Cox Communications

MSA PROJECT WEEK

TEAM INFINITY NORM

Team Members

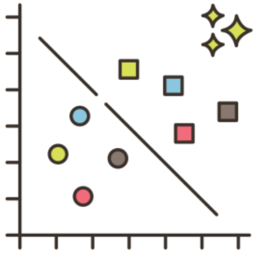
Sittun Swayam Prakash
Atrima Ghosh
Arushi Agrawal

Shyam Krishna Sanapeneni

Project done with social distancing!

Objective 1: Clustering

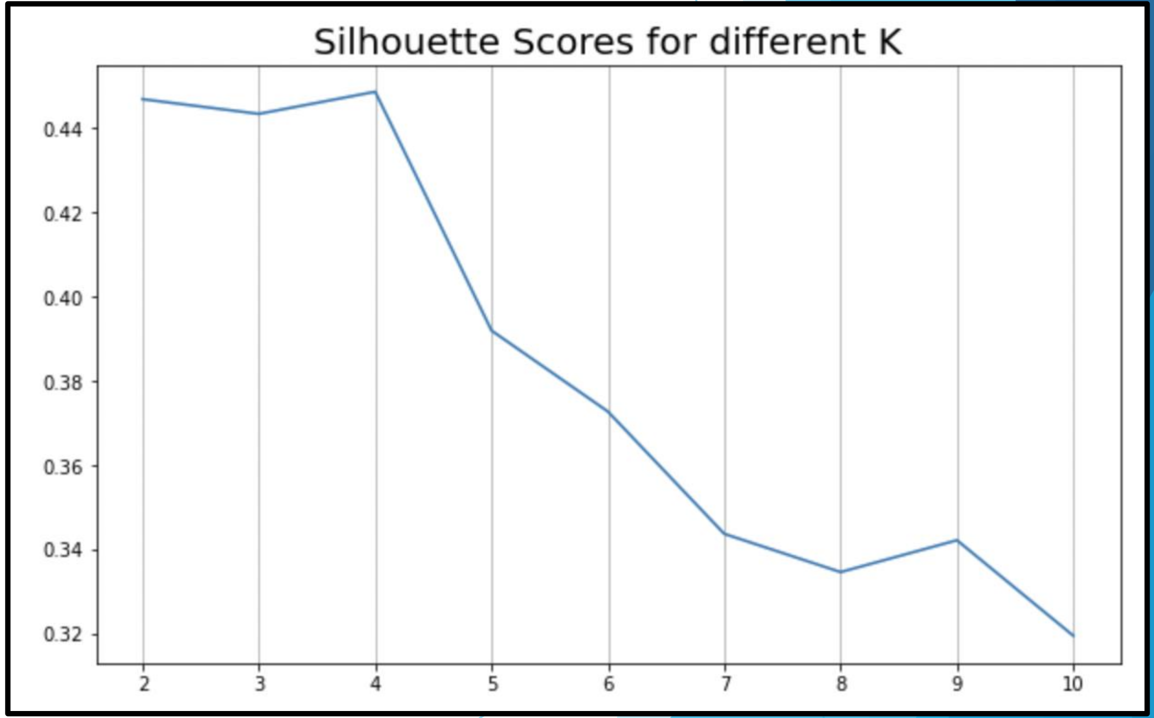
Discover a way to cluster array like FBC strings by Node and measure precision of clustering accuracy.



Algorithm:
K-Means Clustering

- Distance Metric-
  Euclidean Distance
  DTW (Dynamic Time Warping) distance

- Number of Clusters
  Using Silhouette Score



Objective 1: Clustering Performance

Fourier Approach

- Used Inverse Fourier transform to convert waveform from frequency domain to time domain to check for difference in results
- Obtained same number of clusters with similar Silhouette scores.

