ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

A NETWORK ANALYSIS OF

twitch

Bali Niki

Department of Informatics

2020-21

Contents

1. Introduction- Dataset Selection and Creation

Twitch is a live-streaming website geared towards gamers that launched in 2011 and has grown substantially over the years, to what it is today; one of the leading live streaming video services in the world. As of February 2020, it had 3 million broadcasters monthly and 15 million daily active users, with 1.4 million average concurrent users. Users have the freedom to navigate the platform and watch anything they want, from gameplay to music to even educational livestreams. People who livestream content are called "streamers" and their channels usually follow a certain direction - not too often a streamer will fall under many different categories. A big part of the Twitch experience is collaborations between streamers. Like in every big community, it is impossible for people to be absolutely isolated and that is the case for Twitch as well. Streamers will either play games together, collaborate on projects, or even meet up and stream the experience for their respective audiences, garnering more and more viewers in the process. Knowing this, I thought it would be very interesting to dig a little deeper into the Twitch community and analyze the connections between the biggest streamers on the platform.

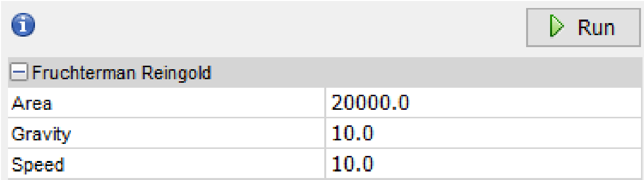
The data used for this analysis are the top 100 streams and all of their viewers during the week of December 6–12th 2020 and have been collected using Twitch's API. For the data visualization I will use Gephi.

The networks is undirected and consists of nodes which represent streamers, and edges which connect two nodes and represent the number of viewers shared between the two streamers. Also, the weight of an edge represents the amount of common viewers.

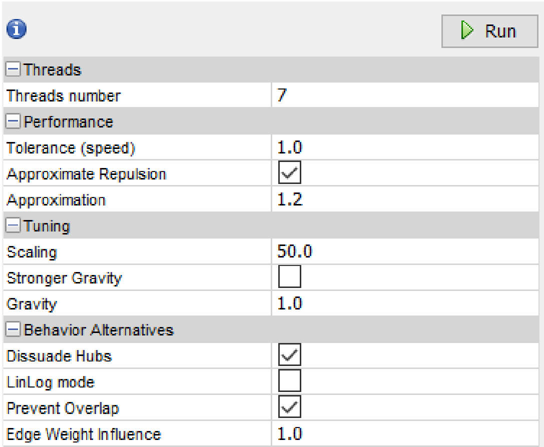
2. Visualization/Graphical representation of the network

Once the data are loaded into the Gephi tool, we can begin the visualization.

Firstly, we apply the Fruchterman Reingold layout algorithm in order to spatialize the data and then the ForceAtlas 2 algorithm so that we disperse groups and give space around larger nodes. Lastly, for the layout, we run Label Adjust to make all the node labels distinguishable and Expansion to make the graph a little bigger and easier to study.



Fruchterman Reingold layout

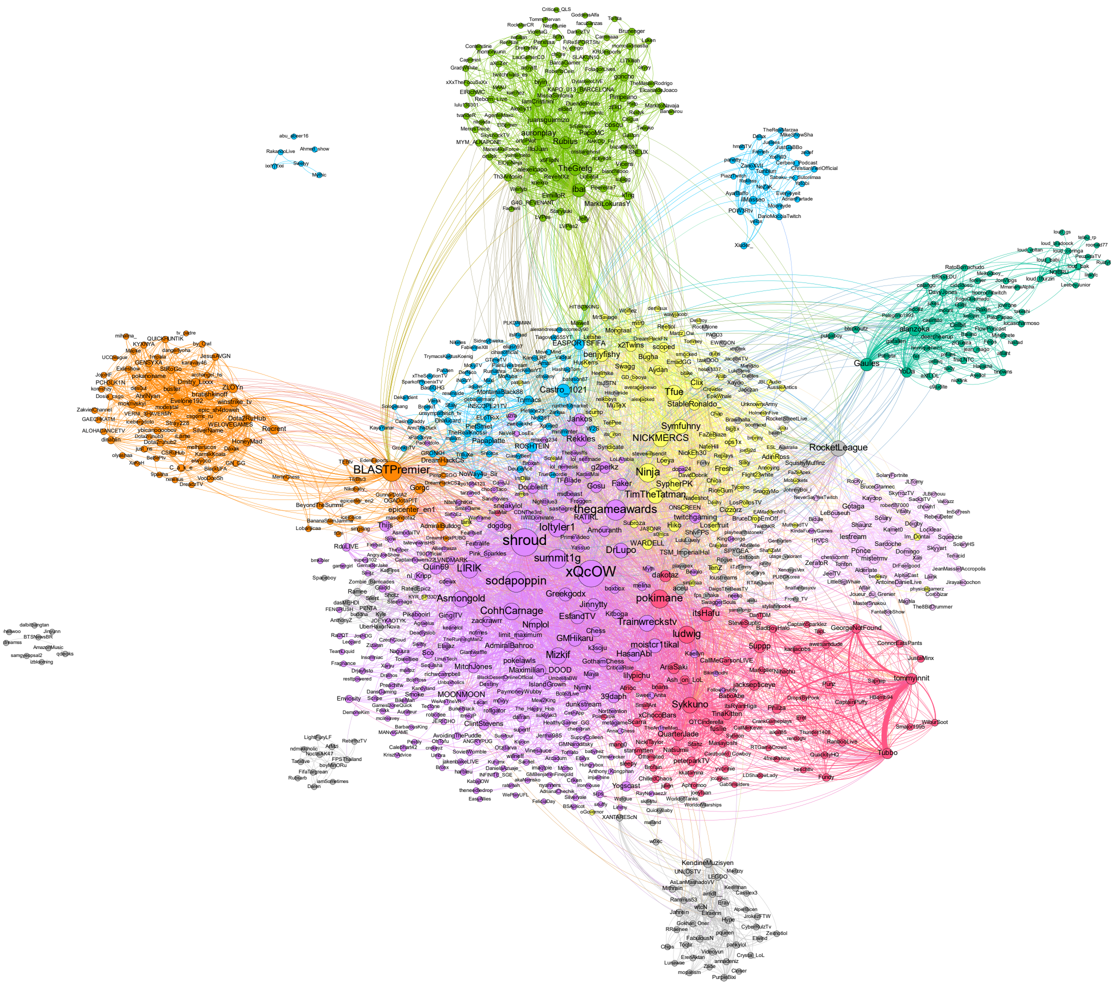


ForceAtlas 2 layout

In order to highlight the significance of weights on the graph, we apply an Edge Weight filter so that edges with weight lower than 860 are not visible and we also tamper a little with the edge weight scale so that the edges with bigger weights are displayed as a thicker line.

Next, we make the nodes' size scale with their degree, and now the nodes with the bigger degrees will also appear bigger on the graph.

Finally, we calculate the modularity and use it to color the nodes accordingly. This creates the final graph, which is shown below.



At first glance, we can already identify communities on the graph but we will further analyze them later. For now, it is safe to assume that the biggest streamers' audiences are closely intertwined and this goes to show how important connections are to streaming.

3. Analysis of the network using Gephi

3. 1. Basic topological properties

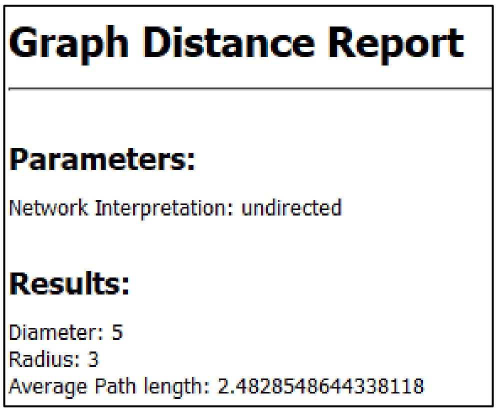
Every analysis must begin with the most basic topological properties which give the rough outline of a network.

Number of nodes or vertices, which we denote with N, represents the number of components in the system. We will often call N the size of the network. In our case, the network consists of 969 nodes, therefore 969 different streamers.

Number of edges or links, which we denote with E, represents the total number of interactions between the nodes. In our network exist 19989 edges.

Next we calculate the **density** of the graph and receive a value of 0.043 as a result. Density can be defined as (number of edges)/(maximum possible number of edges) or $\frac{2E}{N(N-1)}$ where E and N are defined as previously mentioned. Since the calculated value is greatly lesser than 1, we can conclude that our network is significantly sparse, which makes sense since it's virtually impossible for all streamers to have connections to one another and therefore create a dense graph.

