

SOCIAL NETWORK ANALYSIS 2021-2022

ΒΑΣΙΛΕΙΟΣ ΒΛΑΣΟΠΟΥΛΟΣ 3080024

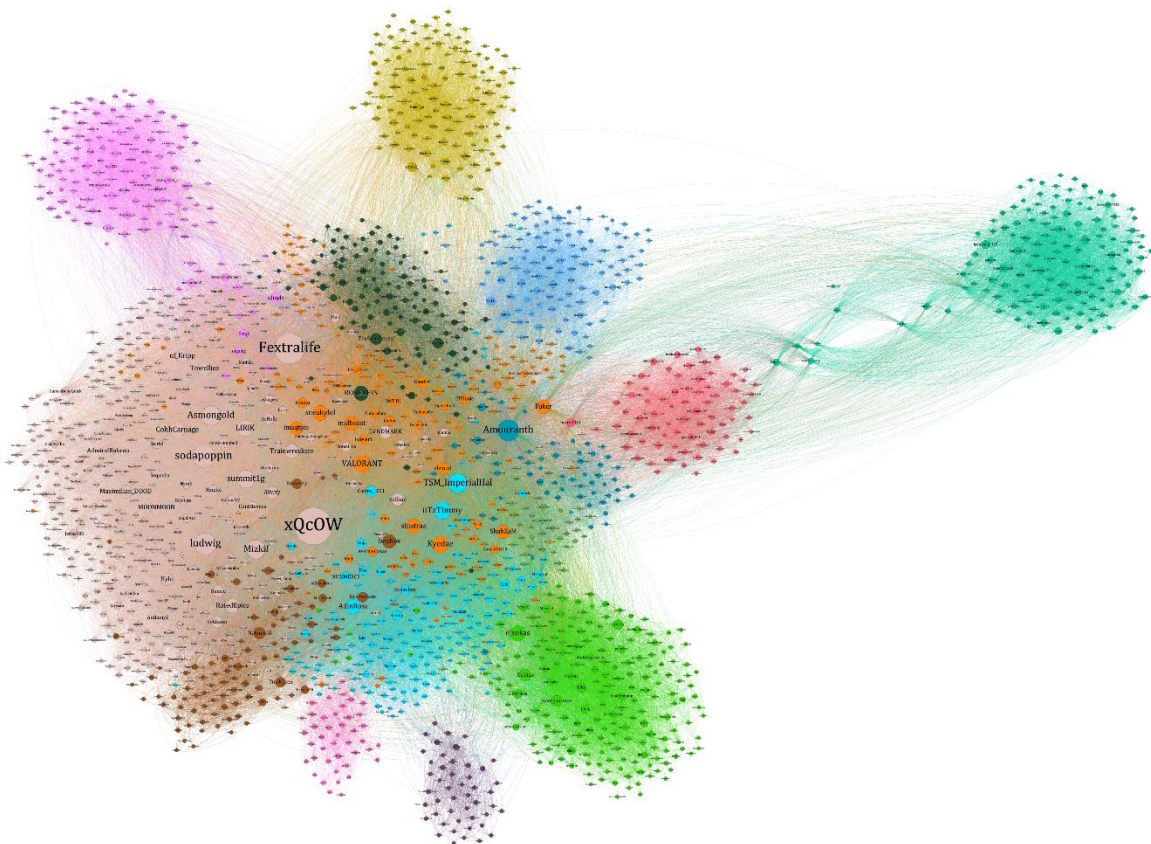
Check README.txt for file information.

Twitch (<https://www.twitch.tv>) is a video live streaming service that focuses on video game live streaming. Any user can create a channel and stream their gameplay live on twitch and have others watch. People can follow or subscribe to a channel and chat with other viewers while watching. It has become a huge social network with millions of active users every moment.

The dataset consists of twitch channels, their viewer count, and the number of common viewers they share with other channels. The source of the dataset is [Kiran Gershenfeld's Visualizing Twitch Communities Project](#). He collected the data by querying the top 100 stream channels and their viewers from the Twitch API every hour for a period of time.

Each node is a twitch channel. The node size is proportional to the channel's viewer count. Every edge is weighted and represents shared viewers from one channel to the other (the weight being the count of shared viewers).

