Paolillo, Structure and Network in the YouTube Core. In HICSS 2008 - 41st Hawaii International Conference on Systems Science