Updates Since Last Week

Done Last Week

- Collecting data sources for both news and equity ohlc
- Adding code for DDTW, clustering, and visualization

Next Week Plans

- Formalizing metrics and criteria for clustering
- Conducting more EDA
- Trying more variants of DTW

Enhancing Equity Predictions Using Informational Signals

Group Members: Bella Macaluso - Elizabeth Yang - Sourav Vemulapalli - Aditiya Palliyil - Joseph Jabbour

Githup repo:: https://github.com/bour278/info-signal-analysis

Table of Contents

1- General Overview
2- Data Sources
3- Methodology
4- Limitations

General Overview

- 🎯 **Target**: Enhance equity predictions using informational signals
- 🛠️ **Methods/Tools**: - Derivative Dynamic Time Warping (DDTW) - Louvain/Leiden Community Clustering - Kalman Filtering - Markov Random Fields

General Overview

DATA COLLECTION

- Kaggle
- Github
- Web Scraping

EDA

- Correlation between news and prices jump
- Optimal number of clusters

MODELS

- DDTW distance graph
- Co-occurrence matrix
- Kalmann Filtering

VISUALIZATION

- Network Clusters
- Time-series Clusters
- Prediction vs Enhanced Prediction using confidence bands

METRICS

- Cut metric / modularity for clustering
- Inter/Intra cluster variance
- MSE for enhanced predictions

Data Sources

- **Kaggle:** Daily OHLC data for US-based equities

Date	Open	High	Low	Close	Volume	OpenInt
1984-09-07	0.42388	0.42902	0.41874	0.42388	23220030	0
1984-09-10	0.42388	0.42516	0.41366	0.42134	18022532	0

Data Sources

- **Github:** Reuters Financial Dataset

```
-- Samsung aims to double its smartphone sales in Africa in 2014
--
-- Wed Nov 13, 2013 2:29am EST
-- http://www.reuters.com/article/2013/11/13/us-africa-samsung-idUSBRE9AC08620131113



 CAPE TOWN  (Reuters) - Samsung Electronics expects to supply half of the smartphones sold in Africa
 this year and aims to double these sales on the continent in 2014, an executive said.
```

