UC San Diego

JACOBS SCHOOL OF ENGINEERING

Video Game Knowledge Graph Analysis

Team: Polina Haryacha
Advisor: Amarnath Gupta

Problem Statement

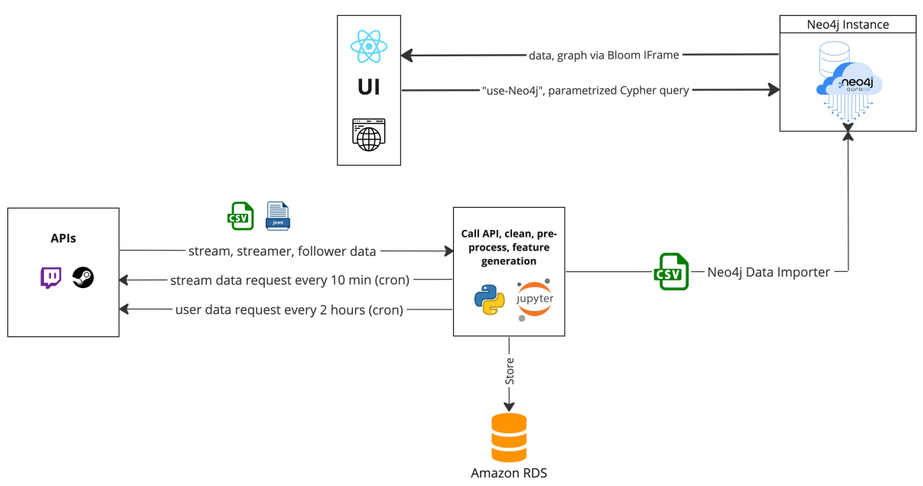
There are 3.09 billion active video gamers worldwide, with a 32% increase in just seven years. Around 81.9% of internet users play games, indicating high gaming penetration.

Game reviews greatly impact sales and are generated by diverse communities and influencers, including gaming journalists, Twitch streamers, YouTube gamers, and Steam Curators. Marketers recognize the immense potential of targeting gaming communities to build brand awareness, drive sales, and foster dedicated player communities. However, modern marketing tools lack the ability to consolidate data from various platforms and overlook the power of gaming communities disregarding the intricate social network dynamics that drive them.

.

The focus of this project is to construct a Knowledge Graph that brings together data from diverse sources and captures crucial information about various actors in the gaming ecosystem. By harnessing the unique public Twitch user data and utilizing Social Network Analysis, marketers can identify key players within gaming communities who possess the influence to effectively communicate their message to the target market.

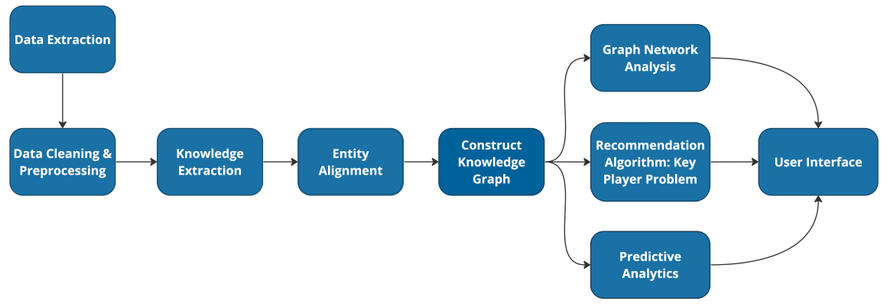
Final Solution Architecture



Key Insights

- High viewership does not necessarily translate to high engagement (high number of chatters) or strong influence (high centrality measure score)
- Users who form communities based on their Chatter relationship do not play the same games.
- Different centrality measures capture different aspects of node importance and identify different Key Players. Combining two or more centralities is required to identify top influencers more accurately

Data Science Pipeline



1 Acquire

Steam API: extract the entire list of games available on Steam
SteamSpy API: acquire additional details about Steam games
Twitch API: cron job to collect all live stream data every 20 minutes for ten days. Separate queries to acquire streamers data and chatters data. Filter streams featuring games available on Steam

2 Prepare

The data was pre-processed and uploaded to a Postgres database. A combination of manual and programmatic access methods was employed for interacting with the database. New features like average viewers and peak viewers were generated by aggregating channels data. Sentiment analysis was applied to Steam curator review data to assess user responses. Sponsored Twitch streams were identified, and sponsoring brand names were extracted using GPT-3. Finally, the data was prepared for upload to Neo4j, enabling further analysis and exploration.

3 Analyze

Knowledge Graph Analysis: use Cypher queries to reveal complex relationships in the graph, providing insights into the gaming ecosystem
Network Analysis**:** 4 centrality measures applied to identify Key Players: In-Degree Centrality, Betweenness Centrality, and PageRank. In-degree centrality is combined with eigenvector centrality using a weighted average approach to consider various aspects of a streamer's influence.

4 Scale

Optimize data extraction and pre-processing by moving to cloud-based services and implementing event-driven architecture for retrieving data from API providers (API Gateway & Lambda)
Uploading pre-processed data directly to Neo4j and utilize AWS S3 for raw data storage
Use Neo4j Community version hosted on AWS
Implement an API layer between the user interface and the Neo4j database

5 Report

The project reporting process involves a series of steps to gather user input, retrieve relevant data on Twitch streamers and Steam curators, and analyze key players in the Twitch network. In step one, users provide information about their game, competition and target influencers. In step two, data about individual influencers is presented in charts and tables, which can be filtered and exported. Step three involves identifying key players based on centrality measures, allowing users to explore network relationships and gain deeper insights.