The **diameter** of a graph is the length of the longest shortest path, or simply the maximum of shortest paths between any two nodes in the network. The diameter of our graph is 5.



Graph Distance Report

Parameters:

Network Interpretation: undirected

Results:

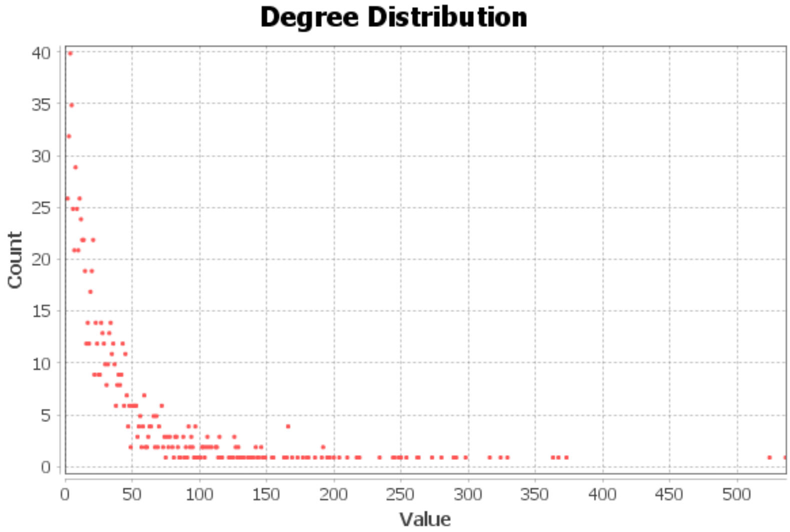
Diameter: 5
Radius: 3
Average Path length: 2.4828548644338118

The **average path length** of a graph is the average number of steps along the shortest paths for all possible pairs of network nodes. In our network it has a value of approximately 2.5, which shows that it is fairly easy for viewers to be led from one streamer to another.

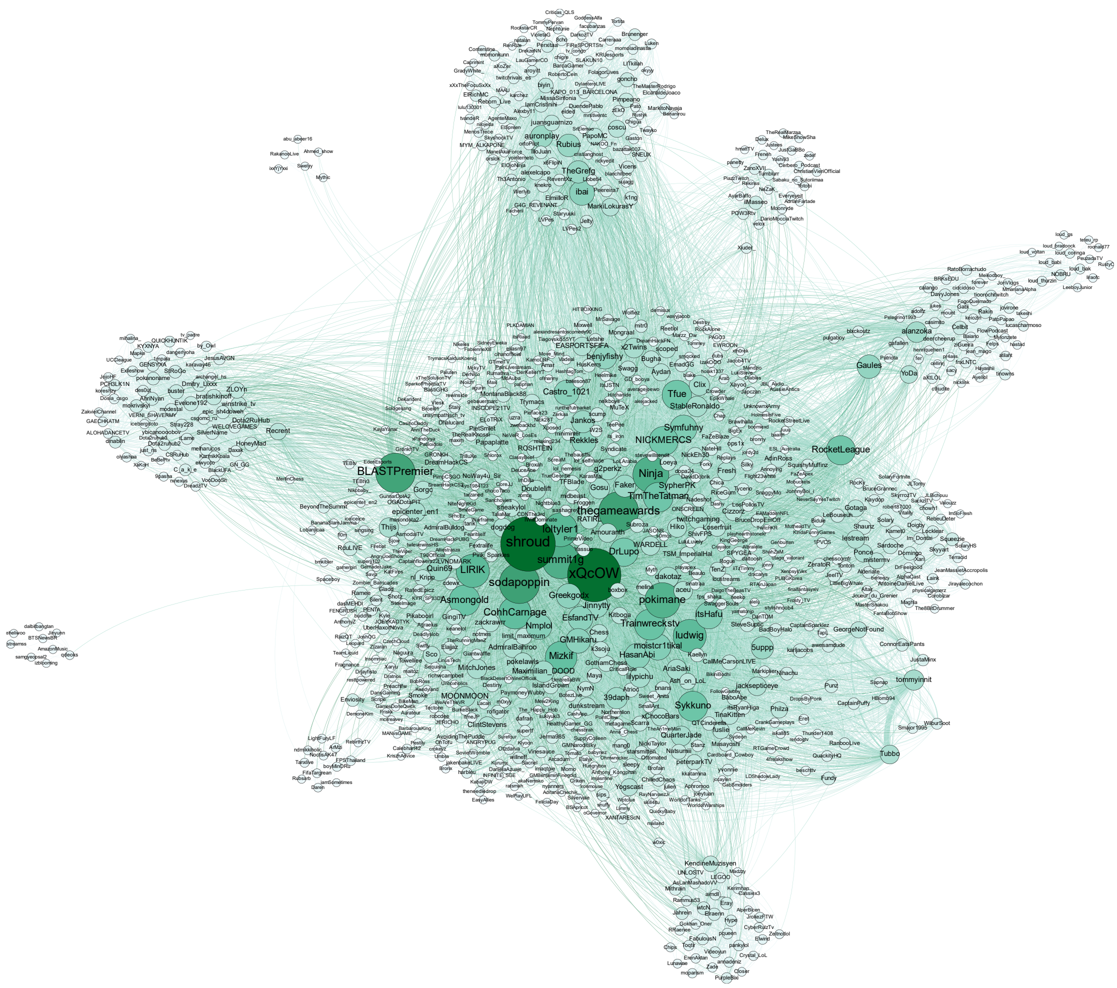
3. 2. Degree measures

The **degree** of a node is the number of relations (edges) it has, independently if it's an in or an out relation. It's the sum of edges for a node. This is an undirected graph therefore we don't have in-degree and out-degree but only (total) degree.

As we request for a degree report on our network, we are informed that the Average degree is 41,257 and the degree distribution is as shown in the diagram below (y axis = how many nodes have this degree, x axis = the degree value):



Our next step is to visualize the variations between degrees among the nodes, using Gephi’s visualizing tools. The result graph is:



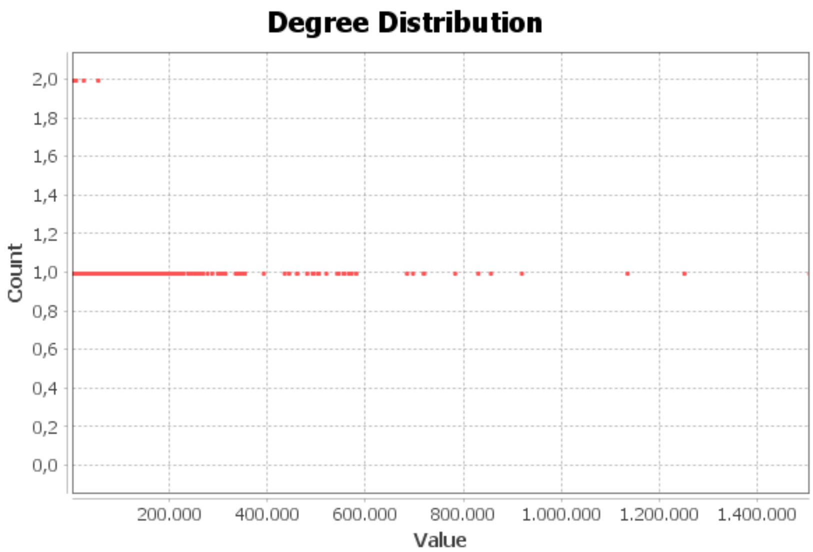
At first glance we notice that there are two nodes – shroud and xQcOW – with the biggest size and darkest shade of green, which indicates that they have the highest degree among all nodes in the graph. When a node has a high degree, it means that it has a lot of relations (edges) or, in our case, that a streamer has common viewership with a lot of other streamers.

Shroud and xQcOW have the highest degrees, therefore they share viewers with many different streamers as well as each other, which is to be expected since they have two of the biggest channels on the platform (shroud with 9 million followers and xQcOW with 7.7 million followers). Plus, their content is almost exclusively gaming which is the most prevalent type of streaming on the platform and they are both professional players in two very popular games – shroud in Counter-Strike: Global Offensive (CS:GO) and xQcOW in Overwatch.