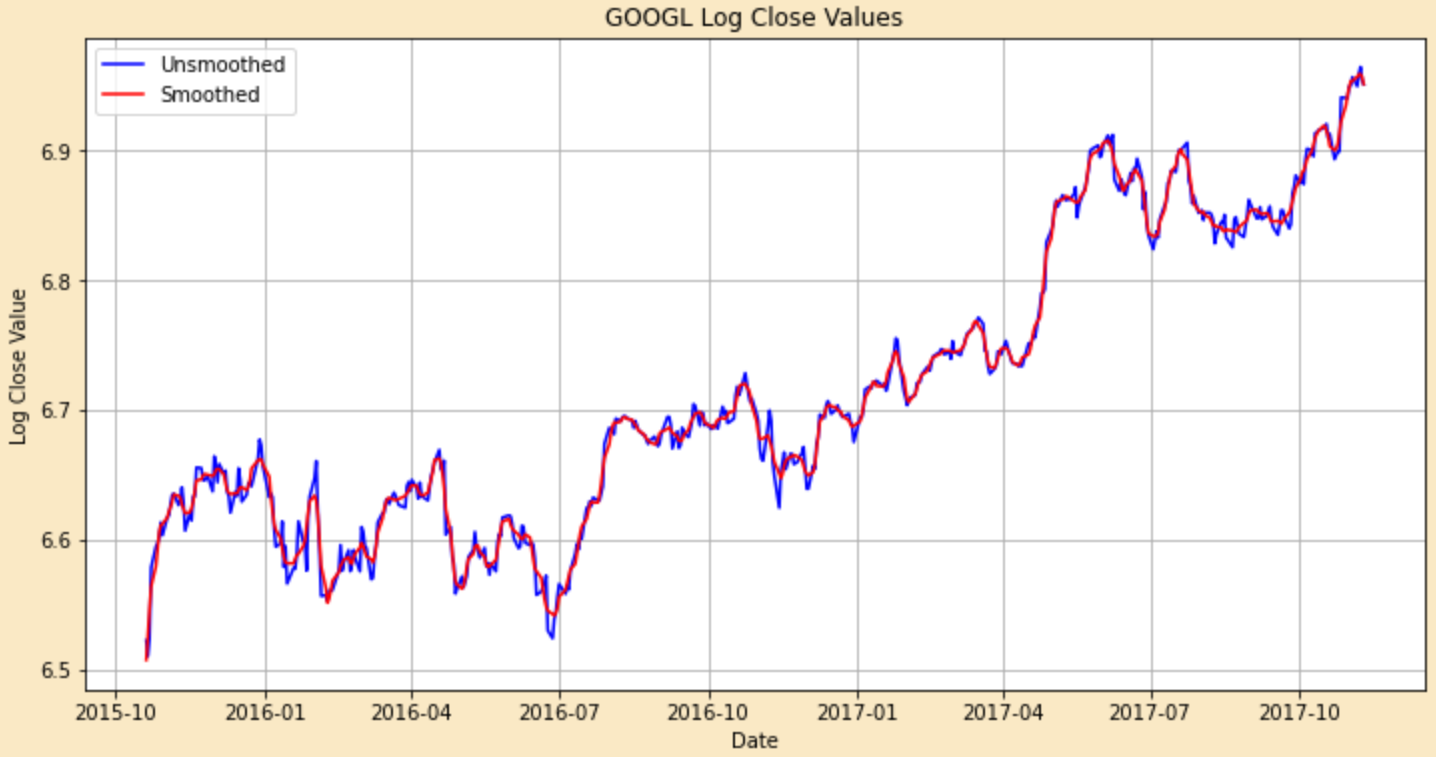
Data Sources

- **Scraping:** New York Times News Archive

```
Chadwick Boseman Played Black Icons, Found Fame With ‘Black Panther’
11:20 PM ET
------------------------------------------------------------------------
Japan
Abe Will Resign as Japan’s Prime Minister, Citing His Health
10:17 PM ET
------------------------------------------------------------------------
Politics
Thousands March on National Mall, Continuing Racial-Justice Push
10:11 PM ET
------------------------------------------------------------------------
```