The dataset consists of 2200 twitch channels (nodes) and the connections between them (edges) but, after culling channels without connections and channels with connections of less than 1000 viewers shared, we're left with the final dataset of 1900 nodes and 78577 edges. The resulting directed graph is:



My history with game livestreaming began in 2008 with the beta of own3d tv, a livestreaming platform centered on gaming that has now ceased to operate, and afterwards with the beginnings of twitch.tv in 2010. Twitch.tv is a spinoff of justin.tv which was a general livestreaming service. The prominence of gaming in justin.tv is what led the owners to create a livestreaming platform just for gaming, the now famous twitch.tv. This is a subject I've grown up with and know the history of, so this is why I chose to analyze a part of it.

Out of this analysis I want to see how twitch viewership behaves, how shared viewership contributes to making a channel big, and the interconnectivity of the game streaming community in general.

Tools used: **Gephi** for the visualization of the graph, **python** with **networkx** for the analysis of the graph.

Analysis of the network:

- **Basic topological properties, such as numbers of nodes and edges, network diameter, and average path length.**

Nodes: 1900

Edges: 78577

Network Diameter: 8

Average Path length: 2.4756

- **Component measures, such as number of connected components, existence of a giant component and component size distribution.**

Number of strongly connected components: 1900

Number of weakly connected components: 1

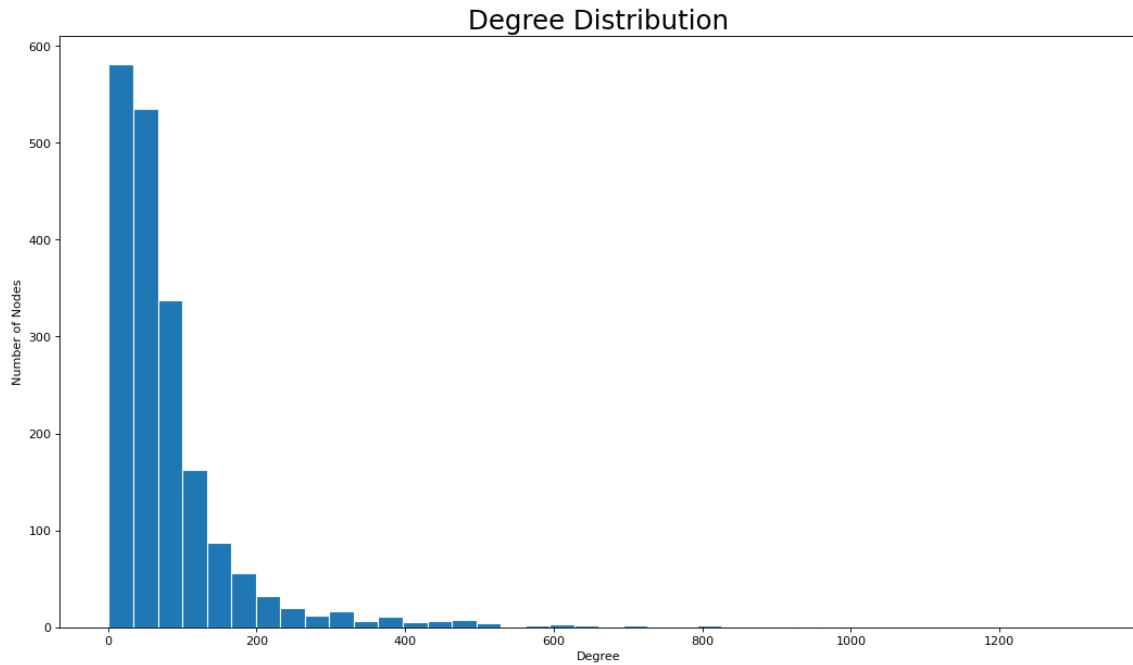
The graph is fully connected therefore there is one giant component, the graph itself.

Component size distribution is 100% for all nodes as the graph is one giant component.

- Degree measures, such as maximum and average node degrees, as well as degree distribution.

