Analysis of Twitch Social Networks

# Abstract

Analysing graphs using different visualisation techniques is a great way to do exploratory analysis on a data set. A lot of information can be extracted just from the size and shape of a given networks, and the use of algorithms that arrange nodes based on their centrality can also help you identify outliers and draw conclusions on both them and the network as a whole. This report contains the initial findings and conclusions from analysing the social network of streamers on the live streaming platform twitch.

# Introduction

Live streaming is a huge part of online entertainment in the digital age and during covid the average views per month doubled. Twitch (recently acquired by amazon) is the oldest of the live streaming platforms and is currently the largest with billions of views each month.

Live streaming typically involves one or more people, playing video games or doing some other activity, live for viewers at home to watch and interact with using the chat feature. A big part of twitch is viewer and streamer interaction through chat which makes the viewers feel like they’re part of a bigger community of friends.

Anyone who wants to can start live streaming on twitch is able to, with the only barrier for entry being an internet connection. This means that there are hundreds of thousands of users which can be represented as nodes and even more edges between them all.

However, what’s interesting about twitch is unlike social circles in person, is that the reach of a streamer is to hundreds of thousands of people which is hard to achieve in person. This is paired with twitches algorithms that suggest that you view streamers who already have a lot of people watching them. The result is that most edges, views and popularity are attributed to the very top 1% of twitch streamers. In this regard the twitch social networks are more akin to celebrity social networks than they are ordinary social circles.

For this reason, I have chosen to only investigate a small portion of the network containing these top streamers as that is where I think the most information is and where we might be able to draw some interesting conclusions. The questions I set out to answer were “Do mature channels mostly interact with one another?”, “Does age of the channel determine friendship networks?”, “Can we predict a user’s partnered status based on their degree?”.

# Materials and Methods

The initial dataset for this analysis was the Twitch Social Networks data set collected by Rozemberczki et al and download from the Stanford Network Analysis Project (SNAP).

The dataset had the users represented as nodes and the edges between them represented them both following each other, following on twitch works the same way as it does on twitter where it doesn’t have to be mutual.

Each user had data about their current total channel views, if they were a mature channel (meaning swearing and violent videogames), if they were partnered or not (like being verified on twitter) and their twitch ID.

The initial exploration of the data raised questions that I could not begin to draw conclusions on without more information about who the channels were.

To get the extra information, I wrote a python script to scrape additional information from the twitch API. The additional information was the channel names, their new current viewer count, and if they were a partner or affiliate (affiliate is a lesser version of partner where they unlock some extra features on your account). The new total view count also allowed me to calculate the view growth since the original data was captured in 2019.

## Visualisation tool

For all the visualization I used an open-source graph analytics tool called Gephi (<https://gephi.org/> ). Due to the large number of nodes in the dataset, the majority of which with very few connections, filtering out nodes by degree was important and done for all the graphs we produced.

## Initial observations

We found that removing all nodes with a degree of less than 30 removed 95% of the nodes and allowed us to inspect more easily the 5% of nodes and edges that contained most of the information in the graph.

Two other techniques we used were scaling and colouring the nodes based on their attributes like their Maturity and their partner status and on their ranking like degree and views, doing both allows you to visualise the interaction between multiple node attributes at once.

To arrange all the related nodes on the graph we used the Force Atlas built into Gephi. This turns your nodes and edges into an N-body problem where the size of the node is the mass of the body and each edge represents and attraction between two nodes. To enable even greater visualisation of these groups, we ran a community detection algorithm (<https://iopscience.iop.org/article/10.1088/1742-5468/2008/10/P10008/pdf>)  in Gephi that separates nodes into groups based on their relative degree to each other.

Once these communities were found we could then colour the nodes to view them.

**Results**

Our initial analysis (Figure 1) revealed that a very small minority of nodes had the most views and the highest degree. In order to create meaningful visualisation all of the graphs we produced were focused on the top ~10% of nodes as this portion had the highest entropy.

Below is a visualisation of The top 5% (ranked by degree) on the left and all the nodes on the right. The nodes are scaled by degree and coloured using a heatmap for view with a logarithmic scale. The nodes are in the same positions in both graphs as you can see with these larger nodes, just a filter was applied.

A picture containing star, outdoor object, light, dark

Description automatically generated

Figure 1: Initial network analysis

### Do mature channels mostly interact with one another?

No. We found that mature channels show no clear preference for interacting with other channels regardless of if they’re family friendly. However, we did clearly see a group of family friendly channels that mostly interacted with other family friendly channels and did not interact with mature channels at all. Our working theory for this is that a mature channel does not risk exposing their viewers to inappropriate content by working with family friendly channels, but family friendly channels would be risking showing their normal family friendly viewers content with cursing and violent video games. This group is visible on graph **S1-B** as the white group on the right.