Methodology - Pre-Processing

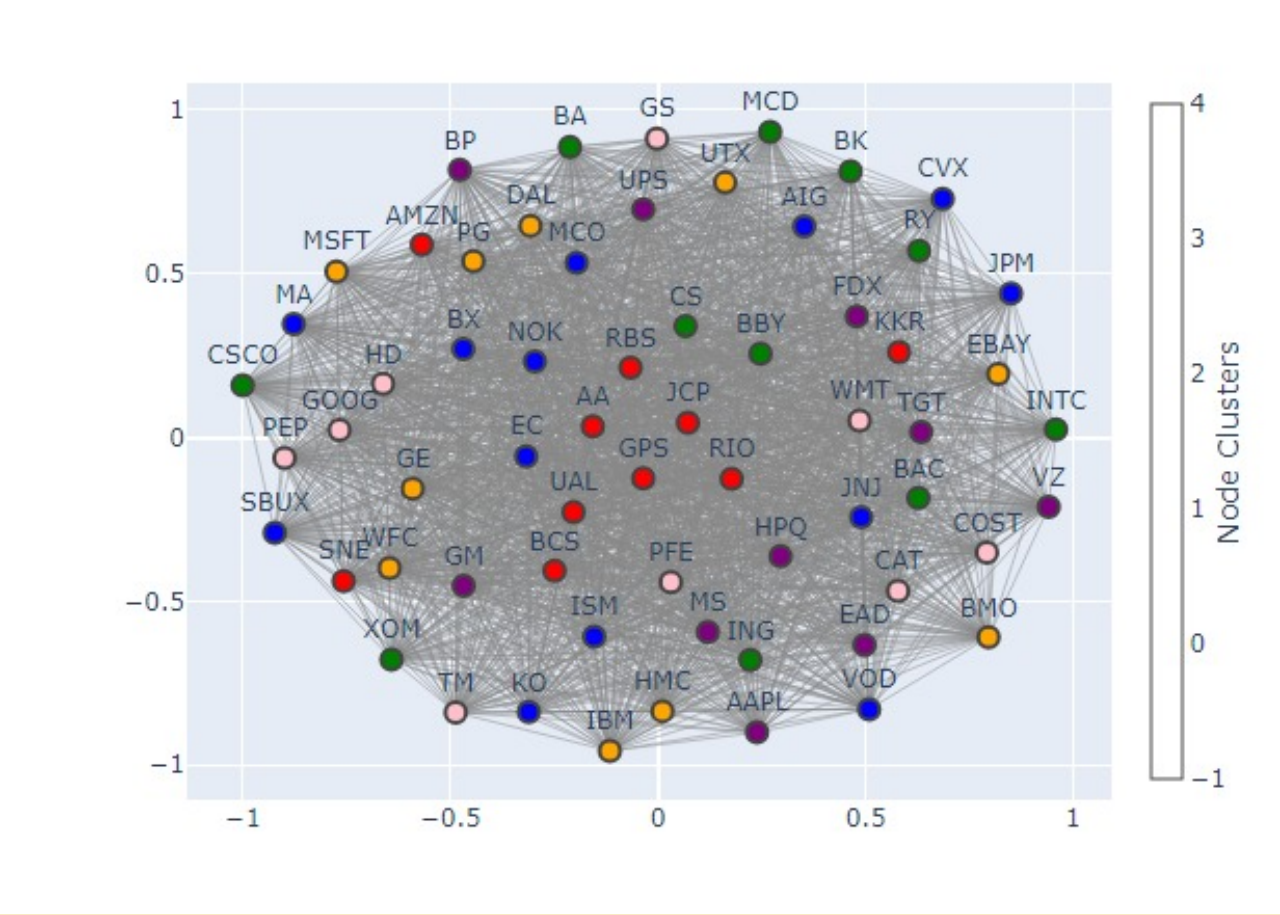
- **Savitzky-Golay Filtering:** removing noise from histoical time series data using polynomial interpolation at fixed-length window



