**Data 620 Week 4 Assignment**

**Project 1 Description**

**Aaron Grzasko**

**9/23/22018**

**Data Source**

For project 1, I will use data furnished through the Star Wars API, or SWAPI. The data are available at <https://swapi.co>. I will rely on helper methods available through a Python library, swapi-python, to make the API calls.

**Graph**

Using SWAPI data, I will construct a network graph as follows:

* Nodes will represent individual Star Wars characters (e.g. Han Solo)
* Edges between a pair of nodes will indicate that two characters appeared in the same Star Wars film. I also plan to weight the edges to indicated the total number of Star Wars films the two characters appeared in together.
* The nodes will also have one or more of the following attributes:
  + Species
  + Home world
  + Gender
  + When character first appeared in series: original trilogy, second trilogy, modern era

**High Level Data and Analysis Plan**

The Star Wars API returns data in JSON format; however, the swapi-python library parses the data and

returns custom objects relating to the films, characters, starships in the movie series. These objects have a variety of attributes, such as character name, height, date of birth, etc. that can be extracted for analytical purposes.

I will initially retrieve data using the swapi-python helper methods. The data will be massaged and refined using loops, built-in Python methods and Regex expressions, if necessary.

Data will be stored in one or more common Python structures: dictionaries, lists, Pandas DataFrames, numpy arrays.

A network graph will then be constructed using functions from the networkx python library. I may produce additional data visualizations using Gephi software.

Finally, I will then calculate degree centrality and eigenvector centrality measures across node categories using networkx functions.

**Hypothetical Outcomes**

The centrality measures from the analysis could be used to predict which Star Wars characters will appear in future movies. Characters with high degree centrality presumably are very popular characters, and could be reasonably expected to appear again. Because the Star Wars series spans a variety of different time periods, the death of a character does not necessarily preclude his or her appearance in future movies.

I could also look at how the mix of character genders varies across the franchise’s history. I suspect that the mix has become more gender inclusive over time. Given the existence of this trend, one could forecast the gender mix of future Star Wars movies. Centrality measures could provide us with a sense of which gender is more influential in the series.