Twitch Social Networks
Data Visualization Project

Darian-Florian Voda
FH Kärnten - Applied Data Science
3^{rd} semester, Engineering & IT Department
Data Visualization II

February 16, 2024

1 Abstract

In this project, a Data Visualization approach was based on Twitch Social Networks dataset, exploring the communities, linkage and different features of the twitch streamers within Twitch platform. The dataset contains numerous communities based on streaming language, but only one was used (the streamers which streamed in english [EN] language). This subset of the dataset contains 7,126 nodes and 35,324 edges. The main tool used for exploring new data visualizations was Gephi, an application based for network visualizations. After many trials and research, several conclusions popped up into beautiful visualizations using Gephi regarding twitch streamers that are having a high number of views and their stream being labeled as mature content (containing swears, horror games, shooting games, etc.). Also, some visualization can show the difference between being a Twitch partner and not, based on the number of views.

2 About the Dataset

These Twitch user-user networks of gamers who stream in a particular language are the datasets utilized for node categorization and transfer learning. The users themselves are the nodes, and the connections between them are their mutual friendships. Based on geography, streaming patterns, and games played and loved, vertex features are retrieved. Because datasets have the same set of node characteristics, transfer learning between networks is feasible. In May 2018, these social networks were gathered. Binary node classification is the supervised task associated with these networks; the task involves predicting whether a streamer utilizes explicit language.

Dataset statistics						
	DE	EN	ES	FR	PT	RU
Nodes	9,498	7,126	4,648	6,549	1,912	4,385
Edges	153,138	35,324	59,382	112,666	31,299	37,304
Density	0.003	0.002	0.006	0.005	0.017	0.004
Transitvity	0.047	0.042	0.084	0.054	0.131	0.049

3 What is Twitch?

The main purpose of the well-known live streaming website Twitch.tv is to allow users to broadcast and watch live video game streams. When it was first released in 2011, the gaming community took to it right away. Users who broadcast their live gaming, commentary, and interactions with viewers in real time are referred to as streamers.

Twitch's content has grown over time to encompass a variety of areas, including music, discussion programs, visual arts, and more, in addition to gaming. Through chat on the platform, viewers may communicate with streamers and ask questions, provide comments, and participate in community discussions. Twitch has developed into a venue for esports competitions, offering a place for broadcasting competitive video game tournaments to a large viewership.

Apart from live broadcasting, Twitch provides features like channel subscriptions, allowing users to pay a monthly fee to support their preferred streamers. Beyond just gamers, the platform's user base is diversified and has grown to be a key element of online entertainment.

3.1 Twitch Partner

For streamers, becoming a Twitch Partner is a big accomplishment because it opens up new features and revenue-generating options. The general procedures to become a Twitch Partner are as follows:

- Meet Eligibility Requirements:
  - Broadcast regularly: You should be consistently streaming on Twitch.
  - Develop a sizeable audience: While there isn't a specific viewer count requirement, having a substantial and engaged viewership is crucial.
  - Adhere to Twitch's community guidelines: Make sure your content complies with Twitch's terms of service and community guidelines.
- Build a Community:
  - Interact with your viewers: Engage with your audience through chat and social media.
  - Create a brand: Develop a unique identity for your channel, including branding elements like logos, overlays, and consistent themes.
- Grow Your Channel:
  - Promote your channel: Utilize social media and other platforms to increase your visibility.
  - Collaborate with other streamers: Networking with other content creators can help you reach new audiences.
- Diversify Your Content:
  - Explore different types of content: While you may start with gaming, consider branching out into other categories that align with your interests and audience.
- Apply for the Twitch Partner Program:
  - Once you meet the eligibility criteria, you can apply for the Twitch Partner Program through your Twitch dashboard. Twitch staff will review your application, considering factors such as concurrent viewership, stream frequency, and content quality.

Once accepted, you'll gain access to features like channel subscriptions, ad revenue sharing, and more.

4 What is Gephi?

Large networks and graphs can be visualized and analyzed using Gephi, an open-source program. It offers an easy-to-use interface for examining and comprehending intricate connections among data sets. Gephi is widely used in many domains where data can be represented as nodes and edges in a graph, including biology, finance, social network analysis, and more.

Key features of Gephi include:

- Graph Visualization: Gephi allows users to create visually appealing representations of networks, making it easier to identify patterns, clusters, and important nodes within the data.
- Data Import: The software supports the import of data from various sources, including spreadsheets, CSV files, and other graph formats, making it versatile for different types of data.
- Analysis Tools: Gephi offers a variety of tools for examining the degree distribution, modularity, clustering coefficient, and other characteristics of networks. Users can better comprehend the structural properties of the networks with the aid of these metrics.
- Layout Algorithms: Gephi includes a variety of layout algorithms that help arrange nodes and edges in the graph, making it easier to interpret and analyze the network structure.
- Filtering and Exploration: Users can apply filters to focus on specific portions of the network, allowing for in-depth exploration and analysis of relevant subgraphs.
- Dynamic Graphs: Gephi supports the visualization of dynamic networks, where changes in the network structure over time can be observed and analyzed.
- Community Detection: The software includes algorithms for detecting communities or clusters within a network, which can reveal groups of nodes that are densely connected to each other.

