

Hannah Choe
Professor Leonidas Kontothanassis
DS210
7 May 2023

Project Report

What my project does:

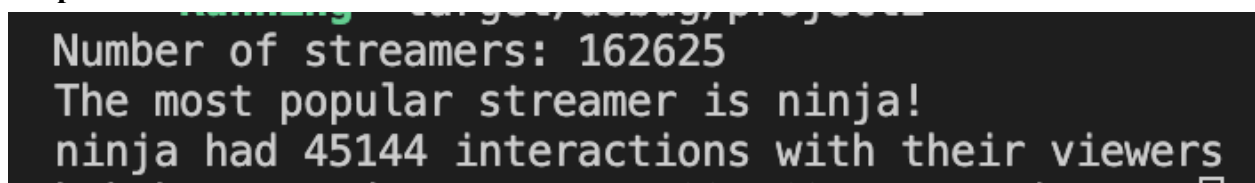
The dataset that I used for my project was a dataset consuming streaming content on Twitch. I found this dataset intriguing because it does not only include details on which streamers viewers are watching but also their interactions with them. With this information, it would be possible to analyze factors such as popular streamers and user behaviors based on the number of interactions they have leading to the goal of the project. My project analyzes user interactions with streamers by reading data from a file and creating a graph. The graph is created where each unique streamer is a node and users are edges. The edges (user) will be pointing towards the node (streamer) in order to represent an interaction that occurred. After building the graph, it calculates the number of interactions each streamer had with their viewers by calculating the in-degree for each node in the graph. By calculating the in-degree of each node, the program would be able to identify the most popular streamer by finding the node with the highest in-degree and displaying their name and the number of interactions they had with their viewers.

To ensure the program's accuracy, it includes unit tests for its three main functions: `read_file`, `calculate_indegree`, and `find_most_popular_streamer`. These tests provide confidence in the program's correctness.

How to Run the Project:

- Download the “100a_k.csv” dataset that is located in google drive through this link <https://cseweb.ucsd.edu/~jmcauley/datasets.html#twitch>
- Ensure that the downloaded dataset is within the same
- Then run the code with “cargo run”

Output:



```
Number of streamers: 162625
The most popular streamer is ninja!
ninja had 45144 interactions with their viewers
```

The output of my project shows the total number of unique streamers present, the most popular streamer, and the number of interactions they had with their viewers. Essentially, it creates a graph using the given dataset and determines the number of nodes in the graph. It also identifies the node with the highest in-degree and calculates the number of edges pointing towards it. This

means that I found that out of the 162,625 streamers within the dataset, the streamer ninja has was the most popular with 45,144 interactions with Twitch users.

To confirm the accuracy of these results, I conducted three tests using the `test_data.txt` file provided. The first test verified the ability of the `read_file` module to read the file and determine the total number of unique streamers. The second test, known as `test_calculate_indgree`, verified the precise calculation of edges pointing towards a node through the `calculate_indegree` function. Finally, the `test_most_pop_streamer` ensured that the `most_pop_streamer` function accurately identifies the node with the highest in-degree.

```
running 3 tests
test tests::test_read_file ... ok
test tests::test_most_pop_streamer ... ok
test tests::test_calculate_indegree ... ok
```

Resources:

- <https://www.geeksforgeeks.org/finding-in-and-out-degrees-of-all-vertices-in-a-graph/>
- <https://docs.rs/petgraph/latest/petgraph/graph/struct.Graph.html>
- <https://smallcultfollowing.com/babysteps/blog/2015/04/06/modeling-graphs-in-rust-using-vector-indices/>
- <https://stackoverflow.com/questions/75349667/how-to-build-a-graph-from-a-vector-of-structs>
- https://doc.rust-lang.org/stable/std/cmp/fn.max_by_key.html
- <https://stackoverflow.com/questions/37127209/using-max-by-key-on-a-vector-of-floats>