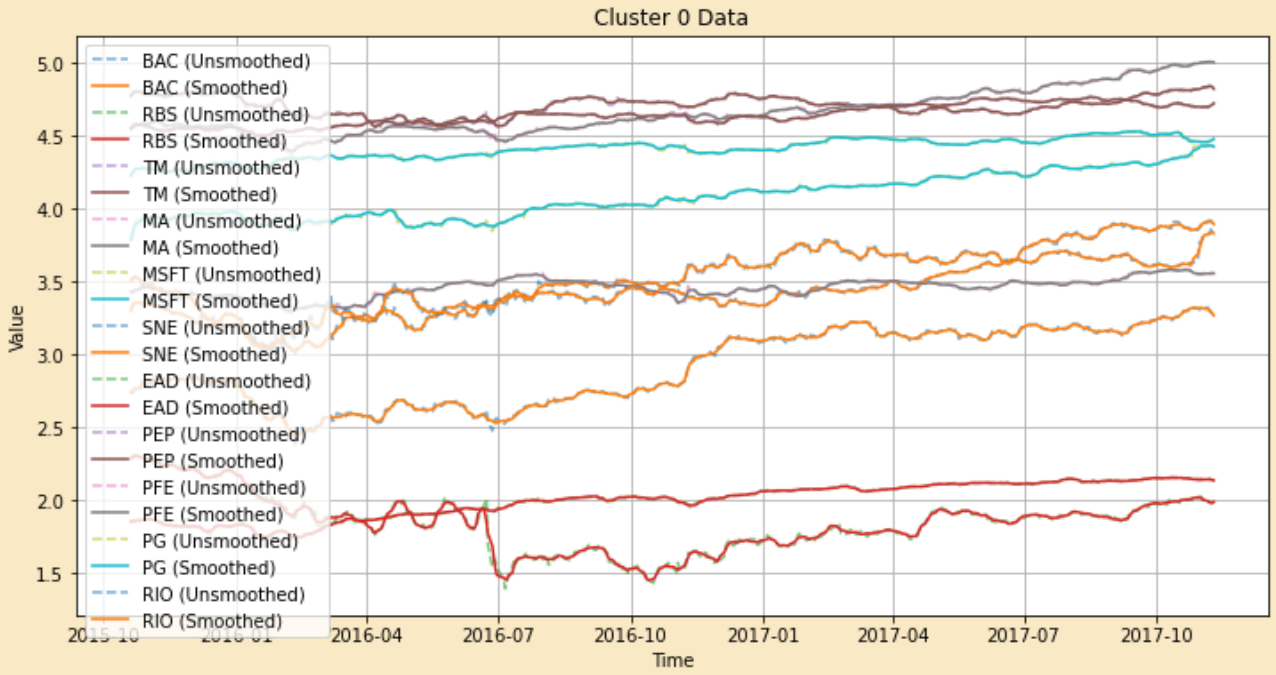
Methodology - DDTW Clustering

- **DDTW:** algorithm finding shortest path distance between 2 time series using dynamic programming approach
- **Graph Representation** Adjacency matrix is built from pairwise DDTW distances between each pairs of equities