Frequency domain	Time Domain
[0.4467485724313497,	[0.44666686993678506,
0.443260156870188,	0.4432057698816877,
0.44850648448156727,	0.44845318662834577,
0.39187864602234446,	0.39186693202907386,
0.3726877311660598,	0.372677654095038,
0.3436932175038725,	0.34369911806865433,
0.334620641304548,	0.33461775547464107,
0.342175954163126,	0.34217926580170677,
0.3195340329507307]	0.3195413786660721]

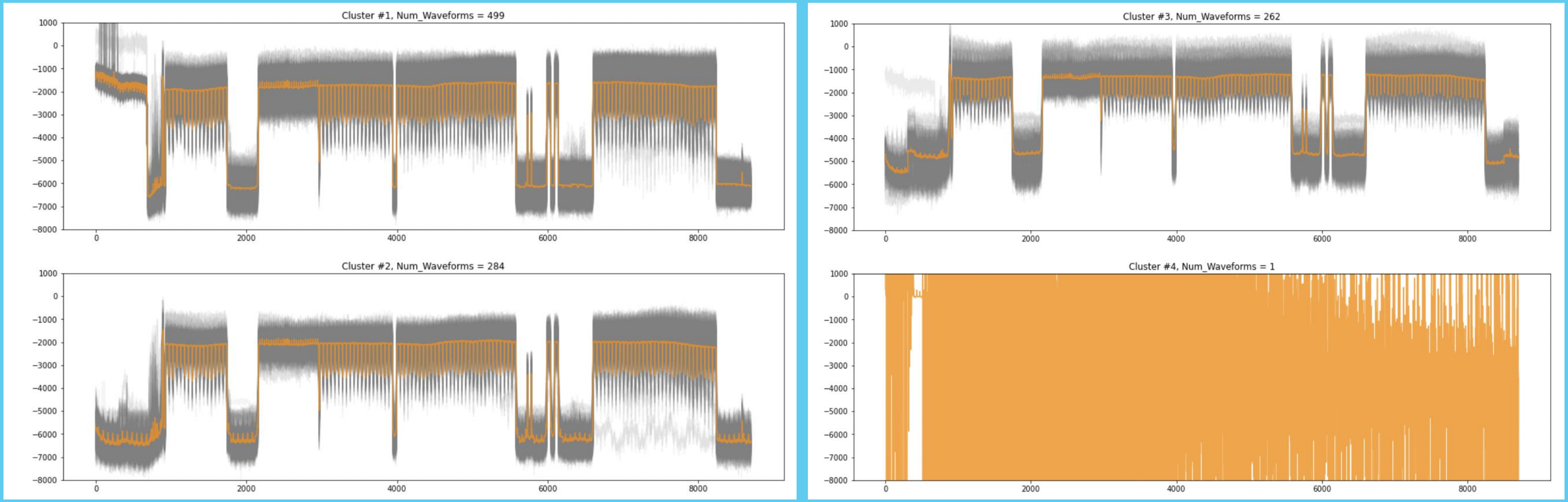
Reconciliation

- Every device ID has multiple entries
- Each entry of a device ID found under a single cluster in our test case