Gephi is extensively utilized in scholarly investigations, data examination, and diverse sectors to acquire understanding of the connections and trends found in intricate data sets depicted as networks. For professionals working with network data, researchers, analysts, and other professionals, it offers an invaluable tool.

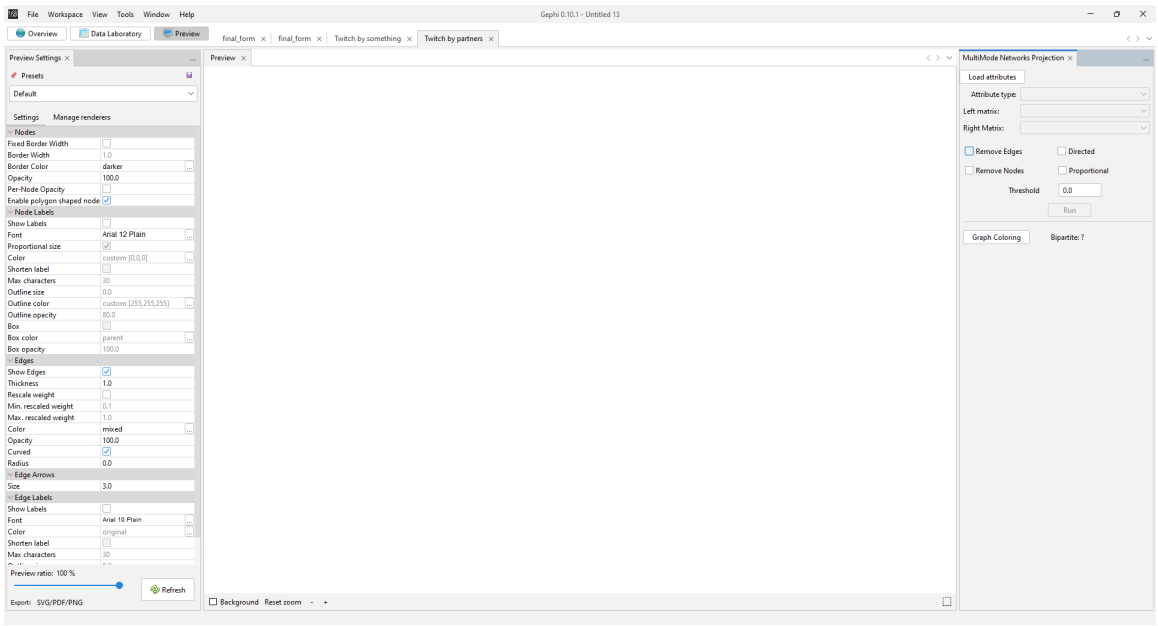

Figure 1: Gephi User Interface

5 Data Visualization on Twitch Social Networks

The main aim of the data visualization task is to find interesting facts regarding Twitch Social Network dataset. This dataset comes in multiple folders based on language streamed, where ENGB was the only one used. Within this folder there are 3 files: two .csv files (one for edges and one for features) and one .json file containing the features.

5.1 Gephi ETL tool

Gephi application comes handy with an ETL tool that allows multiple files to work on the same workspace and merge them together if they have the same dimensions. However, some data preprocessing was needed, since the edges file was not so well recognized by the ETL tool for the merging with the feature files. The features .csv file contains the following attributes:

- id (integer)
- days (integer)
- mature (bool)
- views (integer)
- partner (bool)
- new_id (integer, being the newly reassigned id for the networking visualization task)

The edges .csv file contains the following attributes:

- Source (integer)
- Target (integer)

The main problem represents the fact that Source and Target integers are represented only by the new_id integers, and when merging with the Gephi ETL tool, new IDs are added wrongly.

Thus, for the data preprocessing part, in the features file we had to drop the id column and replace it with the new_id and then merge both files into one Gephi workspace.

The raw result of the ids and their linkage is represented in the next figure.

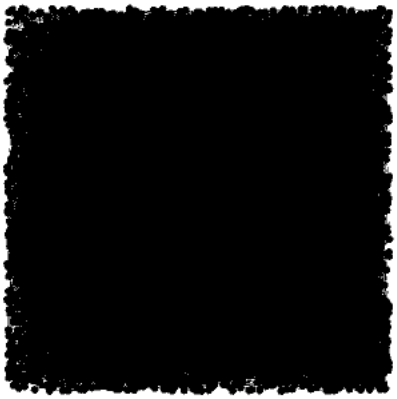

Figure 2: Raw result of Twitch Social Networks

5.2 Statistics with Community Detection

Gephi comes again handy with different features, one which is very important for Data Scientists, such as Statistics. There are several properties that can be achieved using this feature, but for the scope of the project only Community Detection statistics were used.

One of the statistics is called 'Modularity' and uses a specific algorithm written by Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte and Etienne Lefebvre in a paper named "Fast unfolding of communities in large networks". A metric called modularity is used in community detection to assess how well a network has been divided into communities or clusters. A healthy

community structure should have more connections within communities and fewer connections between communities than would be predicted by chance, according to the modularity concept.

