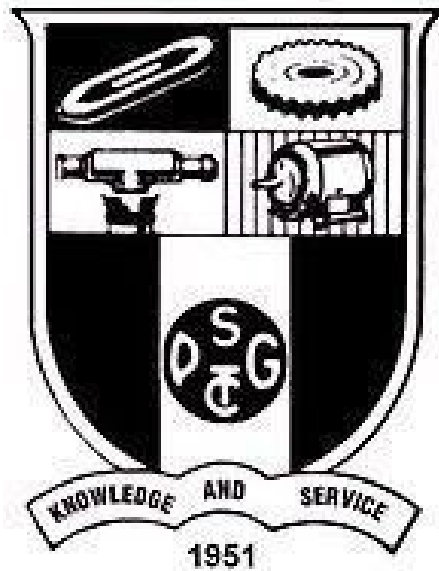# PSG COLLEGE OF TECHNOLOGY, COIMBATORE 641004
Department of Computer Science and Engineering

SOCIAL AND ECONOMIC NETWORK ANALYSIS

ASSIGNMENT PRESENTATION

By

19z301 Kirthic Vishnu

19z302 Adarsh S

19z327 Kumaresh S

19z332 Mridula M

19z353 Swathi S

Introduction

Graph data science focuses on looking at the connections and interactions in data in order to gain insightful knowledge. Every day, enormous amounts of data are generated by various platforms, social media networks being just one of them. However, during data processing, connections between data points are typically overlooked. With the advent of Graph Data Science tools, it is not only the organizations with vast technological resources that are capable of analysing relationships.

Twitch is an American multimedia company specializes in live video game broadcasting, including coverage of esports events. It also provides music streams, original material, and "in real life" feeds. Twitch is a well-known network for live streaming that attracts a sizable audience as well as frequent content producers. One of the largest demographics of audiences on Twitch belong to those who watch content directly or indirectly relating to video games, either console based or PC based. The

Problem Statement

- Import information about current live streams from the Twitch API
- Store this information as a graph in the Neo4j database
- Visualize the data
- Perform exploratory data analysis

Components Required

Neo4j

Neo4j is an open-source, NoSQL, native graph database that provides an ACID-compliant transactional backend with native graph processing and storage. Instead of using a tabular or Json format, it stores and displays the data as a graph. Here, nodes are used to represent the entire set of data, and you can establish connections between them. As a result, the entire database collection will resemble a graph, setting it apart from other database management systems.

Advantages

- Connected data can be represented extremely simply.
- Data from connected sources can be retrieved, traversed, or navigated very quickly.
- It utilizes an effective and simple data model.
- It is simple to express semi-structured data.

APOC

To make Neo4j simple to use and increase productivity, APOC [Awesome Procedure On Cypher] offers a variety of predefined procedures and user-defined functions/views. Its collection of

methods can be used for diverse applications, including concurrent and batch cypher execution, relationship and manual indexing, and graph algorithms.

Seaborn

Seaborn is a Python data visualization package that is built on matplotlib and works well with pandas data structures. It offers a high-level interface for creating visually appealing and informative statistical visuals. Its charting capabilities operate on data frames and arrays containing entire datasets, performing the necessary semantic mapping and statistical aggregation internally to generate useful graphs. Its dataset-oriented, declarative API allows the focus to be directed on what the various aspects of the plots represent rather than how to render them.