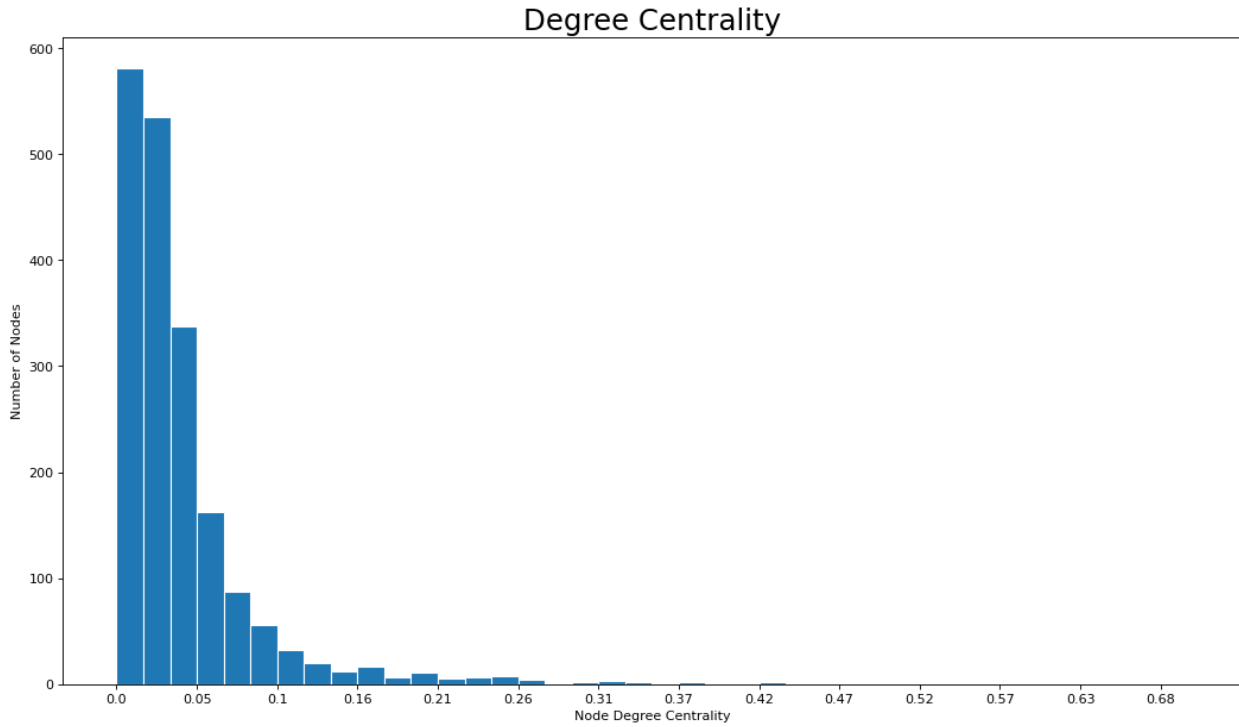
Max Degree: 1320

Average Degree: 82.713

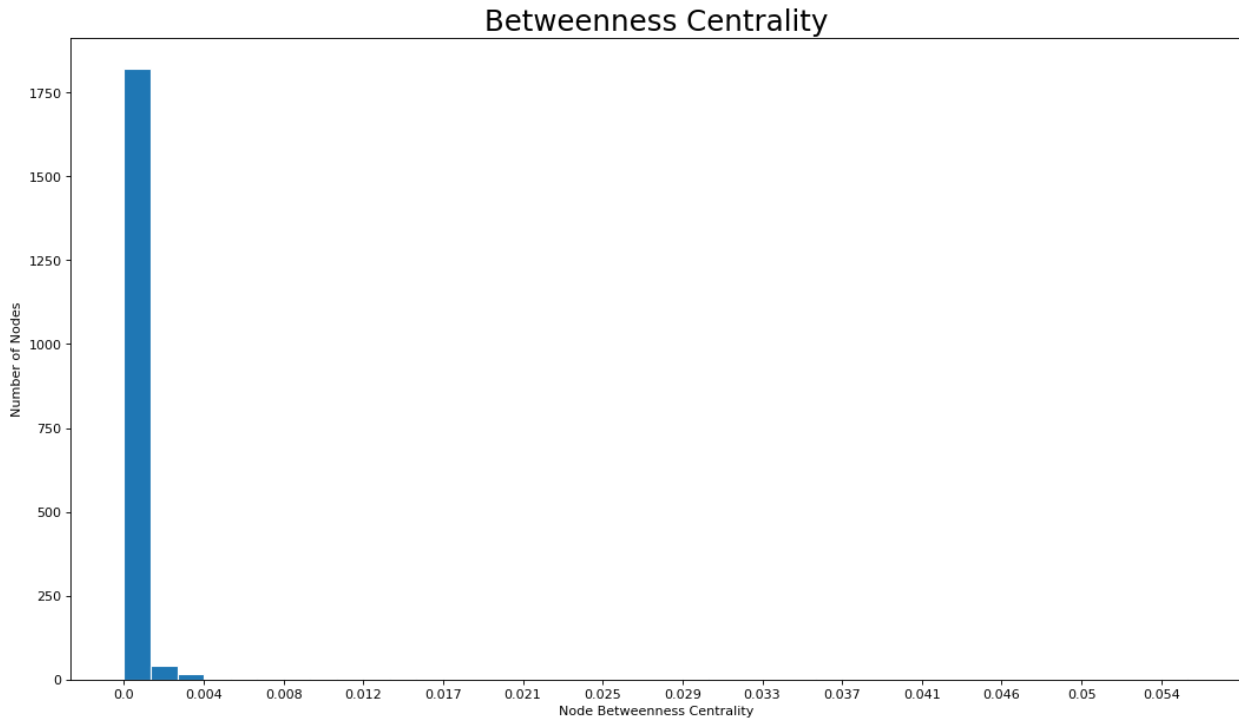


As we can see the average number of channels that share viewership are ~ 82 which is a surprise to me because I was expecting much lower numbers. Also, whilst a big portion of the channels appear