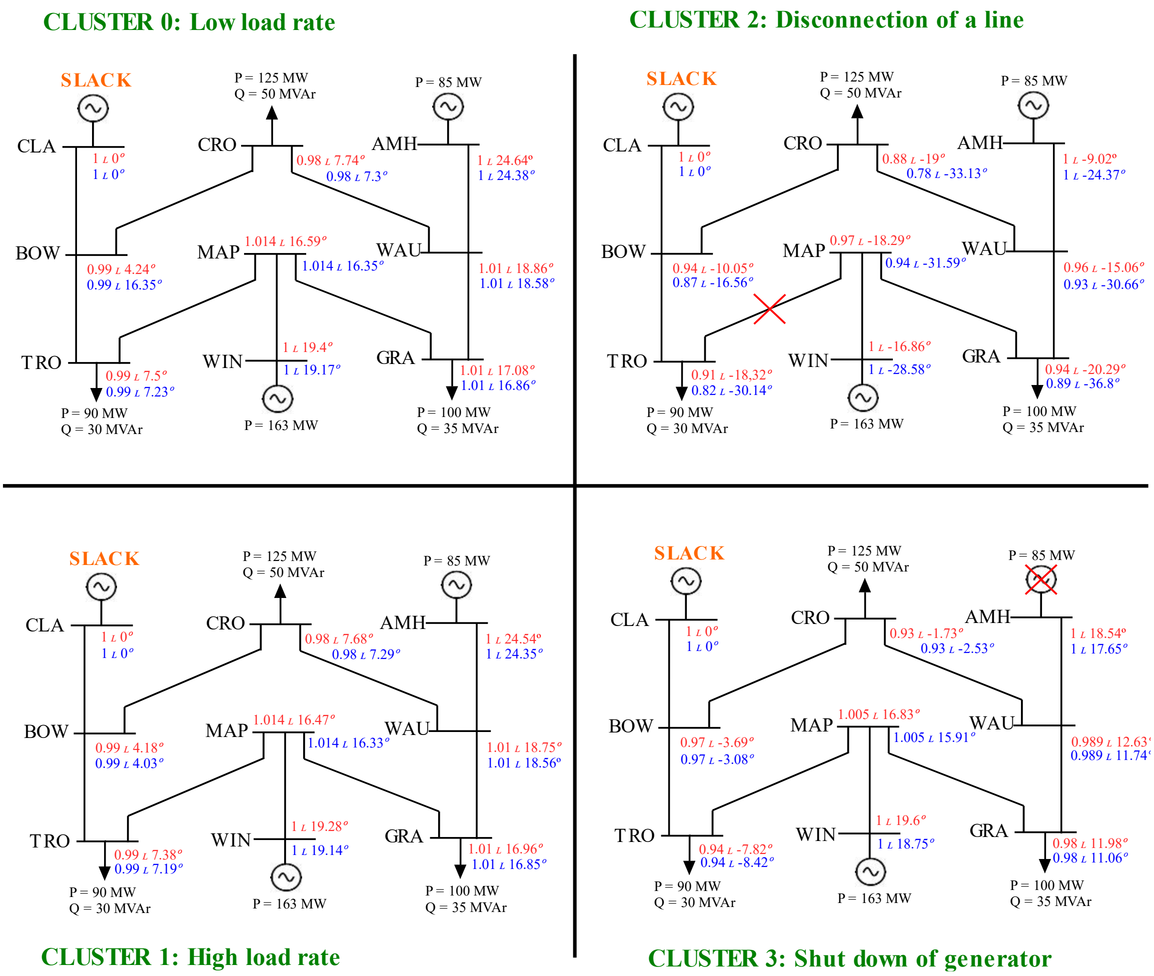
Marvin Killer & Julian Spit

ASSIGNMENT II – CLUSTER LABELING

This section of the project explains how the clusters were identified and labeled from the k-mean output data. Therefore, an analysis approach has been followed in order to understand the different outcomes from an electric point of view. Afterwards, a generic algorithm is outlined which would be recommended to implement in this type of application to automatically identify any type of cluster under the given constraints (four possible states) for any input data.

First of all, the cluster centroids have been identified and labeled to each possible state looking at the output values for the centroids (values in red) at each given dimension (having 18 different dimensions, 9 voltages and 9 voltage angles corresponding to the 9 buses in the system), as well as the furthest object to the the centroid of each class that represents the most extreme conditions (values in blue). In this way, it is possible to ensure that the labeling criteria based on the analyzed data covers all objects within each cluster (from the most extreme to the centroid) without having to look at all data. The average cluster centroid data and furthest object data (voltage and angle) for each bus are graphycally represented in the following figure:



In order to identify the states that each distinct cluster corresponds to, a closer look has to be taken comparing all voltages and voltage angles as well as estimating the power flow in the network. Note that active power always flows from the bus with the higher voltage angle to the bus with lower voltage angle, while reactive power flows from the bus with higher voltage to the bus with lower voltage.

The first very big hint which tells that something is happening in the grid is the value of the voltage and angle itself. As can be seen in cluster 2 and 3, all values are notoriously distorted, having very big power flows on most lines and voltage values that go much below the permitted 5 % voltage variation threshold of most electric networks around the world. Therefore, it is possible to say that in those cases something is wrong in the grid, hence a non-operating generator or line. Next, it is relatively easy to differ between these two states. In order to identify a line that is out of operation, it is enough to compare the voltage angles at each bus. When there are to consecutive buses with very similar voltage angles, there is no active power flow. Since the generators in the grid only produce active power, there is no power transfer in the given line, so most likely it is not working. For the proposed data set, that happens in cluster 2 between bus Troy and bus Maple, for both the centroids as well as the furthest object. Additionally, there is a very big amount of energy drown from the slack bus to feed the load at bus Troy, strengthening the proposed hypothesis. On the other hand, in order to identify irregularities with the generators, the easiest way is to compare the voltage and voltage angle levels of each generator in between the different clusters. In theory, since all generators depict PV buses, under normal conditions (excluding non-working line) they should take similar values. This is true for the generator at bus Winlock, but not at bus Amherst. Here it is possible to see that in Cluster 3, the voltage angle differs from the remaining two clusters on the left.

Next, the two normal conditions need to be distinguished that do not present any changes in the grid topology: low load rate and high load rate. These should usually not suppose major variations in the voltage levels since the network is supposed to be designed for these conditions. As can be observed in Cluster 0 and Cluster 1, all values are very similar to each other comparing both of them, with the only difference that the values in Cluster 1 are slightly smaller at each bus. Therefore, the most likely scenario is that the voltage level and voltage angles lower a little during the high consumption period, hence Cluster 1 depicts high load rate and Cluster 0 low load rate.

Note that all assumptions to label the different clusters are based on electrophysical rules. Since most data just represent average values of what might be happening at many given points in time, the actual clusters might correspond to distinct states. However, following a very generic and simple approach, this proposal depicts a simple solution which is quite reliable to classify the data clusters.

From here on, the proposed decision keys can be developed in an algorithm that automatically identifies and labels the different clusters accordingly to their centroids and furthest objects. This algorithm could be done by identifying extreme low voltage values, high voltage angle differences, consecutive buses with similar voltage angles, comparing generator values and to put the mean data for each cluster in order. With this, a very generic labeling program would work on any possible input grid configuration, always considering that there will only be the four possible proposed states. Due to the scope of the project and time restrictions, this algorithm is only a proposal and has not been implemented in the created Java program.