Twitch API

The Twitch API contains the tools and data required to create Twitch integrations. The data models and systems are intended to give relevant data in a simple, consistent, and dependable manner. For authentication, the API employs OAuth 2.0. Queries are written to include the respective endpoints for retrieving data about streamers, the number of viewers, the language in which the viewers chat, and so on.
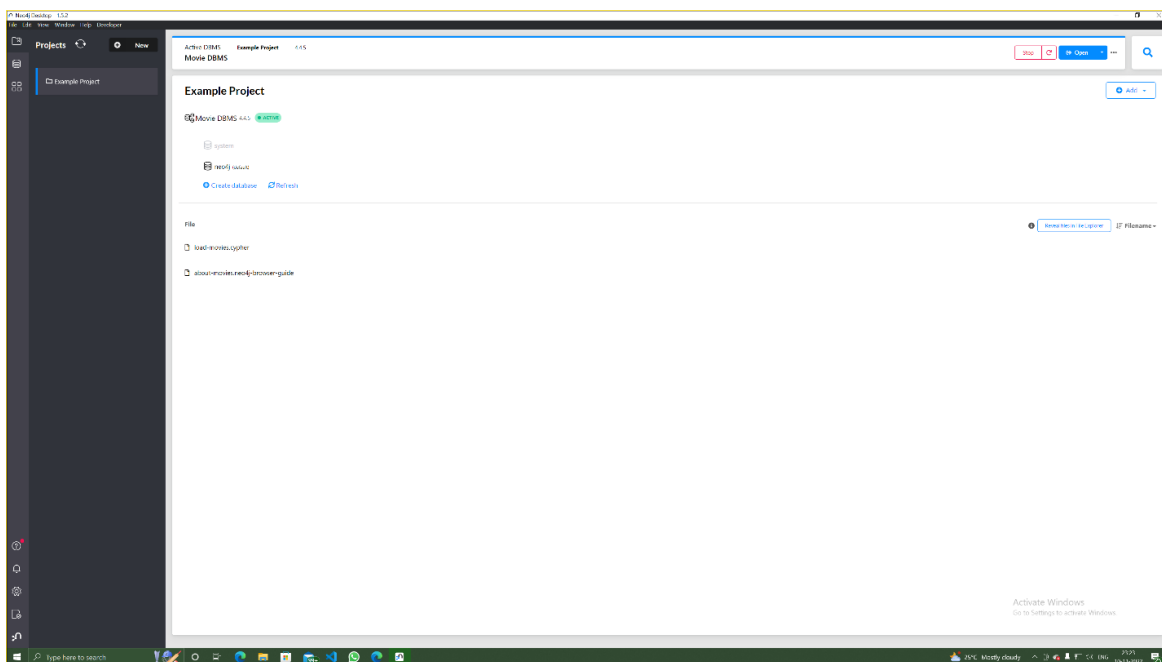
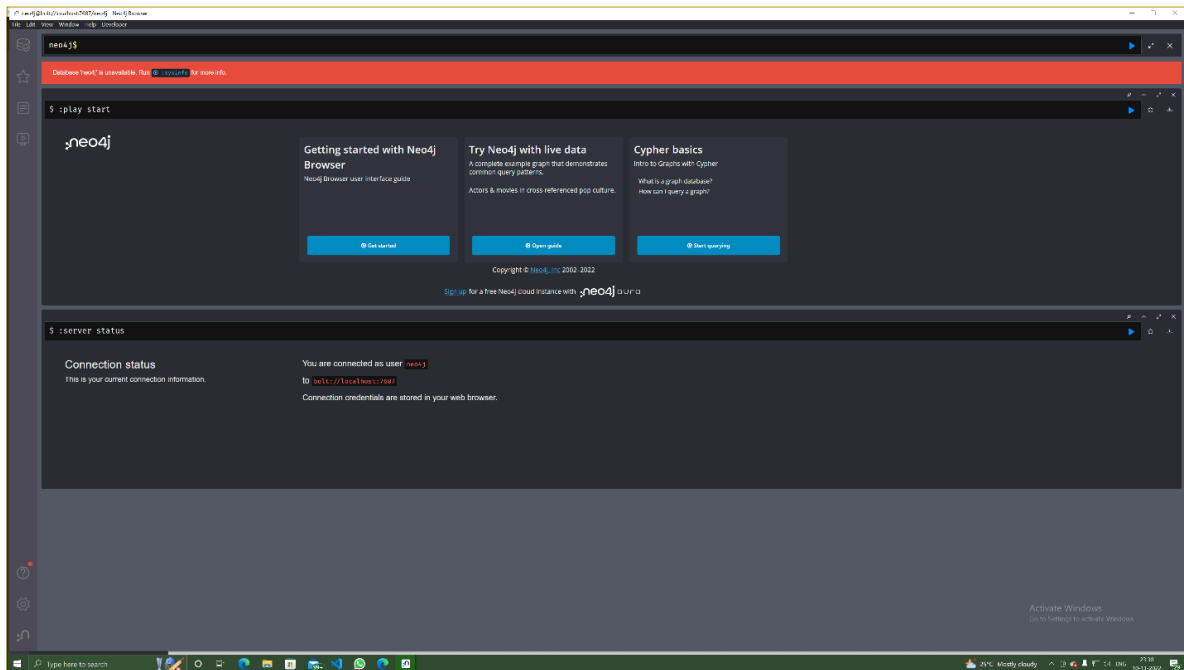## Code and Results

