Problem Representation Team 1

As Team one already proposed, we defined both the Hypervertices V and Hyperedges E as the Sites of the MSA and the Repeat Classes respectively. However we want to note here, that we have to allow multiple of the same hyperedges in our graph representation.This is due to the different repeat classes that can occur on different internal nodes of the phylogenetic tree.

In our representation the weights w_v of the hypernodes are defined as a constant function $w_V : V \to \mathbb{N} : v \mapsto k$, where k represents the number of interior nodes for the given phylogenetic tree. This is later used to calculate the upper bound of calculations needed without the consideration of repeatclasses.

The weights of the hyperedges w_E are defined as $w_e : E \to \mathbb{N}_0 : e \mapsto |e| - 1$. This is used to represent the amount of calculation step savings gained by the repeat class that is represented by e. It maps to $(|e| - 1)$ because we always need to calculate one site of each repeat class at least. This way it is ensured that nothing will be saved if a repeat class contains only one hypernode.

For convenience reasons we will use the same representation of a partition as Team 2. So $\Pi = \{V_1, ..., V_k\}$ The function to be balanced between partitions will be called w_{V_i} for $1 \leq i \leq k$ and is defined as follows:

$$w_{V_i} = \left(\sum_{v \in V_i} w_V(v) \right) - \left(\sum_{e \in Int(V_i)} w_E(e) \right) - \left(\sum_{e \in Ext(V_i)} |e \cap V_i| - 1 \right)$$

$Int(V_i)$ and $Ext(V_i)$ represent the set of internal and external hyperedges of V_i respectively. Internal ones are hyperedges that only contain hypernodes which are included in V_i. External ones have at least one hypernode that is located in a partition other than V_i.

The first sum in this function will represent the full calculation time without consideration of repeat classes. The second one, that is subtracted, represents the full savings of repeat classes of V_i, which are not split. The third one represents the savings that are still left by hyperedges that are cut by the partitioning process. So what we have left would be the calculation time needed for partition V_i while considering the repeat classes.

As for the cut, we would also minimize the connectivity metric given by Team 2.