**Team Name:**

* JLS Analytics

**Team Members:**

* Kim, Justin
* Mitchell, Louis
* Sabih, Syed Muhammad

**Motivation:**

The motivation behind the project topic is our shared interest in the Gaming Industry and domain knowledge regarding Social Network Analysis, which incorporates our analytical and insight for Gaming to produce findings related to the Research Problem.

**Project Overview:**

We have decided to use the concepts learned in this class regarding Graph Theory to do a social network analysis on the Twitch Gamers Social Network Dataset. The dataset is taken from the SNAP data repository which exists in the Stanford University library. The data was collected from the public API in 2018. The Nodes in this dataset represent Twitch users and the edges between them represent mutual follower relationships between them. Altogether, the nodes and edges form a strongly connected component. There are no missing attributes.

**Link to the Dataset:** [DATA](https://snap.stanford.edu/data/twitch_gamers.html)

**2 Files:** Features + Nodes & Edges

**Features for the nodes:**

* views (View count for the streamer)
* mature (Stream is suited for mature viewers)
* life\_time (Days on twitch)
* created\_at (When the twitch account was created)
* updated\_at (How recently they have posted - relative to 2018)
* numeric\_id (Unique ID representing the streamer)
* dead\_account (If the Twitch account is Active/Dead)
* language (Language for the stream)
* affiliate (If the streamer is Twitch Affiliate/Partner)

**Number of Nodes:** 168,114  
**Number of Edges:** 6,797,557  
**Directed or Undirected:** Undirected Graph   
**Weighted:** Unweighted

**Research Problem:**

Investigating the impact of content creators on the Twitch Gaming Streaming platform as it relates to viewership and monetary acquisitions to affiliation.

**Objectives:**

* To produce a graph that shows the relationships between all of the twitch users and identify any communities/clusters between the network based on the features provided.
* To Create a Machine Learning Algorithm that will predict the likelihood of the affiliation of the Twitch streamer based on the aforementioned feature attributes.

**Research Questions:**

* Which Twitch Streamers in the data are most influential? (Betweenness Centrality)
* Which attribute/feature is important in the analysis? (Feature Engineering, F-test)
* How many separate clusters/communities are present in the Network? Which Twitch Users have mutual relationships, or share the common language etc.?
* Can we predict Twitch affiliation based on viewership, type of content provided (mature/PG), Twitch account age, and Language?
* Overall, do Twitch streamers with high count of viewers usually share relationships (mutual followers) with other streamers with high count of viewers, OR is there an inverse relationship?
* Overall, do Twitch affiliate streamers usually share relationships (mutual followers) with other Twitch affiliate streamers, OR is there an inverse relationship?

**Analysis Tools used:**

* Python or Julia