Figure: 1 Example Neo4j project

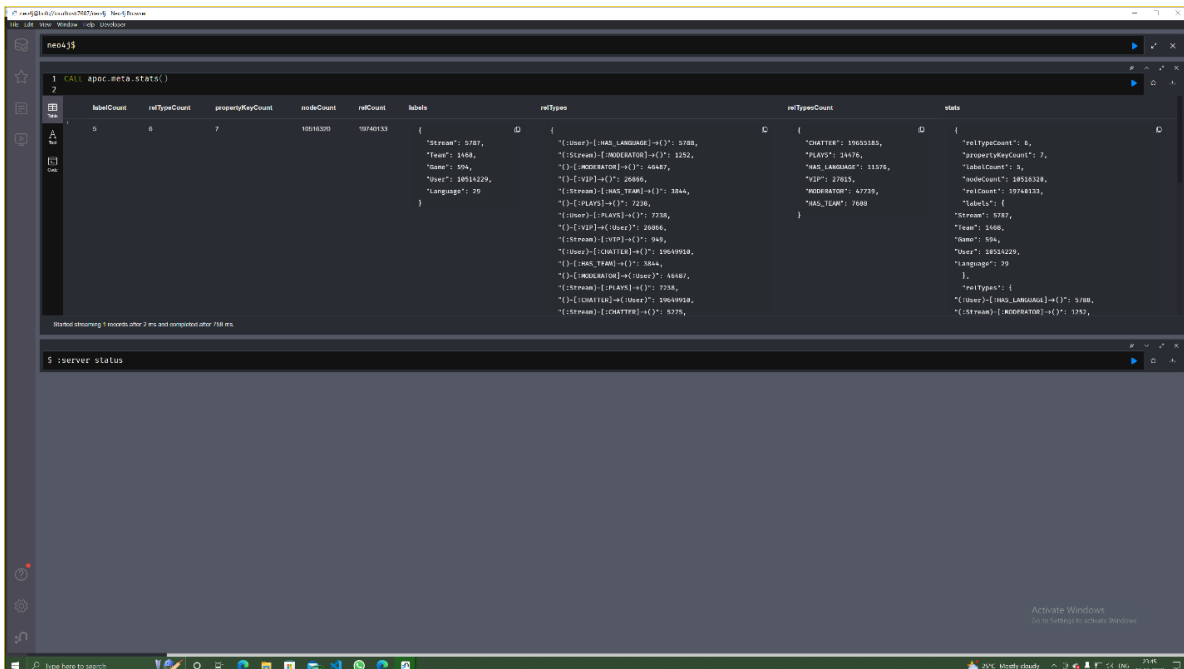*Figure: 2 Establishing connection*
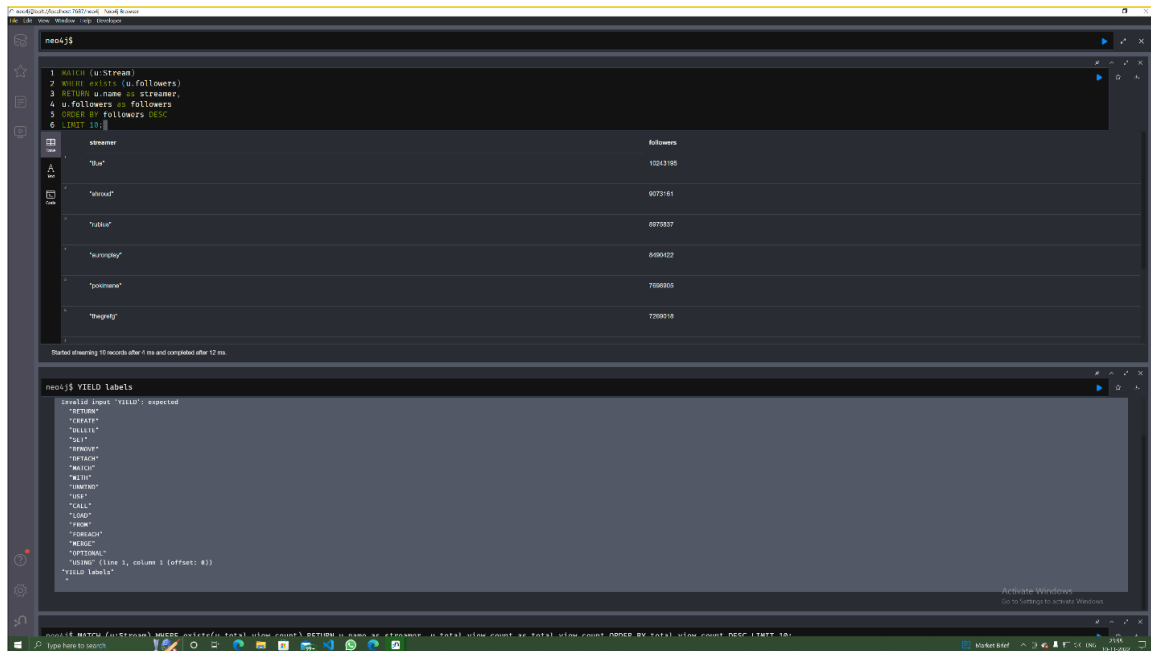

