Visualizing Change in Clustering Data Over Time
Louis Brion – CS 448B

Problem
- Machine learning often accomplishes its intended application (clustering, prediction, classification, recognition)
- Further data analysis can be done on the output of machine learning algorithms to find out more about the nature of the data or its sources

Motivation
- Hearthstone is the largest digital card game and is the first opportunity analysis on a large volume of granular collectible card game data can be analyzed and visualized
- Provides an opportunity to combine the personal and academic interests of the researcher

Objectives
Aim 1.
Use Machine Learning to correctly identify Hearthstone deck archetypes from historic game data

Aim 2.
Convey to viewers the characteristic features of an archetype

Aim 3.
Convey to viewers how deck archetypes change over time

Approach
- Data Normalization
  - Original dataset: CSV turn by turn data of 300,000 Hearthstone games from August 2016 to September 2016
  - Remove noise by filtering out non-card actions and non-drawn cards
- Variable Selection
  - Separate decks by class
  - Analyze games with 12+ cards played
- Cluster Algorithm Configuration
  - Batch clustering with ~200 games per batch
  - Affinity propagation on sparse feature vectors
- Cluster Analysis
  - Trace clusters through previous batches
  - Confirm clustering aligns with common archetypes at the time
- Visualization
  - Arrange genealogies of archetypes sequentially
  - Show card usage rates within archetypes to identify core and tech cards
- Visual Analysis
  - Look at variations in coloring to determine times of less usage for different cards
  - (Future): Align significant events (nerfs, releases, season endings) to data to look for further correlations

Results

Future Work
- Apply this sort of visualization to other clustering datasets to see how behaviors change
- Polish clustering algorithm to be completely unsupervised from input data to visualization
- Create visualizations from live data and get Hearthstone community feedback

Acknowledgements
- Prof. Maneesh Agrwala - CS448B Course Instructor
- Prof. Tzachi Zach and Vicious Syndicate - dataset and ideas