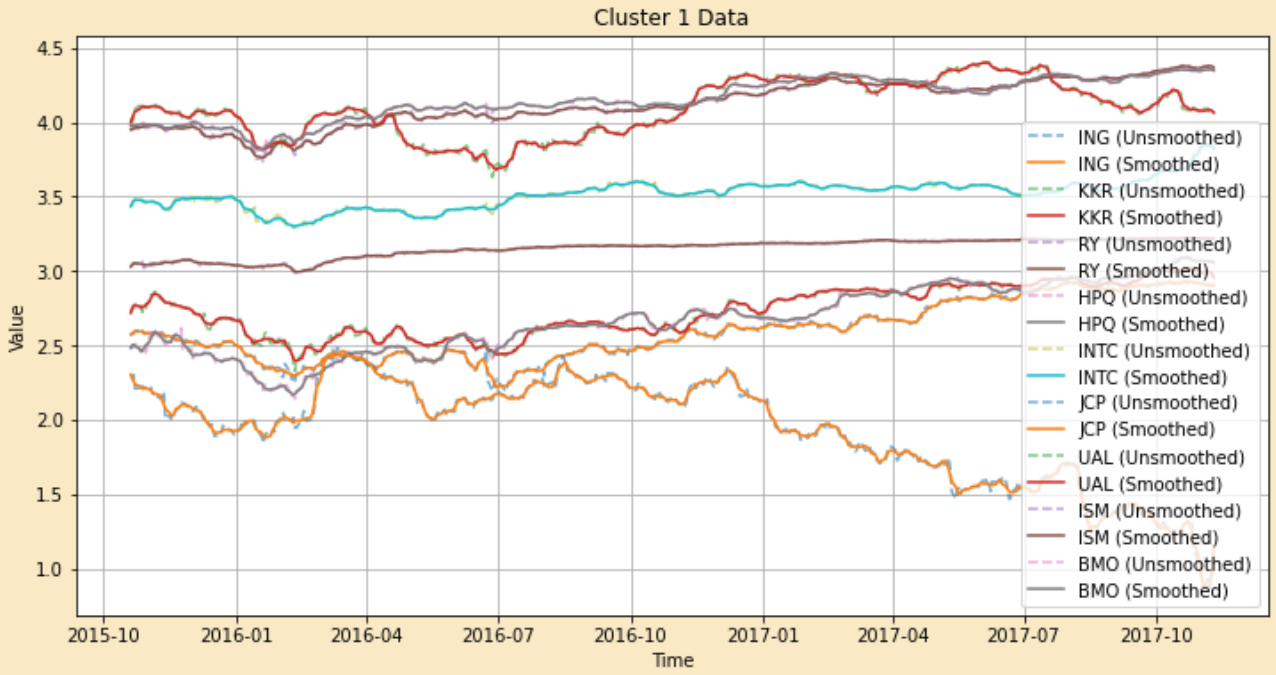
Results - Log Close Graph Cluster



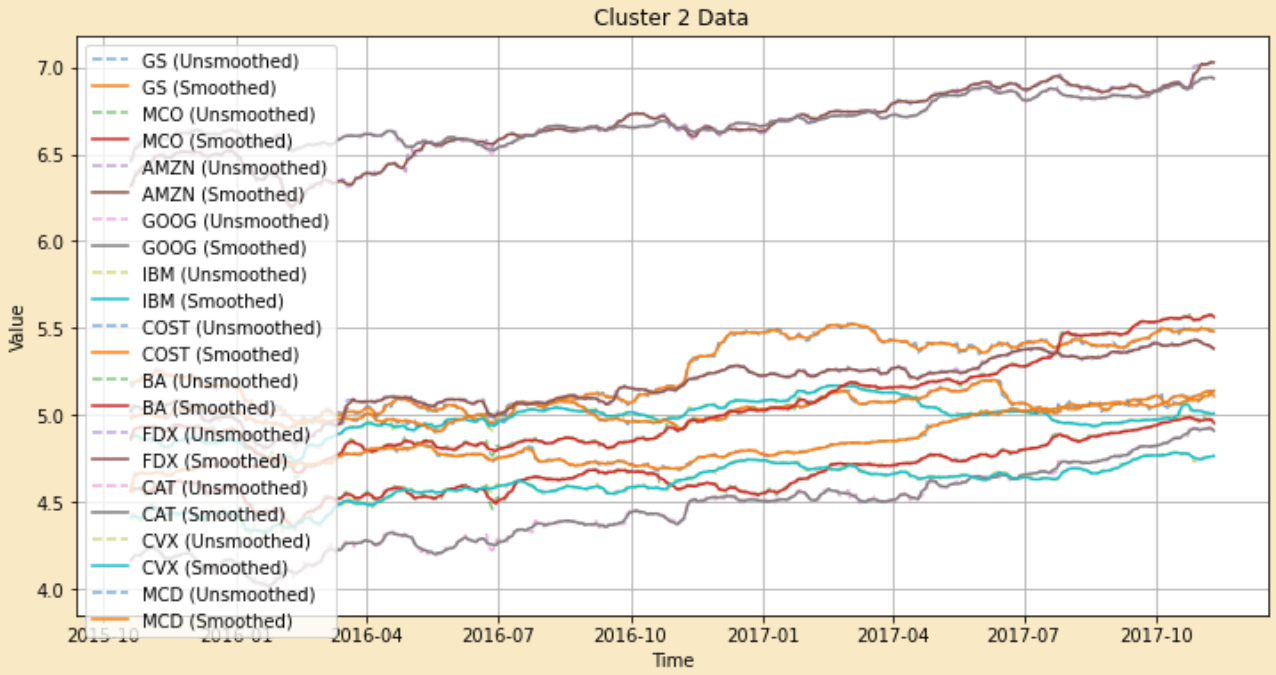
Results - Time Series Cluster 0



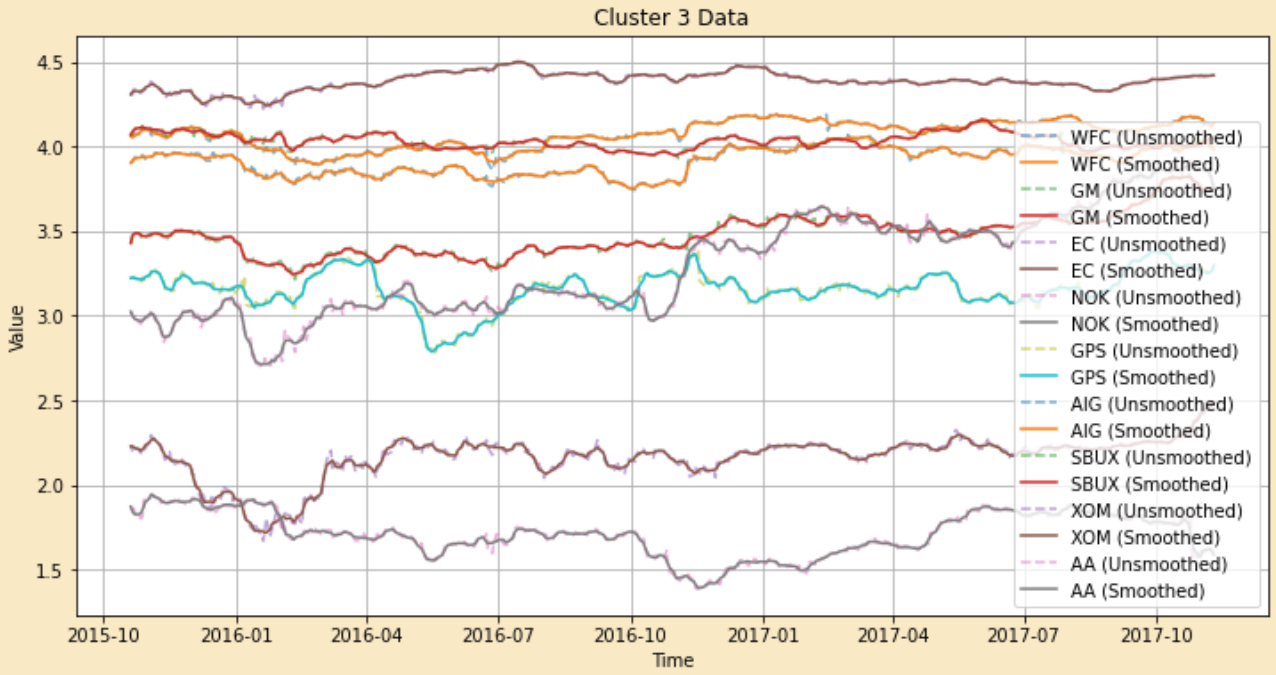