Figure: 3 Statistics of the dataset
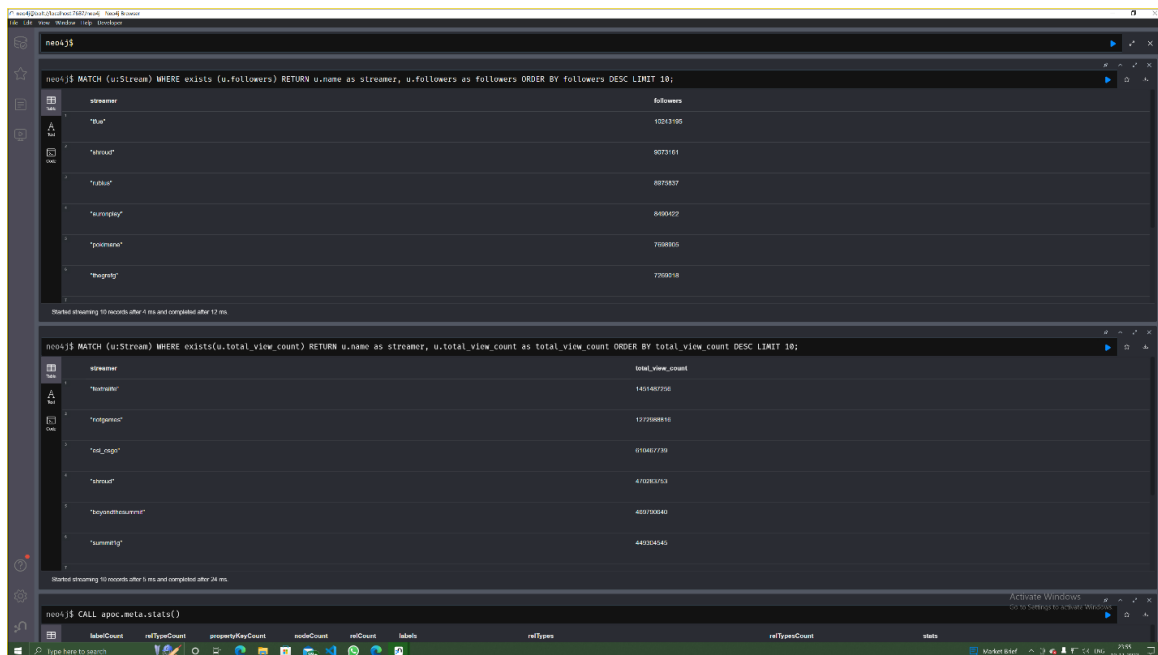
Figure: 4 Top ten streamers with highest followers