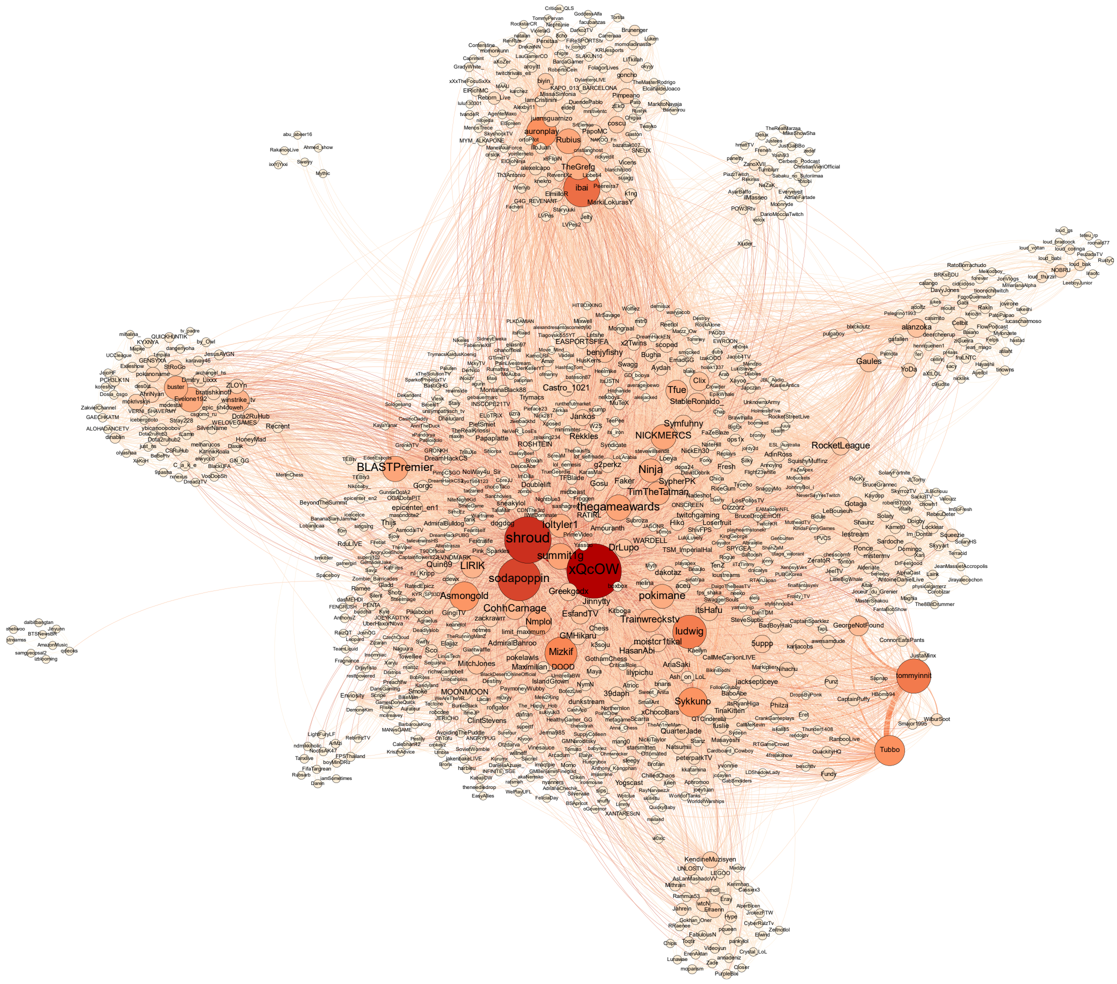
Also worth mentioning are the channels BLASTPremier, Ninja, thegameawards, sodapoppin and pokimane which accumulate very high degrees and common viewership on the platform.

Next, we calculate the **weighted degree** of the nodes. The weighted degree of a node is again based on the number of relations (edges) of said node – just like the degree – but it is ponderated by the weight of each edge.

Using Gephi we see that the average weighted degree is 66705,224 and the weighted degree distribution is as shown in the diagram below:



Now we can provide with a visualization of the weighted degree measure of our graph, with the size and color of each node to be indicating its weighted degree value:



Because the weighed degree factors in the weights of the edges i.e. the number of common viewers shared between two streamers, we can see differences in comparison to the degree visualization.

The node with the biggest weighted degree is inarguably xQcOW as it is the node with the largest size and darkest shade of orange. This means that xQcOW not only shares viewers with the most other streamers, as we already discovered with the help of the degree report, but he also shares the biggest number of viewers with other streamers. This comes to no surprise as xQcOW was the most watched streamer of 2020[1] and he also likes to collaborate with a variety of streamers.

As a close second come shroud and sodapoppin, who are two of the biggest streamers on Twitch and can accumulate tens of thousands of concurrent viewers and millions of total views in each stream.

What's interesting to point out when comparing the weighted degree visualization with the degree visualization is the emergence of nodes which have a great difference between degree and weighted degree values and saw a great boost when we factored in the weights of their edges. Those nodes are ibai, TommyInnit and Tubbo.

[1] https://www.theloadout.com/twitch/most-watched-streamer-2020-xqc

Ibai is a Spanish streamer and esports announcer with a very large Spanish-speaking audience who gained approximately 4.11 million followers in 2020. So even though he might not share viewers with the most streamers since the predominant language in streaming is English, he has still solidified his place as one of the biggest streamers on the platform.

TommyInnit and Tubbo are two British streamers who mostly stream Minecraft gameplay and amass up to hundreds of thousands of concurrent viewers in each stream. They often play and stream together so they have a very large amount of shared viewers and they both have millions of followers. They didn't have the highest degrees as shown from the degree graph, but when we factor in the weights of their edges they gain a very prominent position in the graph. That is because they mainly stream Minecraft and not too many other popular games, as opposed to other big streamers. Therefore, although they have a large audience, their audience is mainly people who play or just enjoy watching Minecraft.

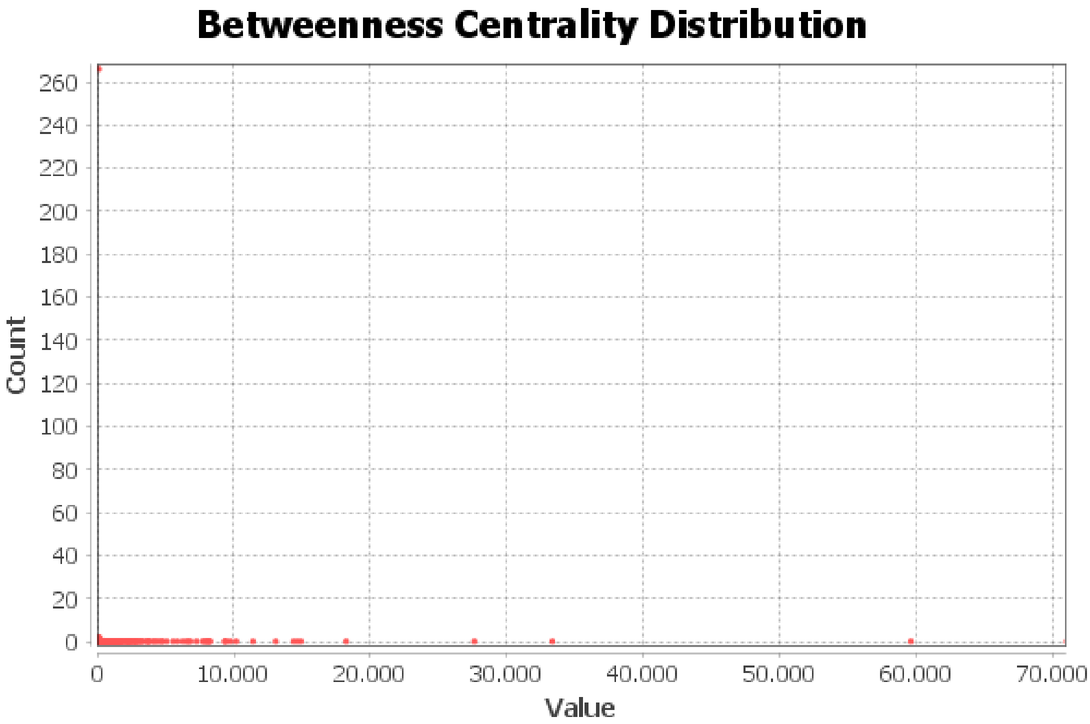
On the opposite side of the spectrum, we notice a certain node that not only didn't grow in size when we took into account the weights, but became smaller instead. The node we are referring to is RocketLeague, which is a channel dedicated to the game Rocket League and exclusively streams Rocket League gameplay. The reason why its weighted degree is smaller than its degree could be that if a viewer doesn't like Rocket League, they will probably not be interested in the channel as it doesn't offer anything else. To put it simply, viewers might discover the channel from a different streamer who plays the game but may not want to watch hours of professional Rocket League gameplay, unless they are a dedicated player of the game.