The following parameters were used for our dataset in order to run this algorithm:

- Randomize: On
- Use edge weights: On
- Resolution: 1.0

The output of the statistics it is also represented in a visualization and written results:

- Modularity: 0,426
- Modularity with resolution: 0,426
- Number of Communities: 180

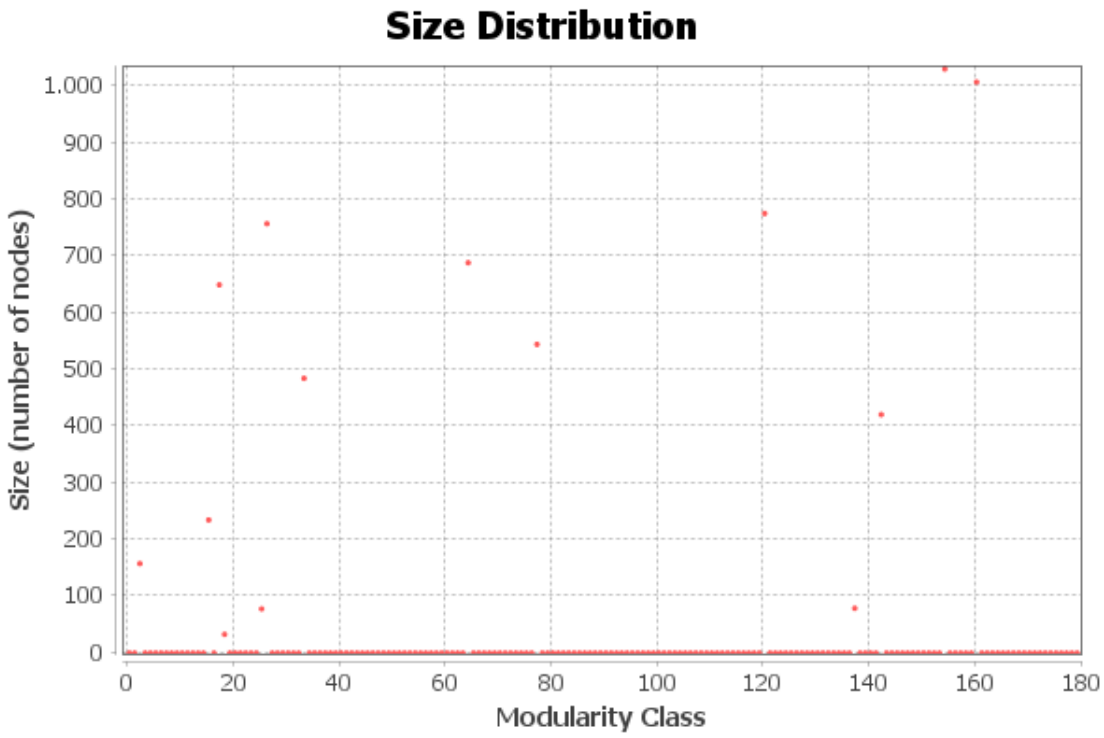

Figure 3: Community Size Distribution using Modularity

Another community statistics using in Gephi was Statistical Inference. This algorithm is written by Lizhi Zhang, Tiago P. Peixoto in "Statistical inference of assortative community structures". Employing statistical techniques to evaluate the existence and importance of assortative patterns within communities in a network is known as statistical inference of assortative community structures. Nodes with similar characteristics tend to connect with each other more frequently than nodes with different characteristics, a phenomenon known as assortativity. Assortative communities, as used in community structures, are those in which nodes in the same community have more connections with one another than with nodes in other communities.

For this algorithm no input parameters were needed, and the results show a significant difference in the number of communities:

- Description Length: 251858,760
- Number of Communities: 834

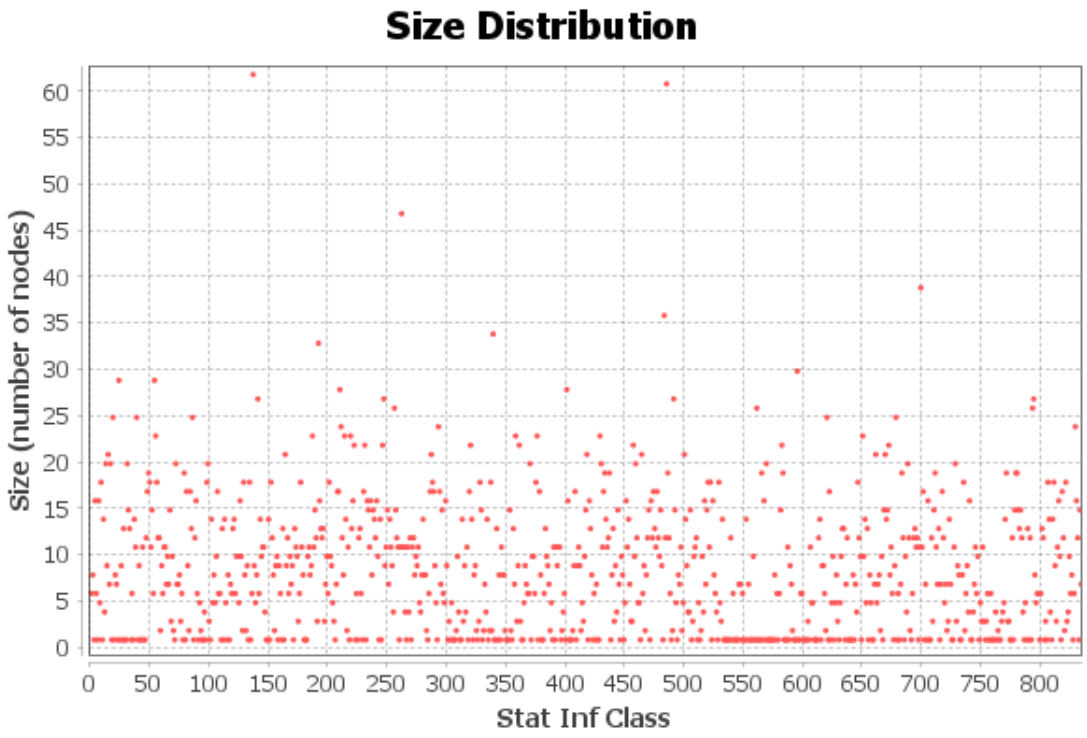

Figure 4: Community Size Distribution using Statistical Inference

5.3 Data Visualization - Ranking by position

Gephi helps the users by providing different plugins for better network visualization. Using specific plugins, such as positioning of the nodes, specific factors could be pointed out.

Before diving deep into the figure, the following preparations of the network were done:

- The nodes were partitioned into two groups by the color: Mature content (red), Not mature content (green)
- The size of the nodes was resized from 1 to 10 based on the number of the days since streaming (1 being small and 10 being big)
- The special plugin regarding the position of the nodes was used based on a ranking on X axis in descending order based on the number of views