Results - Time Series Cluster 1



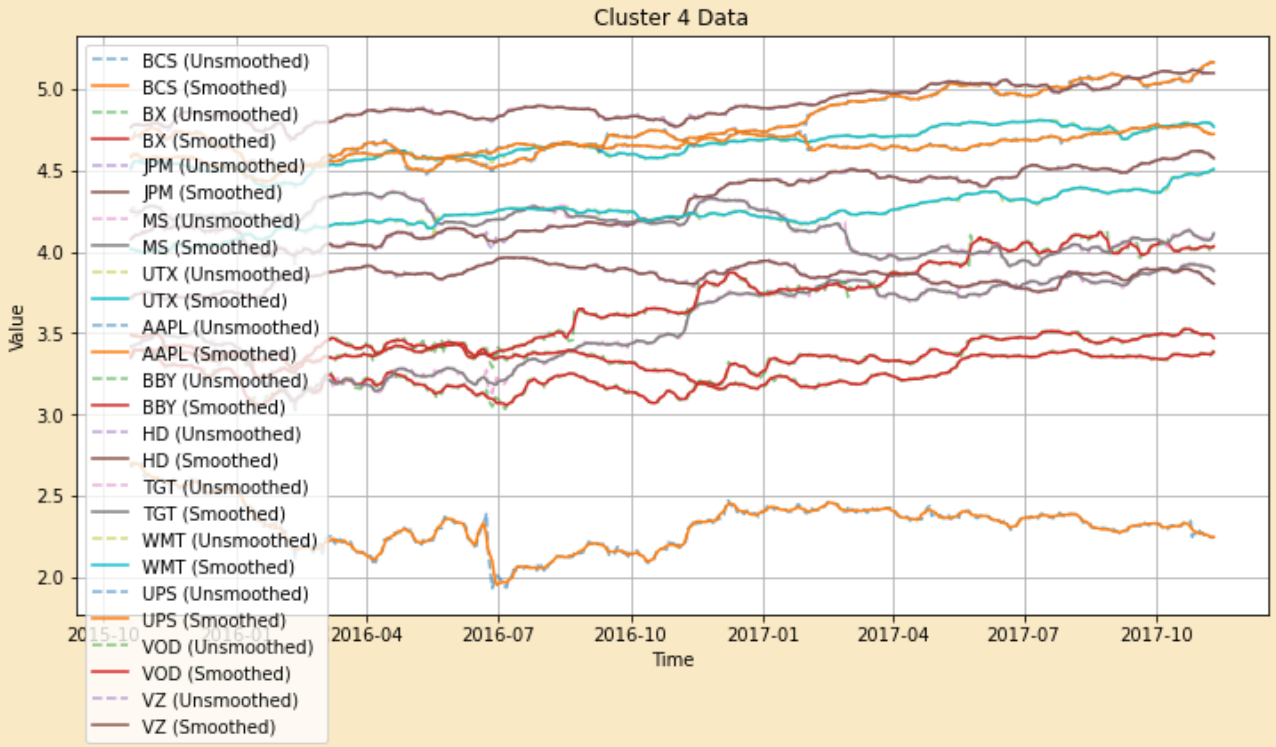
Results - Time Series Cluster 2



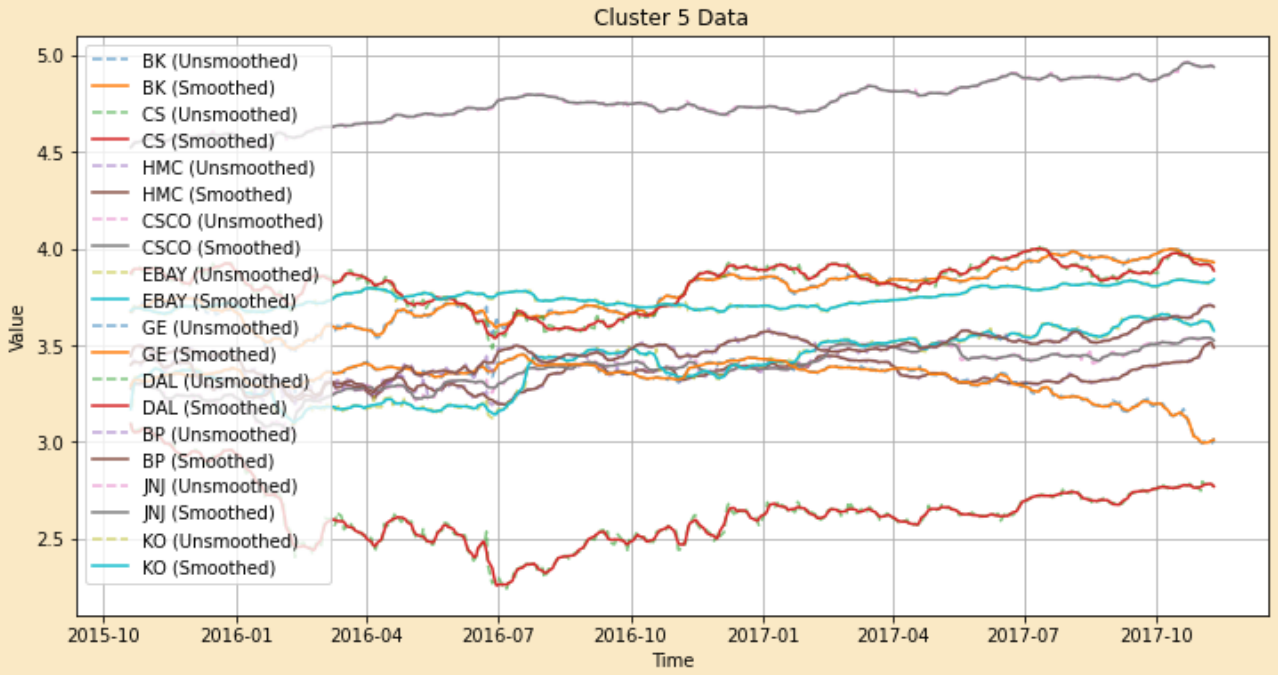
Results - Time Series Cluster 3



Results - Time Series Cluster 4



Results - Time Series Cluster 5



Limitations (More to be found ⏳)

- Limited tick data
- Computationally expensive to build graph for long time series