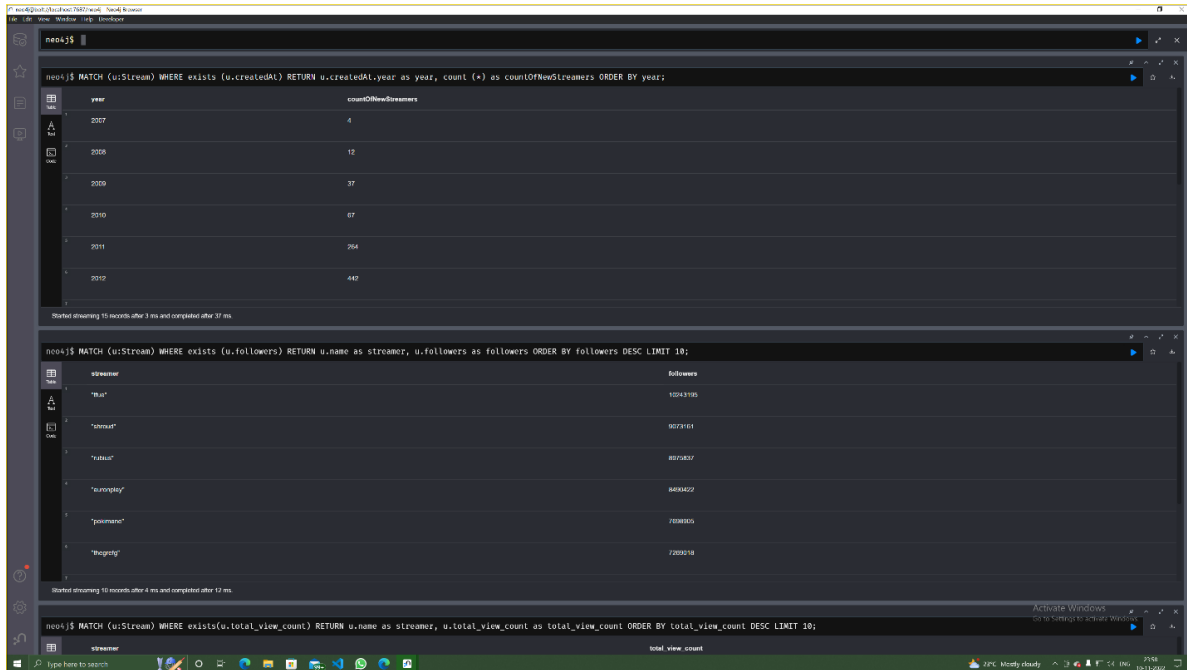*Figure: 5 Top ten viewers by viewers count*