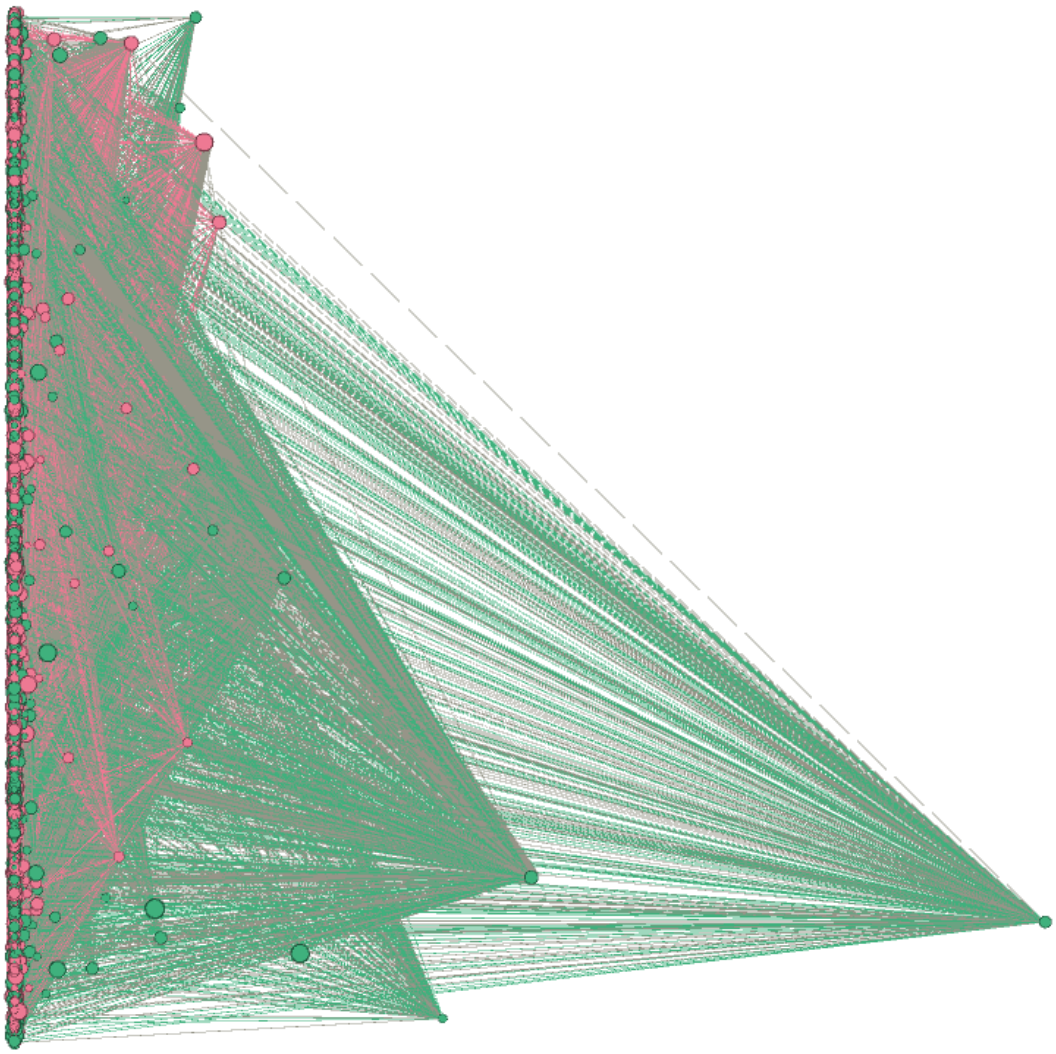

Figure 5: Position by the number of views, mature/not mature and number of days

Gephi comes again handy specific features such as: removing the edges for a better visualization, even though we can conclude from this figure that every big streamer has many friends (since the edges represent the mutual friendship), and also showing the labels on mouse hovering.

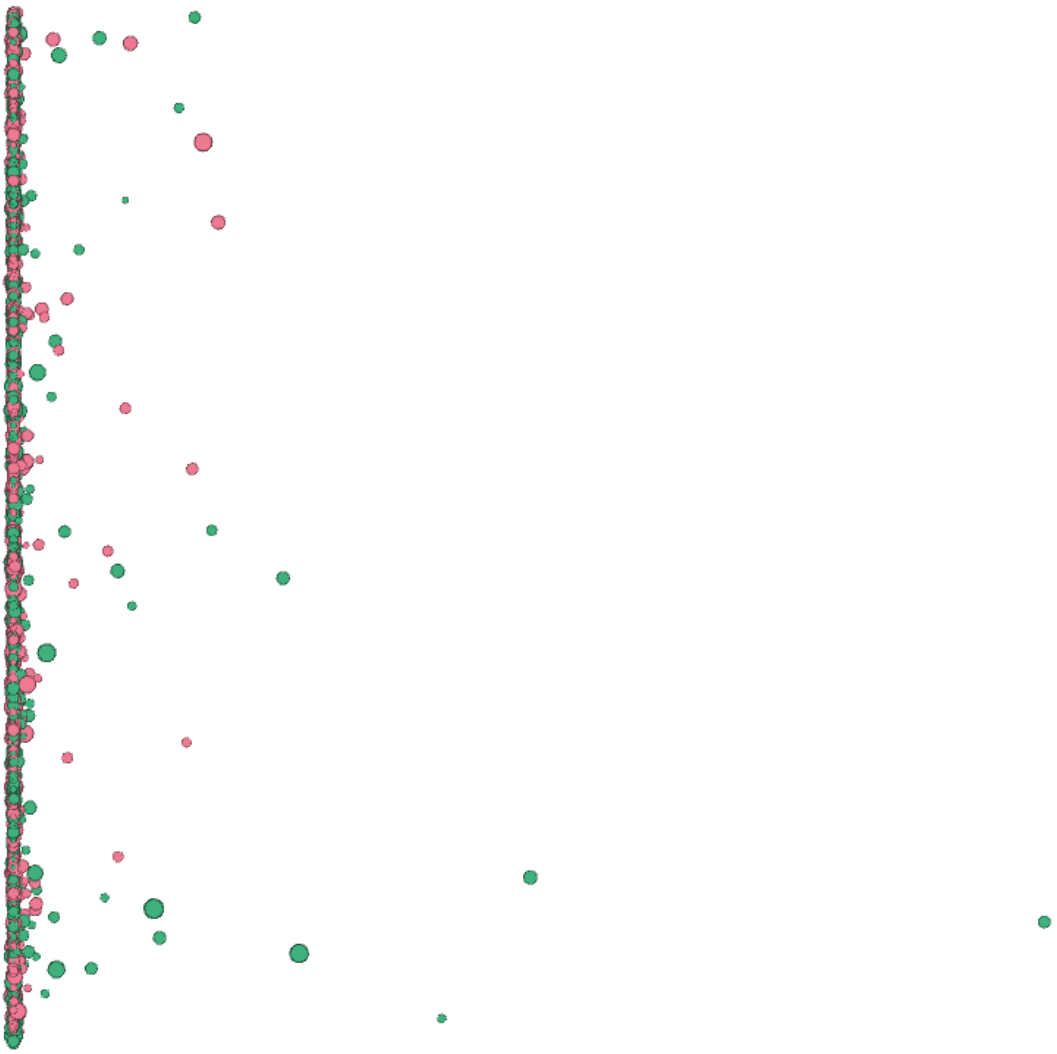

Figure 6: Position by the number of views, mature/not mature and number of days without edges

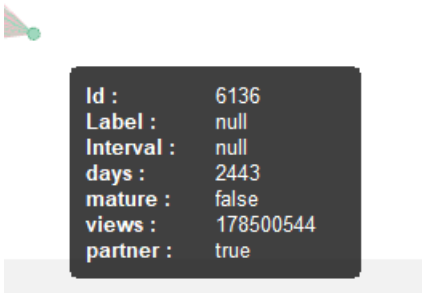

Figure 7: Label of the biggest streamer

Thus, from these Gephi graphs we can conclude easily by visual and quantitative comparison

that the biggest streamer is not streaming mature content, does not have so many days since streaming and **the most of the biggest streamers** do not stream mature content.