### Does age of the channel determine friendship networks?

Maybe, as you can see in graph **S2-A** and **S2-B**, areas of similar color(age) are grouped together by community detection in **S2-B**. But it’s unclear if this grouping is because they’ve all started at a similar time or because none of the newer channels are as popular as the older channels. As you can see in S2-A, the nodes that are most central and have the most inter community connections are the nodes with the largest size (views) and there is no clear relationship to age.

### Can we predict a user’s partnered status based on their degree?

Yes, all channels with a high degree are likely to be partnered. This chance to be partnered is increased if they also have lots of channel views. You can see that all the large and or red nodes in **S3-A** (high degree and or high views) are green (partnered) in **S3-B**.

# Discussions

### Family Friendly Channels

To visualise this, I created two graphs, both share the same edge and node location arranged by degree using a force atlas. This turns the graph into an N-body problem where each node is of equal mass and the edges between nodes represent their attraction to each other. The first graph S1-A was coloured to show pink for mature and green for family friendly and the second S1-B was coloured using the modularity class of the nodes. From the placement of the nodes, it’s clear to see that in general, the channels tend to interact equally with each other regardless of if they were mature channels. However, there is a grouping of nodes to the right of the main mass that seem to be more strongly connected with each other than they are to the whole. This separation becomes even more clear when looking at the modularity class, they can be seen as the white in S1-B. This small group is mostly comprised of family friendly channels that keep to their tight nit social circle, my working theory is that it can be appropriate to show mature audiences family friendly content. But it might not be appropriate to show family friendly audiences of young children and teenagers the mature content. As such we can see that the family friendly group keeps to themselves and doesn’t mix with the mature channels, but the mature channels have no issue with mixing with other family friendly channels.

### Age of channel

From the modularity plot S2-B you can see that there are definite separations of different communities, but the only community that has a significant portion of nodes that interact with nodes outside their group is the Orange group of old channels. This group also has a large amount of channel views and a high degree. It’s hard to determine if friendship networks are a function of channel age, or if streamer friendship networks are just generally tightly coupled with the only channels capable of branching out and making inter-community friends are the older channels. Which also happen to have a lot more viewers and grant them a sort of celebrity status that makes it easy for them to make connection with smaller nodes than themselves. This paired with them being around longer explains how they’ve been able to gain as many intercommunity connections as possible.

### Partnership status

In order to determine if the degree and viewership had any correlation to the user’s partnership status, we made two plots that use the same size and layout of nodes, just coloured differently. Graph **S3-A** and **S3-B** scale the nodes based on their degree and were arranged using a force atlas with each edge being of equal weight. This made the nodes with the higher degree more central. On graph A the colour is a heatmap of views with red being more views. On graph B the green nodes are partners, and the pink nodes are not. As you can see comparing the two, all the larger nodes are green, and all the higher view nodes are likely to be green. However, the inverse does not hold true, being partnered does not necessarily meant that channel has a high degree and or high views. Investigation into some of the low degree outliers showed them to be channels owned by organisations that would stream events like e-sports tournament which garner thousands of views. The edges between the nodes are from mutual connections and the organisations do not tend to follow other channels which leads them to have a disproportionately low degree in relation to their channel views.

# Conclusion

Through our analysis we were able to find satisfying answers to our questions, the channels being family friendly and the age of the channel both influence the communities that form between the nodes. It is also quite clear that the degree and viewership of channels correlated to the partnership status of the channel. These conclusions are limited in that the attributes of the nodes were studied in isolation which makes it possible that the viewership and degree might not be directly correlated as they appear, but are both dependent on something else like the activity the node most commonly engages in. A large part of a channel's viewership can be determined based on what activity they stream most, for example professional e-sports players get much more views when they’re streaming the game they play professionally than anything else. Unfortunately, limitations of the Twitch API makes fetching historical activity data impossible. Further research could be conducted into quantifying exactly how much each of the variables we studied affects the other attributes of the nodes.

# Appendix

S1-A & S1-B

A picture containing chart

Description automatically generated

S2-A & S2-B

A picture containing chart

Description automatically generated

S3-A & S3-B

A picture containing text

Description automatically generated

### Graph key

|  |  |  |  |
| --- | --- | --- | --- |
| Graph | Node scale | Node colour | Filter |
| S1-A |  | Green= PG, Pink = Mature | Degree > 22 (top 11%) |
| S1-B |  | Modularity Class | Degree > 22 (top 11%) |
| S2-A | Views | Heatmap Age | Degree > 22 (top 11%) |
| S2-B | Views | Modularity Class | Degree > 22 (top 11%) |
| S3-A | Degree | Heatmap Views | Degree > 22 (top 11%) |
| S3-B | Degree | Green = Partner, pink = Affiliate | Degree > 22 (top 11%) |
| S4 | Degree | Heatmap views | All nodes |
| S5 | Degree | Heatmap views | Degree > 30 (top 5%) |