Figure: 6 Top ten viewers by age of account

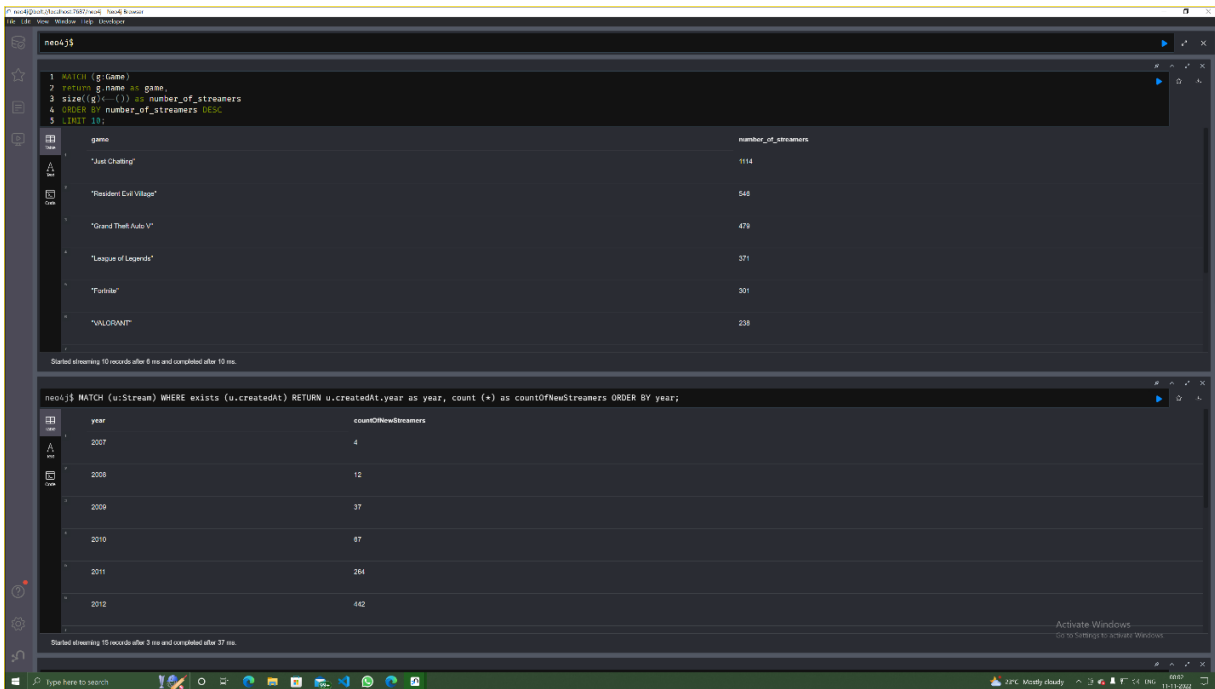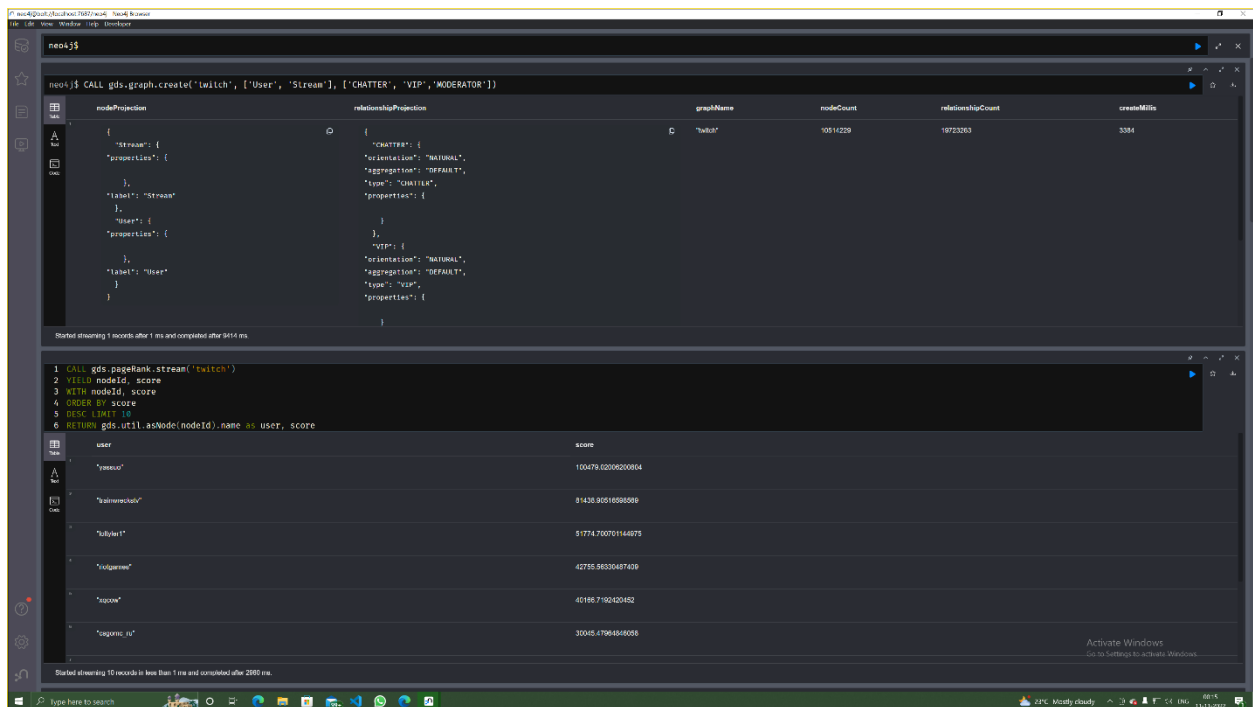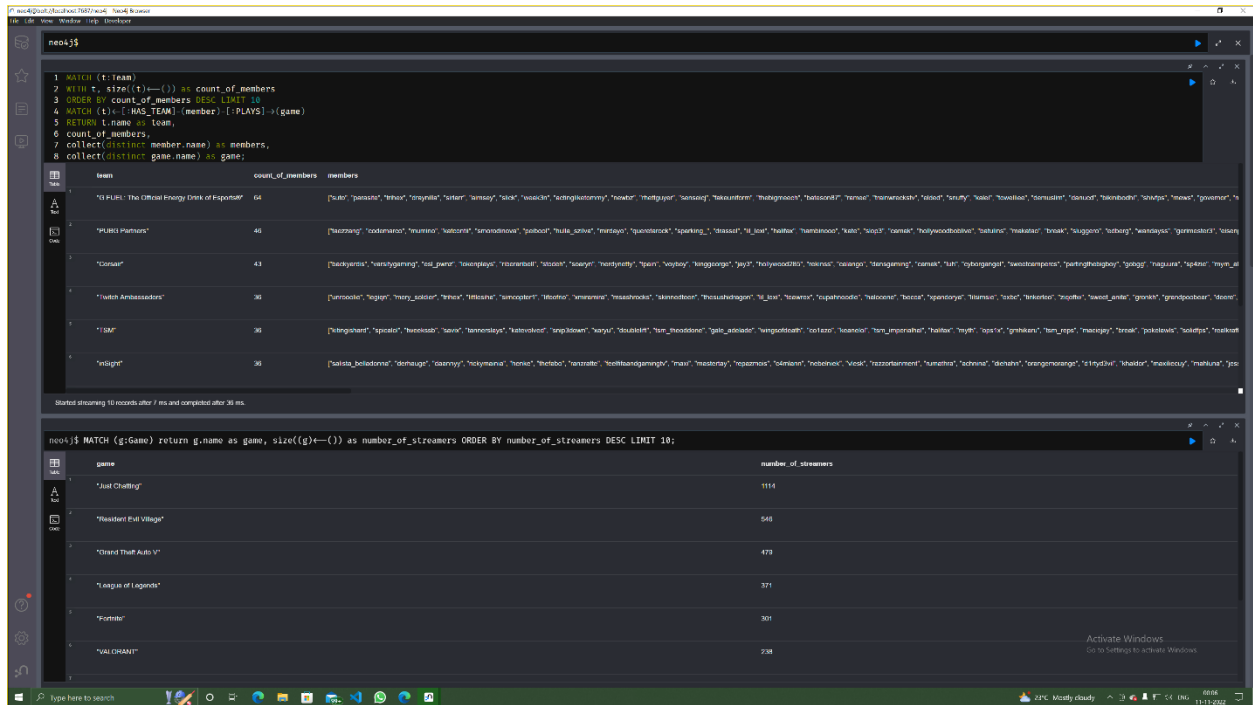*Figure: 7 Top ten games played by most streamers*

Figure: 8 Teams that have the highest member count. and the games that the team members broadcast.

Figure: 9 Running PageRank for all the nodes in the graph

Contribution of Team Members

- Problem statement definition – Mridula m
- Dataset identification – Swathi S, Adharsh S
- Neo4j project initialisation and visualisation – Kumaresh S, Adharsh S
- Exploratory data analysis – Kumaresh S, Swathi S
- Documentation – Mridula M, Kirthic Vishnu A

Conclusion

Neo4j tool provides a new method of performing network analysis. Using the queries to perform network analysis on the Twitch Dataset enabled us to gain a deeper understanding of the media that we consume, and the scale of the communities that we participate in. Further analysis can be done on the dataset to detect communities, and analyse the intersection of interests such as games in those communities.