5.4 Data Visualization - Exploring friendships

Digging further into Gephi functionalities, one important feature represents the **Layout**. This features helps the users to reposition the nodes and its edges based on specific algorithms in this field of Graph Networking. For this project several layouts were used, and many trials of good representations were tried, as we can see a demonstration of using the whole dataset with the algorithm from Yifan Hu.

Figure 8: Twitch Social Networks based on Partners (color) and views (size)

For a better exploration, the dataset had to be subsampled using Gephi Filters feature and exploring the following communities:

- Big streamers (1.000.000 views+) with their content (Mature - red, Not Mature - green) and number of days (small nodes - few days, big nodes - many days)

- Small streamers (under 1.000.000 views) with their content (Mature - red, Not Mature - green) and number of days (small nodes - few days, big nodes - many days)
- Big streamers (1.000.000 views+) with partnership (Partner - green, Not Partner - red) and number of days (small nodes - few days, big nodes - many days)
- Small streamers (under 1.000.000 views) with partnership (Partner - green, Not Partner - red) and number of days (small nodes - few days, big nodes - many days)

Even if we saw in the visualization based on the positioning that most of the big streamers have mutual friendships, when we filter out the data based on only big streamers, the results surprise, somehow, the expectations. For this result, the algorithm "ForceAtlas 2" was used for a better quality in positioning the nodes and its edges. The central nodes show a bigger number in mutual friendships, whereas the outer nodes show no mutual friendship between any other nodes.

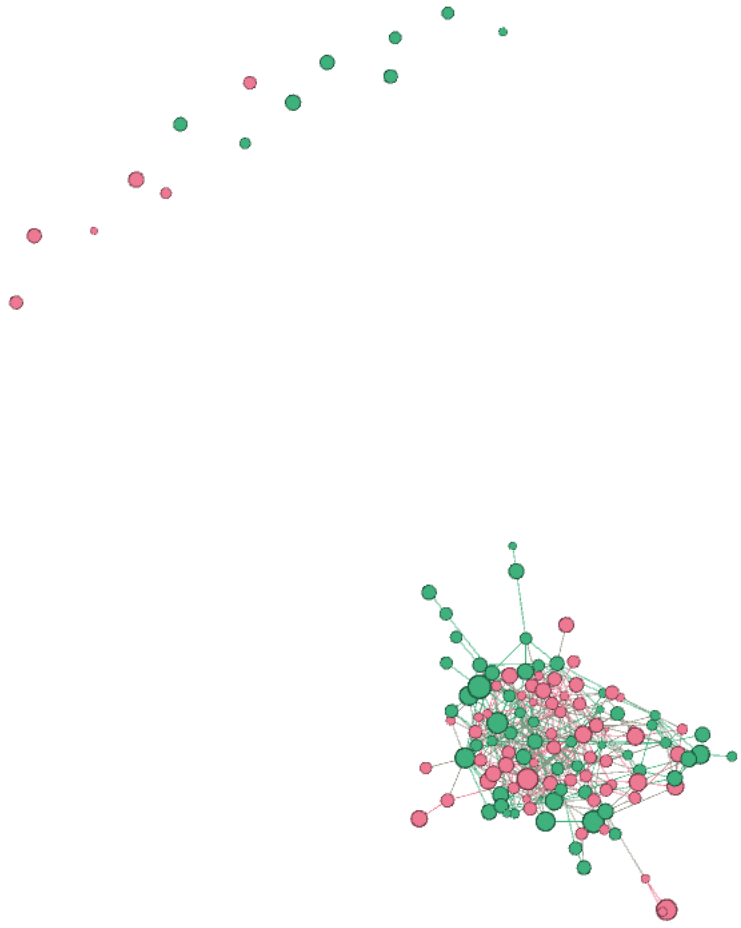

Figure 9: Big streamers based on mature content (color) and number of days (size)

From this figure, we can conclude that not all the big streamers are friendly with other big streamers, and this fact is not based on their streaming content (mature or not).

When it comes to small streamers, the results tend to follow a more inductive way, and due to the bigger number of the resulted nodes, we can easily see that the friendships among the Twitch streamers are not the most important aspect, even though many streamers claim that networking represent a big factor for becoming a big streamer.