sharing viewership with 1-33 other channels there actually is a channel which shares viewership with 1320 other ones!

- Centrality measures (degree, betweenness, closeness, eigenvector)

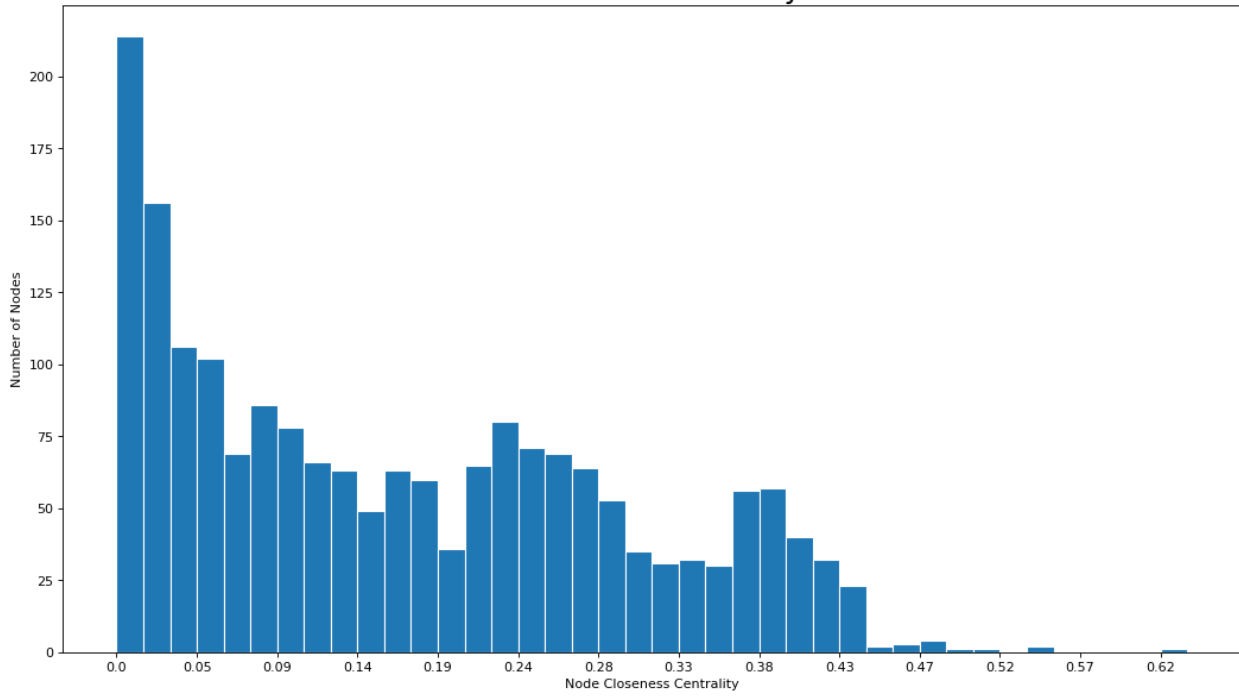


The degree centrality is directly correlated with the degree distribution as expected.