Final Performance

- Obtained a Silhouette score of 0.45

Objective 1: Clustering Results



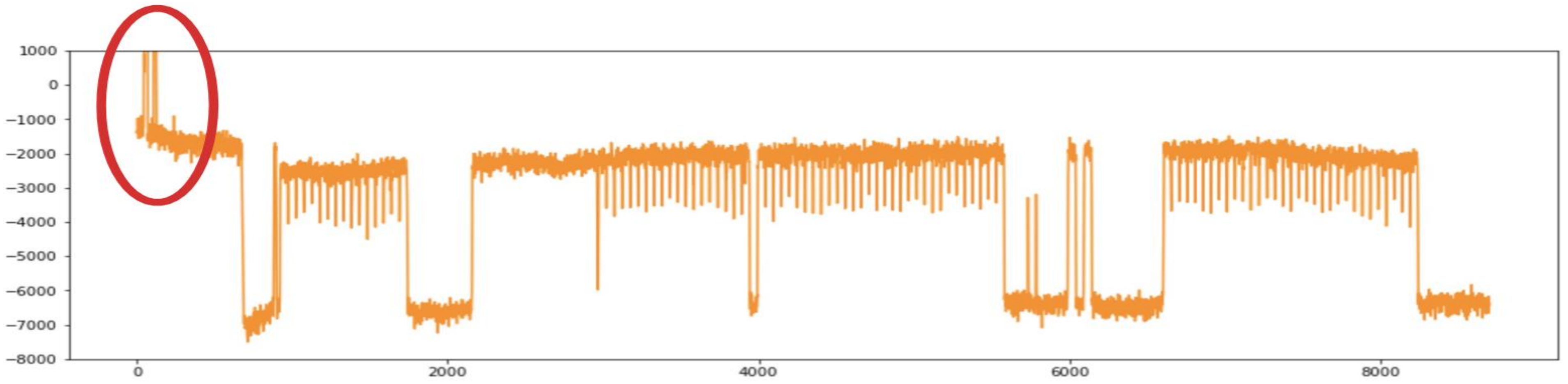
Clusters obtained for file 3VAC1.csv

Objective 2: Locate Wave Impairments

*Using secondary sources, parse the array like FBC strings for **WAVE** and **INGRESS** impairments.*

Type 1 – Above 0db Noise

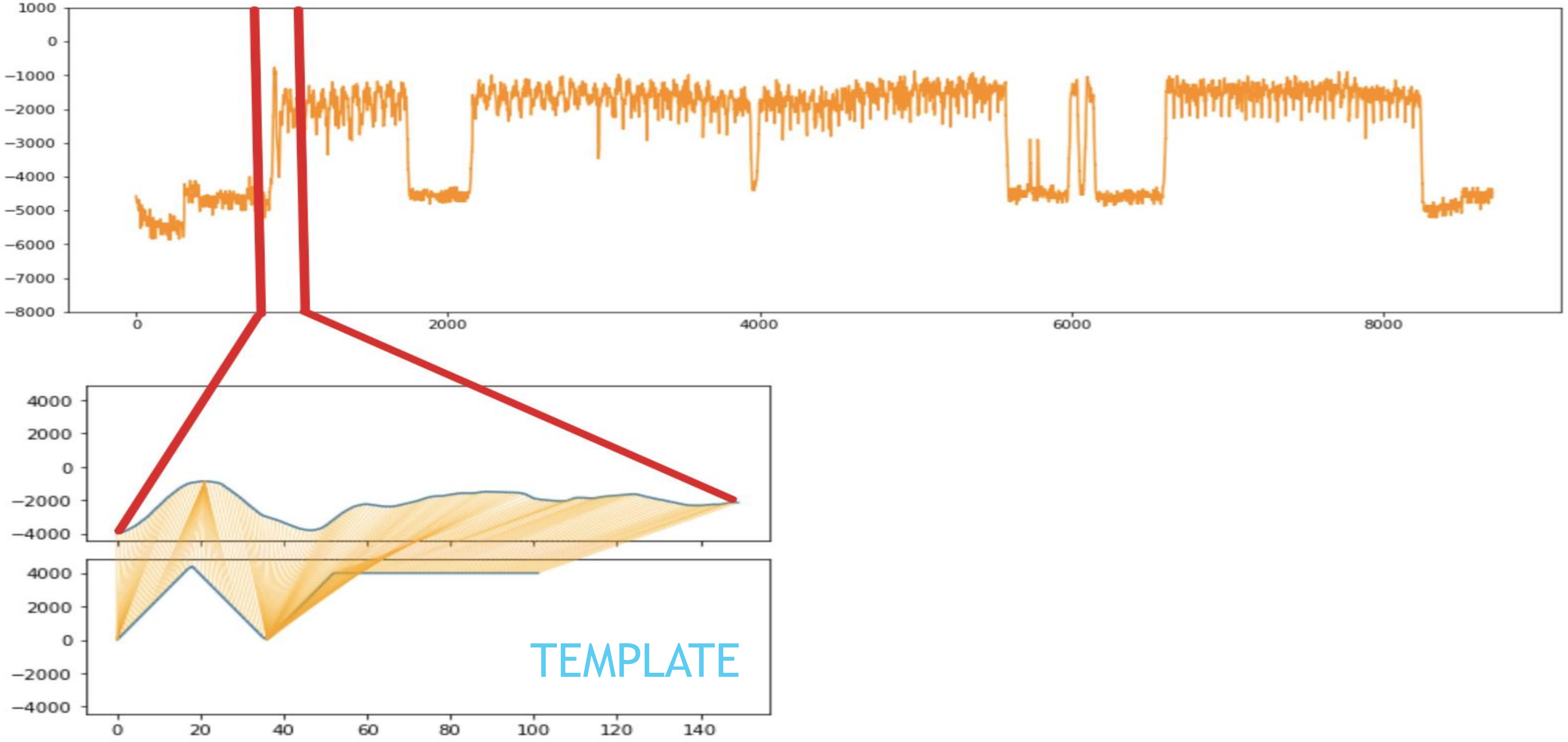
- Approach:
  - The spectral amplitudes are checked for values that go above 0db
  - It is classified as a valid sustaining > 0db noise impulse if it satisfies two criteria:
    - If it spikes at least > 5db (NOISE THRESHOLD)
    - And it sustains for a window of at least 30 points (NOISE WINDOW)
    - Both NOISE WINDOW and NOISE THRESHOLD are parameters that we can tune as per our detection requirements.
- Results:
  - Our algorithm is able to detect above 0 db noise easily. Runs in few seconds for every node.
  - Example of detected > 0db noise:



Objective 2: Locate Wave Impairments

Type 2 – Detect Wave Ingress

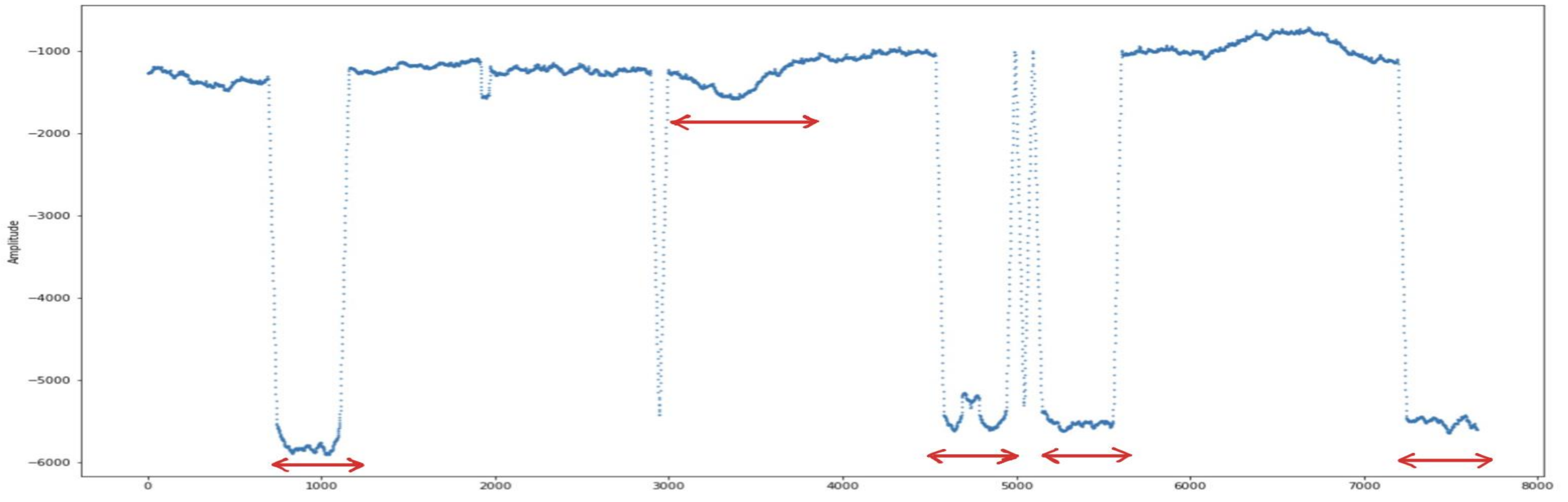
- Approach:
  - All spectral data only in the FM band of interest is extracted.
  - DTW (Dynamic Time Warping) distance score of this extracted segment with respect to a defined template of how the ingress should look is computed.
    - If this distance score is less than a threshold (MIN_DIST_THRESHOLD_FOR_INGRESS_DETECTION), we declare detection.
- Results:



Objective 2: Locate Wave Impairments

Type 3 – Detect Notches/Suckouts

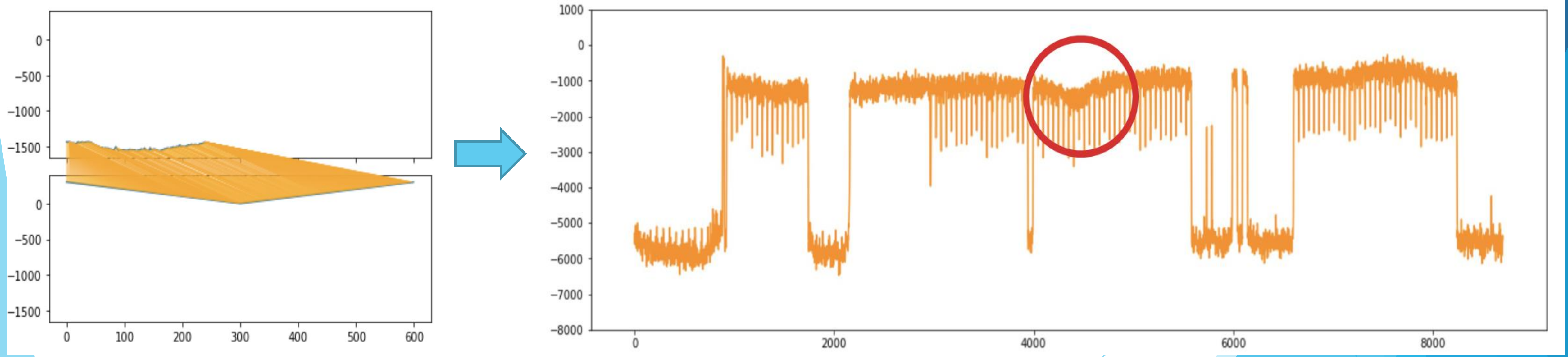
- Approach:
  - First, the input amplitudes are smoothed with a moving average window. All dip segments are extracted for each row's spectral data. This dip is defined w.r.t to the transmission flatband.
  - If these dip segments sustain for a minimum window (MIN_WINDOW_LENGTH_FOR_VALID_DIP), have at least a 3 db dip (MIN_DIP_AMPLITUDE_FOR_VALID_DIP), it is considered a candidate notch.
- Results:



Objective 2: Locate Wave Impairments

Type 3 – Detect Notches/Suckouts

- Approach (continued)
  - DTW (Dynamic Time Warping) distance scores of these candidate segments with respect to a defined template of how the notch/suckout should look are obtained.
    - If this distance score is less than a threshold (MIN_DIST_THRESHOLD_FOR_NOTCH_DETECTION), we declare detection.
- Results:



Objective 3: Most Common Ancestor

Using the network parent to child relationships, discover a way to algorithmically find the most common ancestor of any collection of address network elements and visualize this.

Pre-processing

- Obtain output from Clustering Algorithm
- Merge corresponding Geo-IDs of clustered Mac-IDs with network-parent-child