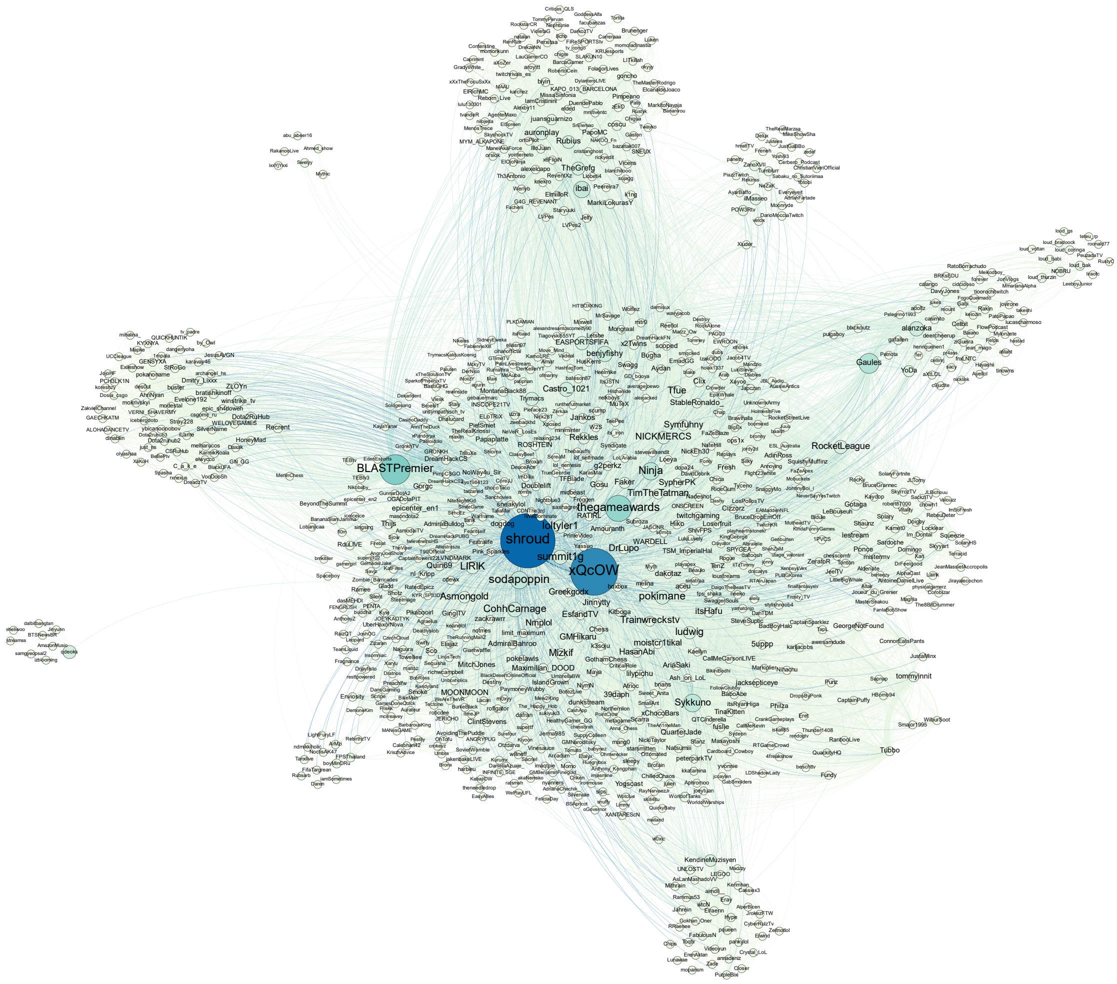
Now that we've discussed about degree measures, it's time we took a look at the centrality measures of our network, more specifically the betweenness centrality, closeness centrality and eigenvector centrality.

3. 3. Centrality measures

The **betweenness centrality** measures all the shortest paths between all pairs of nodes of the network and then counts how many times a node is on a shortest path between two others. In short, betweenness centrality is the answer to the question: how many pairs of nodes would have to go through you in order to reach one another in the minimum number of hops? A node with a high betweenness centrality is most likely to be traversed many times when we want to travel down different paths on the graph, therefore it plays an important role in connecting other nodes. Naturally, if we remove this type of nodes, there is a high probability that our graph will be cut into multiple unconnected components.

The betweenness centrality distribution diagram and visualization are shown next:

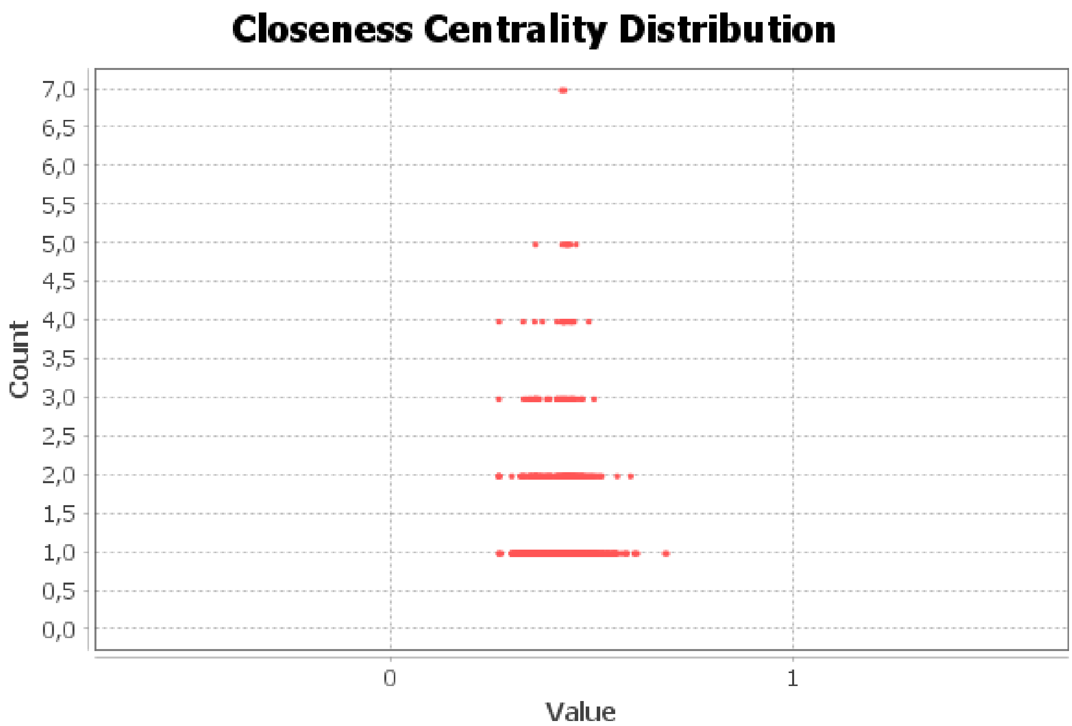


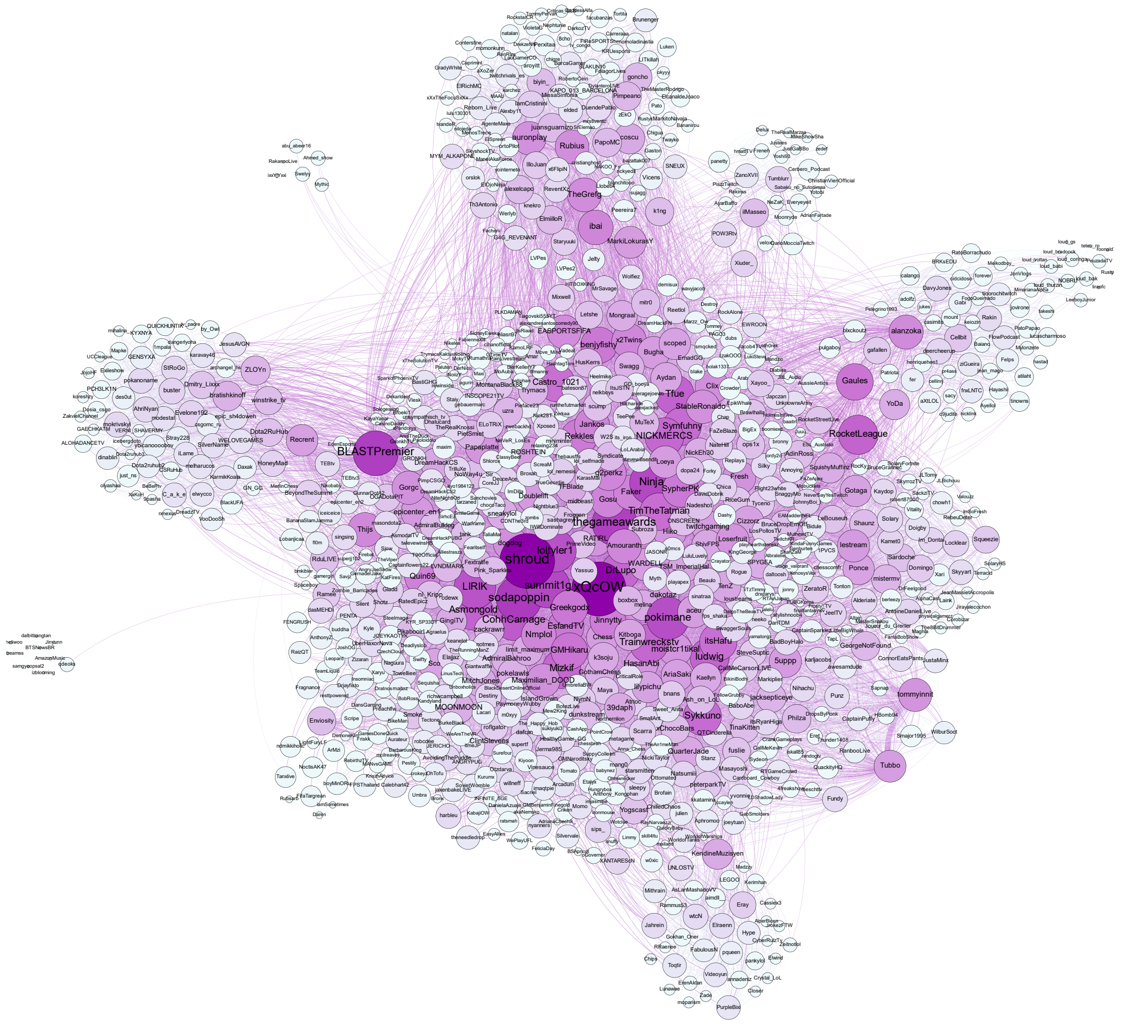


What immediately stands out in our new graph is the node of shroud as it has the biggest size and darkest shade. What this tells us is that shroud plays a vital role in connecting the community of Twitch, as he has links to many different sub-communities and it is almost impossible not to come across his channel while navigating Twitch. This could be due to a combination of factors, from the engaging gameplay he provides to the variety of games he streams and people he streams with. Needless to say, this is a great feat for a streamer and it goes to show why he is one of the most influential streamers on the platform.

Next, we define **closeness centrality** as the measure that shows how close a node is to the center of the network. We calculate it by summing the distances from the node to all others and if the result is a relatively small number, it means the node has high closeness centrality and therefore can easily reach other nodes.

Using Gephi's tools, we produce the closeness centrality distribution diagram and we visualize the results in a new graph.





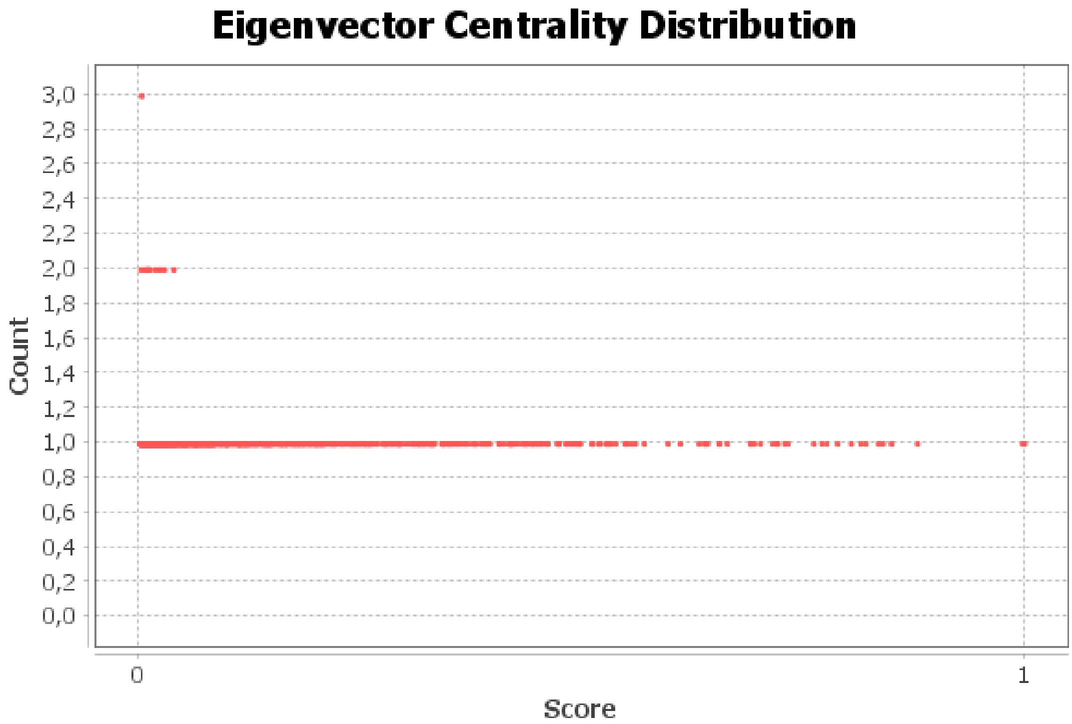
As expected, the nodes with the highest closeness centrality are shroud and xQcOW, closely followed by BLASTPremier which is a channel dedicated to Counter-Strike: Global Offensive (CS:GO) professional esports tournaments. While looking at the distribution diagram we can point out that the majority of the nodes have a low closeness centrality and only very few have achieved higher scores, which makes sense considering that the core of the network can only consist a limited amount of very important nodes – that is precisely what makes it the core.

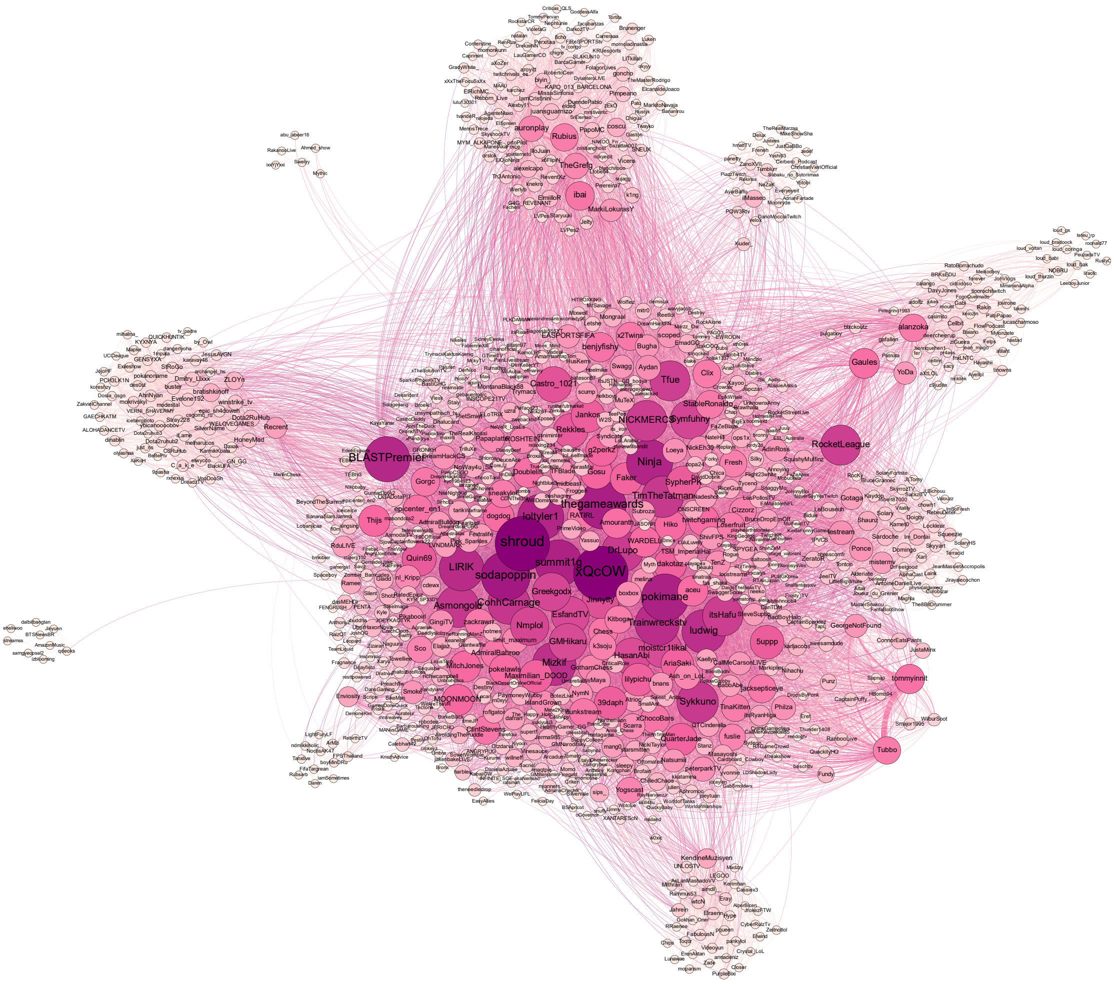
In addition, **eigenvector centrality** is used to measure the level of influence of a node within a network based on how much it is connected to other important nodes. Each node will be given a score: the higher the score the greater the level of influence within the network. This score is relative to the number of connections a node will have to other nodes, but it also awards a number of points proportional to the eigenvector centrality scores of the neighbors. Therefore, connections to high-scoring eigenvector centrality nodes contribute more to the score of the node than equal connections to low-scoring nodes.

