Mid-Term Progress Report COSC 421

Bassim Beshry, Zeyad Elganainy, Jordan Pohr

1. What we have done so far

As of right now, we have focused on analysing a manageable 10,000-node subset of our main dataset due to technical and visualisation constraints associated with the full dataset (discussed further in section 3). We chose this subset to strike a balance between capturing meaningful trends and ensuring the analysis is feasible with our limited available resources. Within this subset, we explored the degree centrality measurement, which allowed us to examine the extent of connections or "popularity" among streamers, and clustering, which gave insights into how tightly knit communities are around individual streamers. These analyses helped us begin to understand how certain structural roles in the network might correlate with streamer characteristics like maturity ratings, which we have detailed in section 2.

2. Results we have

The following results are based on a sample size of 10,000 accounts:

Degree Centrality
How many streamers out of the top 100 largest communities are mature:
49 out of 100

The largest streamer in the sample of 10,000 has a community size of 2,139 but is not mature. For the top 100 largest communities the average degree is larger for non-mature streamers with an average degree of 157.923 compared to 106.573. This means that although 49 out of 100 largest communities are mature most of them are in the lower portion of this 100 streamers meaning non-mature streamers make up a majority of the high follower counts in the network.

Clustering
The averaging clustering coefficient more maturity rating 1 is higher than maturity rating 0:

```
mature avg_clustering
   <int>          <dbl>
1      0         0.0312
2      1         0.0340
```

This likely tells that although out of the 100 most popular streamers mature streamers have less degree centrality on average their communities are more tightly knit compared to non mature. This difference is very small and might not be impactful.

3. Challenges we have faced

Working with this dataset has presented us with several key issues. Firstly, the sheer size of the data has made it difficult to generate visualisations, as rendering complex graphs with a large number of nodes strained both the visualisation tools and our system resources. Additionally, slower or less powerful hardware in our group has limited our ability to process the full dataset efficiently.
To address these issues, we are planning to allocate the remainder of the analysis tasks to group members with stronger PCs that can handle the dataset size more effectively. Another challenge was that the dataset was that it was initially formatted incorrectly, leading to some inaccurate calculations. We are in the process of authenticating our findings by ensuring correct dataset formatting and recalculating where necessary to ensure accuracy before the final report.

4. What we need to do for the final deliverable

For the final deliverable there are a few more tasks that we need to finish up.

Analysis

To complete the project, we have three main analytical tasks remaining, each aimed at deepening our understanding of how maturity rating influences different aspects of Twitch streamer success. Specifically, we need to:

1. Analyse the Effect of Twitch Affiliate Status on Maturity Rating: This analysis will help determine if obtaining affiliate status is correlated with higher maturity ratings among streamers. We will explore whether affiliates are more likely to have a maturity rating, indicating a potential link between affiliation and content type.

2. Investigate the Impact of Maturity Rating on View Counts: We will examine if there is a significant difference in view count between mature and non-mature streamers. This analysis involves comparing view counts across different maturity ratings to see if mature content affects streamer popularity.

3. Explore the Relationship Between Maturity Rating and Follower Count: Here, we will assess if maturity rating influences follower counts by comparing follower metrics across streamers with different ratings. This will provide insights into whether mature content has an effect on attracting or retaining followers.

Timeline for Completion:
- Mid-Late November: Finish analysis on the effect of Twitch affiliate status on maturity rating.
- Late November: Complete analysis on maturity rating and view counts.
- Early December: Wrap up analysis on the relationship between maturity rating and follower count.

At this point, we must complete the analysis for these 3 questions by the end of November. In order to do this, we must fix the dataset to make the analysis easier for all group members. This will be fixed as soon as reading week is over. The three of us being in separate cities has made this a difficult problem to solve for now.

Final Deliverable

Due to our projected completion of the analysis by the end of November, we will use the first week and a half of December for the completion of the final deliverable. Each group member will be incharge of documenting their findings. Zeyad will handle the formatting of the document.