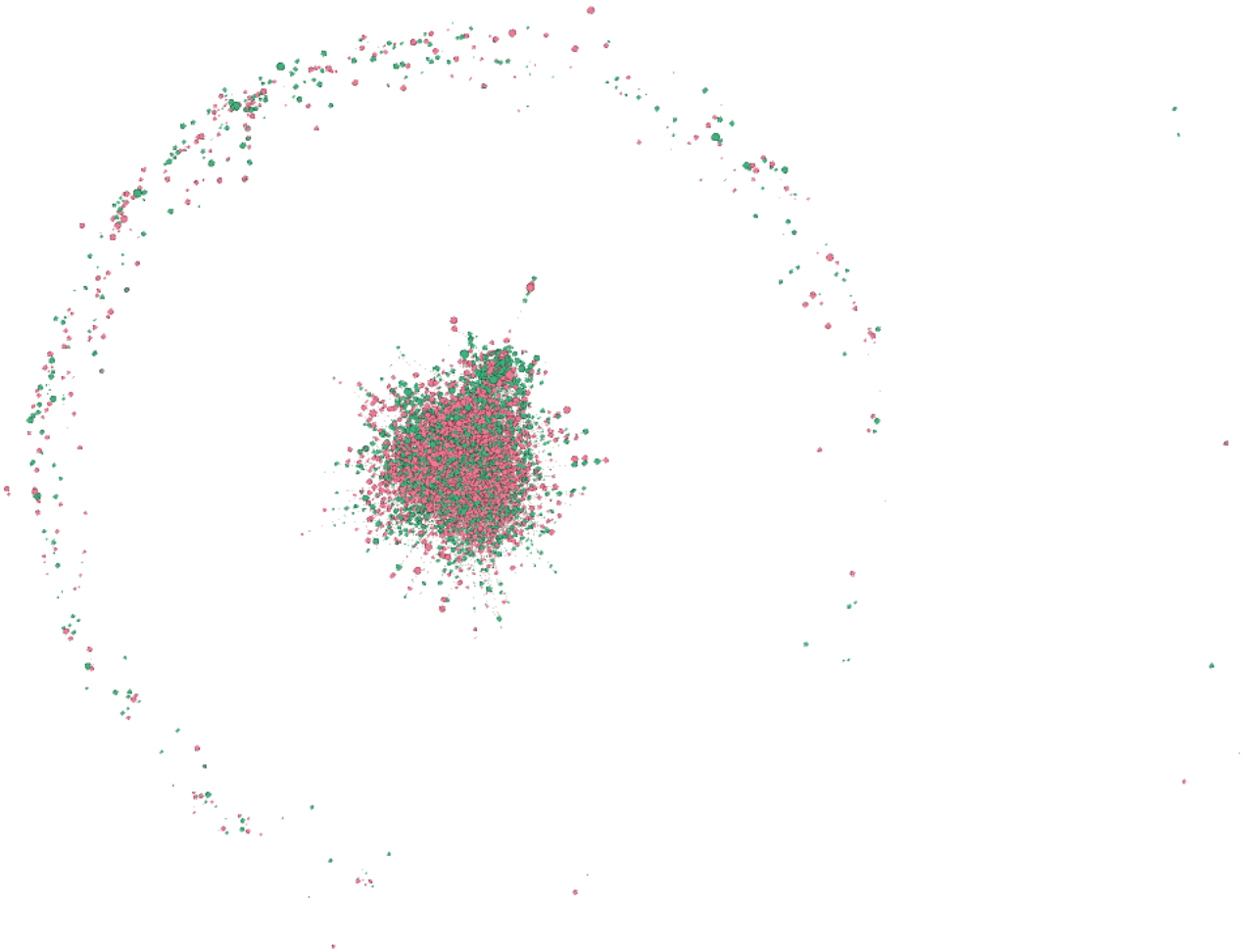

Figure 10: Small streamers based on mature content (color) and number of days (size)

For the last part of the project, which is regarding the partnership of the streamers with respect to other attributes, the results seem to be interesting, since the statistics regarding partnership with Twitch for the streamers is one-handed: most of the streamers (94,61%) are not Twitch partners. In the case of big streamers, the majority is a Twitch partner and also with some streamers without mutual friendships.