To put this into context, a node with a high degree score (i.e. many connections) may only have a relatively low eigenvector centrality score because many of those connections are with similarly low-scored nodes. Also, a node may have a high betweenness score (indicating it connects disparate parts of a network) but a low eigenvector centrality score because it is still distant from the center of the network.

In our case, we will use eigenvector centrality to identify which streamers have a wide-reaching influence within the entire Twitch community.

Of course, we make use of Gephi's tools and we are provided with the eigenvector centrality distribution diagram and the corresponding graph for its visualization. The number of iterations used to calculate the eigenvector centrality was 100.





As we see clearly from the distribution diagram, the vast majority of streamers do not have a high eigenvector centrality and the biggest scores are only achieved by a selective few.

Those with big scores are visible on our new graph and are – not surprisingly – shroud and xQcOW. Their influence in the streaming community is indisputable and the number of connections they have amassed is truly impressive.

Continuing our network analysis, we shall examine the clustering effects in our network; that includes the average clustering coefficient, number of triangles and clustering coefficient distribution.

3. 4. Clustering effects in the network

Before we calculate the average clustering coefficient and draw our conclusions, we must first discuss about a very basic principle in graph theory called triadic closure.

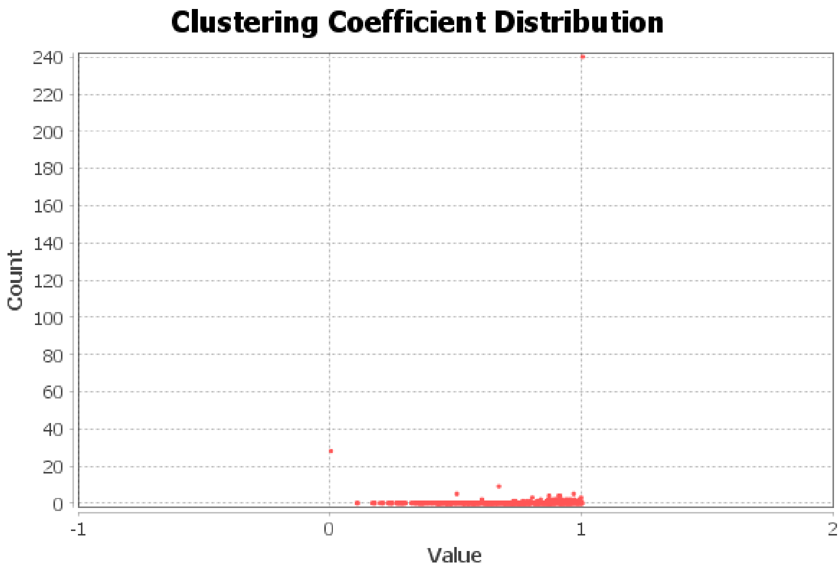
Triadic closure is the property among three nodes A, B, and C (which in our case represent streamers), that if the connections A-B and B-C exist, there is an increased likelihood that a new connection A-C will be formed at some point in the future. Triadic closure can be used to understand and predict how a network grows and evolves over time.

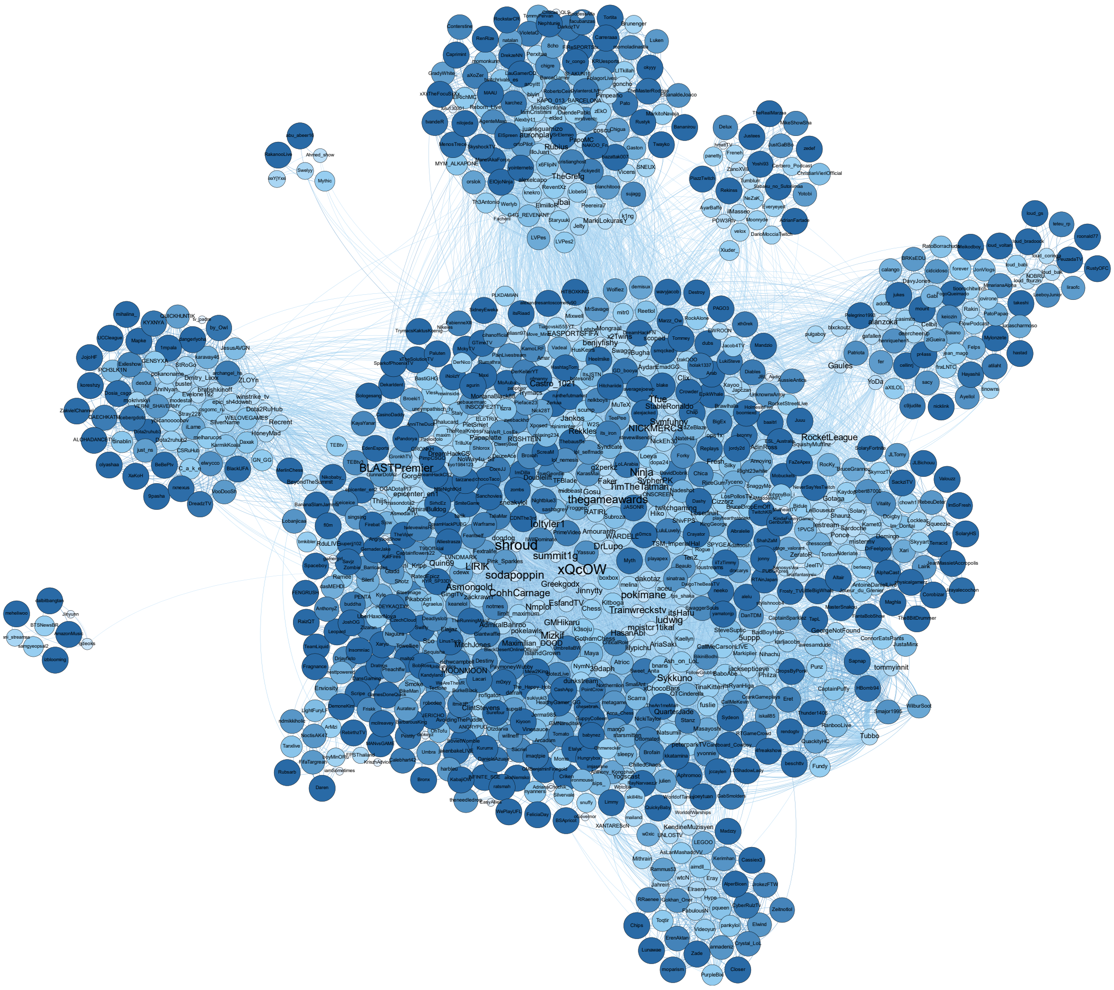
Because of how important the role of triadic closure is in networks, several network measures have been formulated in order to capture its prevalence – one of which is the clustering coefficient.

The **clustering coefficient** of a node A is defined as the probability that two randomly selected nodes that are connected to A are also connected to each other. In other words, it is the fraction of pairs of A's neighbors that are connected to each other by edges.

In general, the clustering coefficient of a node ranges from 0 (when none of the node's neighbors are connected to each other) to 1 (when all of the node's neighbors are connected to each other), and the more strongly triadic closure is operating in the neighborhood of the node, the higher the clustering coefficient tends to be.

In our network, the average clustering coefficient is 0,782 and the total of triangles is 297248, which give us important insight on how connected all the nodes are. A clustering coefficient value of 0,782 indicates that there is an almost 78% chance that two neighbors of a node are linked, which is an impressive percentage and further goes to show just how big of a role connections play in the Twitch community.
The clustering coefficient distribution diagram and visualization graph are shown below.