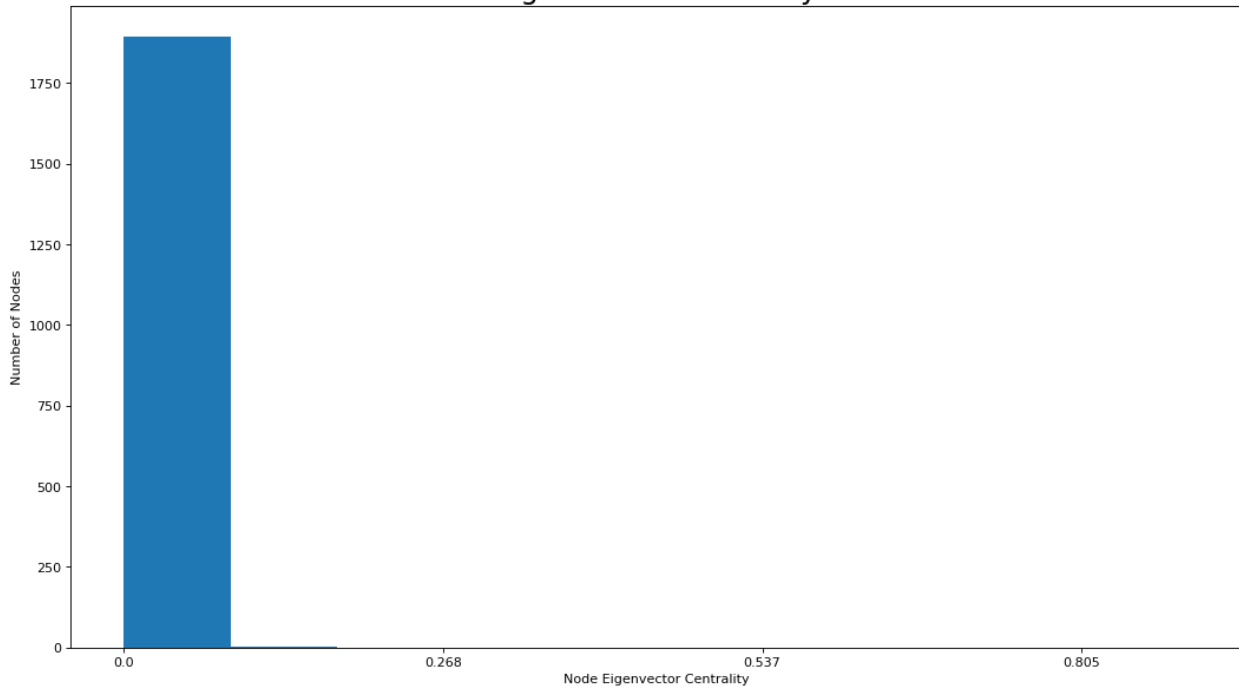
The betweenness centrality is very low because this is a very populated network so there aren't many bridges between the nodes.

Closeness Centrality



Closeness centrality falls mostly below 0.5 with small exceptions which is something to be expected in a network of this size as there aren't many channels that share viewers with every community.