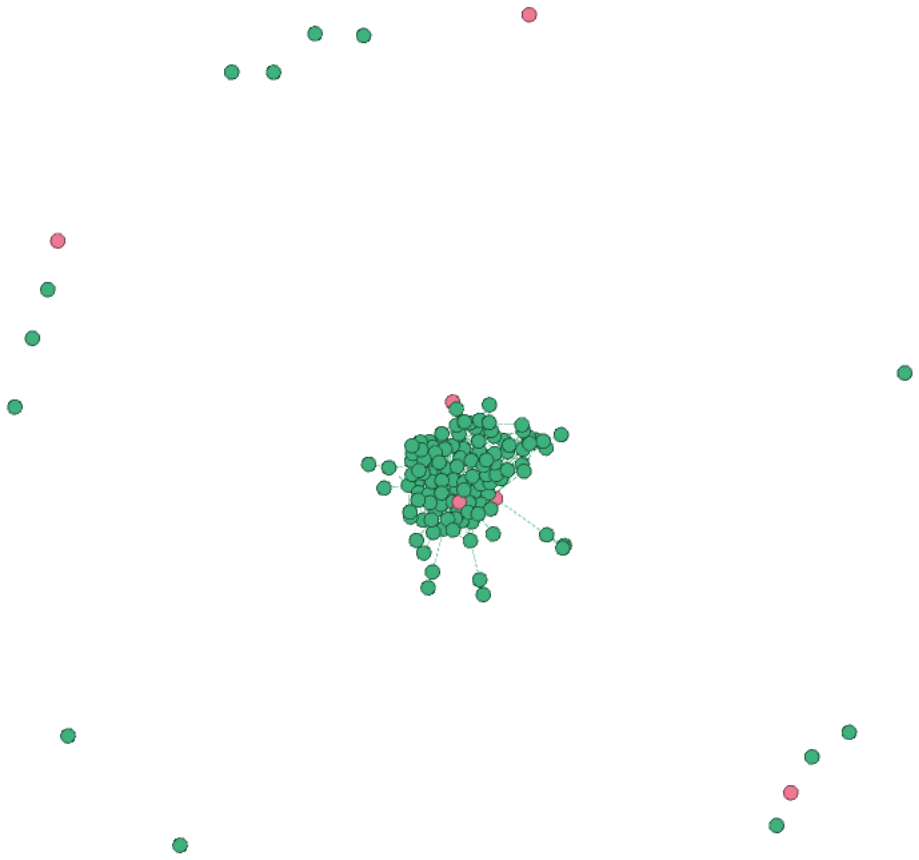

Figure 11: Big streamers based on partnership (color) and number of days (size)

For the smaller streamers, the situation is predictable, since Twitch platform cannot create partnerships with all the Twitch streamers, and, thus, the big majority of it not being a Twitch partner. Also, the mutual friendship of Twitch streamers is not so followed among important facts.

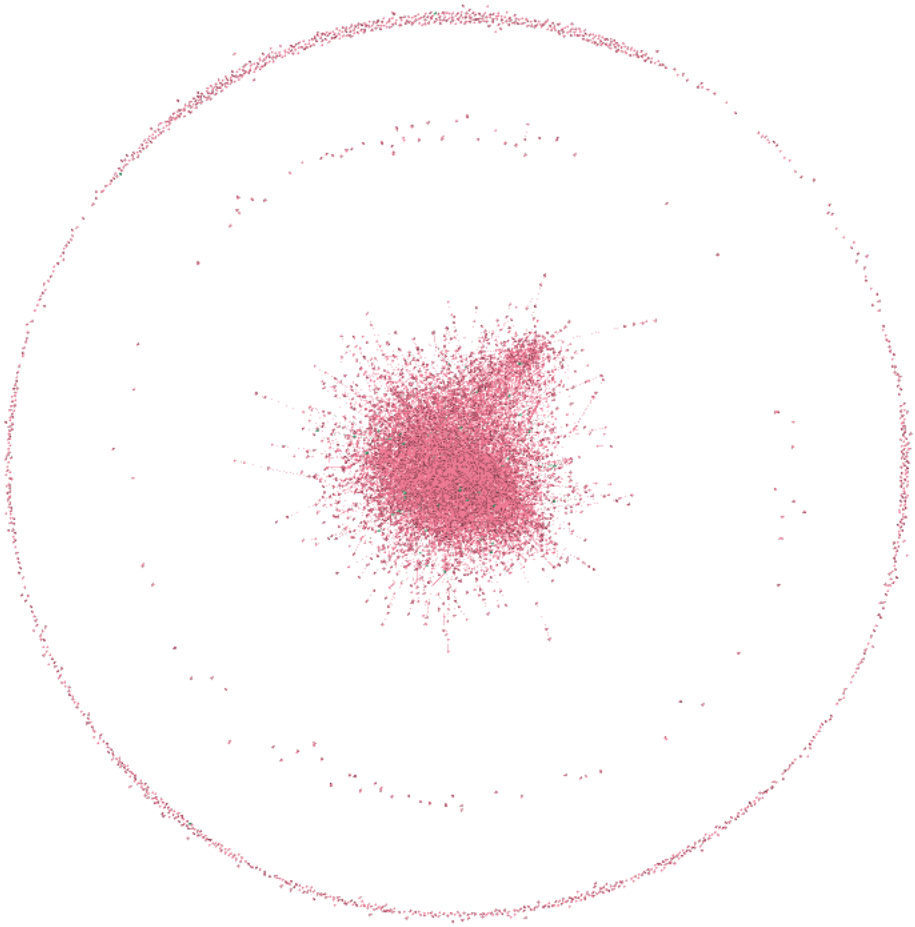

Figure 12: Small streamers based on partnership (color) and number of days (size)

6 Conclusions

In the end, there are many key takeaways from this project, but the most important things are the following:

- Gephi it is a very useful tool for graph visualization, but the ETL tool is not so handy. Thus, the preprocessing part should be done with other tools (e.g. Python/R)
- Twitch Social Networks is a large dataset with many nodes and edges containing much information for potentially important discoveries among the Twitch communities.
- The dataset chosen shows that big streamers use to stream not mature content and their friendship on this famous platform is not the most important aspect for their networking
- The partnership with Twitch is not usual for small streamers, but for big streamers this is a representative fact

- Seniority in terms of Twitch streaming is not a key factor for big streamers. This is represented in the shape of the nodes where many nodes are also small, but with huge amount of views for their interesting content
- Friendship can be obtained through many other social media platforms, and even though Twitch connects people with live streaming, the friendship system do not interest so much the streamers, but maybe only the follow system or other aspects.