As we see both from the diagram and the graph, there are a lot of nodes with a clustering coefficient equal or very near to 1 and they are all scattered in many different communities. What is worth mentioning though is that the streamers who would appear more prevalent in all the previous graphs we made, are now barely noticeable because of their low clustering coefficient scores. This phenomenon can be attributed to the fact that, even though they may have many connections (as we saw from the centrality measures), not all of their neighbors will be connected to each other as they come from different and often very distant parts of the network. On the other hand, streamers who didn't achieve significantly big centrality scores have now big clustering coefficient scores instead, because they tend to have a more closely connected community of people that they share viewers with, as opposed to the very popular streamers that share viewers with a lot of different streamers.

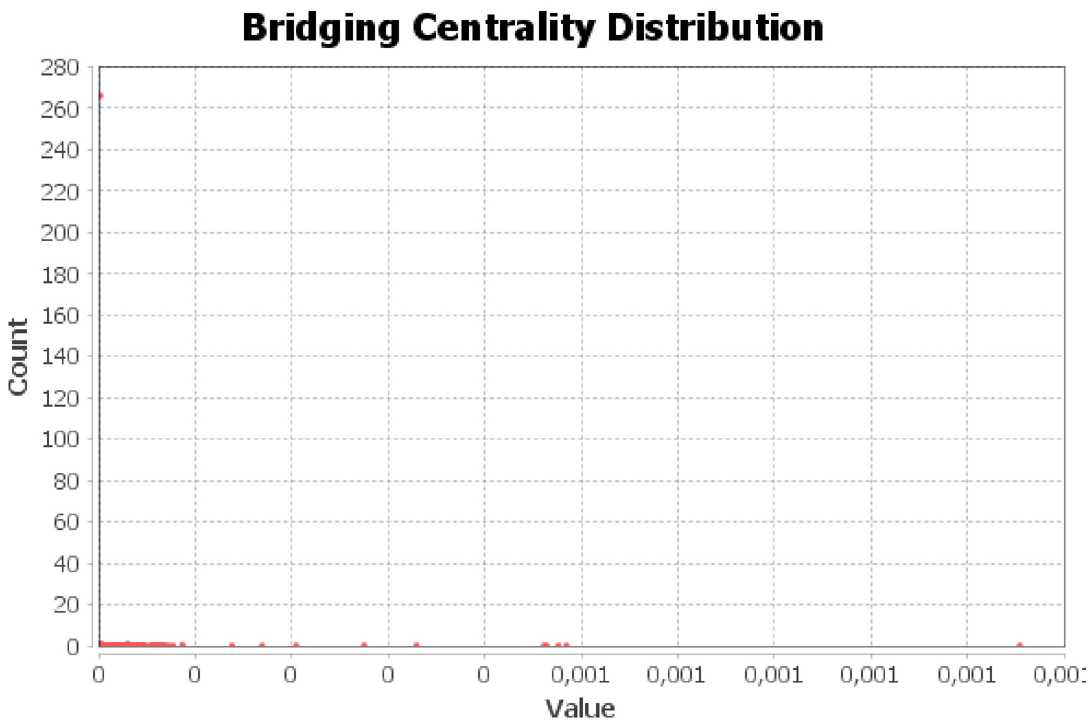
After exploring the clustering effects in our network, we move on to locating possible bridges.

3. 5. Bridges

We define a **bridge** as the edge joining two nodes A and B in a graph which, if deleted, would cause A and B to be part of two different components and disconnect the network.

Bridges are in fact rare to find in real social networks, because usually two nodes will be connected in more than one ways, most likely through longer paths that traverse other nodes.

In order to locate any possible bridges in our network, we installed the Bridging Centrality Plugin that Gephi offers and calculated the bridging centrality. The bridging centrality distribution diagram and visualization are shown below.



As we expected, there are almost no bridges in the network, with the exception of the channel EAMaddenNFL which has a bridging centrality value of approximately 0,001 which is the highest calculated bridging centrality in our entire network. EAMaddenNFL is connected with only two other nodes – MutheadTV and NICKMERCS – and serves as a bridge between them. If EAMaddenNFL was removed from the network, MutheadTV and NICKMERCS would have no other way of connecting to one another.

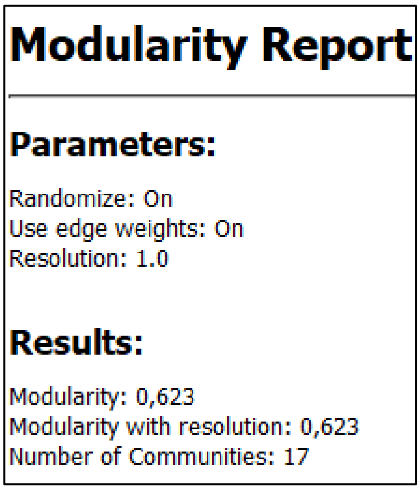
3. 6. Communities

Last but certainly not least, we will analyze the existence and role of communities in our network through the application of the modularity statistic.

In network science we define a **community** as a group of nodes that have a higher likelihood of connecting to each other than to nodes from other communities. The modularity statistic will assess the number of distinct groupings within our network and therefore give us insights on the communities that have been formed.

The algorithm used for calculating the modularity is “Fast unfolding of communities in large networks” by Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte and Etienne Lefebvre, which is a heuristic algorithm based on modularity optimization and detects communities in very little time and with great accuracy. The 17 communities that were detected reflect the communities that we intuitively knew of by examining the data or simply by being a Twitch user, so a search for a different algorithm was not needed.

Our network has a modularity value of 0,623 which is quite high and indicates that there are dense connections between the nodes within communities but sparse connections between nodes in different communities.



Modularity Report

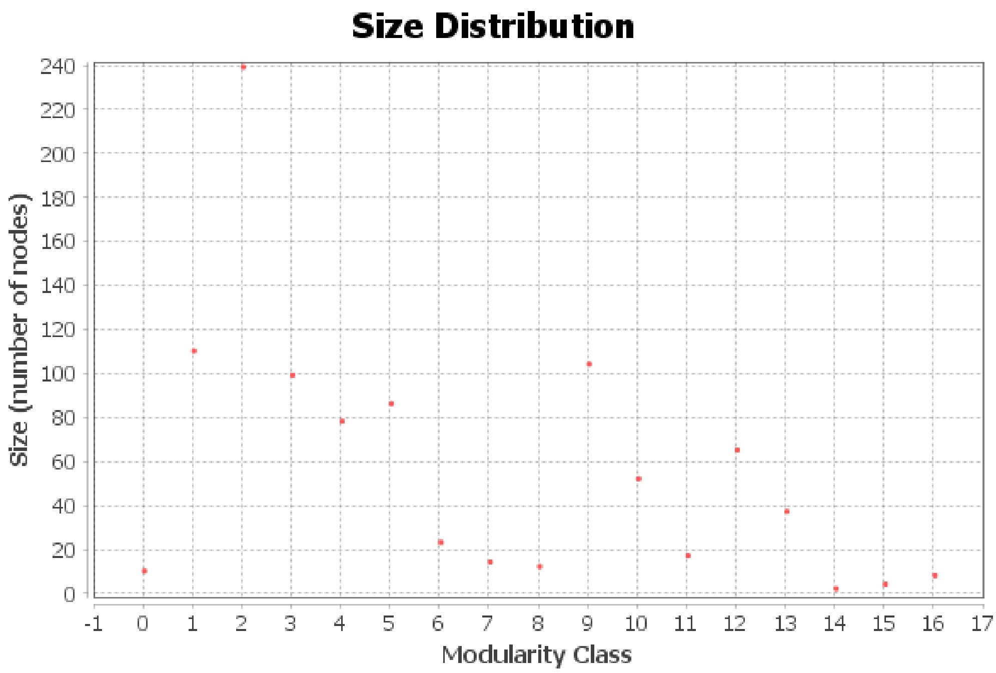
Parameters:

Randomize: On
Use edge weights: On
Resolution: 1.0

Results:

Modularity: 0,623
Modularity with resolution: 0,623
Number of Communities: 17

The size distribution diagram below shows the number of nodes found in each modularity class (community), with the 2nd class accumulating the most nodes out of all of them.