Most Common Ancestor

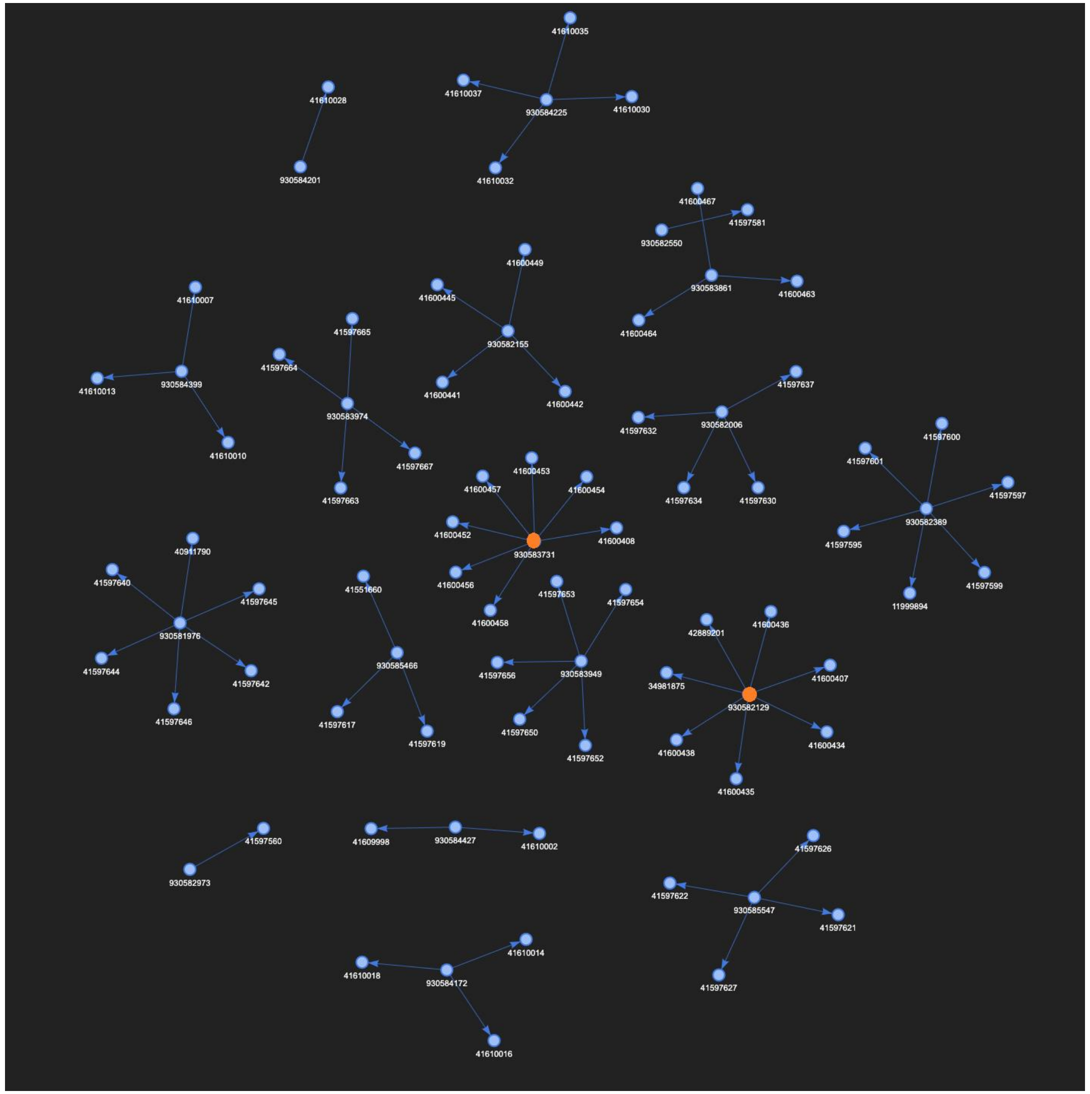
- Trace the path of each Geo-ID to its root node
- Create a dictionary containing counts of all nodes encountered in paths
- Display highest count nodes

Visualization

- Visualize cluster using NetworkX and Pyvis

Objective 3: Most Common Ancestor

Results:



Node_H0182_cluster1

Most common ancestors 930583731 and 930582129 with 7 child nodes each

Objective 4: Lossless Compression

Find a way to losslessly compress the FBC raw signal for easier access in Hive Hadoop.

- Used Gorilla Compression from the gorillacompression library
- Compression Scheme:

  (a) Calculate the delta of delta

  $$D = (t\_n - t\_(n-1)) - (t\_(n-1) - t\_(n-2))$$

  (b) If D is zero, then store a single ‘0’ bit

  (c) If D is between [-63, 64], store ‘10’ followed by the value (7 bits)

  (d) If D is between [-255, 256], store ‘110’ followed by the value (9 bits)

  (e) if D is between [-2047, 2048], store ‘1110’ followed by the value (12 bits)

  (f) Otherwise, store ‘1111’ followed by D using 32 bits

- Results:

	device_id	amplitudes
0	3c:04:61:fe:1e:81	[-6262, -6033, -6128, -6434, -6424, -6254, -62...
1	58:19:f8:2a:ce:a7	[-6612, -6627, -6640, -6914, -7127, -6670, -64...
2	a8:97:cd:41:0e:f1	[-4500, -4510, -4660, -4710, -4710, -4870, -48...

64 MB

	device_id	compressed_waveform
0	3c:04:61:fe:1e:81	{'encoded': b'\xc0\xb8v\x00\x00\x00\x00\x00\xd...
1	58:19:f8:2a:ce:a7	{'encoded': b'\xc0\xb9\xd4\x00\x00\x00\x00\x00...
2	a8:97:cd:41:0e:f1	{'encoded': b'\xc0\xb1\x94\x00\x00\x00\x00\x00...

47 MB

THANK YOU!