Eigenvector Centrality

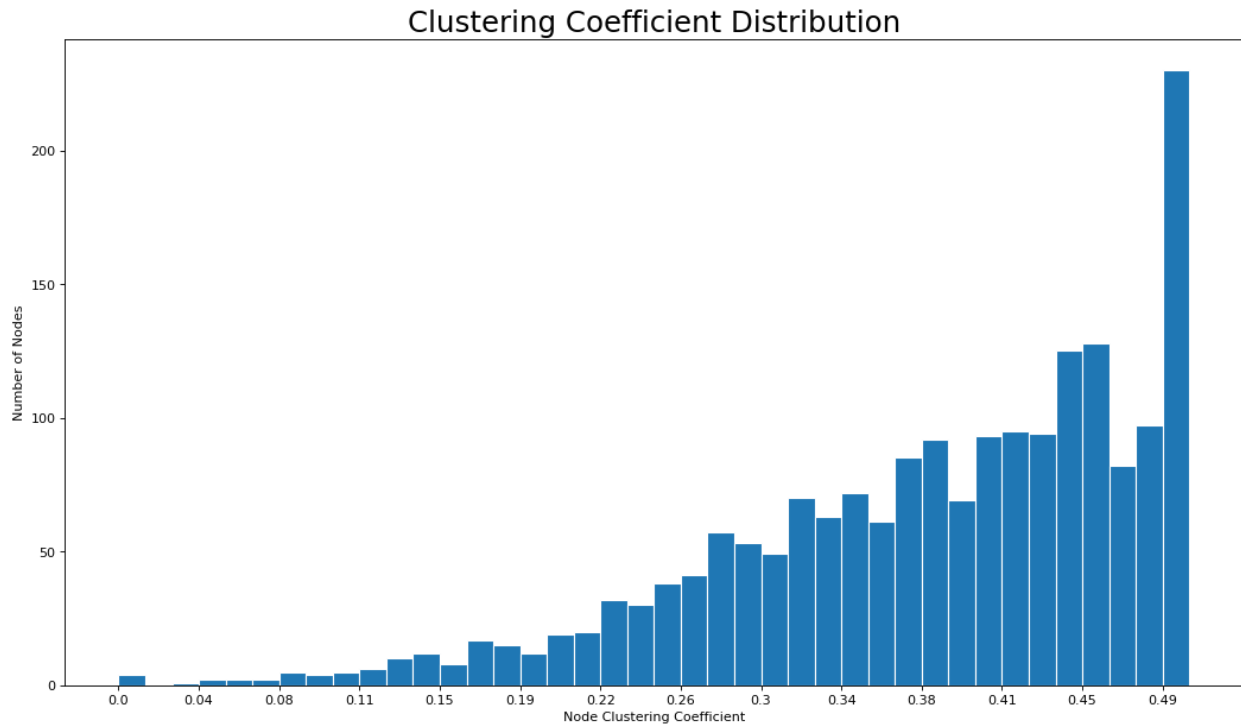


The eigenvector centrality tends mostly towards zero with a few exceptions, so very few channels have influence over the whole of twitch's network.

- **Clustering effects in the network:** average clustering coefficient, number of triangles, clustering coefficient distribution, existence of the triadic closure phenomenon in the friendship neighborhood.

Average Clustering Coefficient: 0.3788

Total triangles in undirected version: 2028366



Transitivity: 0.1833

The clustering coefficient falls below 0.5 and averages at ~ 0.38 while the number of total triangles in the undirected version of the graph is huge, but this is a directed graph so lower clustering coefficient than what the triangles would indicate is to be expected.

- **Bridges and local bridges.**

Bridges:

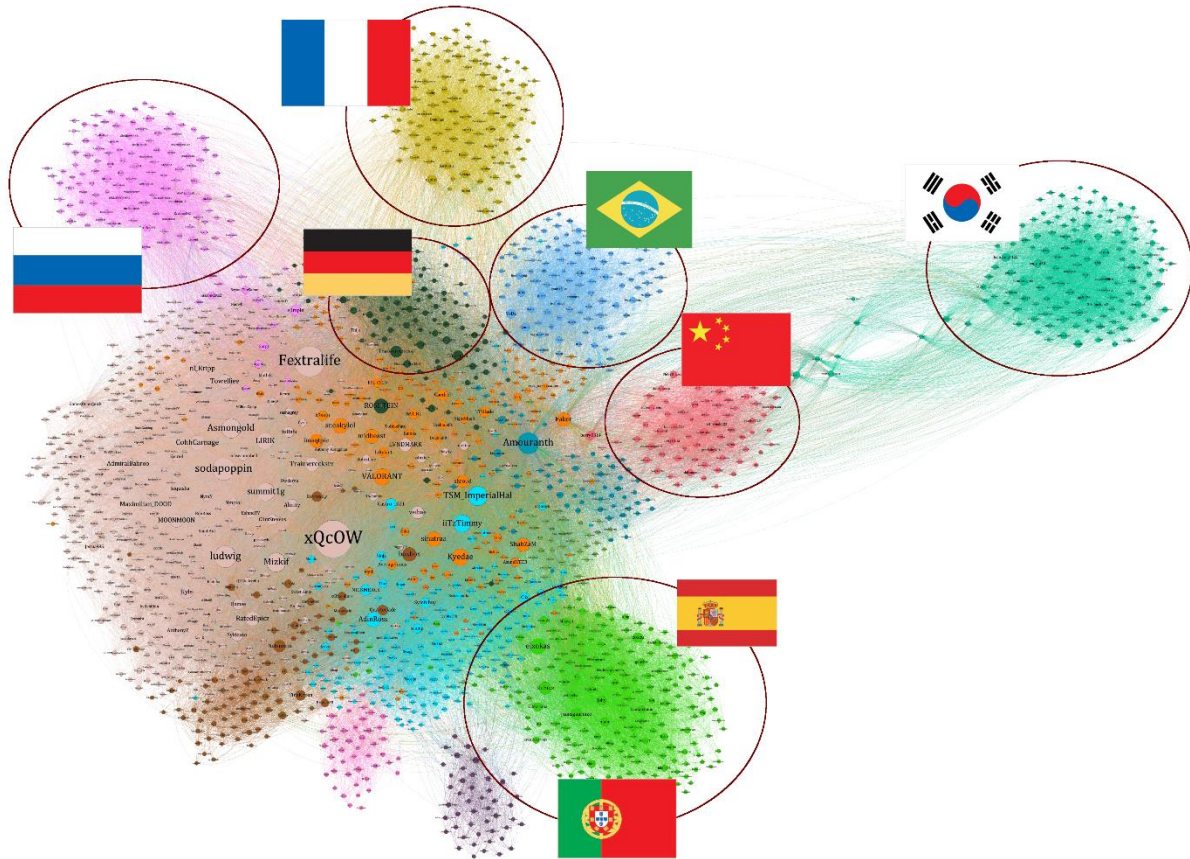
- ('eazykee', 'Ruben')
- ('AbuSwe7l', 'RakanooLive')
- ('Ruben', 'Fyrexxx')
- ('coscu', 'FgKlouf')

Local bridges: (excluding those in Bridges)

- ('Otzdarva', 'Mulder')
- ('RiffTrax', 'MST3K')
- ('Dennsen86', 'BarbarousKing')
- ('TheRealzBlueWater', 'Diegosaurus')

- **Gender and homophily.**

Interestingly homophily appears without the dataset containing any information on ethnicity or spoken language. After further examination of the community structure of the network, we can see that communities based on the spoken language appear. Whilst the majority of twitch’s userbase communicate in English (shown as the biggest and main cluster in the graph), communities that share a common language other than English appear organically. I’ve chosen some examples of non-English speaking communities shown below:



As we can see, there are communities (like the Korean speaking one) that are well connected internally but have less connections towards the English speaking majority, while there are also communities (like the German speaking one) that are not only contained within themselves but have many outward connections (and are therefore closer) towards the English speaking streams.

- **Graph density.**

Graph Density: 0.0218

- **Community structure (modularity), cliques.**

Modularity: 0.640

Number of communities: 15

Graph is directed so no Clique information.

- **PageRank**

Top 20 twitch channels by PageRank:

Scarra : 0.0474

HasanAbi : 0.0287

dogdog : 0.0190

shroud : 0.0167

Klean : 0.0130

Esportmaniacos : 0.0119

Maya : 0.0115

rkdthdus930 : 0.0100

lestream : 0.0098

ibai : 0.0096

knekro : 0.0090

blxckoutz : 0.0087

moistcr1tikal : 0.0085

Manyrin : 0.0080

ESL_CSGO : 0.0078

CDawgVA : 0.0077

Timmac : 0.0075

Aydan : 0.0075

pokimane : 0.0073

Cellbit : 0.0071

The interesting thing about pagerank is that the top 5 channels in viewership do not appear even in the top 50 channels ranked by pagerank score. It would be interesting to analyze the content of the streams ranked by pagerank to understand why they have such a broad appeal and are so interconnected within the network.

It becomes clear that while the english speaking streams are very saturated they also have the greatest potential of having the biggest viewership, but on the other hand native speakers create tighter communities so a moderate viewership might be easier to accomplish. Twitch is a vast network with the biggest channels sharing big portions of their viewership so an aspiring channel will have to break into that sphere to become big. The data missing in this analysis are the hardest to quantify: content and personality, maybe those would give us better answers on what makes a channel “big”.