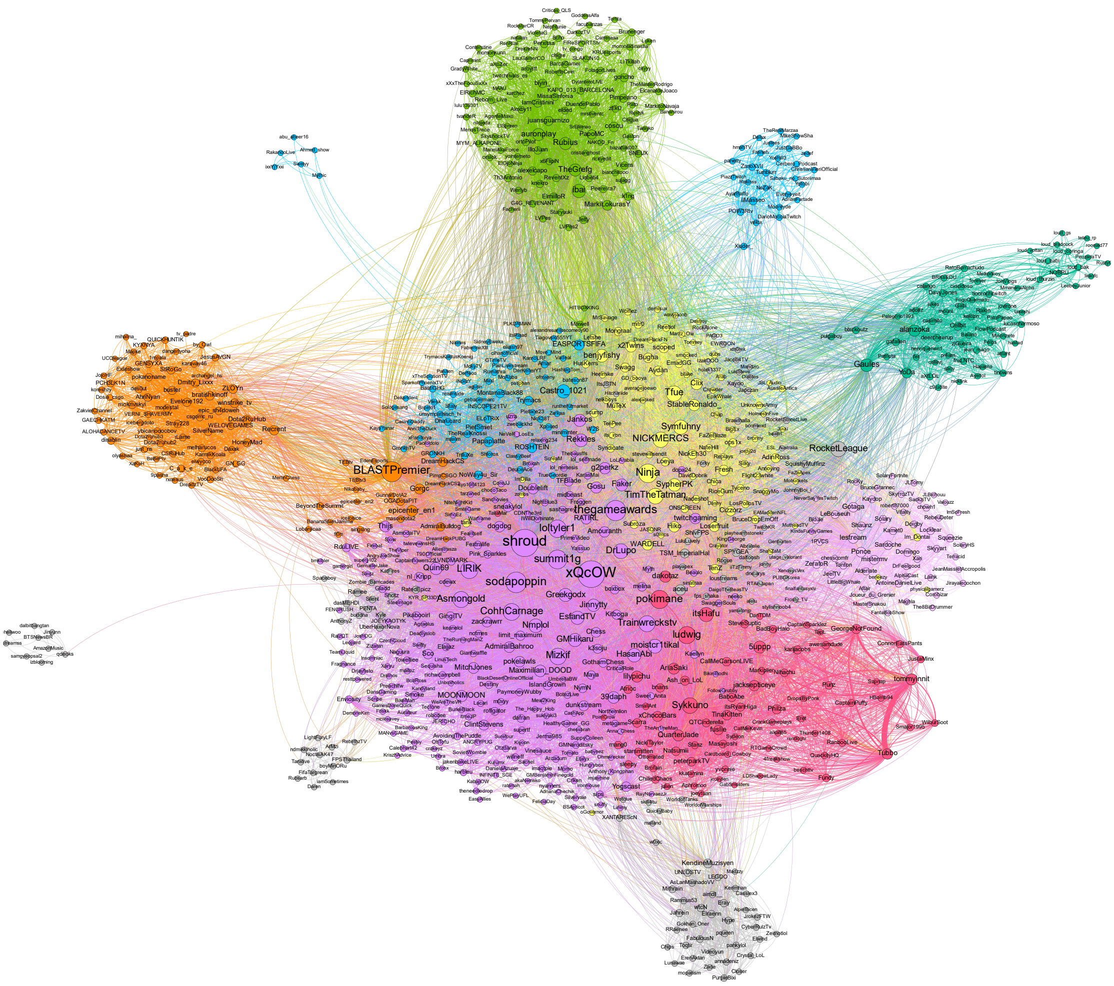


We can also view the percentage of total nodes each modularity class contains and it is very apparent that the 2nd class is the biggest one in the network, but there are also 6 other classes with significant percentages that are worth looking into.

2	(24,77%)
1	(11,46%)
9	(10,84%)
3	(10,32%)
5	(8,98%)
4	(8,15%)
12	(6,81%)
10	(5,47%)
13	(3,92%)
6	(2,48%)
11	(1,86%)
7	(1,55%)
8	(1,34%)
0	(1,14%)
16	(0,93%)

Communities created by Gephi

Before we start diving deeper into the communities, we create a visualization to help us understand the composition of our network. The node sizes scale according to the node's degree.



It is only natural that we start our analysis with the largest community in the graph – the one with modularity class 2 and the color purple. This community's biggest streamers are shroud and xQcOW and the community mostly consists of English-speaking "variety" streamers. "Variety" is a term given by Twitch itself and it describes content creators who do not have a single defining game or category that can label their content. It is understandable that the Variety community would take the most prominent place in the network, as almost every other community shares viewership with it.

But it is important to mention that this isn't the only reason why Variety classifies as a community, because large streamers in Variety very often interact with each other by playing games, competing in tournaments, making content or sometimes living together. The Variety community has garnered an enormous audience over the years and it's fair to say that it is the pillar of Twitch.

The next community we will discuss is the one with modularity class 1 and the color light green, which contains the 11,46% of all nodes in the graph. Upon a closer look, we find out that this community consists of Spanish-speaking streamers, the biggest ones being auronplay, Rubius, TheGrefg and ibai. Not everyone in the community makes the same type of content; for example auronplay is a comedian who provides satirical commentary on current affairs, while Rubius is a gamer. What they all have in common is that they stream in Spanish and that's precisely what sets them apart from the rest of the network and at the same time makes the connections between the streamers even stronger.

The third community to analyze is the one with modularity class 9 and the color blue. The biggest channels in the community are Castro_1021 and EASPORTSFIFA. This community's common denominator is the football video game FIFA and streamers who are dedicated players will be found in this community, such as Castro_1021 who has been playing FIFA since 2013.

The fourth community we will look into has modularity class 3, the color yellow and includes 10,32% of the total nodes in the network. The biggest streamers – in terms of degree – of the community are Ninja and NICKMERCS. Streamers in this community are known for streaming popular first person shooter games such as Fortnite and Call of Duty.

Our next community is the one with modularity class 5 and the color orange in the graph. The biggest streamer in it is BLASTPremier which is a Counter-Strike: Global Offensive (CS:GO) professional esports league launched in 2020. Just by knowing this, we can safely assume that the community consists of CS:GO players. But when looking more closely, we notice that although BLASTPremier is part of the community, its node is fairly away from the main cluster of orange nodes. In order to understand why this happened, we take a look at some nodes within the cluster (e.g. Dota2RuHub, winstrike_tv, evelone192) and come to the realization that all these streamers are Russian and stream in Russian. CS:GO is a game with a very big Russian fanbase and it isn't surprising that a community of Russian CS:GO streamers would form. BLASTPremier's streams are in English and that explains the distance from the rest of the community on the graph.

The fifth community to examine is the one with modularity class 4 and the color pink. We will name this community OfflineTV & Friends, a term coined by fans of OfflineTV, which is a group of content creators including Scarra, Pokimane, lilypichu, yvonnie, and Brofain. The group often plays games and makes content with other streamers in their friend group and so their viewers greatly overlap. The biggest streamers of the community are Pokimane – an OfflineTV member – and Sykkuno – a close friend to the OfflineTV members. Something interesting to note is that streamers who haven't interacted very often with OfflineTV & Friends appear to be in the same community. Such streamers are TommyInnit, Tubbo and GeorgeNotFound, who are all Minecraft streamers and mainly interact with each other. But the reason for their inclusion in the OfflineTV & Friends community could be lying in the date our data was gathered. As mentioned in the introduction, the data was collected during the week of December 6-12th 2020. Around that time, the video game Among Us was at its peak in terms of viewership and almost all streamers would give it a try. On a few occasions, members of Offline TV & Friends played the game with Minecraft streamers and so they naturally began sharing a significant amount of common viewers. It is not clear to say whether this overlap still stands today, as there haven't been many collaborations between the groups but it is very interesting to see how our network reflects real life events.

What is also worth mentioning is the viewership overlap between TommyInnit and Tubbo. As seen from the graph, their nodes have a very thick edge connecting them, meaning that they had a lot of common viewers during the time period the data was collected. Looking at the data we find out that they have the most common viewers – a staggering count of 186677; the most in our network, to be exact. This phenomenon can be attributed to the fact that TommyInnit and Tubbo are actually friends in real life and stream their gameplay together more often than not. In fact, most viewers would argue that you cannot be a fan of one without being a fan of the other – a claim that is definitely supported by our analysis.

The sixth and last community to be analyzed has modularity class of 12 and the color dark green. This community's biggest streamers are Gaules and alanzoka and what separates this community from the rest is that all its streamers stream in Portuguese, therefore their target audience is a tightly knit Portuguese-speaking community.

In the matrix below we can have a clearer view of the biggest communities in our network and their characteristics:

Color in graph	Community name	Biggest streamers
	Variety	shroud, xQcOW
	Spanish	auronplay, Rubius, TheGrefg, ibai
	FIFA	Castro_1021, EASPORTSFIFA
	First person shooter games	Ninja, NICKMERCS
	CS:GO(English/Russian)	BLASTPremier
	OfflineTV & Friends	Pokimane, Sykkuno
	Portuguese	Gaules, alanzoka

4. Conclusion

After exploring all these different aspects of our network with the help of various network analysis metrics, we have gained greater insight into our network and the Twitch community in general. We were able to get information that we couldn't have gotten from just the raw data, we have discovered patterns and interesting connections and, all in all, we definitely have a new understanding of the world named Twitch.